TEHNIČKI GLASNIK

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 15 – BROJ / NUMBER 1

OŽUJAK 2021 / MARCH 2021 – STRANICA / PAGES 1-167

/

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

ISSN 1846-6168 (PRINT) / ISSN 1848-5588 (ONLINE)

TECHNICAL JOURNAL

ISSN 1846-6168 (Print) ISSN 1848-5588 (Online)

TEHNIČKI GLASNIK - TECHNICAL JOURNAL

Scientific-professional journal of University North

Volume 15 Number 1
Varaždin, March 2021 Pages 1–167

Editorial Office:
Sveučilište Sjever / University North – Tehnički glasnik / Technical journal
Sveučilišni centar Varaždin / University Center Varaždin
Jurja Križanića 31b, 42000 Varaždin, Croatia
Tel. ++385 42 493 328, Fax.++385 42 493 333
E-mail: tehnickiglasnik@unin.hr
https://tehnickiglasnik.unin.hr
https://www.unin.hr/djelatnost/izdavastvo/tehnicki-glasnik/
https://hrcak.srce.hr/tehnickiglasnik

Founder and Publisher:
Sveučilište Sjever / University North

Print:
Centar za digitalno nakladništvo, Sveučilište Sjever

All manuscripts published in journal have been reviewed.
Manuscripts are not returned.

The journal is free of charge and four issues per year are published
(In March, June, September and December)
Circulation: 100 copies

Journal is indexed and abstracted in:
Web of Science Core Collection (Emerging Sources Citation Index - ESCI), EBSCOhost Academic Search Complete, EBSCOhost – One Belt, One Road Reference Source Product, ERIH PLUS, CITEFACTOR – Academic Scientific Journals, DOAJ – Directory of Open Access Journals, Hrčak – Portal znanstvenih časopisa RH

Registration of journal:
The journal "Tehnički glasnik" is listed in the HGK Register on the issuance and distribution of printed editions on the 18th October 2007 under number 825.

Preparation ended:
March 2021

Legend:
(1) University North, (2) Mechanical Engineering Faculty in Slavonski Brod, (3) Faculty of Graphic Arts Zagreb, (4) Faculty of Civil Engineering Osijek, (5) Faculty of Engineering Rijeka, (6) Faculty of Mechanical Engineering and Naval Architecture Zagreb, (7) Faculty of Metallurgy Sisak, (8) Tomas Bata University in Zlín, (9) Department of Physics of the University of Josip Juraj Strossmayer in Osijek, (10) Faculty of Humanities and Social Sciences Osijek, (11) Karlovac University of Applied Sciences, (12) University of Applied Sciences Velika Gorica, (13) Department of Polytechnics - Faculty of Humanities and Social Sciences Rijeka, (14) Faculty of Electrical Engineering and Computer Science - University of Maribor, (15) Faculty of Civil Engineering - University of Maribor, (16) University College of Teacher Education of Christian Churches Vienna/Krems, (17) Mechanical Engineering Faculty Tuzla, (18) Mechanical Engineering Faculty Sarajevo, (19) University of Travnik - Faculty of Technical Studies, (20) Higher Education Technical School of Professional Studies in Novi Sad, (21) University of Novi Sad - Faculty of Technical Sciences, (22) Faculty of Mechanical Engineering - University of Montenegro, (23) Brno University of Technology, (24) Odessa State Academy of Civil Engineering and Architecture, (25) Faculty of Civil Engineering - University of Mostar, (26) Faculty of Manufacturing Technologies with the seat in Prešov - Technical University in Košice, (27) Faculty of Mechanical Engineering - University of Maribor, (28) College of Engineering, IT & Environment - Charles Darwin University, (29) Universite Libre de Bruxelles, (30) Vishwakarma Institute of Information Technology (Pune, India), (31) AISSMS Institute of Information Technology (Pune, India), (32) Permtech Research Solutions (India), (33) University of Belgrade, (34) National Dong Hwa University - Taiwan

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210225102651

Convergence of Smart Technologies for Digital Transformation

Bertrand Mareschal, Mandeep Kaur, Vilas Kharat, Sachin Sakhare

Abstract: This foreword is related to the sixteen papers published in this issue, which were presented at IC2ST-2021 - International Conference on Convergence of Smart Technologies. The Aspire Research Foundation organized this conference in Pune, India, January 9-10, 2021. These sixteen papers have not and will not be published anywhere else.

Keywords: Digital Transformation; Smart Technologies

Preface

The convergences of smart technologies have brought digital revolution with a paradigm shift from manual operations to automated operations in the world around us. The convergence of technologies is playing a fundamental role in the growth and evolution of smart and embedded systems in this digital era. We can imagine a world where robots, sensors, automated guided vehicles (AGVs), products, databases communicate and controllers with one another. The smart systems assimilate elements of computing with the physical processes and components. The computing elements communicate with sensors and these sensors monitor physical indicators, which adapt the cyber-physical environment. The convergence of technologies enable the cyber-physical systems (CPSs) to make use of sensors to connect the distributed intelligence for gaining greater knowledge of the environment and for taking more accurate actions and tasks. The design of smart systems for industry 4.0 and future requires joint dynamics of computers, software, networks, and physical processes.

The smart technologies comprise distributed systems, communication systems, computational resources, data resources and control components connected with diverse physical processes. However, the fundamental design requirement for smart systems is that the hardware components, cyber components, distributed systems, communication technologies and computational resources should be able to work altogether to deliver the expected functionality accurately.

The combination of the smart technologies, cyber physical systems, Internet of Things, artificial intelligence and embedded systems will drive innovation and effectiveness in a global marketplace. The convergence of smart technologies will enable us to gather and analyze big data across smart devices and will assist faster, scalable, and efficient processes to produce better-quality goods at the cheapest costs.

The Issues to Be Addressed for the Successful Implementation of Smart Technologies

- **Security**

The most important concern for businesses is security as they embark the 4.0 journey. This includes both cyber and physical security. The industries are worried about the followings:

- Stealing of Intellectual Property Rights
- Stealing of data containing client information
- Hacking of Machines for ransom

The workstations and other systems are at higher risk as the attack surface is increasing day by day and it is easy to attack by using malwares, adwares and phishing scams. The hackers are not only targeting industrial systems but also targeting the IoT-cloud collaborative environments where several users are connected through diverse devices. Usage of smart technologies has introduced new threats to users and proprietary data.

- **Adoption**

The next challenging task for smart technologies is adoption. An automation and evolution of robotics technology is posing a potential threat to the present workforce, which will be replaced by machines, and this issue has to be addressed carefully. The workforce reduction will be resultant from the evolution of smart technologies but it may lead to unemployment around the world. Industry 4.0 may need very skilled workforce. The developing and under-developed countries may be reluctant to adapt the smart technologies to avoid the situation of unemployment and resistance from the current workforce. If the skilled workforce will be required then it may not be possible to train the existing workforce but new workforce can be trained where young generation can certainly contribute. This will drive a solution for the front-line production and second-line functional support. The adaption of latest technologies will remain questionable with respect to adaption of these technologies at global level.

- **Seamless Connectivity**

Connecting workstations, laptops, mobile phones, IoT devices, deploying IoT sensors, through seamless communication channels and collaboration tools will drive a need for highly reliable and secure internet. People around the world will not be able to take advantage of the smart technologies until there is a promising seamless connectivity solutions delivered by the researchers and industrialists.

- **Standardization**

Implementing new techniques, deployment of smart systems will need many disparate technologies to connect and communicate. There is a need to set some standards over the existing standards that can promise seamless connectivity, smooth communication and AI based automated solutions. Industry 4.0 and smart world cannot be created overnight, there are interoperability needs and standard solutions to be developed to permit data from one global position to another global position irrespective of diverse underlying architectures, gateways and transmission media. This implies inclusion of smart technologies in the present world will be a long-term evolution. The standards should be introduced, implemented and accepted worldwide.

The Pillars of Smart Technologies

- **Internet of Things (IoT)**

The physical world has been transformed into the digital world where everything is connected seamlessly. An eruption of IoT devices and latest technologies has permitted human-beings to stay connected irrespective of geographical locations. Internet of Things has generated a sub-segment as industrial Internet of Things, which is known as IIoT. Industry 4.0 with IIoT is anticipated to transform the production line in industries, the way we live, the way we work and the way we interact with the digitized world.

- **Big Data**

The greatest challenge for IIoT is privacy and security. IoT and the seamless connectivity means gathering of real time data from multiple places for enabling quick decisions to enhance productivity and efficiency. The key advantage of IIoT is the data it generates and transforms the data into actionable insights.

- **Digital Twin**

Digital twins (DTs) are gaining extensive consideration from industry engineers, researchers and academicians. Cyber physical systems and digital twins can impact manufacturing and production systems with greater efficacy, intelligence and resilience. There are many revolutionary digital technologies evolving to gear up and restructure the activities to fulfil digital transformation objectives. Digital twin comprises technologies for the upcoming years of automated, knowledge-filled, goal-oriented and human-centric services.

- **Cyber Physical Systems**

Cyber-Physical Systems are assimilations of embedded systems, computation, physical processes and networking. Embedded systems and computer networks control and monitor the physical processes, where physical processes use computations and form the base for futuristic smart devices to improve quality of life in diverse areas. Cyber-physical systems will improve health care systems, emergency services, management of traffic flow, service sector, manufacturing and production line in industries, and many other areas where computer intelligence can be applied. Cyber-Physical Systems will provide the basis of critical infrastructure, and futuristic intelligent services.

- **Cloud Computing**

Cloud computing provides on-demand computational and data-storage resources to the users. Instead of purchasing, possessing, and maintaining data storage centers and servers, the industries can access cloud technology services such as computational power, software instances and data-storage space on as-and-when required basis from the cloud providers. Industries irrespective of size are using the cloud services such as data backup, email, disaster recovery, software development, big data handling, and web based applications. Cloud computing is one of the most important building blocks of smart world as it will help the world with easy access to a wide-ranging of technologies. With the cloud, industries can grow to new geographic areas and deploy internationally in minutes.

- **Artificial Intelligence**

The progressive adoption of Artificial intelligence and machine learning techniques improve the flow of resources through the manufacturing process. In industry 4.0 scenario, an AI becomes an essential factor that assists enterprises in digitization of the manufacturing sector. We can also say that AI is the driving force for this century. AI can assist the world in predicting maintenance, generative designing, manufacturing, delivering quality, human-robot collaboration, supply-chain management, waste reduction, and in production optimization.

- **Simulation**

Nowadays industries are able to manufacture things in a more fascinating way by using simulations. We can say that innovative technologies and artificial intelligence based systems have given a fair share for producing simulations based products in today's era. In simulation, the computer models are capable enough to simulate the operations of any real world system. Perhaps the appropriate simulated systems can assist an organization is to approximation of the return on investment before the product is actually initiated.

- **Cybersecurity**

Smart and intelligent systems require security along with preserving the ideas of scalability and functionality. IoT security solutions must be introduced where all the devices are connected and information can be shared seamlessly. Cybersecurity attacks are increasing more frequently day by

day as attackers exploit vulnerabilities of the systems for financial gain. Cyber attackers have security controls and are attempting to improve their attacking techniques. In order to use the smart technologies optimally, it is important to impose strong cybersecurity controls over the systems.

Digital transformation with the evolution of smart technologies is the profound transformation of the businesses, organizational activities, processes, competencies and models. In order to fully leverage the changes and opportunities of the smart digital technologies and their accelerating impact across society in a strategic way, there is a need to handle the aforementioned issues and to strengthen the pillars of smart technologies. The smart technologies should be embraced by the society and the business world globally to create the capabilities of fully leveraged possibilities & opportunities of new technologies and their impact in a faster, better and in an innovative way in the future.

Authors' contacts:

Dr. **Bertrand Mareschal**, Professor
Solvay Brussels School of Economics & Management,
Avenue F. D. Roosevelt 50, 1050, Brussels, Belgium
bmaresc@ulb.ac.be

Dr. **Mandeep Kaur**, Managing Director
Aspire Research Foundation,
404/405, Pinnacle Prestige, Tilak Road, Pune, India
mandeep.kaur@aspireresearch.in

Dr. **Vilas Kharat**, Professor
Dept. of Mathematics, Savitribai Phule Pune University,
Pune - 411 007, India
vilaskharat@unipune.ac.in

Dr. **Sachin S. Sakhare**, Professor and Head
Dept. of CSE, Vishwakarma Institute of Information Technology,
Bibwewadi, Pune - 411037, Maharashtra, India
sachin.sakhare@viit.ac.in

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210215210742

Original scientific paper

Environmental Adaptation of Construction Barriers under Intuitionistic Fuzzy Theory

Katarina Rogulj*, Jelena Kilić Pamuković

Abstract: The project of construction barriers removal is a comprehensive planning task and it demands a suitable support for identification, and priority ranking of facilities necessary for barriers removal. This paper proposes a multicirteria Intuitionistic Fuzzy (IF) ELECTRE model to support decision makers in the process of managing of removal project of construction barriers for physically disabled in high schools. IF ELECTRE approach is used to deal with complex problems, where decision-makers have ambiguities and dualities in evaluation of considered solution. Hereby 17 high schools are defined and seven criteria are determined by decision-makers. These criteria are further used for the alternatives assessments. Each DM is also evaluated by linguistic and numerical values, assigning them this way an importance according to their background and the years of experience. The Intuitionistic Fuzzy Weighted Average (IFWA) operator is calculated to achieve aggregated alternatives evaluations. Furthermore, concordance and discordance sets and indexes are calculated to obtain dominance matrix and final ranking of schools for the construction barriers removal. The model is validated on high schools in the city of Split. Using IF theory, the given problematic can be operated more effectively by diminishing the inaccuracy of available information.

Keywords: construction barriers; ELECTRE; Intuitionistic Fuzzy Theory; multicriteria decision-making; physically disabled

1 INTRODUCTION

Identifying constriction barriers in the physical world demands taking into account the positioning of necessary ramps, elevators, lifts, adjustment of sanitary elements, etc. To eliminate the construction barriers as much as possible, designers, spatial planners, architects, transportation planners, construction contractors, and many others must have a crucial role in such projects. Before construction even begins, the process requires an effective cooperation between spatial planners, architects, contractors and end-user. Such a cooperation results in constructing the facilities that serve to the public, and are economically practicable, and enables creative design [1]. Managing this type of projects is a complex and poorly structured task, because it includes various aspects which seek a holistic approach. To deal with this problems, systematic and sustainable decision procedures are needed in management activities. Also, there is a significant lack of approaches that deal with this type of problem under the multicriterial decision making (MCDM) environment. According to Kassab et al. [2], MCDM is a decision analysis tool that is beneficial for the assessment and comparison of the alternatives by multiple criteria, and then ranking of these alternatives from most to least preferred. However, multiple-party problems containing various conflicting criteria, various different solutions and multiple decision makers with different opinions and attitudes are more comprehensive and involve a series of actions by participants, and eventually end in failing with decision or indecision.

Therefore, in this paper, a MCDM model is proposed. More particularly presented research refers to construction barriers in public education facilities such are high schools (HS) in urban and suburban areas. The approach is designed to be used at a local governmental level. The aim of the research is to resolve the multicriterial problem dealing with the removal of construction barriers in school facilities by provision of a unique multicirterial model, which stands as a tool for planners dealing with this issue. MCDM is used for solving various complex problems with multiple criteria, solutions and decision-makers to give a support in finding the most appropriate alternative. Although, the problem of construction barriers removal is mucliticirterial, it is not sufficient to use classical MCDM method due to the uncertain information and duality in decision maker's evaluations.

Nowadays, classical MCDM methods are not efficient when dealing with uncertain and vagueness data in decision-making process. Zadeh [3] proposed Fuzzy Set Theory, which was lately integrated in MCDM methods. This tool is very effective when it comes to uncertain data. Since classical Fuzzy Set Theory is shown to be hard for decision makers to quantify opinions between zero and one, an Intuitionistic Fuzzy Theory (IFT), developed by Atanassov [4, 5], has shown to be more applicable. The Intuitionistic Fuzzy Sets (IFS) are characterized by three degrees: membership, non-membership, and hesitancy. In recent years, the IFS has been applied in many fields such as decision-making problems, pattern recognition, health and medical diagnostic, supplier selection, personnel selection, selection of the facility location, and evaluation of renewable energy.

Atanassov, Pasi, and Yager [6] proposed an intuitionistic fuzzy interpretation of multi-stakeholders and MCDM. Each decision maker evaluated the alternatives according to each defined criterion: their evaluations are described as numeric values under the intuitionistic IFT. Hong and Choi [7] developed new functions to measure the degree of accuracy of membership of each alternative evaluate by criteria which are presented as uncertain values. Hung and Yang [8] in their study gave an approach for measuring distance between IFSs which is based on the Hausdorff distance. A new method for solving MCDM problem under IFT is presented by Liu and Wang [9] They firstly defined an evaluation function which served to measure the degrees of satisfaction and non-satisfaction of alternatives. After that, the concept of intuitionistic fuzzy operators is described. Szmidt and Kacprzyk [10], in their study have has determined solutions in group decision using IFS, while in [11] they extended the idea of a fuzzy logic to a state when individual opinions are introduced as IF preference relations. Wang [12] proposed a decision approach under ambiguity information to service

selection application based on Quality of Services. Xu [13] gave a detailed description of preference relations and their properties. In addition, they proposed new preference relations. Xu [14] proposes definition of terms of positive and negative ideal intuitionistic fuzzy set. They applied the measure of similarity to MCDM combined with IFT, while [15] proposed accuracy function for IFS. Another similarity measure between IFS was developed by Xu and Yager [16] which was further applied in group decision making consensus analysis.

Li and Cheng [17] presented a new concept of similarity between IFSs and its application to problems of pattern recognition. Liang and Shi [18] also introduced some new similarity measures of IFSs, and applied them to pattern recognition example to gain more reasonable results than those from existing methods. An information-theoretic approach for IFSs is introduced by Vlachos & Sergiadis [19], and applied to the problems of pattern recognition. Similar study was presented by Wang and Xin [20] who developed some new distance measures of IFSs to pattern recognition applications, and Zhang and Fu [21] proposed a new methodology for measuring the similarity degree between three fuzzy sets and between IFSs.

An IF MCDM methodology is proposed by Boran [22], integrating TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method into the IFT to selection of facility location, while Boran, Genc and Akay [24] proposed an IF group MCDM with TOPSIS method to the supplier selection problem. Comparison of the renewable energy technologies for generating electricity is presented by Boran, Boran, and Menlik [25] using TOPSIS based on IFT, and also, Roy [26] presented methodology under IFT for evaluation of renewable energy.

The ELECTRE method was developed by Roy [27]. It supports the decision making problems with qualitative or quantitative criteria. The ELECTRE stands for ELimination Et Choix Traduisant la REalite [28] and was at first cited ELECTRE for trading reasons [32]. This approach has evolved into different variants, known as ELECTRE I, ELECTRE II [29], and ELECTRE III [30, 31]. The idea of ELECTRE concerning concordance, discordance and outranking concepts originate from real-world applications, and it uses concordance and discordance to analyze outranking relations among solutions [32]. The method is based on a pair-wise solutions' comparison. These comparisons are compounded of evaluated information obtained from decision maker. The decision maker uses concordance and discordance sets and indices to analyze relations among different alternatives and to choose the best one. The ELECTRE method includes evaluation information, which are fuzzy and mostly not applicable to the real-life decision-making problems [23]. Hereby ELECTREE method is proposed with IF Theory to gain more accurate and precise results. Furthermore, IFT can be used to evaluate different solutions and to classify different kinds of concordance and discordance sets to fit the real decision-making process. In this paper, the model of IF ELECTRE is proposed to managing the project of removal of construction barriers in high schools' facilities. The model is validated on the high schools in the city of Split, Croatia.

1.1 Literature Review

In the previous studies authors performed qualitative and quantitative research on construction barriers. They gave a list of construction barriers of facilities; access measures for them; and ways to improve through renovation, remodeling, and removing barriers across the urban landscape. Woolley [34] in their study identified the barriers to the outdoor play spaces for children with disabilities. Agarwal et al. [35] studied disabled students who faced various structural barriers, such as lack of ramps or elevators in school facilities, heavy doors, inaccessible washrooms, and inaccessible transportation to and from school, lack of automatic doors, etc. The photo voice method was used to gather images with construction barriers. Hammel et al. [36] developed a framework to describe how environmental factors influence the participation of people with disabilities, highlighting domains of facilitators and barriers.

Church and Marston [37] measured an access for people with physical impairments within urban areas that extends beyond the standard-based approach. Martin [38] gave a qualitative and quantitative overview on physical activity engagement of the disabled from a social relational model perspective. Hannon [39] analyzed the general physical accessibility of facilities in the Munster region, while Kroll et al. [40] made an exploratory study to investigate how the physically disabled struggle with access barriers, and to define strategies to increase access to needed services. Kayes et al. [41] adopted a qualitative research on physical barriers as an interference for physical activity for people with multiple sclerosis. Yuker et al. [42] gave a comprehensive source of information pertinent to the education of physically disabled, with an accent on the construction of facility barriers. Leigh Hill [43] examined the level of physical accessibility for students with disabling conditions in universities across Canada, while Klinger [44] examined the evidence of the physical accessibility of schools for students with mobility impairments and provided an overview of the barriers and facilitators. Burton et al. [45] presented findings from a project that examined the environmental, and institutional barriers faced by disabled people. Martin Ginis et al. [46] made a review of published studies of factors related to leisure-time physical activity among people with physical disabilities.

None of the mentioned studies dealt with the construction barriers removal as a multicriterial problem and none has developed any type of model to solve vagueness and duality in the observed problems.

2 METHODOLOGY

Atanassov [4] developed the IFS theory to deal with uncertainty. In this section, a review of some necessary concepts related to intuitionistic fuzzy sets is given.

Definition 1: Let X be a finite set, and let $A \subset X$ be a fixed set [5]. Where X can be described as:

$$X = \left\{ a, \mu_x(a), \nu_x(a) \middle| a \in A \right\} \quad (1)$$

Where

$$\mu_x(a):\mu_x(a)\in[0,1], S\to[0,1]$$
$$\nu_x(a):\nu_x(a)\in[0,1], S\to[0,1] \quad (2)$$

Membership and non-membership functions can be explained as:

$$0\le\mu_x(a)+\nu_x(a)\le 1 \quad \forall s\in S, R\to[0,1] \quad (3)$$

Definition 2: The hesitation degree of IFS is $\pi_x(a)$ and can be described as:

$$\pi_x(a)=1-\mu_x(a)-\nu_x(a) \quad (4)$$

where $\pi_x(a)$ is degree of uncertainity of x to A.

Let Z and Y be IFSs of the set A, then multiplication operators are [4]:

$$Z+Y=$$
$$=\left\{\mu_z(a)\cdot\mu_y(a),\ \nu_z(a)+\nu_y(a)-\nu_z(a)\cdot\nu_y(a)\middle| s\in S\right\} \quad (5)$$

In this section, an IF ELECTRE approach is presented to the project management of construction barriers removal in HS objects. In proposed algorithm, the judgments provided by different decision makers are given as well as the quantitative and the qualitative data [32, 33]. Three groups of experts are asked to compare each alternative by each criterion. The algorithm is described in an eight-step, and presented in Fig. 1.

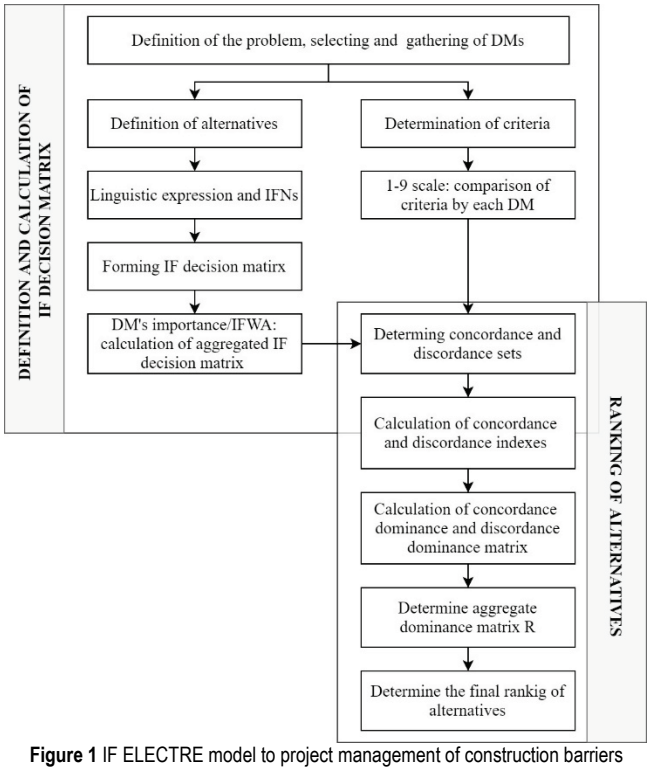


Figure 1 IF ELECTRE model to project management of construction barriers

Algorithm:

Step1. Criteria weight determination by pariwise comparions using 1-9 scale developed by Saaty [47].

Step2. DMs importance calculation using Boran et al. [48] expression:

$$\lambda_l=\frac{\left(\mu_l+\pi_i\left(\frac{\mu_l}{\mu_l+\nu_l}\right)\right)}{\sum_{l=1}^{k}\left(\mu_l+\pi_i\left(\frac{\mu_l}{\mu_l+\nu_l}\right)\right)} \quad (6)$$

where $\lambda_l\in[0,1]$ and $\sum_{l=1}^{k}\lambda_l=1$.

Step3. Calculation of aggregated intuitionistic fuzzy decision matrix using DMs' importance as the base to equation of IFWA operator [49]. Each DM's opinion is merged into single opinion.

Let $P^{(l)}=(p_{ij}^{(l)})_{m*n}$ be the IF decision matrix of each DM. $\lambda=\{\lambda_1,\lambda_2,\ldots,\lambda_k\}$ is the importance of the DM.

$$P=(p_{ij})_{m'*n'}$$

Where

$$p_{ij}=IFWAp_\lambda\left(p_{ij}^{(1)},p_{ij}^{(2)},\ldots,p_{ij}^{(k)}\right)=\lambda_1p_{ij}^{(1)}+\lambda_2p_{ij}^{(2)}+\ldots+\lambda_kp_{ij}^{(k)}=$$
$$=\left[1-\prod_{i=1}^{k}\left(1-\mu_{ij}^{(l)}\right)^{\lambda_l},\prod_{i=1}^{k}\left(\nu_{ij}^{(l)}\right)^{\lambda_l},\prod_{i=1}^{k}\left(1-\mu_{ij}^{(l)}\right)^{\lambda_l}-\prod_{i=1}^{k}\left(\nu_{ij}^{(l)}\right)^{\lambda_l}\right] \quad (7)$$

Hereby, decision problem is described as:

$$P_{ij}=\begin{bmatrix}p_{11}&p_{12}&\cdots&p_{1n}\\p_{21}&p_{22}&\cdots&p_{2n}\\\vdots&\vdots&\ddots&\vdots\\p_{m1}&p_{m2}&\cdots&p_{mn}\end{bmatrix} \quad (8)$$

$P_{ij}=(\mu_{ij},\nu_{ij},\pi_{ij})$ ($i=1,2,\ldots,m; j=1,2,\ldots,n$) is an element of aggregated intuitionistic fuzzy decision matrix.

Step4. Calculation of the concordance and discordance sets where C_{zy} shows the degree of confidence in the pairwise comparison of the z and y alternatives ($X_z\to X_y$, $x, y=1,2,\ldots,m$; $z\ne y$). The concordance set $C_{zy}^{(1)}$ of Z_k and Y_l is composed of all criteria for which Z_k is preferred to Y_l. The concordance set is defined as:

$$C_{zy}^{(1)}=\left\{l\middle|\mu_{zl}\ge\mu_{ly},\ \nu_{zl}<\nu_{ly},\ \pi_{zl}<\pi_{ly}\right\} \quad (9)$$

The midrange concordance set is defined as:

$$C_{zy}^{(2)}=\left\{l\middle|\mu_{zl}\ge\mu_{ly},\ \nu_{zl}<\nu_{ly},\ \pi_{zl}\ge\pi_{ly}\right\} \quad (10)$$

The weak concordance set is defined as

$$C_{zy}^{(3)} = \left\{ l \middle| \mu_{zl} \geq \mu_{ly},\ \nu_{zl} \geq \nu_{ly} \right\} \tag{11}$$

The discordance set is composed of all criteria for which Z_k is not preferred to Y_l. The degree of disagreement in ($X_z \rightarrow X_y$) is constructed as follows:

$$D_{zy}^{(1)} = \left\{ l \middle| \mu_{zl} < \mu_{ly},\ \nu_{zl} \geq \nu_{ly},\ \pi_{zl} \geq \pi_{ly} \right\} \tag{12}$$

The midrange discordance set is defined as follows:

$$D_{zy}^{(2)} = \left\{ l \middle| \mu_{zl} < \mu_{ly},\ \nu_{zl} \geq \nu_{ly},\ \pi_{zl} < \pi_{ly} \right\} \tag{13}$$

The weak discordance set is defined as follows:

$$D_{zy}^{(3)} = \left\{ l \middle| \mu_{zl} < \mu_{ly},\ \nu_{zl} < \nu_{ly} \right\} \tag{14}$$

Step5. Determination of the concordance index C_{zy} and the discordance index D_{zy} for the proposed model using IFS is defined as follows:

$$C_{zy} = w_c^{(1)} * \sum_{i \in c_{zy}^1} w_l + w_c^{(2)} * \sum_{i \in c_{zy}^2} w_l + w_c^{(3)} * \sum_{i \in c_{zy}^3} w_l \tag{15}$$

The concordance index is equal to the sum of the weights of criteria that are contained in the concordance sets, where $w_c^{(1)}$, $w_c^{(2)}$, and $w_c^{(3)}$ are the weights of the concordance, midrange concordance, and weak concordance sets, respectively. The assessments of a Z_k are worse than assessments of a competing Y_l. Hereby, the discordance index is defined as follows:

$$D_{zy} = \frac{max_{j \in D_{ij}} W_D \times d(x_{il}, x_{lj})}{max_{j \in D_{ij}}\, d(x_{il}, x_{lj})} \tag{16}$$

Where W_D is equal to $w_D^{(1)}$, $w_D^{(2)}$ or $w_D^{(3)}$. These sets integrate the weight of discordance, midrange discordance, and weak discordance sets, respectively. The distance between X_{il} and X_{lj} is shown as:

$$d(x_{il}, x_{lj}) = \sqrt{\frac{1}{2n} \sum_{j=1}^{n} \left((\mu_{il} - \mu_{lj})^2 + (\nu_{il} - \nu_{lj})^2 + (\pi_{il} - \pi_{lj})^2 \right)} \tag{17}$$

Where $d(x_{il}, x_{lj})$ is Euclidian distance between X_{il} and X_{lj}.

Step6. In the concordance dominance matrix calculation process, the chosen alternative has the shortest distance from the positive ideal solution. Hence, the concordance dominance matrix $\boldsymbol{K}$ can be defined as:

$$K_{ij} = (C_{zy})^* - (C_{zy})_{ij} \tag{18}$$

where $(C_{zy})^*$ is the maximum value of $(C_{zy})_{ij}$, and $C_{zy} \geq C$, where $C = \dfrac{\sum_{z=1,\, z \neq y}^{m} \sum_{y=1,\, y \neq z}^{m} C_{zy}}{m(m-1)}$, which refers to the separation of each alternative from the positive ideal solution. A higher value of K_{ij} indicates that Z_k is less favorable than Y_l.

In the discordance dominance matrix calculation, the chosen alternative has the longest distance from the negative ideal solution. Hence, the discordance dominance matrix $\boldsymbol{L}$ is defined as follows:

$$L_{ij} = (D_{zy})^* - (D_{zy})_{ij} \tag{19}$$

where $(D_{zy})^*$ is the minimum value of $(D_{zy})_{ij}$, and $D_{zy} \geq D$, where $D = \dfrac{\sum_{z=1,\, z \neq y}^{m} \sum_{y=1,\, y \neq z}^{m} D_{zy}}{m(m-1)}$, which refers to the separation of each alternative from the negative ideal solution. A higher value of L_{ij} indicates that Z_k is preferred to Y_l.

Step7. To determine aggregate dominance matrix, the distance of each alternative to both positive and negative ideal alternatives should be calculated to determine the ranking. Hence, the aggregate dominance matrix $\boldsymbol{R}$ is defined as follows:

$$R_{ij} = \begin{bmatrix} - & r_{12} & \cdots & \cdots & r_{1m} \\ r_{21} & - & r_{23} & \cdots & r_{2m} \\ \cdots & \cdots & - & \cdots & \cdots \\ r_{(m-1)m} & \cdots & \cdots & - & r_{(m-1)m} \\ r_{m1} & r_{m2} & \cdots & r_{m(m-1)} & - \end{bmatrix} \tag{20}$$

Where $r_{kl} = \dfrac{l_{kl}}{k_{kl} + l_{kl}}$, and l_{kl} and k_{kl} are defined in (18) and (19), respectively. r_{kl} refers to the relative closeness to the ideal alternative, with a range from 0 to 1. Higher the value of r_{kl}, closer is the alternative X_k to the positive ideal and more distant form the negative ideal solution then the alternative X_l. Hence, it is better solution.

Step8. In final ranking of the alternatives, determination of matrix $\boldsymbol{T}$ is needed. T_{ij} is the final value of assessment, and is defined as follows:

$$T_{kl} = \frac{1}{m-1} \sum_{l=1,\, l \neq k}^{m} r_{kl} \quad k = 1, 2, ..., m \tag{21}$$

Alternatives are ranked according to T_{ij}. The best alternative X^*, which is the one with the shortest distance to the positive ideal point and the longest distance from the negative ideal point, and is defined as:

$$X^* = \max(T_{ij}) \tag{22}$$

Where X^* is the best atlernative.

3 RESULTS AND DISCUSSION

In this section, the proposed model is applied on the HS facilities. There are 17 HSs in the City of Split that need construction barriers removal for the physically disabled. These construction barriers are mostly related to the external access of school buildings and other facilities. Furthermore, the existence of a ramp and elevator outside and inside schools and school facilities is missing. Also, an appropriate front door width, and the existence of access to sanitary facilities that are adapted to people with disabilities are crucial to embed. The Department of Construction and Urban Planning is continuously taking measures to adapt facilities for disabled students, urgently fulfilling the needs of schools [50]. In Tab. 1, a list of HSs necessary for construction barriers removal are presented.

Table 1 List of HSs for construction barriers removal project

HS	High School name
HS1	II. Grammar School
HS2	IV. Grammar School
HS3	V. Grammar School
HS4	Science-technical sch.
HS5	Construction-geodetic sch.
HS6	Electrotechnic sch.
HS7	Industrial sch.
HS8	Trade sch.
HS9	Technic sch.
HS10	Touristic-hospitality sch.
HS11	Art sch.
HS12	Design, graphics and sustainable construction sch.
HS13	Trade-technic sch.
HS14	Commercial-trade sch.
HS15	Maritime sch.
HS16	Music sch. Josip Hatze
HS17	Technical-traffic sch.

Furthermore, to support final decision-makers in managing the projects of removal of constrain barriers in HSs, which is a problematic with a high uncertainty and duality, all HSs must be evaluate by certain number of criteria. Decision-makers, three in this case, define these criteria. A list of criteria is given in Tab. 2, and only the crucial one are presented to avoid extensive calculations and data presentation.

Table 2 Final list of criteria

Criterion	Criterion name
C1	Number of construction barriers
C2	External access to school's facilities
C3	Ramp or elevator inside school's facilities
C4	Adjusted door width
C5	Adjusted sanitary access in the building
C6	Cost of project documentation
C7	Amount of investment

After the relevant cirteria are defined, the comparison of all criteria is done using 1-9 number scale defined by Saaty [47]. Each decision-maker (DM) made a comparison. The 1-9 scale is given in Tab. 3, while calculated criteria weights by each DM and aggregated weight are presented in Tab. 4.

Table 3 The basic 1-9 scale [47]

Importance	Definition
1	Same significance
3	Average significance of one over another
5	Powerful significance of one over another
7	Very powerful significance of one over another
9	Extreme significance of one over another
2, 4, 6, 8	Intermediate values

Table 4 Calculated criteria weights

	DM1	DM2	DM3	W
1	0.20	0.25	0.20	0.22
C2	0.05	0.06	0.03	0.05
C3	0.15	0.06	0.03	0.08
C4	0.05	0.06	0.03	0.05
C5	0.05	0.06	0.03	0.05
C6	0.25	0.35	0.13	0.24
C7	0.25	0.15	0.53	0.31

The linguistic expressions of DM's importance with IF numbers are given in Tab. 5. Each DM is evaluated according to his/her background and years of experience. Using Eq. (6) the importance for each DM is calculated, and presented in Tab. 6 with linguistic and numerical value.

Table 5 Linguistic expressions and IFNs for DM's importance

Linguistic expression	IFNs
Very Important (VI)	(0.8, 0.1, 0.1)
Important (I)	(0.6, 0.3, 0.1)
Medium (M)	(0.5, 0.5, 0.0)
Bad (B)	(0.3, 0.6, 0.1)
Very Bad (VB)	(0.1, 0.8, 0.1)

Table 6 Calculated importance of DMs

	DM1	DM2	DM3
Linguistic expression	VI	I	M
λ	0.41	0.34	0.25

Furthermore, the linguistic expressions for the alternatives assessments are given in Tab. 7 with appurtenant IF numbers, and are further used to evaluate each HS by each criterion. The alternatives assessments are calculated by IFWA operator, defined in Eq. (7). The membership, non-membership and hesitation degree are determined as common values of all DMs' opinions that were included in the evaluation process. The aggregated IF decision matrix is presented in Tab. 8.

Table 7 Linguistic expression and IFNs for alternatives assessment

Linguistic expression	IFNs
Very Good (VG)	(0.9, 0.05, 0.05)
Good (G)	(0.7, 0.2, 0.1)
Medium (M)	(0.5, 0.5, 0.0)
Bad (B)	(0.2, 0.7, 0.1)
Very Bad (VB)	(0.05, 0.9, 0.05)

Table 8 IFWA: aggregated IF decision matrix

Alt.	C1	C2	C3	C4	C5	C6	C7
HS1	(0.20,0.70,0.10)	(0.70,0.20,0.10)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.70,0.20,0.10)
HS2	(0.50,0.50,0.00)	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.20,0.70,0.10)	(0.50,0.50,0.00)	(0.50,0.50,0.00)
HS3	(0.50,0.50,0.00)	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.20,0.70,0.10)	(0.56,0.40,0.04)	(0.56,0.40,0.04)
HS4	(0.05,0.90,0.05)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.90,0.05,0.05)	(0.90,0.05,0.05)
HS5	(0.50,0.50,0.00)	(0.70,0.20,0.10)	(0.20,0.70,0.10)	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.56,0.40,0.04)	(0.56,0.40,0.04)
HS6	(0.20,0.70,0.10)	(0.70,0.20,0.10)	(0.20,0.70,0.10)	(0.70,0.20,0.10)	(0.05,0.90,0.05)	(0.70,0.20,0.10)	(0.70,0.20,0.10)
HS7	(0.05,0.90,0.05)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.05,0.90,0.05)	(0.90,0.05,0.05)	(0.90,0.05,0.05)
HS8	(0.05,0.90,0.05)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.70,0.20,0.10)	(0.20,0.70,0.10)	(0.90,0.05,0.05)	(0.90,0.05,0.05)
HS9	(0.50,0.50,0.00)	(0.70,0.20,0.10)	(0.50,0.50,0.00)	(0.50,0.50,0.00)	(0.20,0.70,0.10)	(0.56,0.40,0.04)	(0.50,0.50,0.00)
HS10	(0.05,0.90,0.05)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.05,0.90,0.05)	(0.90,0.05,0.05)	(0.90,0.05,0.05)
HS11	(0.05,0.90,0.05)	(0.70,0.20,0.10)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.90,0.05,0.05)	(0.90,0.05,0.05)
HS12	(0.50,0.50,0.00)	(0.70,0.20,0.10)	(0.20,0.70,0.10)	(0.50,0.50,0.00)	(0.70,0.20,0.10)	(0.56,0.40,0.04)	(0.56,0.40,0.04)
HS13	(0.05,0.90,0.05)	(0.20,0.70,0.10)	(0.05,0.90,0.05)	(0.70,0.20,0.10)	(0.20,0.70,0.10)	(0.90,0.05,0.05)	(0.90,0.05,0.05)
HS14	(0.05,0.90,0.05)	(0.20,0.70,0.10)	(0.05,0.90,0.05)	(0.50,0.50,0.00)	(0.70,0.20,0.10)	(0.90,0.05,0.05)	(0.90,0.05,0.05)
HS15	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.05,0.90,0.05)	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.70,0.20,0.10)
HS16	(0.05,0.90,0.05)	(0.20,0.70,0.10)	(0.05,0.90,0.05)	(0.50,0.50,0.00)	(0.05,0.90,0.05)	(0.90,0.05,0.05)	(0.90,0.05,0.05)
HS17	(0.05,0.90,0.05)	(0.20,0.70,0.10)	(0.20,0.70,0.10)	(0.70,0.20,0.10)	(0.20,0.70,0.10)	(0.90,0.05,0.05)	(0.90,0.05,0.05)

Applying step 4, and using aggregated IF values of HS evaluations, the concordance sets are defined. The concordance set, applying (9), is defined as follows:

$$c_{zy}^{1} = \begin{Bmatrix} - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - \\ 1 & - & - & 1 & - & 1 & 1 & 1 & - & 1 & 1 & - & 1 & 1 & 1 & 1 & 1 \\ 1 & - & - & 1 & - & 1 & 1 & 1 & - & 1 & 1 & - & 1 & 1 & 1 & 1 & 1 \\ 6,7 & - & - & - & - & 6,7 & - & - & - & - & - & - & - & - & - & - & - \\ 1 & - & - & 1 & - & 1 & 1 & 1 & - & 1 & 1 & - & 1 & 1 & 1 & 1 & 1 \\ - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - \\ - & - & - & - & - & 6,7 & - & - & - & - & - & - & - & - & 6,7 & - & - \\ 6,7 & - & - & - & - & 6,7 & - & - & - & - & - & - & - & - & 6,7 & - & - \\ 1,3,4 & - & - & 1,3,4 & - & 1,3 & 1,3,4 & 1,3 & - & 1,3,4 & 1,3,4 & 3 & 1,3 & 1,3,4 & 1,3 & 1,3,4 & 1,3 \\ - & - & - & - & - & 6,7 & - & - & - & - & - & - & - & - & 6,7 & - & - \\ 6,7 & - & - & - & - & 6,7 & - & - & - & - & - & - & - & - & 6,7 & - & - \\ 1,4 & - & - & 1,4 & - & 1 & 1,4 & 1 & - & 1,4 & 1,4 & - & 1 & 1,4 & 1 & 1,4 & 1 \\ 6,7 & - & - & - & - & 6,7 & - & - & - & - & - & - & - & - & 6,7 & - & - \\ 6,7 & - & - & - & - & 6,7 & - & - & - & - & - & - & - & - & 6,7 & - & - \\ - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - \\ 6,7 & - & - & - & - & 6,7 & - & - & - & - & - & - & - & - & 6,7 & - & - \\ 6,7 & - & - & - & - & 6,7 & - & - & - & - & - & - & - & - & 6,7 & - & - \end{Bmatrix}$$

Then, the midrange concordance set, applying (10), is:

$$c_{zy}^{2} = \begin{Bmatrix} - & 5,6,7 & 5,6,7 & 1,2,5 & 6,7 & 5 & 1,2,5 & 1,2,5 & 5,6,7 & 1,2,5 & 1,5 & 6,7 & 1,2,3,5 & 1,2,3,4 & 1,2,3 & 1,2,3,4,5 & 1,2,5 \\ 3,4 & - & 5 & 5 & 2,3,4,5 & 3 & 3 & 2,3,4,5 & 2,3,5 & 3,4,5 & 2,3,4,5 & 3,4,5 & 3,4 & 2,3,5 & 2,3,4 & 2,3 & 2,3,4,5 \\ 3,4 & 5,6,7 & - & 5 & 2,3,4,5 & 3 & 3,5 & 2,3,4,5 & 2,3,5 & 3,4,5,7 & 2,3,4,5 & 3,4,5 & 3,4, & 2,3,5 & 2,3,4 & 2,3 & 2,3,4,5 \\ - & 6,7 & 6,7 & - & 2 & 6,7 & 6,7 & 5 & 2,5 & 6,7 & 2,5 & - & 6,7 & 2,3 & 2,3,4 & 2,3 & 2,3,4,5 \\ 4 & 5,6,7 & 5 & 2,4,5 & - & 5 & 2,4,5 & 2,5 & 4,5,7 & 2,4,5 & 4,5 & 4 & 2,3,5 & 2,3,4 & 2,3, & 2,3,4,5 & 2,5 \\ 4 & 6,7 & 6,7 & 1,2,4 & 6,7 & - & 1,2,4 & 1,2 & 4,6,7 & 1,2,4 & 1,4 & 4,6,7 & 1,2,3 & 1,2,3,4 & 2,3 & 1,2,3,4 & 1,2 \\ - & 6,7 & 6,7 & - & 6,7 & - & - & - & 6,7 & - & - & 6,7 & 3 & 3,4 & 3 & 3,4 & - \\ 4 & 6,7 & 6,7 & 4 & 6,7 & 5 & 4,5 & - & 4,6,7 & 4,5 & 4 & 4,6,7 & 3 & 3,4 & 3 & 3,4 & - \\ - & 6 & - & 2 & - & 5 & 2,5 & 2 & - & 2,5 & - & - & 2 & 2 & 2 & 2,5 & 2 \\ - & 6,7 & 6,7 & - & 6,7 & - & - & - & 6,7 & - & - & 6,7 & 3 & 3,4 & 3 & 3,4 & - \\ - & 6,7 & 6,7 & 2 & 6,7 & 5 & 2,5 & 2 & 6,7 & 2,5 & - & 6,7 & 2,3 & 2,3,4 & 2,3 & 2,3,4,5 & 2 \\ - & 5,6,7 & 5 & 2,5 & - & 5 & 2,5 & 2,5 & 5,7 & 2,5 & 5 & - & 2,3,5 & 2,3 & 2,3 & 2,3,5 & 2,5 \\ 4 & 6,7 & 6,7 & 4 & 6,7 & 5 & 4,5 & - & 4,6,7 & 4,5 & 4 & 4,6,7 & - & 4 & - & 4,5 & - \\ - & 5,6,7 & 5,6,7 & 5 & 6,7 & 5 & 5 & 5 & 5,6,7 & 5 & 5 & 6,7 & 5 & - & - & 5 & 5 \\ 4 & 5,6,7 & 5,6,7 & 1,4,5 & 6,7 & 5 & 1,4,5 & 1,5 & 4,5,6,7 & 1,4,5 & 1,4,5 & 4,6,7 & 1,5 & 1,4 & - & 1,4,5 & 1,5 \\ - & 6,7 & 6,7 & - & 6,7 & - & - & - & 6,7 & - & - & 6,7 & - & - & - & - & - \\ 4 & 6,7 & 6,7 & 4 & 6,7 & 5 & 4,5 & - & 4,6,7 & 4,5 & 4 & 4,6,7 & 3 & 3,4 & 3 & 3,4,5 & - \end{Bmatrix}$$

And the weak concordance set, applying (11), is defined as:

$$c_{zy}^{3}=\left\{\begin{array}{cccccccccccccccccc}
- & 2 & 2 & 3,4 & 2,3,5 & 1,2,3,6,7 & 3,4 & 3 & 2 & 3,4 & 2,3,4 & 2,3,5 & - & 5 & 1,5,6 & - & 3 \\
2 & - & 1,2,3,4 & - & 1,2,4 & 2,4 & - & 4 & 1,2,7 & - & 2 & 1,2 & 4 & - & 4 & - & 4 \\
2 & 1,2,3,4 & - & - & 1,2,4,6,7 & 2,4 & - & 4 & - & 2 & 1,2,6,7 & - & 4 & - & 4 & - & 4 \\
3,4 & 5 & 5 & - & 3 & 3 & 1,3,4,6,7 & 1,3,5,6,7 & 5 & 1,3,4,6,7 & 1,3,4,5,6,7 & 3 & 1,5,6,7 & 1,6,7 & - & 1,6,7 & 1,5,6,7 \\
2,3 & 1,2,4 & 1,2,4,6,7 & 3 & - & 2,3,4 & 3 & 3,4 & 1,2,6 & 3 & 2,3 & 1,2,3,5,6,7 & 4 & 5 & 4,5 & - & 4 \\
1,2,3 & 2,4 & 2,4 & 3 & 2,3,4 & - & 3,5 & 3,4 & 2 & 3,5 & 2,3 & 2,3 & 4 & - & 1,4,6,7 & 5 & 3,4 \\
3,4 & - & - & 1,2,3,4,6,7 & 3 & 3,5 & - & 1,2,3,6,7 & - & 1,2,3,4,5,6,7 & 1,3,4,6,7 & 3 & 1,2,6,7 & 1,2,6,7 & 2 & 1,2,5,6,7 & 1,2,3,6,7 \\
3 & 4,5 & 4,5 & 1,2,3,5,6,7 & 3,4 & 3,4 & 1,2,3,6,7 & - & 5 & 1,2,3,6,7 & 3,5,6,7 & 3 & 1,2,4,5,6,7 & 1,2,6,7 & 2,4 & 1,2,6,7 & 1,2,3,4,5,6,7 \\
2 & 1,2,5,7 & 1,2,5,6 & 5 & 1,2,6 & 2 & - & - & - & - & 2,5 & 1,2,4,6 & 5 & - & - & - & 5 \\
3,4 & - & - & 1,2,3,4,6,7 & 3 & 3,5 & 1,2,3,4,5,6,7 & 1,2,3,6,7 & - & - & 1,3,4,6,7 & 3 & 1,2,6,7 & 1,2,6,7 & 2 & 1,2,5,6,7 & 1,2,3,6,7 \\
2,3,4 & 2,5 & 2,5 & 1,3,4,5,6,7 & 2,3 & 2,3 & 1,3,4,6,7 & 1,3,5,6,7 & 2,5 & 1,3,4,6,7 & - & 2,3 & 1,5,6,7 & 1,6,7 & - & 1,6,7 & 1,5,6,7 \\
2,3,5 & 1,2 & 1,2,6,7 & 3 & 1,2,3,5,6,7 & 2,3 & 3 & 3 & 1,2,4,6 & 3 & 2,3 & - & - & 5 & 5 & - & - \\
- & 4,5 & 4,5 & 1,2,5,6,7 & 4 & 4 & 1,2,6,7 & 1,2,4,5,6,7 & 5 & 1,2,6,7 & 1,5,6,7 & - & - & 1,2,3,6,7 & 2,3,4 & 1,2,3,6,7 & 1,2,4,5,6,7 \\
5 & - & - & 1,2,6,7 & 5 & - & 1,2,6,7 & 1,2,6,7 & - & 1,2,6,7 & 1,6,7 & 5 & 1,2,3,6,7 & - & 2,3,5 & 1,2,3,4,6,7 & 1,2,6,7 \\
1,5,6,7 & 4 & 4 & 2 & 4,5 & 1,4,6,7 & 2 & 2,4 & - & 2 & - & 5 & 2,3,4 & 2,3,5 & - & 2,3 & 2,4 \\
- & - & - & 1,2,6,7 & - & 5 & 1,2,5,6,7 & 1,2,6,7 & - & 1,2,5,6,7 & 1,6,7 & - & 1,2,3,6,7 & 1,2,3,4,6,7 & 2,3 & - & 1,2,6,7 \\
3 & 4,5 & 4,5 & 1,2,3,5,6,7 & 3,4 & 3,4 & 1,2,3,6,7 & 1,2,3,4,5,6,7 & 5 & 1,2,3,6,7 & 1,3,5,6,7 & 3 & 1,2,4,5,6,7 & 1,2,6,7 & 2,4 & 1,2,6,7 & — \\
\end{array}\right\}$$

The discordance set is determined using (12), and given as follows:

$$D_{zy}^{1}=\left\{\begin{array}{cccccccccccccccccc}
- & 1,3,4 & 1,3,4 & - & 4 & 4 & 6,7 & 4,6,7 & 1,3,4 & 6,7 & 6,7 & 1,4 & 4,6,7 & 6,7 & 4 & 6,7 & 4,6,7 \\
5 & - & - & - & 5 & 4 & - & - & - & - & - & 5 & - & 5 & 5 & - & - \\
5 & - & - & - & 5 & - & - & - & - & - & - & 5 & - & 5 & 5 & - & - \\
2,5 & 1,2,3,4 & 1,2,3,4 & - & 1,2,4,5 & 1,2,4 & - & 4 & 1,2,3,4 & - & 2 & 1,2,4,5 & 4 & 5 & 5 & - & 4 \\
- & 3 & 3 & - & - & - & - & - & 3 & - & - & - & - & - & - & - & - \\
- & 1,3 & 1,3 & 6,7 & 1 & - & 6,7 & 6,7 & 6,7 & 6,7 & 6,7 & - & 6,7 & 6,7 & - & 6,7 & 6,7 \\
2 & 1,2,3,4 & 2,3,4 & 2,3 & 1,2,4 & 2,4 & - & 4 & 2,3,4 & - & 2 & 4 & - & 4,5 & 4 & 4 & 4 \\
2 & 1,2,3 & 1,2,3 & - & 1,2,3 & 2 & - & - & 1,2,3 & 4 & 2 & 2,5 & - & 5 & 4,5 & - & - \\
5 & - & - & - & 5 & - & - & - & - & - & - & - & - & 5 & - & - & 5 \\
2 & 2,3 & 3,4 & - & 2,4 & 2,4 & - & 4 & 1,2,3,4 & - & 2 & 1,2,4 & 4 & - & 4 & 4,5 & 4 \\
- & 1,3,4 & 1,3,4 & - & 1 & 4 & - & - & 1,4 & - & - & 1,4,5 & 4,5 & 2,3 & 2,4,5 & 2,4 & 2,5 \\
- & 3 & 3 & - & - & - & - & - & 3 & - & - & - & - & - & - & - & - \\
2 & - & 2 & - & 2 & 2 & - & - & 1,2 & - & 2 & 2,5 & - & 5 & 5 & - & - \\
2 & 1 & - & - & 1,2 & 2 & - & - & 1 & 2,3 & 2 & 1,2 & - & - & - & - & - \\
- & 1,2 & 1,2 & 6,7 & 1,2 & - & 6,7 & - & 1,2,3 & 6,7 & 2,6,7 & 6,7 & 6,7 & 6,7 & - & 6,7 & 6,7 \\
2 & 1,2 & 1,2 & - & 1,2 & 2 & - & - & 2 & - & 2 & 1,2 & - & - & - & - & - \\
2,5 & 1,2,3 & 1,2,3 & - & 1,2,5 & 2 & - & - & 1,2,3 & - & 2 & - & - & 5 & 2,5 & - & — \\
\end{array}\right\}$$

The midrange discordance set, applying (13), is:

$$D_{zy}^{2}=\left\{\begin{array}{cccccccccccccccccc}
- & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - \\
7 & - & 7 & 7 & 7 & 7 & 7 & 7 & - & 7 & 7 & 7 & 7 & 7 & 7 & 7 & 7 \\
6,7 & - & - & 6,7 & - & 6,7 & 6,7 & 6,7 & - & 6,7 & 6,7 & - & 6,7 & 6,7 & 6,7 & 6,7 & 6,7 \\
- & - & - & - & - & 1 & - & - & - & - & - & - & - & - & 1 & - & - \\
6,7 & - & - & 6,7 & - & 6,7 & 6,7 & 6,7 & - & 6,7 & 6,7 & - & 6,7 & 6,7 & 6,7 & 6,7 & 6,7 \\
5 & 5 & 5 & 5 & 5 & - & - & 5 & 5 & - & 5 & 5 & 5 & 5 & 5 & - & 5 \\
1 & - & - & - & - & 1 & - & - & - & - & - & - & - & - & 1 & - & - \\
1 & - & - & - & - & 1 & - & - & - & - & - & - & - & - & 1 & - & - \\
6,7 & 3,4 & 3,4,7 & 6,7 & 4,7 & 4,6,7 & 6,7 & 4,6,7 & - & 6,7 & 6,7 & - & 4,6,7 & 6,7 & 4,6,7 & 6,7 & 4,6,7 \\
1 & - & - & - & - & 1 & - & - & - & - & - & - & - & - & 1 & - & - \\
1 & - & - & - & - & 1 & - & - & - & - & - & - & - & - & 1 & - & - \\
6,7 & 4 & 4 & 6,7 & 4 & 4,6,7 & 6,7 & 4,6,7 & - & 6,7 & 6,7 & - & 4,6,7 & 6,7 & 4,6,7 & 6,7 & 4,6,7 \\
1,3 & 3 & 3 & 3 & 3 & 1,3 & 3 & 3 & - & 3 & 3 & 3 & - & - & 1 & - & 3 \\
1,3,4 & 3,4 & 3,4 & 3,4 & 3,4 & 1,3,4 & 3,4 & 3,4 & - & 3,4 & 3,4 & 3 & 4 & - & 1,4 & - & 3,4 \\
3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & - & 3 & 3 & 3 & - & - & - & - & 3 \\
1,3,4 & 3,4 & 3,4 & 3,4 & 3,4 & 1,3,4 & 3,4 & 3,4 & - & 3,4 & 3,4 & 3 & 4 & - & 1,4 & - & 3,4 \\
1 & - & - & - & - & 1 & - & - & - & - & - & - & - & - & 1 & - & — \\
\end{array}\right\}$$

The weak discordance set is defined by (14), and is presented as: $D_{zy}^{3}=\{-\}$.

The concordance and discordance dominance matrices ***K*** and ***L*** are calculated applying step 6, then the aggregate dominance matrix ***R*** defined by (20) is calculated and presented as follows:

For the final ranking calculation (21) is used, to determine matrix ***T***. According to ***T***, alternatives are ranked where the best alternative is the one that has the shortest distance from the positive ideal point and the longest distance from the negative ideal point.

The values of *T* are determined as follows: $T_1 = 0.86$; $T_2 = 0.69$; $T_3 = 0.85$; $T_4 = 0.54$; $T_5 = 0.85$; $T_6 = 0.85$; $T_7 = 0.82$; $T_8 = 0.59$; $T_9 = 0.93$; $T_{10} = 0.55$; $T_{11} = 0.60$; $T_{12} = 0.86$; $T_{13} = 0.51$; $T_{14} = 0.63$; $T_{15} = 0.91$; $T_{16} = 0.67$; $T_{17} = 0.65$.

Hence, the final ranking of HS according to the necessary for the construction barriers removal is achieved as: HS9 > HS15 > HS12 >HS1 > HS3 > HS5 >HS6 > HS2 > HS16 > HS17 > HS14 > HS8 > HS11 > HS7 > HS10 > HS4 >HS13. Accoridng to final ranking Technic school ihas the highest priory for the caontruction barriers remova, and Trade-technic school the last priory. HS9 has the most constrction barriers , all four of them while HS13 has only one. This way, the porposed model is shown to be applicable and usefull in dealing with this type of civile engineering

problematics.

$$R_{zy} = \begin{Bmatrix} - & 1 & 1 & 0.48 & 0.35 & 0.50 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0.36 & 1 & 1 \\ 1 & - & 0.54 & 0.54 & 1 & 1 & 0.54 & 0.41 & 0.48 & 0.54 & 0.41 & 1 & 0.54 & 1 & 1 & 0.54 & 0.41 \\ 1 & 0.29 & - & 1 & 0.34 & 1 & 1 & 1 & 0.48 & 1 & 1 & 0.50 & 1 & 1 & 1 & 1 & 1 \\ 0.28 & 1 & 1 & - & 1 & 0 & 0.22 & 0.50 & 1 & 0.22 & 0.37 & 1 & 0.24 & 0.50 & 0.54 & 0.48 & 0.35 \\ 1 & 0.31 & 0.50 & 1 & - & 1 & 1 & 1 & 0.29 & 1 & 1 & 0.48 & 1 & 1 & 1 & 1 & 1 \\ 0.49 & 0.39 & 0.39 & 1 & 1 & - & 1 & 1 & 1 & 1 & 1 & 0.34 & 1 & 1 & 0.49 & 1 & 1 \\ 0.54 & 1 & 1 & 0.52 & 1 & 1 & - & 0.50 & 0.41 & 0.48 & 0.50 & 0.37 & 0.48 & 0.51 & 0.29 & 0.37 & 0.50 \\ 0.29 & 1 & 1 & 0.48 & 1 & 0.28 & 0.36 & - & 1 & 0.38 & 0.50 & 0.37 & 0.48 & 0.50 & 1 & 0.48 & 0.48 \\ 1 & 0.36 & 1 & 1 & 1 & 1 & 1 & 1 & - & 1 & 1 & 0.48 & 1 & 1 & 1 & 1 & 1 \\ 0.54 & - & - & - & - & 1 & - & - & - & - & - & - & - & - & 1 & - & - \\ 0.28 & 1 & 1 & 0.48 & 0.41 & 0.28 & 0.48 & 0.48 & 1 & 0.36 & - & 1 & 0.51 & 0.39 & 1 & 0.37 & 0.51 \\ 1 & 0.34 & 0.52 & 1 & 0.49 & 1 & 1 & 1 & 0.36 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.32 & 0.36 & 0.38 & 0.50 & 0.38 & 1 & 0.50 & 0.50 & 1 & 0.50 & 0.51 & 0.39 & - & 0.50 & 0.29 & 0.48 & 0.50 \\ 1 & 1 & 0.37 & 0.51 & 1 & 1 & 0.51 & 0.51 & 0.41 & 0.55 & 0.52 & 1 & 0.49 & - & 0.29 & 0.48 & 0.51 \\ 0.50 & 1 & 1 & 1 & 1 & 0.50 & 1 & 0.50 & 1 & 1 & 1 & 1 & 1 & 1 & - & 1 & 1 \\ 1 & 1 & 1 & 0.51 & 1 & 1 & 0.51 & 0.51 & 0.38 & 0.51 & 0.52 & 1 & 0.49 & 0.48 & 0.30 & - & 0.51 \\ 1 & 1 & 1 & 0.48 & 1 & 0.29 & 0.48 & 0.48 & 1 & 0.48 & 0.50 & 0.34 & 0.48 & 0.50 & 1 & 0.35 & — \end{Bmatrix}$$

4 CONCLUSION

The IF ELECTRE method is provided for solving multicriterial problem with IFS information. The IFS data are used instead of single values in the evaluation process of the ELECTRE method. With these data, different sets of concordance and discordance are classified to fit a real life decision. IF ELECTRE algorithm is proposed to support final decision makers in managing the project of removal of construction barriers in HSs. There were 17 HSs defined for evaluation by seven criteria. Three DMs are included in the process according to their background and the years of experience to lower a partiality in the decision-making process. Hereby, proposed model uses determined concordance and discordance sets to construct concordance and discordance matrices. Defining these matrices, an aggregated matrix is calculated and the final ranking of HSs is achieved. Technic school is rank with the highest priority for the removal of construction barriers, and the least necessary is the Trade-technic school. These school are ranked according to the number and complexity of barriers that need to be remove. Only the crucial criteria are included in the assessment process, to lower the comprehensives of the calculations and results. Since the problem is multicirterial with large amount of uncertain data, therefore the proposed IF ELECTRE is an effective approach because fuzzy theory can precisely resolve the natural duality associated with the DM's definition of uncertain data. Furthermore, the approach enables DMs to select the best alternative by determine the shortest distance form positive ideal solution and negative ideal solution. The IF ELECTRE hereby, is used to define the HS that is the most necessary for the construction barriers removal and then by the final ranking of all HSs, defined the plan for the removal projects. For distance measure calculation of discordance index, the Euclidian distance is used. This approach gives more systematic description of the decision process, and it removes the ambiguity and vagueness in collected data. It has the ability to solve other complex problems with the high degree of uncertainty and hesitation. For the future study, more criteria described by interval valued IF information will be defined to achieve a detailed evaluation of alternatives, integrating users' opinions.

Acknowledgment

This research is partially supported through project KK.01.1.1.02.0027, a project co-financed by the Croatian Government and the European Union through the European Regional Development Fund - the Competitiveness and Cohesion Operational Programme.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

5 REFERENCES

[1] Gray, D. B., Gould, M., & Bickenbach, J. E. (2003). Environmental barriers and disabilities. *Journal of Architectural and Planning Research, 20*(1), 30-37.

[2] Kassab, M., Hegazy, T., & Hipel, K. (2010). Computerized DSS for construction conflict resolution under uncertainty. *J. Constr. Eng. Manage. 136*(12), 1249-1257. https://doi.org/0.1061/(ASCE)CO.1943-7862.0000239

[3] Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

[4] Atanassov, K. T. (1986). Intuitionistic fuzzy sets. *Fuzzy sets and Systems, 20*(1), 87-96. https://doi.org/10.1016/S0165-0114(86)80034-3

[5] Atanassov, K. T. (1999). *Intuitionistic fuzzy sets: Theory and applications*. New York: Physica-Verlag. https://doi.org/10.1007/978-3-7908-1870-3

[6] Atanassov, K., Pasi, G., & Yager, R. (2005). Intuitionistic fuzzy interpretations of multi-criteria multi-person and multi-measurement tool decision making. *International Journal of Systems Science, 36*(14), 859-868. https://doi.org/10.1080/00207720500382365

[7] Hong, D. H. & Choi, C. H. (2000). Multicriteria fuzzy decision-making problems based on vague set theory. *Fuzzy Sets and Systems, 114*(1), 103-113. https://doi.org/10.1016/S0165-0114(98)00271-1

[8] Hung, W. L. & Yang, M. S. (2004). Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance. *Pattern Recognition Letters, 25*, 1603-1611. https://doi.org/10.1016/j.patrec.2004.06.006

[9] Liu, H. W. & Wang, G. J. (2007). Multi-criteria decision-making methods based on intuitionistic fuzzy sets. *European Journal of Operational Research, 179*(1), 220-233. https://doi.org/10.1016/j.ejor.2006.04.009

[10] Szmidt, E. & Kacprzyk, J. (2002). Using intuitionistic fuzzy sets in group decision making. *Control and Cybernetics, 31*, 1037-1053.

[11] Szmidt, E. & Kacprzyk, J. (2003). A consensus-reaching process under intuitionistic fuzzy preference relations. *International Journal of Intelligent Systems, 18*, 837-852. https://doi.org/10.1002/int.10119

[12] Wang, P. (2009). QoS-aware web services selection with intuitionistic fuzzy set under consumer's vague perception. *Expert Systems with Applications, 36*(3), 4460-4466. https://doi.org/10.1016/j.eswa.2008.05.007

[13] Xu, Z. S. (2007a). A survey of preference relations. *International Journal of General Systems, 36*(2), 179-203. https://doi.org/10.1080/03081070600913726

[14] Xu, Z. S. (2007b). Some similarity measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making. *Fuzzy Optimization and Decision Making, 6*(2), 109-121. https://doi.org/10.1007/s10700-007-9004-z

[15] Xu, Z. S. (2007c). Intuitionistic fuzzy aggregation operators. *IEEE Transactions on Fuzzy Systems, 15*(6), 1179-1187. https://doi.org/10.1109/TFUZZ.2006.890678

[16] Xu, Z. & Yager, R. R. (2009). Intuitionistic and interval-valued intuitionistic fuzzy preference relations and their measures of similarity for the evaluation of agreement within a group. *Fuzzy Optimization and Decision Making, 8*(4), 123-139. https://doi.org/10.1007/s10700-009-9056-3

[17] Li, D. F. & Cheng, C. T. (2002). New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions. *Pattern Recognition Letters, 23*, 221-225. https://doi.org/10.1016/S0167-8655(01)00110-6

[18] Liang, Z. & Shi, P. (2003). Similarity measures on intuitionistic fuzzy sets. *Pattern Recognition Letters, 24*, 2687-2693. https://doi.org/10.1016/S0167-8655(03)00111-9

[19] Vlachos, I. K. & Sergiadis, G. D. (2007). Intuitionistic fuzzy information? Applications to pattern recognition. *Pattern Recognition Letters, 28*, 197-206. https://doi.org/10.1016/j.patrec.2006.07.004

[20] Wang, W. & Xin, X. (2005). Distance measure between intuitionistic fuzzy sets. *Pattern Recognition Letters, 26*, 2063-2069. https://doi.org/10.1016/j.patrec.2005.03.018

[21] Zhang, C. & Fu, H. (2006). Similarity measures on three kinds of fuzzy sets. *Pattern Recognition Letters, 27*, 1307-1317. https://doi.org/10.1016/j.patrec.2005.11.020

[22] Boran, F. E. (2011). An integrated intuitionistic fuzzy multi criteria decision making method for facility location selection. *Mathematical and Computational Applications, 16*, 487-496. https://doi.org/10.3390/mca16020487

[23] Rouyendegh, B. D. (2018). The Intuitionistic Fuzzy ELECTRE model. *International Journal of Management Science and Engineering Management, 13*(2), 139-145. https://doi.org/10.1080/17509653.2017.1349625

[24] Boran, F. E., Genc, S., Kurt, M., & Akay, D. (2009). A multi-criteria intuitionistic fuzzy group decision making for selection of supplier with TOPSIS method. *Expert Systems with Applications, 36*(8), 11363-11368. https://doi.org/10.1016/j.eswa.2009.03.039

[25] Boran, F. E., Boran, K., & Menlik, T. (2012). The evaluation of renewable energy technologies for electricity generation in Turkey using intuitionistic fuzzy TOPSIS. *Energy Sources Part B, 7*, 81-90. https://doi.org/10.1080/15567240903047483

[26] Roy, B. (1968). Classement et a choix en presence de points de vue multiples (la methode Electre) [Ranking and choice in the presence of multiple points of view]. *RIRO, 8*, 57-75. https://doi.org/10.1051/ro/196802V100571

[27] Benayoun, R. & Billsberry, J. (2007). *Experiencing recruitment and selection*. Hoboken, NJ, Wiley &Sons.

[28] Hashemi, S. S., Hajiagha, S. H. R., Zavadskas, E. K., & Mahdiraji, H. A. (2016). Multicriteria group decision making with ELECTRE III method based on interval-valued intuitionistic fuzzy information. *Applied Mathematical Modelling, 40*, 1554-1564. https://doi.org/10.1016/j.apm.2015.08.011

[29] Roy, B. (1978). ELECTRE III: Un algorithme de classements fonde sur une representation floue des preferenceen presence de criteres multiples [An algorithm of rankings based on a fuzzy representation of the preference in the presence of multiple criteria]. *Cahiers de CERO, 20*, 3-24.

[30] Beccali, M., Cellura, M., & Ardente, D. (1998). Decision making in energy planning: The ELECTRE multicriteria analysis approach compared to a fuzzy-sets methodology. *Energy Conversion and Management, 39*(16-18), 1869-1881. https://doi.org/10.1016/S0196-8904(98)00053-3

[31] Chen, T. Y. (2016). An IVIF-ELECTRE outranking method for multiple criteria decision-making with interval-valued intuitionistic fuzzy sets. *Technological and Economic Development of Economy, 22*, 416-452. https://doi.org/10.3846/20294913.2015.1072751

[32] Wu, M. C. & Chen, T. Y. (2011). The ELECTRE multicriteria analysis approach based on Atanassov's intuitionistic fuzzy sets. *Expert Systems with Applications, 38*, 12318-12327. https://doi.org/10.1016/j.eswa.2011.04.010

[33] Rouyendegh, B. D. & Erol, S. (2012). Selecting the best project using the Fuzzy ELECTRE method. *Mathematical Problem in Engineering, 2012*, 1-12. https://doi.org/10.1155/2012/790142

[34] Woolley, H. (2013). Now being social: The barrier of designing outdoor play spaces for disabled children. *Children Soc., 27*, 448-458. https://doi.org/10.1111/j.1099-0860.2012.00464.x

[35] Agarwal, N., Moya, E. M., Yasui, N. Y., & Seymour, C. (2015). Participatory action research with college students with disabilities: Photovoice for an inclusive campus. *J. Postsecondary Educ. Disability, 28*(2), 243-250.

[36] Hammel, J. et al. (2015). Environmental barriers and supports to everyday participation: A qualitative insider perspective from people with disabilities. *Arch. Phys. Med. Rehabil., 96*(4), 578-588. https://doi.org/10.1016/j.apmr.2014.12.008

[37] Church, L. R. & Marston, J. R. (2003). Measuring accessibility for people with a disability. *Geog. Anal., 35*(1), 83-96. https://doi.org/10.1111/j.1538-4632.2003.tb01102.x

[38] Martin, J. J. (2013). Benefits and barriers to physical activity for individuals with disabilities: A social-relational model of disability perspective. *Disability Rehabil., 35*(24), 2030-2037. https://doi.org/10.3109/09638288.2013.802377

[39] Hannon, B. (2014). Exploring the accessibility of leisure facilities in Munster for people with physical disabilities. *CARL research project*, Univ. College Cork, Cork, Ireland.

[40] Kroll, T., Jones, G. C., Keh, M., & Neri, M. T. (2006). Barriers and strategies affecting the utilisation of primary preventive services for people with physical disabilities: A qualitative inquiry. *Health Soc. Care Commun., 14*(4), 284-293. https://doi.org/10.1111/j.1365-2524.2006.00613.x

[41] Kayes, N. M., McPherson, K. M., Taylor, D., Schlüter, P. J., & Kolt, G. S. (2011). Facilitators and barriers to engagement in physical activity for people with multiple sclerosis: A qualitative investigation. *Disability Rehabil., 33*(8), 625-642. https://doi.org/10.3109/09638288.2010.505992

[42] Yuker, H. E., Revenson, J., & Fracchia, J. F. (1968). *The modification of educational equipment and curriculum for maximum utilization by physically disabled persons*, Human Resources Center, Albertson, NY.

[43] Leigh Hill, J. (1992). Accessibility: Students with disabilities in universities in Canada. *Can. J. Higher Educ., 22*(1), 48-83.

[44] Klinger, L. E. (2014). Scoping review—Physical accessibility and postsecondary education. *Occupational Therapy Publications*, Paper 7. hhttp://ir.lib.uwo.ca/otpub/7i (Nov. 17, 2016).

[45] Burton, G., Sayrafi, I., & Srour, S. A. (2013). Inclusion or transformation? An early assessment of an empowerment project for disabled people in occupied *Palestine. Disability Soc., 28*(6), 812-825. https://doi.org/10.1080/09687599.2013.802223

[46] Ginis, K. A. M., Ma, J. K., Latimer-Cheung, A. E., & Rimmer, J. H. (2016). A systematic review of review articles addressing factors related to physical activity participation among children and adults with physical disabilities. *Health Psychol. Rev., 10*(4), 478-494. https://doi.org/10.1080/17437199.2016.1198240

[47] Saaty, T. L. (1996). *Decision making with dependence and feedback, the analytic network process*. Pittsburgh, PA: RWS Publications.

[48] Boran, F. E., Genc, S., Kurt, M., & Akay, D. (2009). A multi-criteria intuitionistic fuzzy group decision making for supplier selection with TOPSIS method. *Expert Systems with Applications, 36*, 11363-11368. https://doi.org/10.1016/j.eswa.2009.03.039

[49] Szmidt, E., & Kacprzyk, J. (2000). Distances between intuitionistic fuzzy sets. *Fuzzy Sets and Systems, 114*(3), 505-518. https://doi.org/10.1016/S0165-0114(98)00244-9

[50] Rogulj, K. & Jajac, N. (2018). Achieving a construction barriers free-environment: Decision Support to Policy Selection. *Journal of Management in Engineering (ASCE)*. 04018020-1-18. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000618

Authors' contacts:

Katarina Rogulj, Assist. Prof.
University of Split,
Faculty of Civil Engineering, Architecture and Geodesy,
Matice hrvatske 15, 21000 Split, Croatia
e-mail: katarina.rogulj@gradst.hr

Jelena Kilić Pamuković, Assist. Prof.
University of Split,
Faculty of Civil Engineering, Architecture and Geodesy,
Matice hrvatske 15, 21000 Split, Croatia
e-mail: jelena.kilicj@gradst.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
httpa://doi.org/10.31803/tg-20210204161210

Original scientific paper

Tracking Keypoints from Consecutive Video Frames Using CNN Features for Space Applications

Janhavi Borse*, Dipti Patil, Vinod Kumar

Abstract: Hard time constraints in space missions bring in the problem of fast video processing for numerous autonomous tasks. Video processing involves the separation of distinct image frames, fetching image descriptors, applying different machine learning algorithms for object detection, obstacle avoidance, and many more tasks involved in the automatic maneuvering of a spacecraft. These tasks require the most informative descriptions of an image within the time constraints. Tracking these informative points from consecutive image frames is needed in flow estimation applications. Classical algorithms like SIFT and SURF are the milestones in the feature description development. But computational complexity and high time requirements force the critical missions to avoid these techniques to get adopted in real-time processing. Hence a time conservative and less complex pre-trained Convolutional Neural Network (CNN) model is chosen in this paper as a feature descriptor. 7-layer CNN model is designed and implemented with pre-trained VGG model parameters and then these CNN features are used to match the points of interests from consecutive image frames of a lunar descent video. The performance of the system is evaluated based on visual and empirical keypoints matching. The scores of matches between two consecutive images from the video using CNN features are then compared with state-of-the-art algorithms like SIFT and SURF. The results show that CNN features are more reliable and robust in case of time-critical video processing tasks for keypoint tracking applications of space missions.

Keywords: artificial intelligence; convolutional neural network; feature descriptor; machine learning; space missions

1 INTRODUCTION

Although many space missions have been successfully conquered by national and international government bodies, automation in the space related tasks is still developing. Many research challenges related to space exploration are still in their early stage of development. One of the reasons is that space applications are time critical and decision making in constrained time span is really very important. While revolving around a target planet a spacecraft always keeps on taking videos of a scene ahead using on board cameras for study purpose. These real time videos are needed to be processed within time constraints for different purposes. Extracting features useful for further space exploration and navigation tasks is at primary stage. The motion of a spacecraft result in spatially transformed images of the same scene majority of times. Detecting the most informative keypoints from videos in real-time is a challenging task. Further tracking keypoints between two consecutive video frames is a next important task for many flow estimation algorithms. In this paper, tracking keypoints between consecutive video frames is achieved using CNN features. We propose a methodology for keypoints tracking which will be suitable for time critical space applications.

The existing state-of-the-art feature detectors & descriptors [1-8] are efficient enough but are computational expensive for real-time applications. These algorithms are used for detecting and describing the most informative points of interest from an image. Important keypoints are extracted from an image and these keypoints are described in a way to suit an application of interest. Certainly, applications define that how the features must be described and represented. Statistical terms like mean, standard deviation of image intensity and yet any higher moments may serve as an efficient means of feature description. Few functions like energy, entropy or any other complex frequency transforms may also serve as feature descriptors. A keypoint extraction algorithm, Harris Corner detector [9] used the combination of corner points and edge points to describe the features but it fails to describe surfaces or an object as a whole. Scale Invariant Feature Transform (SIFT) [10] algorithm was introduced in 2004, which showed a huge paradigm shift in feature extraction and description. SIFT addressed the challenge of invariance to affine transformations along with being the most efficient descriptor. It used difference of gaussian function for detecting potential keypoints and used im-age gradient magnitude, direction from local neighbourhood for keypoint description. It is prevalently used for object detection and image matching tasks. But it is computation-ally intensive and hence not suitable for time critical appli-cations. Speeded procedure for keypoint extraction is brought up to by Speeded up robust features (SURF) [11, 12], Features from accelerated segment test (FAST) [13, 14], Binary robust independent elementary features (BRIEF) [15] and Oriented FAST and rotated BRIEF (ORB) [16] eventually. Amongst all of these algorithms SURF is found to deliver good quality and also computationally efficient features. ORB algorithm is a combination of FAST & BRIEF and works faster than SURF but its features are not found suitable for image matching tasks. Moreover, theses algorithms are standalone versions and cannot be trained for real time functioning.

With advancements computational resources and data sources few deep learning techniques [17, 2] are also developed recently. These are meant for object detection, recognition and other computer vision tasks. To name a few, deep neural network models like Inception [18], VGG [19], XCeption [20], ResNet [21] are already being implemented and tested on variety of datasets. Few new techniques have been proposed, which use transfer learning by finetuning few parameters of these already built models for completing their tasks.

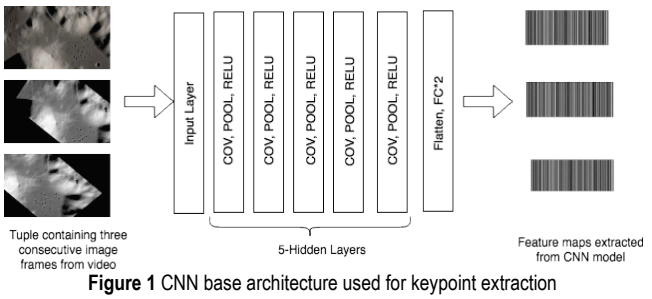


Figure 1 CNN base architecture used for keypoint extraction

Glancing through the literature we arrive at a decision of devising a model that will be suitable for time critical space applications and does not involve any complexity of detection, description of keypoints and further it must be useful for keypoints tracking operation. The solution to this problem can be using a convolutional neural network (CNN) model for feature extraction, which it does seamlessly and autonomously without any overhead of complex computations and seems reliable in real time functioning.

To address the aforementioned problem, we propose a 7-layer CNN model for keypoint extraction from three consecutive video frames. The paper discusses the complete meth-od of feature extraction from the proposed model, the model parameters, and its application to image tracking. Further to test the robustness and reliability of this method for space applications, it is compared with existing SIFT and SURF algorithms.

The organization of this paper is as follows: Section 2 discusses proposed CNN architecture, methodology. Subsection 2.1 explains experimental setup and dataset used. The performance metrics are stated in subsection 2.2. Section 3 discusses results and comparisons with state-of-the-art algorithms and finally Section 4 discusses conclusion.

2 PROPOSED METHODOLOGY

The proposed CNN architecture consists of five convolutional layers, one flattened layer and two fully connected (FC) layers as shown in Fig. 1. First layer being the input layer accepts image frames from input video sequence. The model outputs F number of 2-dimensional feature maps which are further used for feature matching and fed to Nearest Neighbor algorithm. The parameters of the model are set empirically. The flatten layer is used to convert 3-dimensional input to 1-dimensional tensor for faster processing of images. Two fully connected layers are introduced at the end to extract features. The number of computational units in last layer are equal to the feature vector dimensions. There is no need of nonlinear function at the output as the job is to just extract features not the recognition or classification task.

Each input image $I^{i}_{(M,N)}$ is padded to preserve the size of original image to get a padded image, $I^{p}_{(M+p,N+p)}$. An image is convolved (*) with f_i number of filters, each of which size is ($m_i \times n_i$) and then passed through a non-linear functional unit. Then the padded image is passed into the convolution layer to get an output image as,

$$I^{o}_{(M,N)} = MaxPool\left[Re\,LU\left(I^{p}_{(M+p,N+p)} * f_{i(m,n)} \right) \right] \quad (1)$$

Rectifier Linear Unit (ReLU) is used as an activation function for all hidden layers. Max pooling is applied on each successive output after padding. In short, each hidden unit in the model is a combination of convolution layer, a ReLU unit and a pooling layer as in any general case. Detailed procedure for keypoints matching through CNN features is detailed in Fig. (2).

Using proposed methodology shown in figure (2), CNN features can be used to track the keypoints in two consecutive images, which can be further useful in flow detection algorithms. A lunar descent video (credit: https://svs.gsfc.nasa.gov/) is used for experimentation purpose. From the video with known frame rate, images are extracted. To train the model for all kinds of transformations, each image from the dataset is undergone 100 different transformation operations. Such transformed dataset contains both reference image and its transformed versions and is known as augmented dataset D_{aug}. Subsection 2.1 will discuss this in more details. The features extracted from reference image I^{ref} works as reference values for the features extracted from its transformed versions, $I^{\text{trans}} = (I_R^{\text{ref}} + I_T^{\text{ref}} + I_S^{\text{ref}})$. The reference descriptors will be used to perform the matching between the consecutive descriptors of its transformed versions. These features represent the key points of interest from each image.

$$f_{\text{loss}} = \frac{\sum_{\text{ref, aug} \in D_{\text{aug}}} \underset{\boldsymbol{W}}{min} \left\| D_p^{\text{ref}} - D_p^{\text{aug}} \right\|}{D_{\text{aug}}} \quad (2)$$

In (2) $\boldsymbol{W}$ is the weight vector of the model. $\boldsymbol{W}$ is adjusted during each epoch to minimize loss function, f_{loss}. The algorithm starts with initial weights fetching from pretrained model VGG net and further finetuning the base model with these starting weights.

Steps for keypoint matching are described below:

For each image tuple (I^{ref}, I^{1}, I^{2}) in D_{aug} repeat the following steps:

- Supply this tuple to CNN Model to extract keypoints $(K_p^{\text{ref}}, K_p^2, K_p^3)$
- Apply image matching technique to find matched keypoints between two consecutive images,

$$\boldsymbol{K}_p^{\text{matched}} := NearestNeighbour(K_p^{\text{ref}}, K_p^1)$$

- Pass vector $\boldsymbol{K}_p^{\text{matched1}}$ to compute matching score between two images.
- Repeat step 2 and 3 for next augmented image i.e. the pair $(K_p^{\text{ref}}, K_p^2)$ and get $\boldsymbol{K}_p^{\text{matched2}}$

The procedure shown in Fig. 2 is adopted for computing matching score for all test samples through studied

algorithms for unbiased evaluation. Initially features were extracted from reference image and then from its transformed version. An efficient matching algorithm called Fast Library for Approximate Nearest Neighbours (FLANN) is used to match keypoints from both images. Those keypoints are matched whose nearest neighbors from both images have equal contribution in representing that keypoint.

All the key points in the common region of reference image and its transformed image are called correspondences between the two images. After computing the maximum correspondences, the algorithm tries to find the correct matches using some threshold.

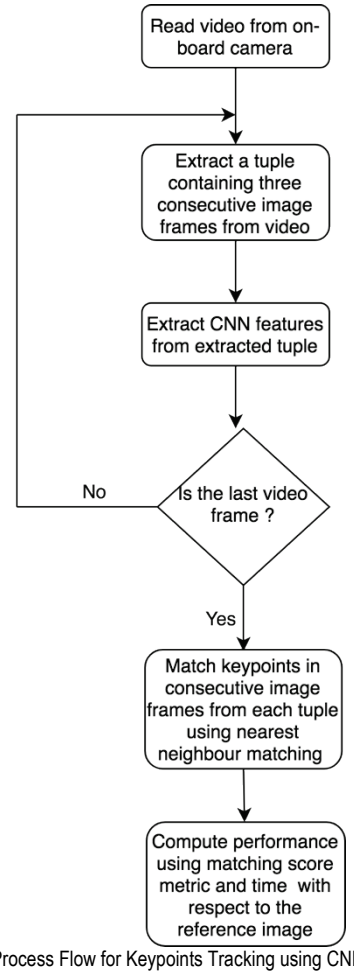


Figure 2 Process Flow for Keypoints Tracking using CNN features

2.1 Experimental Setup

The CNN model was implemented and trained on Intel core i7 processor with 16 GB RAM with NVDIA GTX 1650 graphics card for performance boost while training. The implementation is tested on 2.4 GHz Intel Core i5 processor with 10 GB DDR3 RAM. Programming language Python 3.7 in tensor flow environment is used for implementation of this work.

A video freely available on the website https://svs.gsfc.nasa.gov/ is used for generation of augmented dataset. A python script was written for extracting image frames from a spacecraft landing video. This video is an animated view of landing site of Apollo 17 mission. This video was created by the sources from Lunar Reconnaissance Orbiter (LRO) photographs and elevation maps. Total 915 images were extracted from this video to create a raw image dataset. From raw image dataset few images were selected at random, and 100 known affine transformations were applied to generate an augmented image dataset which contains total 3489 images. Size of each image is 640 × 360 pixels. Sample images from the dataset are shown in Fig. 3. Training with such augmented data will increase the robustness in the decision for feature matching algorithms. The known transformation matrices would help in finding the exact matches. All these efforts are based on the pre assumption that while descending, a satellite captures the same scene with different orientations, scales and very little spatially translated versions of it. So ultimately the maximum number of descriptors must match with its reference image.

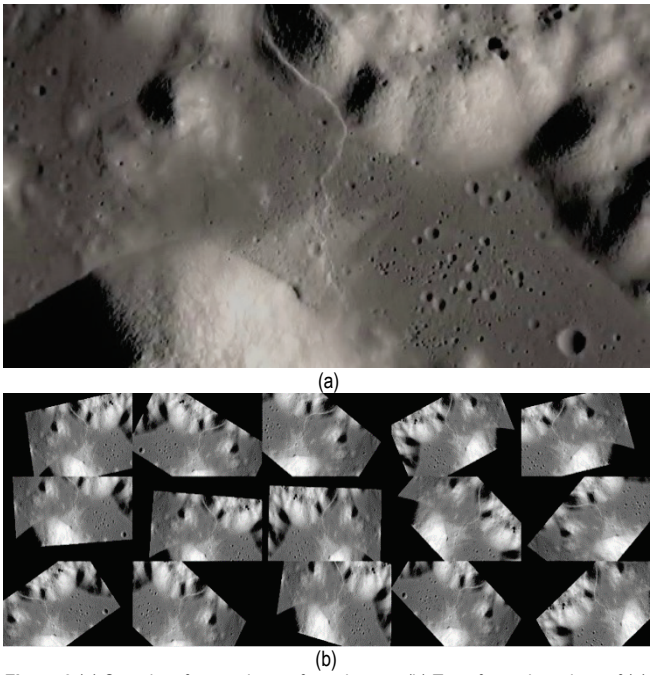


Figure 3 (a) Sample reference image from dataset, (b) Transformed versions of (a) from augmented dataset

For training CNN, padding is used to avoid loss of input dimensions. Stride of 2 is employed to use all parts of image evenly while performing convolution.

The performance of proposed model is evaluated using matching score [3, 6, 22-24], a metric which is widely being used for feature tracking. To check the reliability of our model, it is compared with the state-of-the art algorithms like SIFT [10] and SURF.

2.2 Performance Metrics

Performances of the system is evaluated using matching score [3] and time taken for feature extraction.

Matching score is computed through image correspondences which are matched key points in common regions of two images for which homography is known. It is

the measure of accuracy as it matches descriptors of logically same key points from two different images. Matching score is calculated by (3).

$$Matching\ Score = \frac{C^{+} \bigcap C^{*}}{F_{\text{ref}}} \tag{3}$$

Where, C^{+} is maximum image correspondences, C^{*} is number of correct matches and F_{ref} being the number of descriptors of reference image.

3 RESULTS & DISCUSSION

3.1 Keypoints Tracking using CNN features

Tab- 1 shows matched keypoints between reference and transformed image frames. The first column contains few transformed sample images from the augmented database. Descriptors of all these test samples are compared with reference image shown in Fig. 3(a). The second column contains the images showing visualizations of this matching. Only initial 10 matching points are shown inside the image pairs. Last column shows the corresponding matching score computed through CNN descriptors. It can be observed that, visualizations show the matched keypoints between the two images.

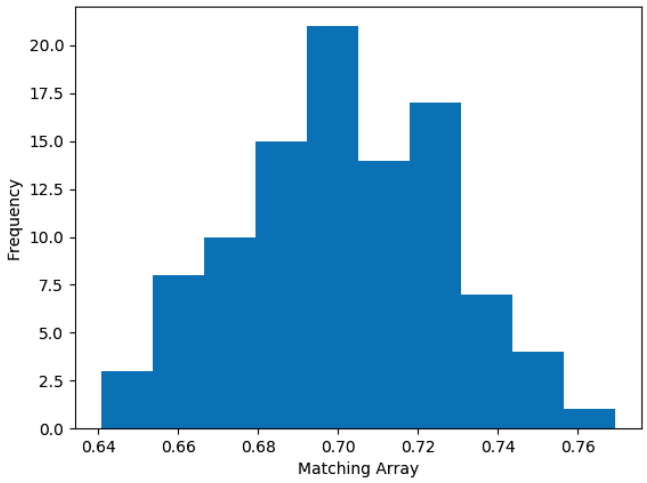
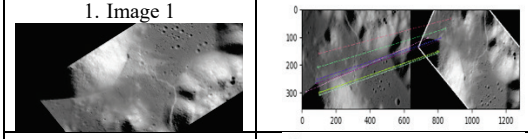
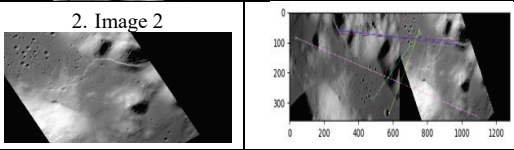
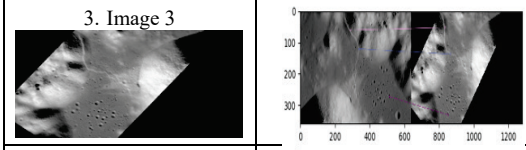
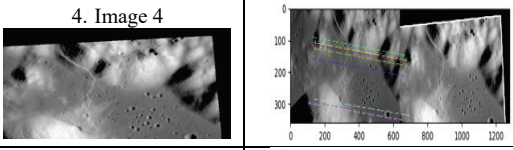
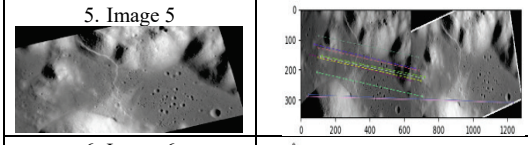
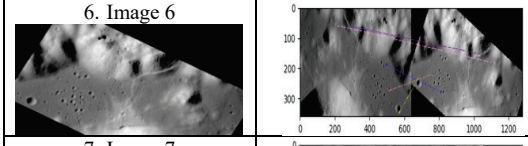
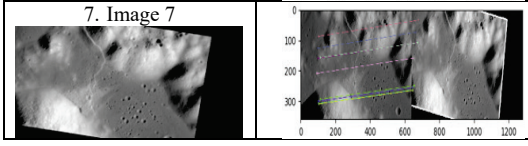


Figure 4 Frequency Histogram of Matching Score

Matching score coveys the correct matches out of total matched keypoints. It is the measure of accuracy in the image tracking algorithms. As all the test samples are spatially transformed versions of the same reference image, ideally matching score should be on the higher side i.e. nearer to 1. Image 5 in the Tab. 1 shows the highest matching score, while Image 3 has the lowest matching score. In image 5, the keypoints below $y \leq 300$ are incorrectly matched, but rest shown points are accurately matched. It can be observed that the test image is slightly ~30° rotated version of reference image and hence maximum scene of reference image is present inside the test image. Estimated matching score of 0.75 guarantees that. On the other hand, image 3 is almost ~180° vertical flip along with quite a large translated version of reference image. So, the chances of finding match between the two are lesser than the previous case. This is explained by its estimated matching score of 0.6582. Rest all images and their visual matching with reference image are self - explanatory.

Fig. 4 shows the frequency distribution of matching score obtained by applying matching algorithm on the 100 test samples. Maximum samples are found in the range of 0.69 to 0.71.

Table 1 Keypoints Tracking using CNN Features
(Note: For better visualization please zoom in the images)

Input Test Image	Visual Matching between reference image and test image	Matching Score
1. Image 1		0.7207
2. Image 2		0.6953
3. Image 3		0.6582
4. Image 4		0.7246
5. Image 5		0.7500
6. Image 6		0.6992
7. Image 7		0.6895

3.2 Comparison with State-of-the-Art Methods

Matching score of the CNN model is compared with SIFT and SURF descriptors. Tab. 2 shows the range and variation of matching score obtained from SIFT, SURF and CNN descriptors. The average value of matching score is highest equal to 0.7008 for CNN descriptors and lowest i.e. 0.3079 for SURF descriptors. SIFT value is 0.4744. As the test images are transformed versions of reference image, the visual matching between two images should be more than 0.50. It can be seen that only CNN descriptors are able to reach that visual threshold. CNN descriptors are showing

visual perfection in computing matching score as compared with the other two descriptors. The variations in the readings are least and almost negligible in case of CNN descriptors. On the contrary, there are much variation is shown in the performance of SIFT and SURF descriptors. This shows robust performance of CNN model for test samples. Tab. 2 also shows the maximum and minimum values of matching score for all the three algorithms. The minimum values shown by SIFT and SURF do not match with visual matching context. The minimum value computed by CNN model is also more that our visual threshold. Hence CNN model seems to be more reliable than other two models for our problem. These statistics are visually represented in Fig. 5. These visual representations would further clarify the discussions in the previous section.

Table 2 Keypoints Matching Score Statistics from Lunar Descent Video Frames for Different Algorithms

Algorithms	SIFT	SURF	CNN
Mean Value	0.4744	0.3079	0.7008
Max Value	0.6357	0.5663	0.7695
Min Value	0.0145	0.0156	0.6406
Variance	0.0079	0.0119	0.0006

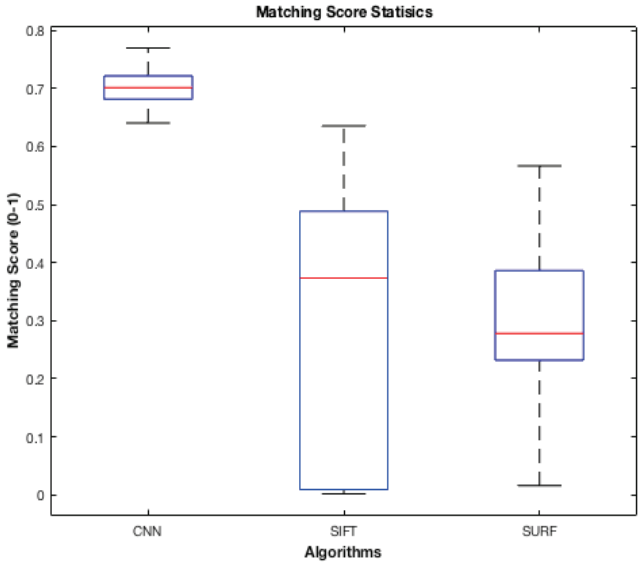


Figure 5 Statistical analysis of matching score obtained from conventional algorithms and CNN model

The visual comparison of matching score computed by three algorithms is shown in Tab. 3. As per human perception of visual context, image 1 is obtained by horizontal flipping and 30° rotation of reference image (Fig. 3(a)). So maximum context of the original image is saved in the test image. The matching score should reflect that visual matching. CNN descriptors show maximum matching as shown in table. The same interpretation is suitable for image 4 and image 5, where CNN model outperforms. Image 2, 6, and 7 are spatially rotated and translated versions of reference image and hence the less area is matched.

This is contributed by decrease in matching score of all the descriptors. But here also as per visual perception, it should be more that threshold 0.50 which is shown by CNN model. For image 7 performance of SIFT and SURF is also good almost near threshold but for image 2 it is quite low. For image 6, SIFT and SURF failed to find correct matching between reference and test image. But CNN shows a robust performance. The same interpretation is for image 3.

Table 3 Visual Comparison of Matching Score Based on SIFT, SURF and CNN Descriptors
(Note: For better visualization please zoom in the images)

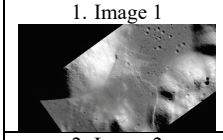
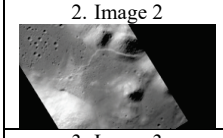
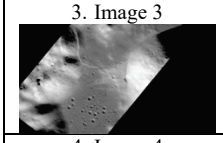
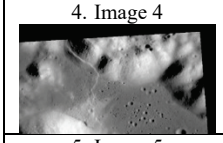
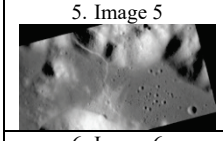
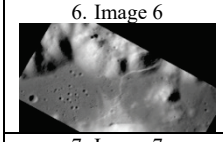
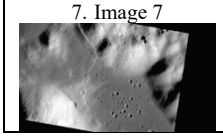
Input Test Image	Matching Score by		
	SIFT descriptors	SURF descritpors	CNN descritpors
1. Image 1 	0.5282	0.2838	0.7207
2. Image 2 	0.2185	0.1823	0.6953
3. Image 3 	0.1473	0.0982	0.6582
4. Image 4 	0.5318	0.4564	0.7246
5. Image 5 	0.5573	0.4322	0.7500
6. Image 6 	0.0982	0.0543	0.6992
7. Image 7 	0.5774	0.4721	0.6895

Table 4 Processing Time of SIFT, SURF and CNN MODELS

Evaluation Criterion	SIFT	SURF	CNN
Average Time (seconds)	8.1573	5.3321	**0.0348**

Tab. 4 shows the average processing time enquired for all the algorithms. Processing time is computed by adding descriptor extraction time and feature matching time. As CNN is trained model, it requires very less time for processing of a single test image which is only 34.8 milliseconds. The highest processing time is needed for SIFT descriptors as it requires lots of computations. As our problem requires real time processing of space videos, CNN seems more convenient and reliable model as far as time constraints of space applications are considered.

Overall performance of CNN model is more reliable and robust as compared with state-of-the-art methods and hence it seems more suitable for real time video processing of space missions.

4 CONCLUSION

Real time space mission tasks are time critical and hence for such tasks processing time plays an important parameter of evaluation. Keypoints, which are special points of interest inside an image must be tracked between consecutive image frames of a real time video captured by on board spacecraft cameras. It is useful for many flow detection algorithms and other space applications. In this paper a methodology using CNN descriptors is proposed for such time critical applications. A new 7-layer CNN model is developed and features thus extracted are used for keypoints matching between consecutive image frames of a lunar descent video. The CNN is trained using pre-trained VGG model parameters and fine-tuned for the new data. Total 100 test image samples for a single reference image were used for evaluating the performance of the CNN model. It is observed that, CNN descriptors are time efficient, robust and hence reliable for image tracking applications. Statistical analysis of matching score shows less variations in the CNN descriptors and is ideal for real time performance.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

5 REFERENCES

[1] Awad, A. I. & Hassaballah, M. (2016). *Image Feature Detectors and Descriptors: Foundations and Applications*. Studies in Computational Intelligence 630, Springer International Publishing. https://doi.org/10.1007/978-3-319-28854-3

[2] Fischer, P., Dosovitskiy, A., & Brox, T. (2014). Descriptor Matching with Convolutional Neural Networks: a Comparison to SIFT. *ArXiv*, abs/1405.5769.

[3] Boyraz, P. & Bayraktar, E. (2017). Analysis of Feature Detector and Descriptor Combinations with a Localization Experiment for Various Performance Metrics. *Turkish Journal of Electrical Engineering and Computer Sciences, 25*(3), 2444-2454. https://doi.org/10.3906/elk-1602-225

[4] Lenc, K. & Vedaldi, A. (2018). Large scale evaluation of local image feature detectors on homography datasets. https://arxiv.org/pdf/1807.07939.pdf

[5] Li, S. (2017). A review of feature detection and match algorithms for localization and mapping. *IOP Conference Series: Materials Science and Engineering, 231*, 012003. https://doi.org/10.1088/1757-899X/231/1/012003

[6] Mikolajczyk, K., Tuytelaars, T., Schmid, C. et al. (2005). A Comparison of Affine Region Detectors. *Int J Comput Vision, 65*(1-2), 43-72. https://doi.org/10.1007/s11263-005-3848-x

[7] Moura, G. M. & Silva, R. L. S. (2017). Analysis and Evaluation of Feature Detection and Tracking Techniques using OpenCV with Focus on Markerless Augmented Reality Applications. *Journal of Mobile Multimedia, 12*(3-4), 291-302. https://doi.org/10.26421/JMM12.3-4

[8] Salahat, E. & Qasaimeh, M. (2017). Recent advances in features extraction and description algorithms: A comprehensive survey. *2017 IEEE International Conference on Industrial Technology (ICIT)*, Toronto, ON, 1059-1063, https://doi.org/10.1109/ICIT.2017.7915508

[9] Harris, C. & Stephens, M. (1988). A Combined Corner and Edge Detector. In C. J. Taylor, editors, *Proceedings of the Alvey Vision Conference*, 23.1-23.6. Alvey Vision Club. https://doi.org/10.5244/C.2.23

[10] Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision, 60*, 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[11] Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. *Proceedings, Part I, 9th European Conference on Computer Vision - ECCV 2006*, Graz, Austria, 404-417. https://people.ee.ethz.ch/~surf/eccv06.pdf

[12] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding, 110*(3), 346-359. https://doi.org/10.1016/j.cviu.2007.09.014

[13] Rosten, E. & Drummond, T. (2006) Machine Learning for High-Speed Corner Detection. In: Leonardis A., Bischof H., Pinz A. (eds) *Computer Vision – ECCV 2006. Lecture Notes in Computer Science, vol 3951*, Springer, Berlin, Heidelberg, 430-443. https://doi.org/10.1007/11744023_34

[14] Rosten, E., Porter, R., & Drummond, T. (2010). Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Anal. Mach. Intell., 32*(1), 105-119. https://doi.org/10.1109/TPAMI.2008.275

[15] Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011). BRISK: Binary Robust invariant scalable keypoints. *Proc. IEEE Int. Conf. Comput. Vis.*, 2548-2555. https://doi.org/10.1109/ICCV.2011.6126542

[16] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. *Proc. IEEE Int. Conf. Comput. Vis.*, 2564-2571. https://doi.org/10.1109/ICCV.2011.6126544

[17] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234-241.

[18] Szegedy, C. et al. (2015). Going deeper with convolutions. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 1-9. https://doi.org/10.1109/CVPR.2015.7298594

[19] Simonyan, K. & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 1-14.

[20] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, 1800-1807. https://doi.org/10.1109/CVPR.2017.195

[21] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 770-778. https://doi.org/10.1109/CVPR.2016.90

[22] Ehsan, S., Kanwal, N., Clark, A. F., & McDonald-Maier, K. D. (2010). Improved repeatability measures for evaluating performance of feature detectors. *Electron. Lett., 46*(14), 998-1000. https://doi.org/10.1049/el.2010.1442

[23] Mouats, T., Aouf, N., Nam, D., & Vidas, S. (2018). Performance Evaluation of Feature Detectors and Descriptors beyond the Visible. *Journal of Intelligent & Robotic Systems, 92*, 33-63. https://doi.org/10.1007/s10846-017-0762-8

[24] Schönberger, J., Hardmeier, H., Sattler, T., & Pollefeys, M. (2017). Comparative Evaluation of Hand-Crafted and Learned

Local Features. *Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, Honolulu, HI, 6959-6968. https://doi.org/10.1109/CVPR.2017.736

Authors' contacts:

Janhavi H. Borse, Assistant Professor
(Corresponding author)
Sandip Institute of Technology & Research Centre,
Affiliated to Savitribai Phule Pune University,
Ganeshkhind Road, Pune, Maharashtra 411007, India
Tel.: 9225367554, E-mail: borse.janhavi@gmail.com

Dipti D. Patil, Associate Professor
MKSSS's Cummins College of Engineering for Women,
Karve Nagar, Pune, Maharashtra 411052, India
E-mail: diptivt@gmail.com

Vinod Kumar, Division Head - Control Dynamics Design Group
U R Rao Satellite Centre,
Old Airport Road, Vimanapura, Bangalore, Karnataka 560017, India
E-mail: vinodkkaushik@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
httpa://doi.org/10.31803/tg-20210204155601

Original scientific paper

Fuzzy Logic Approach for Routing in Internet of Things Network

Reena Pingale*, S. N. Shinde

Abstract: A performance of network is evaluated by considering different parameters. The network lifetime depends on many factors Residual energy, Link lifetime and Delay. The Major Challenge in IoT is to the increased lifetime of low power and lossy network (RPL).The process considering input and output to evaluate Network performance by considering the above factors. The proposed system makes use of FIS (Fuzzy Inference System) for selecting the best path to maximize network lifetime. The outcome obtained by using MATLAB and Network performance is increased. The excellent route is selected if Residual Energy is 194, Link quality is 51.2 and Delay is 1.05 then excellent route quality is 73.4%.

Keywords: fuzzy logic; Internet of Things (IoT); link lifetime; reliability; residual energy

1 INTRODUCTION

The word [1] "Internet of Things (IoT)" performances as an umbrella word that shields the several structures. The deployment of different embedded devices having to sense capabilities to communicate with embedded devices and linking between physical and digital devices .The IoT is providing smarter services and always changing technology [2]. In [3] Internet of Things (IoT) has delivered a encouraging opportunity to form prevailing developed structures and applications by leveraging the increasing ubiquity of radiofrequency identification (RFID), and wireless, mobile, and sensor devices. As an energy to comprehend the enlargement of IoT in productions novelists evaluation the recent exploration of IoT, key-enabling skills, foremost IoT solicitations in productions, and recognizes exploration leanings and encounters. The key involvement of authors is that they brief the recent advanced IoT and its use in businesses analytically.

RPL is measured [4] ordinary for improving the routing structure for congregating troupe movement design. Preliminary from a boundary router, RPL paradigms a DODAG by one or numerous metrics. The DODAG is produced by seeing the concerning budgets, node restriction and multiobjective purpose. Rank group for each node on the DODAG is completed by the detached function. It cares numerous kinds of circulation such as MTP, points to multipoint and points to point. For consuming free topology, the rank essential severely growth from the root near plants of the DODAG. In composite situations lossy relation system is separated into several screens reliant on the request's situation. So in conditions, it might procedure numerous awkward DODAG's with autonomous roots. RPL having many occurrences and it can be route synchronously on the system device and in RPL if nodes need to contribute in DODAG by using different direction-finding procedures for conclusion the greatest way to transporting data. In this paper, we suggest three main limitations residual energy (RE), Link Lifetime (LT), Delay to choice the greatest direction.

The main objective of this research is to invent and enlarge routing algorithm for IoT Network by proposing Node selection algorithm. The goal of the research is to develop a novel routing strategy based node selection algorithm. The selection of best route is based on residual energy, link quality and delay.

The major contributions of the paper to proposed Node Selection algorithm to acquire best route quality to improve network performance for IoT network.

The section of the paper is planned as follows: Literature survey is considered in section 3, Section 4 cover the problem definition, section 5 deals with the proposed factors affecting to route. Section 7 with proposed algorithm. The results, with outcome in section 8 and, section 9 deliberates the conclusion.

2 RELATED WORKS

2.1 RPL Overview

RPL routing protocol remains to exploit the complete generation of the system by attractive maintenance of the most energy-constrained nodes. RPL planned the Expected ELT for meaning the outstanding instance of the node. They created a DODAG constructed on the ELT metric for precisely approximating the period of all the routes near the boundary router and envisioned a device for observing bottlenecks designed for dispersion the circulation load to numerous parents.

RPL [4] has mostly four control messages, DODAG Information Solicitation (DIS), Information Object (DIO), Advertisement Object (DAO) and Advertisement Object – Acknowledgement (DAO-ACK). Firstly, the DODAG request is carried out in two ways

- Applicant node directs the DIS demand to DODAG
- DODAG directs the DIO demand messages to all contributor nodes.

The DODAG permits the drop timer and the contributor node wants to transmit DAO controller communication to DODAG inside the time intermission. Then, the DODAG direct DAO-ACK controller communication to entirely contributor nodes.

2.2 Challenges

1.1 The steady system is conserved by decreased the overhead and end-end delay [5].
1.2 The routing in the system in serious condition due to convergence problems [6].
1.3 The main factors related to security tasks are network topology [7].
1.4 The IoT used mainly the relay function for proper functioning of sensor node.

3 LITERATURE SURVEY

Many types of researches have areas completed work on energy-aware routing in RPL and in this, it will minimize energy consumption and increase network lifetime.

In [8] this offerings the routing protocols for the Internet of Things which is supportive in transporting the data into the vapors or to the operators. Several of the general direction-finding protocols are studied in this laterally with the submissions of IoT. In this paper stretches a short-term opinion of the tasks which originate when by IoT for real-time. Here IPv6, CoAP, MQTT and RPL routing protocols are conferred and enlarged. IoT consumes the possible to yield a huge quantity of facts into the folders and the data will be transmitted proficiently.

Secure Multi-hop Routing Protocol (SMRP)[9] protocol attentions on collective the security of the data by avoiding spiteful outbreaks. This direction-finding protocol allows the IoT strategies to confirm previously starting a novel network or construction a standing one. The confirmation uses multilayer restrictions such as User-Controllable ID, user's pre-agreed submission(s) and list of allowable strategies into routing algorithms for joining the confirmation and routing procedures without suffering substantial expenses.

As per observation by Sharief M. A et al. [10], given that IoT system fits to dissimilar holders, PAIR protocol announce a estimating perfect for assistances the transitional nodes to acquire the economic assistance as they apply their properties for transmitting. As estimating perfect of PAIR protocol is based on many restrictions like Residual energy and power consumption, recent weight and buffer space, Distance to neighbours.

The persistence of the routing network designed for IoT (AOMDV-IoT) [11] is to find and generate the linking among expected nodes and the Internet nodes. The protocol defines as reactive protocol that defines the pathway on request. In this paper, the author contributions an expansion of AOMDV improved used for IoT, which can choice a steady Internet broadcast pathway energetically through informing the Internet linking the table. Using reproductions authors presented that the package defeat is better-quality then the end to end delay is reduced.

The main detached of the Energy-aware Ant Routing Algorithm (EARA) is to adjust the routing process for exploiting the lifetime of network [12]. It defines as the swarm intelligence algorithm and reflects the similarly equal number of nodes. As the remaining drive in the IoT strategies deviations finished phase, the authors had announced the instrument near appraise energy evidence. Routing protocol originated on link and residual energy (REL) [13] usages the linkage excellence of remaining energy and wireless network throughout the pathway collection procedure to growth organizations dependability then offers QoS towards the various IoT requests. The load balancing device of this protocol circumvents the extreme use of a solitary track or solitary knot which can additional support in dropping the spots or energy hovels in the system. The energy application will be unchanging in the system. In this paper [14], the authors spoke the network lifespan optimization for the wireless sensor system. The Authors defined the strategy and investigation of numerous energy complementary methods. For a consistent grid topology, we resulting an ideal explanation. The authors demanded that the location of the base position (in the corner) streamlines the optimization problem. They presented that variable the base station location presents new dissimilarities restrictions to the problematic.

Authors in [15] reflect together energy and delay metric to discovery and best pathway with lowest energy ingesting and a lowest end to end delay for real-time circulation in wireless sensor systems. This total is calculated as a linear grouping of the broadcast delay and node's energy on the pathway.

4 PROBLEM DEFINITION

Internet of things having an increase in the number of devices due to this strategy traffic will increase which is beyond the capacity of the network. The outcome will be to decrease the performance of the network. It is necessary to find proper routing paths that will give good network performance.

5 THE PROPOSED WORK

We suggest an enhanced type of RPL network. The fuzzy logic approach to excellent the finest direction to transmission the facts proficiently. The proposed algorithm finds out the quality of the selected node and it compares with the set of nodes and then selects the finest node in DODAG and the remaining nodes send data through the finest node

The factors consider as below.

5.1 Residual Energy Consumption

Residual Energy ingestion of node is calculated after every time interval *t*. With the following equation, it is possible to find out the value of every node with some time interval [16].

$$EN_t = (N_t \cdot E_t) + (N_r \cdot E_r), \tag{1}$$

$$R_{e_t} = \frac{E_{\text{initial}} - E_{\text{spent energy}}}{E_{\text{initial}}}, \tag{2}$$

Where: EN_t - After time t energy spent by node N, N_t - Total of transferred packets, N_r - Total of expected packets, E_t -

Energy of transferred packet, E_r - Energy of acceptance the packet.

The remaining energy is intended by the variance among primary energy and consumed energy

5.2 End to End Delay

As per specified [17] as average interval occupied by data packets to effectively communicating messages crossways the system from source to destination

$$Delay = \sum_{i=1}^{n} \frac{(R_i - S_i)}{n}. \quad (3)$$

5.3 Link Lifetime

The system link lifespan is predicated from the quantity of transmissions. It represents forward and reverses data delivery.

$$N_i(X) = \frac{1}{F_d \cdot R_d}. \quad (4)$$

The Link quality of the path can be calculated by

$$N_q = \sum_{i=1}^{n} N_i(X). \quad (5)$$

Where: N_i - Link Lifetime, F_d - Represents data packet reach to the destination successfully, R_d - Represents acknowledge packets are received by the sender successfully [18].

6 FUZZY LOGIC BASED ROUTING ALGORITHM IN RPL

The fuzzy logic applies completed routing to excellent the greatest route for transporting data effectively with attention of three-parameter Residual energy consumption, Delay and Link lifetime. The fuzzy logic set was presented in 1965 as a scientific way to denote linguistic vagueness (Zadeh, 1965) [21]. Allowing to the fuzzy logic impression, features and measures can be secret without certain bounds. Fuzzy logic is actual valuable for lecturing real-world difficulties, which typically contain a grade of vagueness.

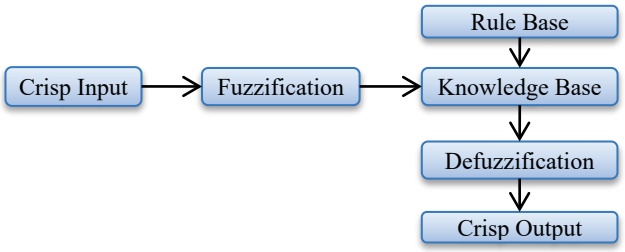


Figure 1 Fuzzy inferences System

The FIS takes linguistic inputs (as stated for simplification), procedures the evidence and outputs the presentation [19].

6.1 Fuzzification

Fuzzification takes input fuzzy value from crisp value. The input values are (Residual Energy, Link Lifetime, Delay) its convert these values in linguistic variable and membership function

Table 1 Fuzzy Input Variables

Linguistic Expression	Input Range
Low Residual Energy	0 - 100
Average Residual Energy	50 - 200
High Residual Energy	150 - 250
Low Link Lifetime	0 - 25
Average Link Lifetime	15 - 80
High Link Lifetime	60 - 100
Low Delay	0 - 2
Average Delay	1 - 4
High Delay	3 - 5

Table 2 Fuzzy output Variables

Linguistic Expression	Output Range
Awful Route Quality	0 - 25
Bad Route Quality	10 - 35
Degraded Route Quality	25 - 50
Average Route Quality	40 - 60
Acceptable Route Quality	50 - 75
Good Route Quality	60 - 90
Excellent Route Quality	75 - 100

6.2 Linguistic Variable

The variable represents the input and output of the variable. In this residual energy having three linguistic variables High, Average and Low. The output variable also define linguistic variable Awful, Bad, Degraded, Average, Acceptable, Good, Excellent

6.3 Membership Function

It is a mapping of membership function values to the real world measurement values, so that the actions can be functional to them. This function evaluates the linguistic variable. Membership function values are in-between range 0 to 1.

$$\mu_A(z) = \begin{cases} 1 & \text{if} \quad z \le a \\ \dfrac{z-a}{x-a} & \text{if} \quad a < z \le x. \\ 0 & \text{if} \quad z \ge x \end{cases} \quad (6)$$

6.4 Fuzzy Rule Base

The effect which the FIS types is resulting from the instructions which are kept in the record. These are kept as a set of instructions. The rules are 'If-Then' declarations that are in-built and informal to appreciate meanwhile they are unknown but public English declarations

6.5 Defuzzification

It is the procedure of changing the fuzzy input into a crisp set. The value ranges by MF in between 0 and 100 and it delivers single crisp value. We require certain weighted average technique for Defuzzification [20].

$$X = \frac{\sum_{i=1}^{n} Z_i \cdot \mu \cdot B(Z_i)}{\sum_{i=1}^{n} \mu \cdot B(Z_i)}. \quad (7)$$

The fuzzy inference system to determine the optimal path from a basis node to the endpoint node. This will progress the performance of the network.

7 PROPOSED WORK BASED ON RANK CALCULATION

The rank of the node computes from the root node and at each level increases the rank by 1. The Equation value can be calculated by using the Defuzzification process. The rank equation can be defined as

Rank1 (N) = Root_Rank (N) + Increase_Rak1 (8)
Increase_Rak1 = Equation + MinHop_Rank_increase (9)

7.1 Node Selection Process

The node selection process based on construction of MF using rule based system. The node selection process using FIS system.

Algorithm1 - Node Selection

1. INPUT: No.of Node N, Node_parennt_ID, Sender_parentID, Excellent Route=∞
2. Output : Selected Route
3. For Selected_Node ∈ List_Node do
4. Rank (No. of. Node N) <– Rank_Selected_Node(N) + Increase_Rank
5. Increase_Rak1 <– Equation + MinHop_Rank_increase
6. Intialize variable Energy, Link_quality, Delay
7. Consider fuzzy input and output
8. Input : No. of Node N, Node_parennt_ID, Sender_parent_ID, Excellent Route=∞
9. Output : Selected Route
10. Contruct fuzzy membership_ function
11. Check fuzzy rule base in fuzzy inference system
12. Apply Defuzzification
13. Equation $= \frac{\sum_{i=1}^{n} Z_i \cdot \mu \cdot B(Z_i)}{\sum_{i=1}^{n} \mu \cdot B(Z_i)}$.
14. If Excellent Route >= Selected Route then
15. Excellent Route <– Selected Route
16. end if
17. end

7.2 Results and Discussion

This segment demonstrates the evaluation of anticipated system with fuzzy inference over feigning the routing for IoT network. The analysis is done by Fuzzy rule based system.

The study of presentation built on the suggested Node Selection algorithm using residual energy, delay and link quality factors is estimated in this segment with output parameter as Route quality.

The analysis is performed by selecting rule based system using FIS to generate the result. The analysis is done by varying parameters

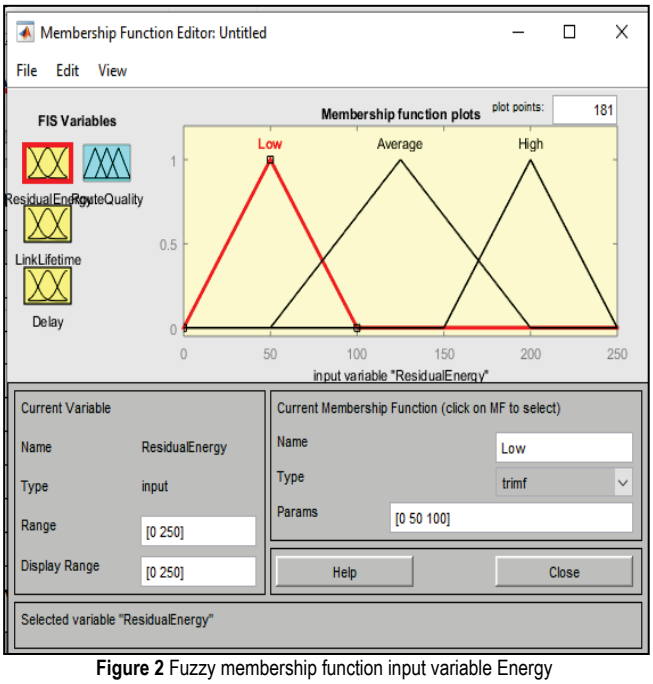


Figure 2 Fuzzy membership function input variable Energy

The Residual energy values are represented in between 0 to 250. The linguistic parameters are Low, Average and High. The membership function collections from 0 to 1.If the Network lifetime increases then the network is energy efficient. The fuzzy membership function can be represented of input variable delay, residual energy and link quality and output variable representation using a fuzzy rule-based system. The delay values are represented in between 0 to 5 and Link quality values are 0 to 100.

The fuzzy set is grouping of dissimilar metrics, every metric cover specific fuzzy variable. The rule constructed contains of $3^3 = 27$ fuzzy based rules. This is constructed on the input variable and membership function. We can describe the fuzzy based rule which characterizes the first column as the count of total number of rules and next 2 to 4 column signifies input fuzzy logic variable and the last column characterizes output variable in the form of Route Quality. The output follows max operator as combination and min operator as configuration function.

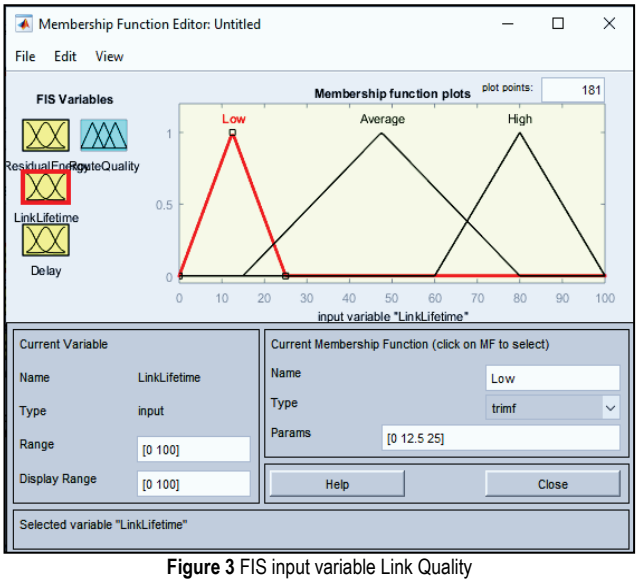


Figure 3 FIS input variable Link Quality

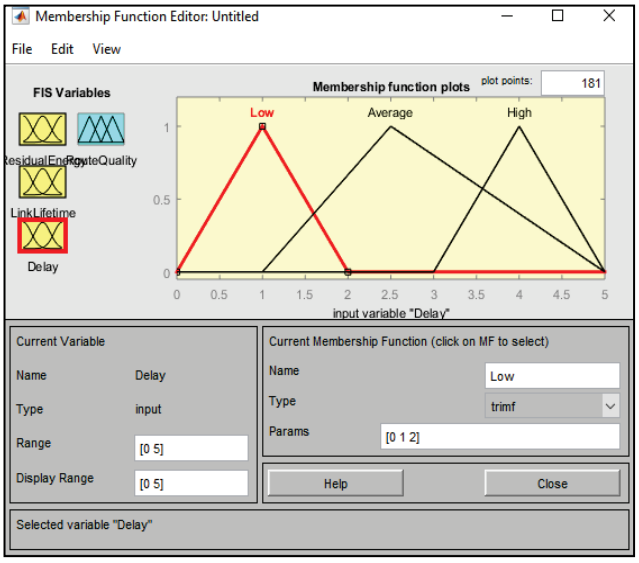


Figure 4 FIS input variable Delay

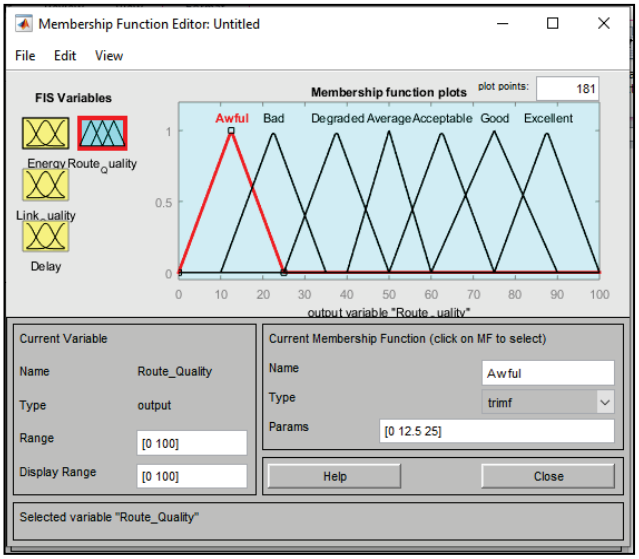


Figure 5 FIS output variables Route Quality

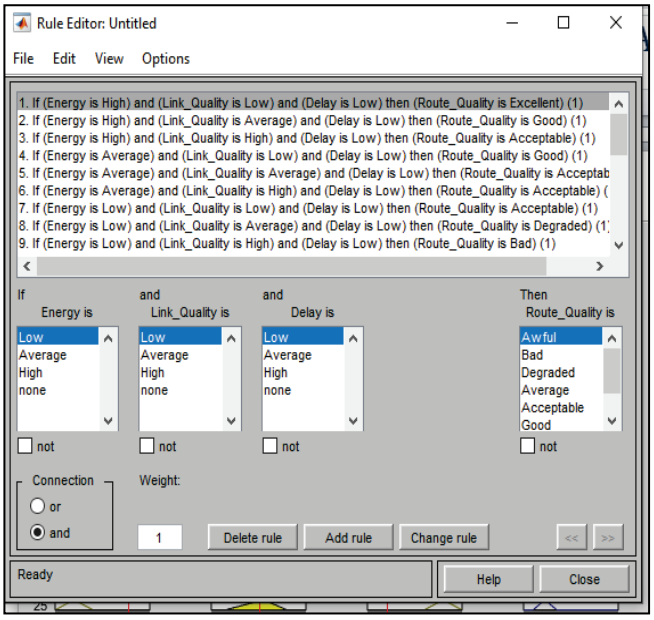


Figure 6 Fuzzy Rule-based system

Figure 7 Outcome of Route Quality

From Fig. 7, we observed that Route_ quality is above 73.4% means its excellent route selected from this we first declare the variable as Residual Energy is 194, Link Quality is 51.2, and Delay is 1.05. As per fuzzy membership function, the linguistic variable Energy value is "Average and high" and MF values are 0.5 and 0.5. The Link Quality of the linguistic variable significance is "Average" with MF value is 1, linguistic variable significance of delay is "Low" with the MF value is 1. From rule number (2) and (5) its process outcome in the form of route quality parameter as acceptable and good. The Defuzzification process can be applied and calculated using a formula

$$X = \frac{(0.5 \cdot 68 + 0.5 \cdot 79)}{0.5 + 0.5} = 73.4 \qquad (10)$$

From the outcome excellent route selected with proper selection of input variable. The surface view can be represented with parameter detail.

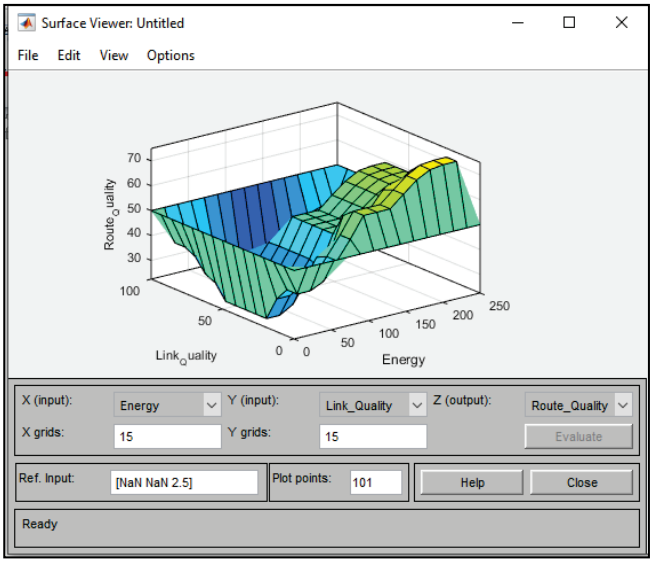


Figure 8 Surface Residual Energy, Link Quality, Delay

8 CONCLUSION

In this paper fuzzy logic approach for RPL network utilised for the IoT network system. Considering the input and output parameter in a FIS to generate the required outcome in the form of route quality. The selection of the route is constructed on three factors Residual Energy, Link Lifetime, Delay to generate proper route selection to increase network lifetime. The suggested algorithm allows the operative presentation and collection of achievable and excellent path. The yield of anticipated algorithm is calculated by selecting excellent route ,if Residual Energy is 194, Link quality is 51.2 and Delay is 1.05 then excellent route quality is 73.4%. The Matlab simulation gives the outcome in the form of route quality and future work will consider the deployment of the node in the real-time environment.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

9 REFERENCES

[1] Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I. (2012). Internet of things: Vision, applications and research challenges. *Ad Hoc Networks, 10*(7), 1497-1516. https://doi.org/10.1016/j.adhoc.2012.02.016

[2] Coetzee, L. & Eksteen, J. (2011). The Internet of Things -- Promise for the future? An introduction. In *IST-Africa Conference Proceedings, 2011*, 1-9.

[3] Li Da Xu, Wu He, & Shancang Li. (2014). Internet of Things in Industries: A Survey. *IEEE Transactions on Industrial, Informatics, 10*(4), 2233-2243. https://doi.org/10.1109/TII.2014.2300753

[4] Thubert, P., Winter, T., Brandt, A., Hui, J., Kelsey, R., Levis, P., Pister, K., Struik, R., Vasseur, J., & Alexander, R. (2012). RPL: IPv6 Routing Protocol for Low power and Lossy Networks. IETF. RFC 6550.

[5] Lim, W. H. & Isa, N. A. M. (2014). Particle swarm optimization with adaptive time-varying topology connectivity. *Applied Soft Computing, 24*, 623-642. https://doi.org/10.1016/j.asoc.2014.08.013

[6] Quan Le, Thu Ngo-Quynh, & Magedanz, T. (2014). RPL-based multipath routing protocols for internet of things on wireless sensor networks. In *2014 International Conference on Advanced Technologies for Communications (ATC 2014)*, 424-429.

[7] Al-Turjman, Fadi. (2017). Energy-aware data delivery framework for safety-oriented mobile IoT. *IEEE Sensors Journal, 18*(1), 470-478. https://doi.org/ 10.1109/JSEN.2017.2761396

[8] Channamma, S., Bhuvaneshwari, M., Pooja, M., & Jayalakshmi, J. (2016). Routing, Application and Research challenges in IoT. *Special Issue on ICRIET-2016*, Published by Research Trend, Website: www.researchtrend.net

[9] Loh, P., Chze, R., & Leong, K. S. (2014). A Secure Multi-Hop Routing for IoT Communication. *IEEE World Forum on the Internet of Things (WF-IoT)*, 428-432. https://doi.org/10.1109/WF-IoT.2014.6803204

[10] Oteafy, S. M. A., Al-Turjman, F. M., & Hassanein, H. S. (2012). Pruned Adaptive Routing in the Heterogeneous Internet of Things. *Global Communications Conference (GLOBECOM), 2012 IEEE*, 214-219. https://doi.org/10.1109/GLOCOM.2012.6503115

[11] Tian, Y. & Hou, R. (2010). An Improved AOMDV Routing Protocol for the Internet of Things. *2010 International Conference on Computational Intelligence and Software Engineering (CSE 2010)*, Wuhan, 1-4. https://doi.org/10.1109/CISE.2010.5676940

[12] Frey, M., Grose, F., & Gunes, M. (2014). Energy-aware Ant Routing in Wireless Multihop Networks. *2014 IEEE International Conference on Communications (ICC 2014)*, 190-196. https://doi.org/10.1109/ICC.2014.6883317

[13] Machado, K., Rosário, D., Cerqueira, E., Loureiro, A. A. F., Neto, A., & de Souza, J. N. (2013). A Routing Protocol Based on Energy and Link Quality for Internet of Things Applications. *Sensors, 13*(2), 1942-1964. https://doi.org/10.3390/s130201942

[14] Kacimi, R., Dhaou, R., & Beylot, A.-L. (2013). Load-Balancing Strategies for Lifetime Maximizing in Wireless Sensor Networks. *Ad Hoc Networks, 11*(8), 2172-2186. https://doi.org/10.1016/j.adhoc.2013.04.009

[15] Mohajerzadeh, A. H. & Yaghmaee, M. H. (2009). An efficient energy-aware routing protocol for real-time traffic in wireless sensor networks. *International Conference on Ultra Modern Telecommunications & Workshops, ICUMT '09*, 1-9. https://doi.org/10.1109/ICUMT.2009.5345536

[16] Hu, X., Wang, J., & Wang, C. (2011). Mobility-adaptive routing for stable transmission in mobile ad hoc networks. *Journal of Communications, 6*(1), 79-86.

[17] Douglas, S. J., de Couto, D., Aguayo, J., & Bicket, R. M. (2003). A High-Throughput Path Metric for Multi-Hop Wireless Routing. *Wireless Networks, 11*(4), 419-434. https://doi.org/10.1145/938985.939000

[18] The Mathworks. (2009). *Fuzzy Logic Toolbox User's Guide*, The Mathworks Inc. Retrieved: September 10, 2009 from http://www.mathworks.com/access/helpdesk/help/pdf_doc/fuzzy/fuzzy.pdf.
[19] Alhumaidi Hanouf, M. (2016). Fuzzy weighted average approach to ranking projects in contractor initial bidding. *Fuzzy information processing Society, 2016 Annual conference of the North American, IEEE*, 1-7.
[20] El Alami, H. & Najid, A. (2017). Fuzzy Logic Based Clustering Algorithm for Wireless Sensor Networks. *International Journal of Fuzzy System Applications (IJFSA), 6*(4), 63-82. https://doi.org/10.4018/IJFSA.2017100105
[21] Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

Authors' contacts:

Ms. **Reena P. Pingale,** Assistant Professor
(Corresponding author)
Sinhgad College of Science,
S. No. 9/1/5 and 9/2/4, Off West. Bypass Highway,
Ambegaon (BK.), Pune - 411041, Maharashtra, India
reena.pingale@gmail.com

Dr. **S. N. Shinde,** Principal
C.M.C.S. College,
Udoji Maratha Boarding Campus, Near, Gangapur Road,
D. K. Nagar, Nashik, Maharashtra 422013, India
sns110@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210204153751

Original scientific paper

Spatial Correlation Based Clustering with Node Energy Based Multi-Hop Routing Scheme for Wireless Sensor Networks

Panchikattil Susheelkumar Sreedharan*, Dnyandeo Jageshwar Pete

Abstract: Major points of concern in implementing a wireless sensor network (WSN) are the network lifetime and energy utility within any delay tolerant network. Both these parameters define the success of the sensor network. The higher the expectancy of network Lifetime, the higher is the probability of acceptance of the network. Similarly, better the energy utilization in the network, better are the chances of success and implementation of the sensor network. Clustering is one such scheme adopted in WSN towards harnessing the best of above specified parameters for the network implemented. Most popular clustering techniques are the variants of LEACH protocol that facilitate cluster formation based on the proximity of an individual node to other nodes in the sensor network. These protocols are based on a single hop structure from the selected cluster heads in the network. This paper embarks on a multi-hop clustering algorithm that takes into consideration the spatial correlation between the nodes to form clusters and implements a highly energy efficient routing scheme which selects the multi-hop path in the network in a dynamic fashion.

Keywords: clustering; multi-hop; network lifetime; single-hop; spatial correlation

1 INTRODUCTION

Wireless Sensor Networks find its application in almost all fields of study like agriculture, environment monitoring, seismological study, defense and security and even extending into today's IoT [1]. It generally consists of high density sensor nodes that are spatially distributed in a given area of interest. Since these nodes are not easily humanly accessible, we have to be very particular about the energy utility in these nodes as its not possible to replace or recharge the batteries in them. Hence, many algorithms have adopted a clustering approach [2-11] with the cluster heads choosing a single hop data transfer mechanism to the sink. This mechanism is effective in case of small network area. But as the area of the sensor network grows in size, this particular mechanism can only support shorter network life time with reduced throughput. Standard algorithms like LEACH and its recent variants [4-15] are the examples of this particular approach. Here in this paper, we have proposed a cluster based multi-hop data transfer mechanism from the cluster head to the sink based on spatial correlation [16-23] between the scattered nodes. The routing path is arrived at dynamically based on the distance between the next possible hop satisfying a certain specified distance criterion and the minimum energy requirement to be satisfied by the next-hop cluster head. The performance of our proposed algorithm is compared with that of existing standard algorithm LEACH [24] and its recent variant Enhanced LEACH [7]. Therefore, here we have tried to implement an algorithm, which involves three phases .i.e. the clustering phase, cluster-head selection phase and the data transfer phase. The rest of the paper follows standard structure i.e. section II is dedicated to related work, section 3 & 4 covers the models used while section 5 details the methodology involved. Section 6 covers the algorithms implemented, section 7 details the findings and results and lastly, section 8 concludes the findings of the paper.

2 RELATED WORK

Reference [3] introduces the standard clustering algorithm LEACH wherein the authors R. Sinde et al. (2020) uses a randomized way of selecting cluster head and also supports the data fusion or aggregation within the cluster. Here the operation during each round of data transfer is divided into two phases namely the set-up phase and the data transfer phase. During the setup phase, the cluster head is selected along with the cluster formation. In addition, during the data transfer phase, the sensed data flows from the cluster member nodes to their respective cluster head and the cluster heads aggregates the received data from all its members and transfers the aggregated data to the sink in a single hop. However, here the drawback lies in the fact that the cluster head is chosen in a randomized way not taking into consideration the residual energy of the candidate nodes for cluster head function. In addition, the effective communication path between the cluster member nodes and the sink through the cluster head is not considered for cluster formation process. These drawbacks have been taken into account and addressed to some extent in the recent variations of LEACH algorithm to some extent. In [13], the authors D. Mahmood et al. (2013) have altered the selection of cluster head in such a way that the cluster head continues to be the head till it's energy falls below a required threshold level. In addition, as far as data communication is concerned, they have envisaged intra cluster communication, inter cluster communication and cluster-head to sink communication and accordingly they have chosen two levels of amplification. The choice of these levels can be made dynamic thereby not fixing the number of levels at two. In [15], Zhidong et al. (2018) have devised a new strategy to arrive at the best optimum number of clusters balancing the energy need between the clusters as well as within each cluster. In their proposed scheme, they have introduced the distance variance factor in the initial set-up phase, and effectively implemented the node dormancy to save energy. Again, as the network

area increases in size, the effectiveness of this algorithm also reduces since propagation of data involves two to three hops only. In [7], the authors Amer O. Abu Salem and Noor Shudifat (2019) have proposed an enhanced LEACH protocol, which takes into consideration the effective distance between the cluster members to the cluster head and from cluster head to the sink while implementing the cluster formation process. They have successfully shown that their scheme facilitates an extended network lifetime and reduced power usage in comparison to the standard protocol LEACH. But as the network area gets larger in size, even this protocol suffers due to higher levels of energy expended in transmitting the data to sinks which are very far located. Considering these shortcomings, our algorithms implements a multi hop mechanism for data propagation from the sensing nodes to the sink to have better energy efficiency and enhanced network lifetime.

3 PROPOSED CORRELATION MODEL

Sensing range or coverage area of any sensor node is a typical characteristic, which influences its region of sensing. Wireless Sensor networks are network of sensor nodes that typically consists of multiple nodes, which are more closely located. Since these nodes are closely located to each other in many applications, we can derive a spatial correlation between these nodes dictated by the overlapping region of coverage or the sensing region.

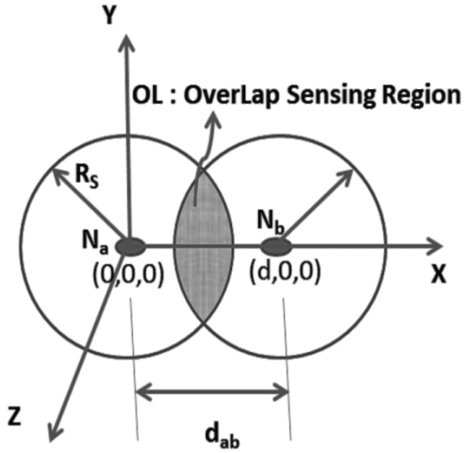


Figure 1 Over-lapped sensing region of sensor nodes [20]

Fig. 1 represents two nodes that are closely located to each other, separated by a distance d whose region of coverage is shown by the two circles with the nodes located at the Centre of each circle as depicted. We can observe that there is a region of overlap between the coverage areas of these two nodes that is highlighted by the shaded area. Now as the region of overlap increases the similarity in the data sensed by these nodes increases. Depending of the specific application requirement, we can have a trade-off between the data accuracy and the sensor nodes selected for communicating their sensed data instead of every node communicating their data to the sink. Hence, we try to make use of this spatial correlation between the sensor nodes in the development of our algorithm wherein clustering or cluster formation also takes into account the spatial correlation between these nodes.

Fig. 1 represents a two dimensional model of the sensor node, its coverage area and the region of overlap between their coverage region [25]. In real time the regions correlate to a three dimensional structure wherein each region represents a specific volume. Here, we define spatial correlation between two sensor nodes by the amount of overlap in their coverage area. We can say that, in a homogeneous sensor network, spatial correlation exists if the distance of separation between the two sensor nodes is less than twice the sensing radii of each node, which is quite evident from the study of geometry of the figure represented above. Moreover, as the distance between these two nodes increases beyond twice the sensing radii of the nodes, we say that there is no spatial correlation between these two nodes. In addition, the two nodes are 100% spatially correlated if they are co-located i.e. the distance of separation between them becomes zero. Considering these facts, we can define the overlap coefficient of correlation [29] between any two nodes N_a and N_b as:

$$\sigma_{ab} = \frac{\text{Volume of Overlap - of - Sensing Regions of Nodes } N_a, N_b}{\text{Combined Volume - of - Sensing Regions of Nodes } N_a, N_b}$$

$$\sigma_{ab} = \frac{V_{OL}}{V_{Comb}} \tag{1}$$

where V_{OL} and V_{Comb} is the volume of the overlapping region of sensing of the two nodes and combined volume of sensing regions of the two nodes considered respectively.
We also define the region of overlap (ROL) as:

$$(\%ROL)_{ab} = \sigma_{ab} \times 100 \tag{2}$$

Using spherical geometry as given in [26], the numerator i.e. the dividend and denominator which is the divisor in Eq. (1) is expressed as:

$$V_{OL} = \frac{\pi}{12}\left(2SR - d_{ab}\right)^2\left(d_{ab} - 4SR\right) \tag{3}$$

$$V_{Comb} = \frac{8\pi \times SR^3}{3} - \frac{\pi}{12}\left(2SR - d_{ab}\right)^2\left(d_{ab} + 4SR\right) \tag{4}$$

Combining (1), (3) and (4), we express the overlap coefficient of correlation between the two nodes as:

$$\sigma_{ab} = \frac{\left(2SR - d_{ab}\right)^2\left(d_{ab} + 4SR\right)}{32SR^3 - \left(2SR - d_{ab}\right)^2\left(d_{ab} + 4SR\right)} \tag{5}$$

where SR is assumed to be the uniform node sensing range, N_a, N_b are the two nodes considered at a distance dab from each other.

Thus, it can be expressed that $\sigma_{ab} = 0$ for any two nodes separated by a distance greater than twice the sensing range

of each node i.e. 2*SR*. Hence, we can sum-up the *SR* correlation coefficient between two nodes separated by a distance d as:

$$\sigma = \left\{ \begin{array}{cc} \dfrac{(2SR-d)^2(d+4SR)}{32SR^3-(2SR-d)^2(d+4SR)} & \text{if} \quad 0 \le d < 2SR \\ 0 & \text{if} \quad d \ge 2SR \end{array} \right\} \quad (6)$$

Expression (6) represents the sensing region correlation model [20].

4 ENERGY MODEL

The standard reference energy model is taken from ref. [9] which gives a detailed analysis of two main types of energy associated with WSN namely the propagation energy which is the energy involved in the transmission of data from the sensor nodes to the sink and energy used for various electronics involved before transmission after reception of data in the nodes. The propagation distance decides whether the propagation energy is influenced by free space propagation model or multi-path propagation model. If the propagation distance is less than the crossover distance it is the free space propagation that defines the propagation energy and if the propagation distance is beyond the crossover distance, the propagation energy is defined by the multi space propagation model as given in ref. [9, 26, 27]. Using this model, the energy involved in the transmission of l-bit message is expressed as:

$$E_{\mathrm{T}} = l \times E_{\mathrm{EX}} + l \times p_{\mathrm{l}} \times d_{\mathrm{T}}^{(n)} \quad (7)$$

where E_{EX} is the energy/bit involved in the electronics, $l \times p_{\mathrm{l}} \times d_{\mathrm{T}}^{(n)}$ is the propagation-energy for l-bit message for covering a transmission distance d with a propagation loss exponent n.

Ref. [26] states for any transmission distance less than the crossover distance d_0, the free space model is applied and the expression for the transmission energy is given as:

$$E_{\mathrm{T1}} = l \times E_{\mathrm{EX}} + l \times \varepsilon_{\mathrm{fs}} \times d_{\mathrm{T}}^{(2)} \quad (8)$$

where $\varepsilon_{\mathrm{fs}} = p_{\mathrm{l}}$, i.e. the propagation loss for free space.

And for any transmission distance greater than the crossover distance d_0, the multi-path model is applied and the transmission energy is stated as:

$$E_{\mathrm{T2}} = l \times E_{\mathrm{EX}} + l \times \varepsilon_{\mathrm{mp}} \times d_{\mathrm{T}}^{(4)} \quad (9)$$

where $\varepsilon_{\mathrm{mp}} = p_{\mathrm{l}}$, i.e. the propagation loss in multi-path transmission. Here cross-over distance d_0 is given by:

$$d_0 = \sqrt{\frac{\varepsilon_{\mathrm{fs}}}{\varepsilon_{\mathrm{mp}}}} \quad (10)$$

5 PROPOSED METHOD

Here we have selected a wireless sensor network area of 200 by 200 m^2 with a dense spread of sensor nodes in the given area. The total number of nodes are 1000 that are initially given a random spread across the network. The initial random locations of all these nodes are maintained the same throughout the simulation for varying parameters and algorithms under study in this paper. The sink is assumed to be centrally located and aware of all the GPS enabled node's location detail. The network is supposed to be a static homogeneous network that is the nodes are stationary in nature and uniformly energized at the beginning of implementation. All the nodes transmits their location detail to the sink and the sink implements the centralized clustering algorithm to form cluster groups based on the spatial correlation criterion chosen which is maintained throughout the network lifetime. The cluster information is relayed by sink to all the nodes which is now aware of its cluster details that is member ids and member locations. All the nodes keep a record of its neighboring nodes detail that are separated by a maximum specified distance (of 5*SR*) from itself which is the limitation placed for next hop node. Thus cluster formation phase is only implemented once in the lifetime of network. After the cluster formation phase, then we implement a cluster head election phase through a distributed algorithm implemented at the individual nodes in the network. This algorithm estimates a fitness function value for the node based on the leftover energy after the previous rounds of data transfer, its distance from sink and the sum of propagation distance from other member nodes to the candidate CH node within the cluster. Each node relays this value to its other cluster members and the node with largest valued fitness function is chosen as the cluster head for the cluster, thus presenting a dynamic CH selection process. The CHs then relay their election as CH to other member nodes as well as other CHs located nearby which are in the range of 5*SR* from itself. Initially all the CHs are by-default the main cluster heads. Few amongst these cluster heads that satisfy the defined minimum spatial correlation factor of 0.1 with their neighboring cluster head and whose fitness value is lower than neighbor cluster head are labeled as secondary cluster head. Now, here all the cluster members that satisfy a minimum spatial correlation factor of 0.1 with its cluster head are made to sleep, thereby avoiding any redundant data as defined in the application requirement based on the tradeoff chosen between data accuracy and number of active sensing nodes transmitting their sensed data i.e. energy conservation. All the members of associated secondary cluster are also checked for the satisfaction of minimum spatial correlation factor with the associated main cluster head too and are either made dormant or live depending on the conditions prevalent. Immediately after the segregation of the cluster heads as main or secondary cluster heads, each of the main cluster head finds the next hop node (which is also a designated main cluster head) to chart out the route forward towards the sink. These next hop nodes are selected only if they satisfy the minimum and maximum propagation distance fixed for being eligible which is between 2*SR* and 5*SR*. All these

intermediate routing nodes performs the role of a repeater for other main cluster head's relayed packets. In the absence of finding a suitable intermediate routing node, the cluster heads are programmed to directly send the message to the sink. Speaking of the data flow in the network, the secondary cluster heads will remove the redundant data received from its cluster members and forwards the aggregated data to the associated main cluster head for further transmission. The main cluster head also aggregates the data received from its cluster members and the associated secondary cluster heads and relays the aggregated data either to the sink directly or to the next hop cluster head node as estimated by the distributed algorithm running at the cluster head node itself. The sinks keeps a record of all the dead nodes or renders any node as dead if it does not receive any information from that node in the aggregated packet message relayed by its main cluster head in three consecutive rounds of data transfer. The sink is aware of the dead nodes and therefore has an estimate of the total number of packets to be received from the network. On receiving messages from all the active clusters or waiting for pre-specified buffer-time, it relays the start of next round to all nodes in the network. All the cluster heads on receiving the round-start message from the sink drops the data if any received by it and pending for further relay and starts initial phase of estimating the fitness function value for the next round under consideration, further followed up by the subsequent stages as defined in the implementing distributed algorithm at the node. The entire implementation is simulated using MATLAB@2016a wherein the cluster formation is carried out for various values of Spatial correlation factor ranging from 0.1 to 1 (one). Here we have chosen five values of spatial correlation factor i.e. 0.1, 0.2, 0.3, 0.5 and 0.7 for simulation purpose and studied the effect of the variation in spatial correlation coefficient on the network lifetime and throughput in our proposed approach along with the comparison of each results (specific values of correlation coefficient) with standard Algorithm namely LEACH and its recent variant called Enhanced LEACH.

6 ALGORITHM

The proposed scheme is implemented using a combination of centralized algorithm run at the sink and a distributed algorithm run at each node. The centralized algorithm is taken from paper [28, 29] that facilitates the cluster formation at the sink side and then relaying the cluster information to the sensor nodes in the wireless sensor network area. The distributed algorithm accomplishes three tasks namely the cluster head selection and designating them as main cluster head or secondary cluster head followed by the next hop node estimation and finally the data transfer phase from the sensing node to the sink. The centralized cluster forming algorithm is shown in the flowchart in Fig. 2:

The cluster formation process is followed by the distributed algorithm that accomplishes the three tasks mentioned earlier namely the cluster head selection and designating them as main cluster head or secondary cluster head followed by the next hop node estimation and finally the data transfer phase. The distributed algorithm implemented at the node level is depicted in the flow chart given in Fig. 3. The fitness-function (*ff*) value is expressed using the following expression [28, 29]:

$$ff = \left(\frac{W_a \times N_{LOE} + W_b \left(\frac{1}{SPD^2} \right)}{N_{LOE} + \left(\frac{1}{SPD^2} \right)} \right) \tag{11}$$

where *ff* represents fitness-function value for the node supporting its candidature for the role of CH, LOE represents its left-over energy (residual energy) after the previous round, W_a and W_b are the proportional weights expressed as:

$$W_a = \left(\frac{D_{NS}}{D_{NS} + SPD} \right), W_b = \left(\frac{SPD}{D_{NS} + SPD} \right) \tag{12}$$

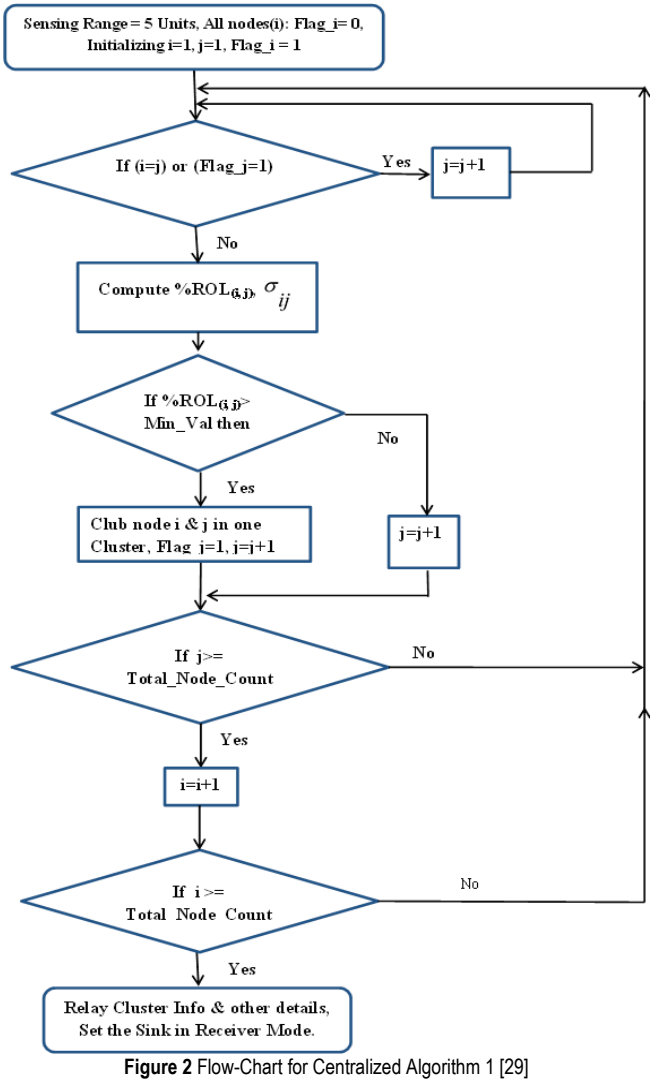


Figure 2 Flow-Chart for Centralized Algorithm 1 [29]

Here D_{NS} represents distance between the node and sink, *SPD* represents sum of propagation distance. *SPD* is the total propagation distance encountered in a cluster owing to the

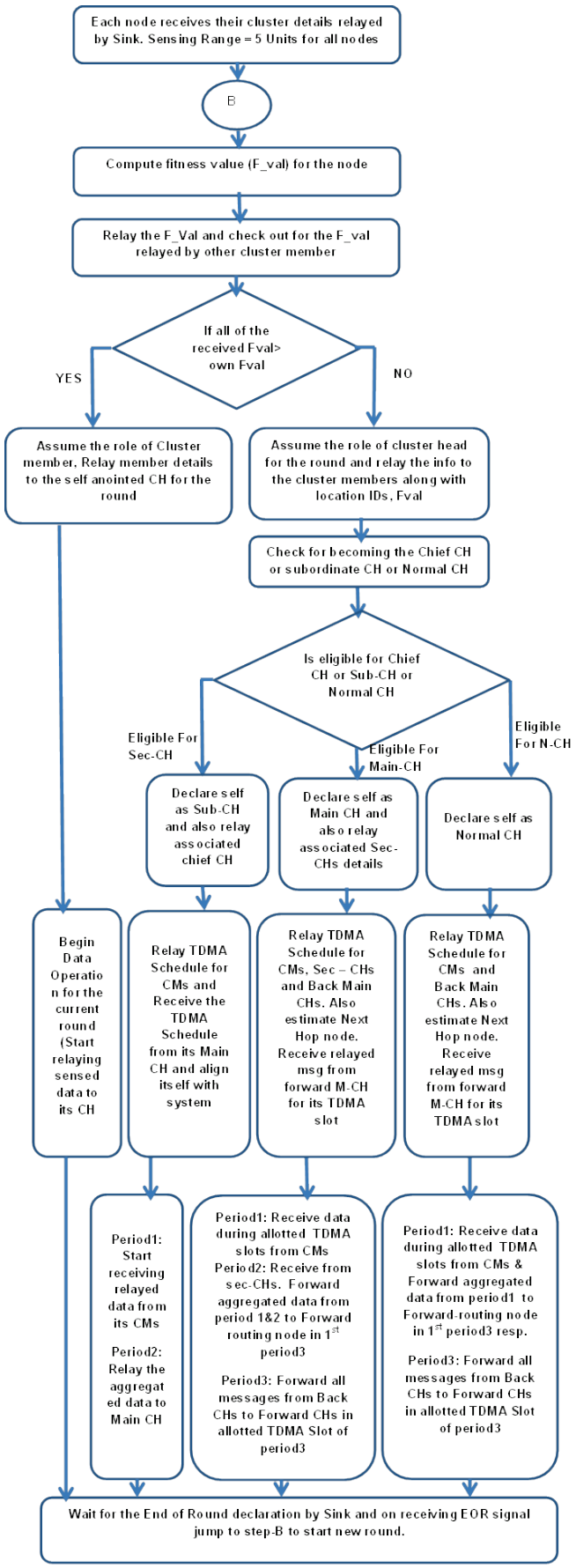


Figure 3 Flow-Chart for Algorithm 2

relay of messages from the cluster member to the candidate cluster head in the cluster for the round under consideration. W_a and W_b are dynamic proportional weights used to estimate the fitness function of the candidate node for cluster head election taking into consideration both the LOE of the node and the expendable energy in the cluster by the member nodes for intra-cluster communication (which should be minimum) proportionately. If *SPD* is maximum, that means more energy needs to be expended cumulatively by the cluster members and if *SPD* is minimum less energy needs to be expended cumulatively by the cluster members. Hence, ideally when *SPD* is lower, *ff* value should be higher and when *SPD* is higher, *ff* value should be lower which is satisfied by Eq. (10). Using the fitness function value, the distributed algorithm enables the selection of main cluster heads, secondary cluster heads and the further routing paths for the round under consideration. Each cluster head maintains a TDMA schedule for each of its cluster members for message transfers. The first period of each round is solely dedicated for cluster members to relay their data to their cluster heads. The second period is for secondary cluster heads to relay their data to the main cluster heads using a TDMA schedule dictated by its main cluster head. The third period is for the movement of aggregated data from the main cluster heads to their first forward routing node. A TDMA Schedule is also maintained by the forward (towards sink) main cluster head for all its neighboring back main cluster head (which are further away from the sink) that are in the range of 5*SR*. The third period is repeated till sink signals the end of round. Thus, using the defined TDMA schedules the data flow is simulated from the sensing nodes to the sink through various intermediary routing nodes. The sink receives the forwarded messages from various clusters and regularly checks for the reception of messages from all the active clusters. If the sink has received the aggregated data message from all the active clusters, it signals the end of round otherwise it waits for a specific period of time which is more than sufficient for the message from the farthest cluster to reach the sink before it signals the end of round to all the cluster members.

7 RESULT

In the simulation, we have enforced the above parameter values as specified in Tab. 1. We have implemented the simulation with varying degrees of correlation value chosen for clustering. Here the correlation value is varied between 0 and 100 percent and typically for our simulation we have selected a correlation value of 0.7, 0.5, 0.3, 0.2, and 0.1 (i.e. 70%, 50%, 30%, 20% and 10%) as the sample correlation values for which we have implemented our multi-hop routing algorithm. Under each correlation value, we have three instances of simulation with a minimum node energy parameter value set to each node at $10E_{min}$, $500E_{min}$, 1000Emin respectively to support multi-hop routing of data. Here E_{min} is taken as 0.0002 Joules and hence $10E_{min}$, $500E_{min}$ and $1000E_{min}$ which corresponds to a minimum nodal residual energy of 0.002 Joules, 0.1 Joules and 0.2 Joules respectively set for each node to participate in multi-hop

routing during each of the three instances of simulation respectively. The results for the various instances of simulations are represented below graphically. The results are expressed in terms of network energy utilization, throughput for the various algorithms under consideration and the network life-time reflected for all the algorithms under study. The network life-time is assumed to be the round of data transfer supported by the respective algorithm till the death of 70% of the sensing nodes present in the network. And the maximum throughput of the networks is taken as the total number of message packets arriving at the sink during the life-time of network. For the purpose of comparison in regards with the network energy consumption, we have taken the reference of LEACH and Enhanced-LEACH algorithm's network lifetime and it is observed that the network energy is depleted in case of LEACH and enhanced-LEACH before data transfer round number 1500. The simulation results encompasses the results for the standard algorithm LEACH, its advanced version Enhanced-LEACH and three instances of our proposed algorithm with varying values of correlation fixed for Spatial correlation based clustering.

Table 1 Parameters taken for simulation (@MATLAB 2016a)

No	Design Para-meters	Value/ Symbol
1	Total Sensor Nodes	1000
2	WSN Area	200×200 m^2
3	Initial energy of all sensor nodes	0.5 J
4	Sensing Range (SR) of Node:	5 m
5	Free-Space factor for propagation distance less than cross-over distance d_0 $(d < d_0)$	10 nJ/bit/m^2
6	Multi path factor for longer distance $(d > d_0)$	0.0013 pJ/bit/m^4
7	Energy required for reception & transmission of signals by Electronics involved	50 nJ/bit
8	Energy expended for data aggregation	5 nJ/bit
9	No of bits per message packet	4000 bits

The simulation for a correlation value of $\sigma = 0.7$ is depicted below with the three instances for the three minimum nodal energy parameter value set in our proposed algorithm. The graphical results of the three instances of simulation for our proposed algorithm with the set parameter values along with the simulation results for existing standard LEACH and Enhanced LEACH protocol is given below in Fig. 4(a), Fig. 4(b) and Fig. 4(c).

Fig. 4(a) represents the Network's energy utilization curve with respect to the rounds of data transfer rounds supported by the network for standard algorithms LEACH and Enhanced-LEACH in comparison to the three instances of our proposed algorithm with $\sigma = 0.7$.

Fig. 4(b) represents the Network's throughput curve with respect to the rounds of data transfer supported for all the existing protocols and instances of our proposed algorithm with $\sigma = 0.7$ while Fig. 4(c) represents the comparative network lifetime with respect to the rounds of data transfer supported for all the existing protocols and instances of our proposed algorithm with $\sigma = 0.7$.

Similarly, we have implemented the simulation for varying values of correlation and the graphical results for the same are presented below. For correlation value of $\sigma = 0.5$, we have the following results.

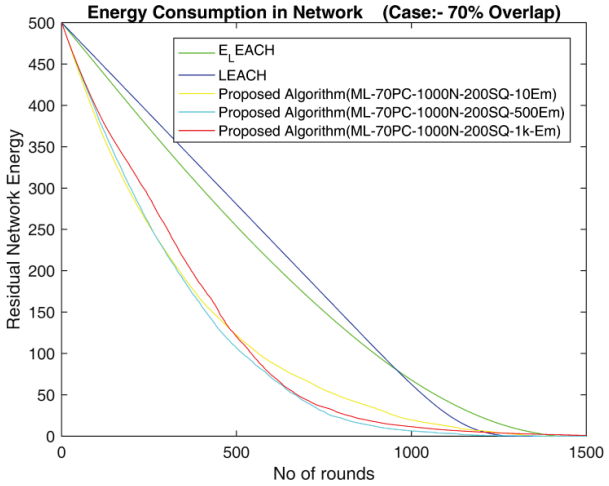


Figure 4(a) Network Energy Utilization for LEACH, Enhanced LEACH and three instances of our proposed Algorithm ($\sigma = 0.7$)

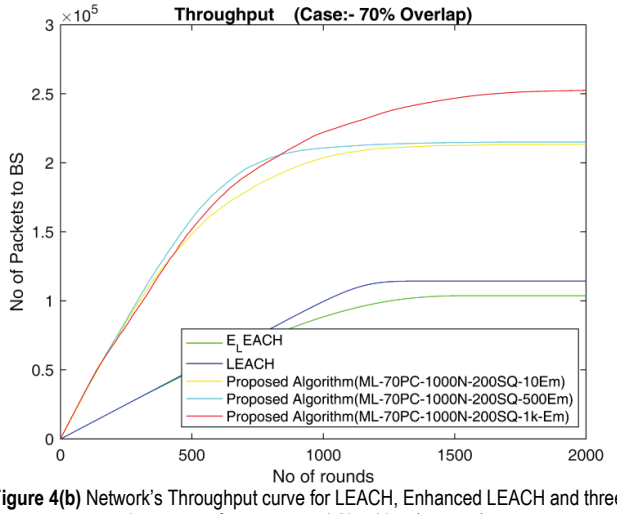


Figure 4(b) Network's Throughput curve for LEACH, Enhanced LEACH and three instances of our proposed Algorithm ($\sigma = 0.7$)

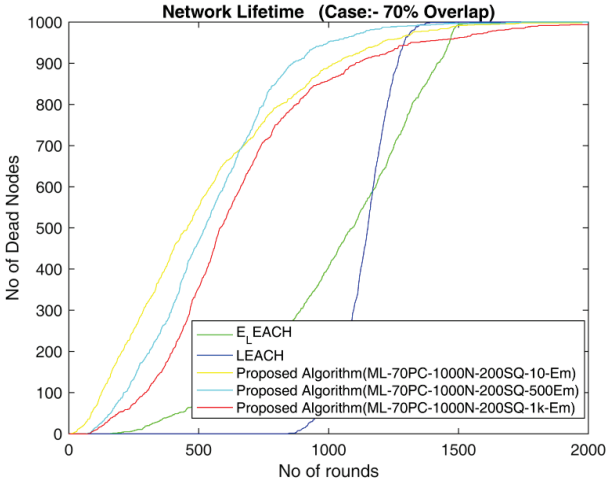


Figure 4(c) Network Lifetime for LEACH, Enhanced LEACH and three instances of our proposed Algorithm ($\sigma = 0.7$)

Fig. 5(a) represents the Network's energy utilization curve with respect to the rounds of data transfer rounds supported by the network for standard algorithms LEACH

and Enhanced-LEACH in comparison to the three instances of our proposed algorithm with $\sigma = 0.5$.

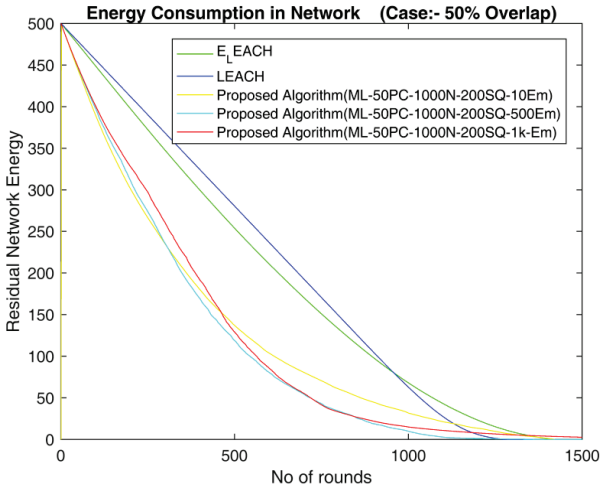


Figure 5(a) Network Energy Utilization for LEACH, Enhanced LEACH and three instances of our proposed Algorithm ($\sigma = 0.5$)

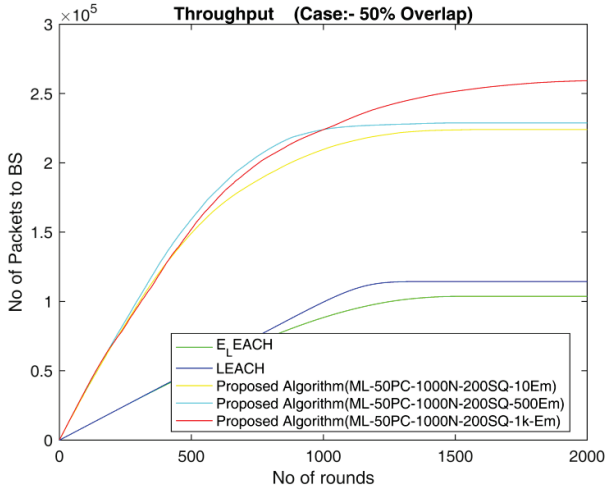


Figure 5(b) Network's Throughput curve for LEACH, Enhanced LEACH and three instances of our proposed Algorithm ($\sigma = 0.5$)

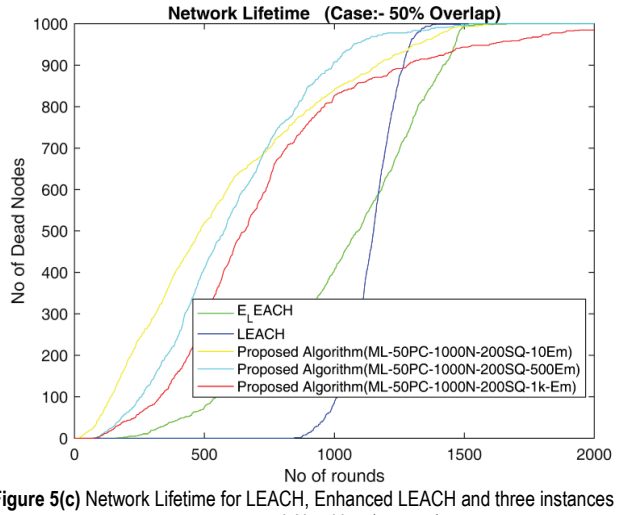


Figure 5(c) Network Lifetime for LEACH, Enhanced LEACH and three instances of our proposed Algorithm ($\sigma = 0.5$)

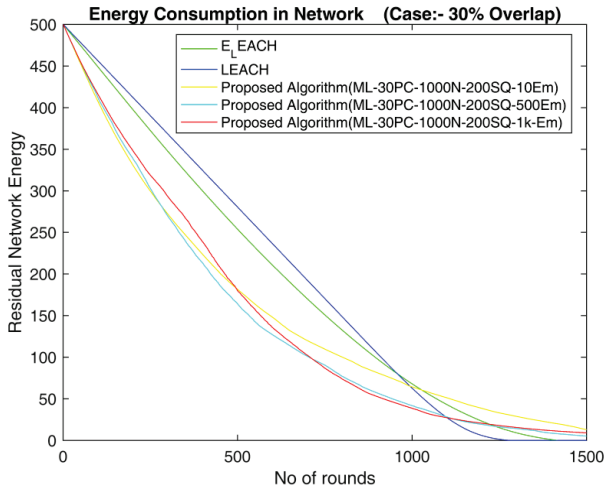


Figure 6(a) Network Energy Utilization for LEACH, Enhanced LEACH and three instances of our proposed Algorithm ($\sigma = 0.3$)

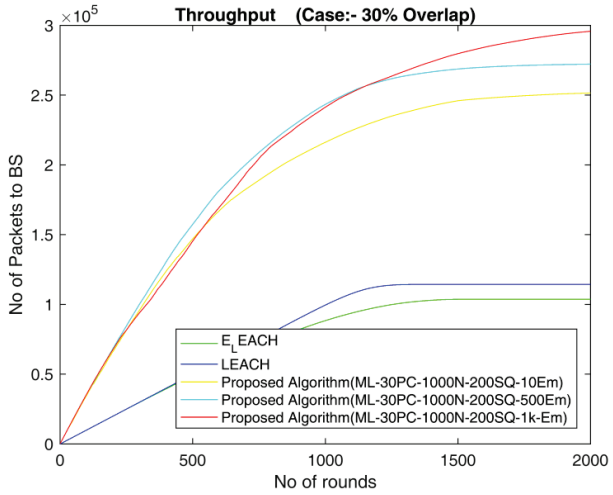


Figure 6(b) Network's Throughput curve for LEACH, Enhanced LEACH and three instances of our proposed Algorithm ($\sigma = 0.3$)

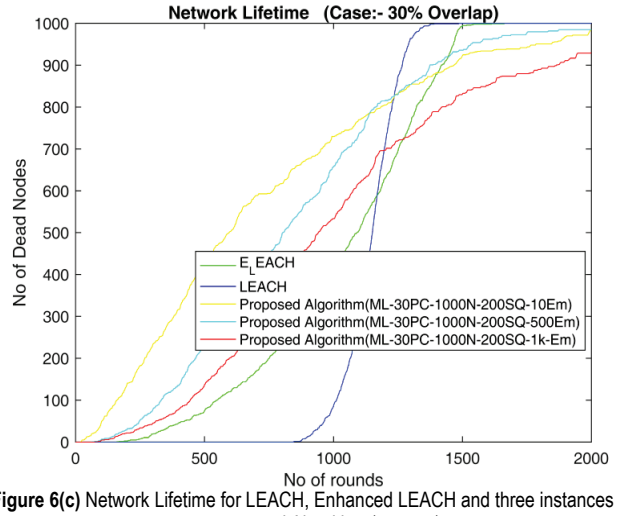


Figure 6(c) Network Lifetime for LEACH, Enhanced LEACH and three instances of our proposed Algorithm ($\sigma = 0.3$)

Fig. 5(b) represents the Network's throughput curve with respect to the rounds of data transfer supported for all the existing protocols and instances of our proposed algorithm

with σ = 0.5 while Fig. 5(c) represents the comparative network lifetime with respect to the rounds of data transfer supported for all the existing protocols and instances of our proposed algorithm with σ = 0.5.

For correlation value of σ = 0.3, we have the following results:

Fig. 6(a) represents the Network's energy utilization curve with respect to the rounds of data transfer rounds supported by the network for standard algorithms LEACH and Enhanced-LEACH in comparison to the three instances of our proposed algorithm with σ = 0.3.

Fig. 6(b) represents the Network's throughput curve with respect to the rounds of data transfer supported for all the existing protocols and instances of our proposed algorithm with σ = 0.3 while Fig. 6(c) represents the comparative network lifetime with respect to the rounds of data transfer supported for all the existing protocols and instances of our proposed algorithm with σ = 0.3.

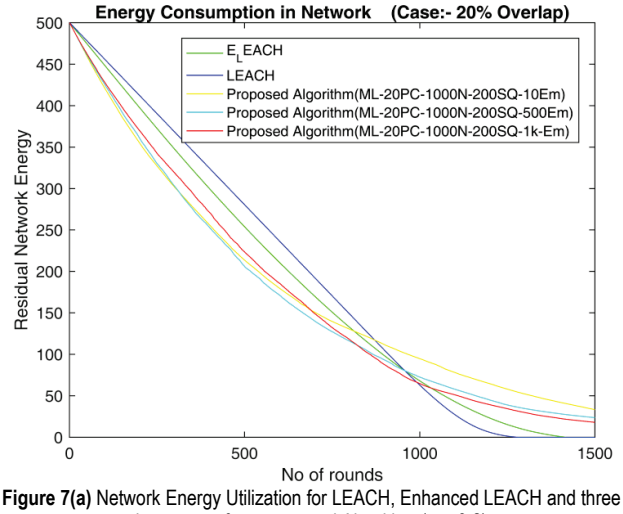


Figure 7(a) Network Energy Utilization for LEACH, Enhanced LEACH and three instances of our proposed Algorithm (σ = 0.2)

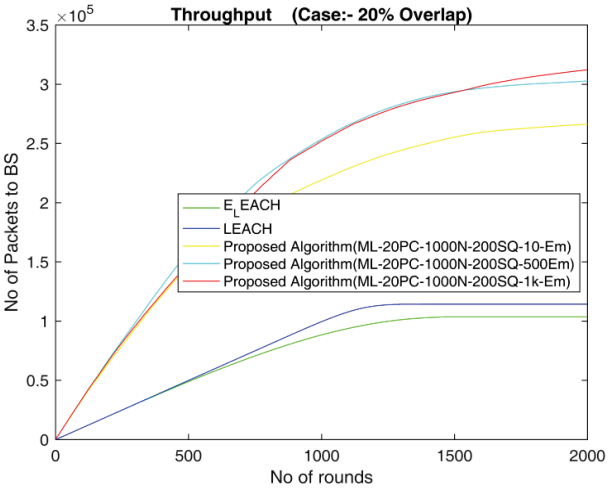


Figure 7(b) Network's Throughput curve for LEACH, Enhanced LEACH and three instances of our proposed Algorithm (σ = 0.2)

For correlation value of σ = 0.2, we have the following results:

Fig. 7(a) represents the Network's energy utilization curve with respect to the rounds of data transfer rounds supported by the network for standard algorithms LEACH and Enhanced-LEACH in comparison to the three instances of our proposed algorithm with σ = 0.2.

Fig. 7(b) represents the Network's throughput curve with respect to the rounds of data transfer supported for all the existing protocols and instances of our proposed algorithm with σ = 0.2 while Fig. 7(c) represents the comparative network lifetime with respect to the rounds of data transfer supported for all the existing protocols and instances of our proposed algorithm with σ = 0.2.

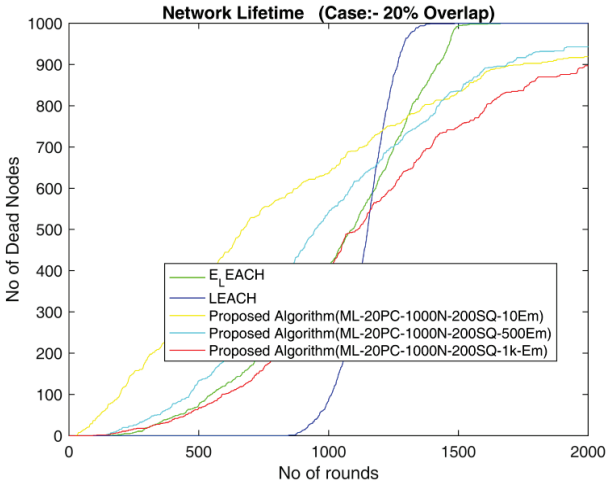


Figure 7(c) Network Lifetime for LEACH, Enhanced LEACH and three instances of our proposed Algorithm (σ = 0.2)

For correlation value of σ = 0.1, we have the following results:

Fig. 8(a) represents the Network's energy utilization curve with respect to the rounds of data transfer rounds supported by the network for standard algorithms LEACH and Enhanced-LEACH in comparison to the three instances of our proposed algorithm with σ = 0.1.

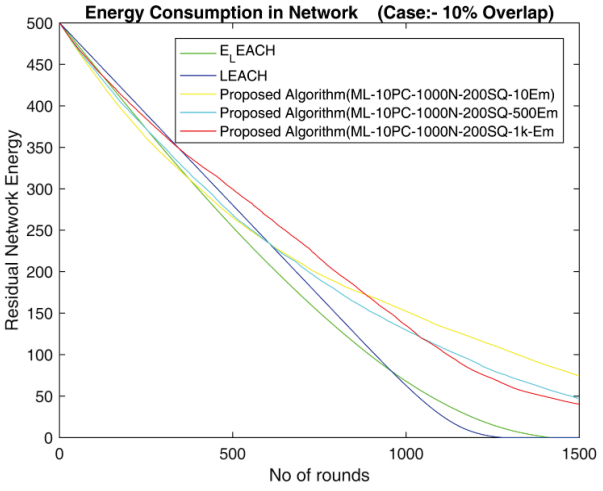


Figure 8(a) Network Energy Utilization for LEACH, Enhanced LEACH and three instances of our proposed Algorithm (σ = 0.1)

Fig. 8(b) represents the Network's throughput curve with respect to the rounds of data transfer supported for all the

existing protocols and instances of our proposed algorithm with σ = 0.1 while Fig. 8(c) represents the comparative network lifetime with respect to the rounds of data transfer supported for all the existing protocols and instances of our proposed algorithm with σ = 0.1.

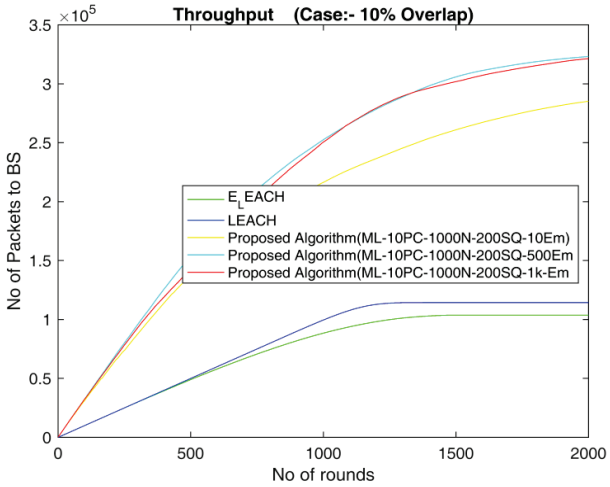


Figure 8(b) Network's Throughput curve for LEACH, Enhanced LEACH and three instances of our proposed Algorithm (σ = 0.1)

The above graphical results are represented in tabular form in Tab. 2 and Tab. 3. Tab. 2 represents the output values corresponding to the results obtained during the simulation of existing algorithms standard LEACH and Enhanced-LEACH. Tab. 3 on the other hand represents the various output values of the various instances during the simulation of our proposed algorithm with σ = 0.7, 0.5, 0.3, 0.2 and 0.1.

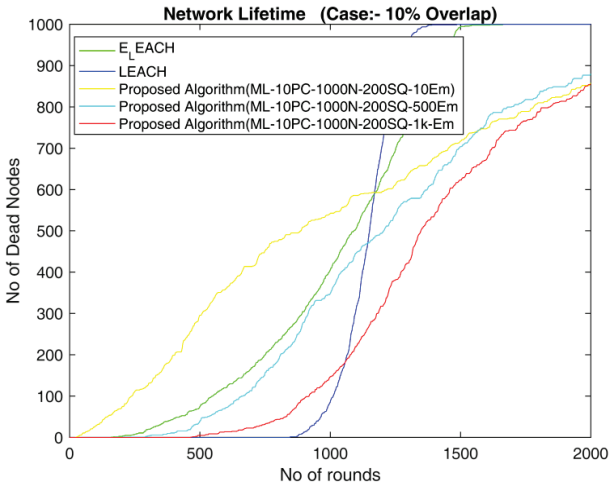


Figure 8(c) Network Lifetime for LEACH, Enhanced LEACH and three instances of our proposed Algorithm (σ = 0.1)

Table 2 Parameters observed during simulation of LEACH, ENHANCED LEACH algorithms

Parameters/ Algorithms	1000-Sensor Nodes in 200×200 m²	
	LEACH	E-LEACH
Clusters formed Roundwise between:	1 to 128	1 to 134
1st Node Death (ND) Round	845	160
10% ND Round	1011	546
20% ND Round	1065	750
50% ND Round	1152	1091
70% ND Round (LifeTime)	1198	1252
Throughput	112795	100009
Balance Network Energy after 1500 rounds	0	0

Table 3 σ and corresponding results for 1000 Nodes spread in a WSN Area of 200×200 m²

σ	0.7			0.5			0.3			0.2			0.1		
E_{min} Set for Forwarding Agent	E_{min} = 0.002 J	E_{min} = 0.1 J	E_{min} = 0.2 J	E_{min} = 0.002 J	E_{min} = 0.1 J	E_{min} = 0.2 J	E_{min} = 0.002 J	E_{min} = 0.1 J	E_{min} = 0.2 J	E_{min} = 0.002 J	E_{min} = 0.1 J	E_{min} = 0.2 J	E_{min} = 0.002 J	E_{min} = 0.1 J	E_{min} = 0.2 J
No of Clusters present	953	953	953	847	847	847	634	634	634	522	522	522	374	374	374
1st ND Round	9	68	74	9	74	74	21	64	74	27	93	93	23	175	465
10% ND Round	126	212	293	139	256	331	161	332	445	197	466	590	237	659	917
20% ND Round	205	311	399	217	354	439	269	466	596	321	627	798	404	810	1077
50% ND Round	457	524	582	484	576	656	595	803	941	671	948	1113	885	1205	1348
70% ND Round	674	670	740	750	737	816	952	1050	1205	1130	1241	1395	1454	1493	1621
Throughput (K)	175.48	191.02	194.92	187.30	202.94	207.06	211.74	248.28	260.44	231.42	278.97	287.72	257.82	305.42	308.29
Balance Energy	0 J	0 J	1.04 J	0 J	0 J	2.79 J	13.1 J	5.36 J	9.24 J	33.4 J	23.7 J	18.1 J	74.5 J	47.1 J	40.5 J
	Balance Energy in the network after 1500 rounds of data transfer in each of the algorithms considered														

The step-wise analysis of the above results is given below:

Here we have taken the life-time of networks as the time or the round till the number of active sensor nodes in the network falls below 30% of the actual nodes present at the beginning. In other words, it is the time or round till which the number of dead nodes touches or surpasses 70% of the total number of nodes present in the network.

7.1 Observation-1

From Tab. 2, it is clear that standard LEACH protocol supports 1198 rounds of data transfer with a throughput of

1,12,795 packets while Enhanced-LEACH supports 1252 rounds of data transfer with a throughput of 1,00,009. Thus we observe that Enhanced-LEACH gives a better network lifetime in comparison to Standard LEACH protocol but its throughput is reduced in comparison to Enhanced-LEACH protocol.

7.2 Observation-2

From Tab. 3, it is observed that for $\sigma = 0.7$ and $\sigma = 0.5$, our algorithm is not at all competitive with either LEACH or Standard LEACH is concerned as far as network-lifetime is concerned. But as you decrease the value of σ to 0.2 and 0.1, we observe that when we increase the minimum residual energy required for all the sensor nodes to participate in the data routing process as a forwarding agent, there is a measurable improvement in the network lifetime which is better than that observed with standard LEACH and Enhanced LEACH. For the simulation of our proposed algorithm with $\sigma = 0.2$ and the minimum node residual energy of 0.1 J, the network lifetime observed is 1241 almost equal to that observed in LEACH with 1252 rounds of data transfer. But when minimum node residual energy (E_{min}) is fixed at 0.2 J, the network lifetime in our proposed algorithm increases drastically to 1395 rounds of data transfer.

Similarly the simulation of our algorithm with $\sigma = 0.1$ with various values of minimum residual energy set for each node to be allowed to participate in the data routing process shows much better network lifetime than that observed with LEACH and Enhanced LEACH simulation. The network lifetime for our proposed algorithm with $\sigma = 0.1$ and a minimum data-routing-participating residual energy for nodes set at 0.002 J, 0.1 J and 0.2 J is at 1454, 1493 and 1621 respectively which reflects a drastic improvement over the measurements seen in LEACH and Enhanced LEACH.

7.3 Observation-3

As far as throughput measurements are concerned, it is reflected from Tab. 2 and Tab. 3 that throughput in our proposed algorithm increases from 1,75,487 packets to a maximum of 3,08,290 packets for correlation value of $\sigma = 0.7$ to $\sigma = 0.1$ which are comparatively much higher values than that observed with the simulation of LEACH and Enhanced LEACH each with a throughput value of 1,12,795 and 1,00,009 respectively.

7.4 Observation-4

As far as the energy utility parameter is concerned, it is seen from the graphs in Fig. 4(a) and Fig. 5(a) that for $\sigma = 0.7$ and $\sigma = 0.5$, energy consumption in LEACH and Enhanced LEACH is better than all the instances of proposed Algorithm. It is observed that the energy in the network is exhausted before round number 750 in all instances of simulation using our proposed algorithm with $\sigma = 0.7$ and $\sigma = 0.5$ which is approximately half the Network-Lifetime observed in LEACH and Enhanced LEACH.

But as the value of σ is further decreased, we observe a gradual improvement in Network Lifetime using our proposed algorithm over LEACH and Enhanced LEACH. It is observed that for $\sigma = 0.3$, 0.2 and 0.1 even after round number 1500, which is taken as a reference, there is network energy balance in the network and the amount of network energy balance in the network increases as the value of σ decreases from 0.3 to 0.1, thereby increasing the network-lifetime.

For $\sigma = 0.2$ and $\sigma = 0.1$, we observe the best performance for our proposed algorithm with reference to the performance of LEACH and Enhanced LEACH, wherein even after 1500 rounds of data transfer, there is a network energy balance ranging from 18 J to 74 J respectively for the various instances considered for σ. This improvement in network energy consumption is reflected in the enhanced lifetime of the network using our proposed algorithm having $\sigma = 0.1$ with a network lifetime of 1454 rounds, 1493 rounds, 1621 rounds for the three instances of our algorithm defined by $E_{min} = 0.002$, 0.1 and 0.2 respectively which is much higher than the network lifetime seen in LEACH and Enhanced LEACH. Similarly for $\sigma = 0.2$, the improvement in network energy consumption is seen from the observed data represented in Tab. 2 and Tab. 3. In this case of our proposed algorithm we arrive at a network lifetime of 1130 rounds, 1241 rounds, 1395 rounds associated with the three instances of our algorithm defined by $E_{min} = 0.002$, 0.1 and 0.2 respectively. Thus the observed lifetime in this case with $E_{min} = 0.2$ shows drastic improvement over LEACH and Enhanced LEACH.

8 CONCLUSION

From the above findings presented in section 7, we come to the following conclusions:

For $\sigma = 0.7$ and 0.5, the results achieved are inferior to standard LEACH and Enhanced LEACH algorithms as far as network lifetime is concerned. But throughput is better using our proposed algorithm which is greater than LEACH and Enhanced LEACH by about 1.2 to 1.4 times.

But as we decrease the value of σ to 0.3, we see throughput increases from 1.8 to 2.3 times the throughput observed using LEACH and Enhanced LEACH. But the network Lifetime lags behind LEACH and Enhanced LEACH very narrowly.

Further as we decrease the value of σ to 0.2, we can conclude from the findings and results that the throughput is enhanced to around 2.05 to 2.5 times the throughput observed in LEACH and Enhanced LEACH. As far as the network lifetime is concerned, the lifetime is almost comparable to that achieved in Enhanced LEACH in case of second instance ($E_{min} = 0.1$ J) of our Proposed Algorithm. Similarly for the third instance ($E_{min} = 0.2$ J) of our proposed algorithm, the network lifetime surpasses the lifetime as seen with Enhanced LEACH by around 150 rounds.

Lastly when the value of σ is decreased to 0.1, from the findings and results of the previous section we can conclude that the throughput using our proposed algorithm is enhanced to around 2.28 to 2.73 times the throughput available with LEACH and Enhanced LEACH. And as far as the network

lifetime is concerned, our proposed algorithm gives a better network lifetime which shows a rise of around 200 to 350 additional rounds of data transfer in comparison to standard algorithms LEACH and Enhanced LEACH which converges to better energy utilization in our proposed algorithm.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

9 REFERENCES

[1] Sohraby, K., Minoli, D., & Znati, T. (2007). *Wireless Sensor Networks: Technology, Protocols and Applications*. John Wiley and Sons, Hoboken, New Jersey, 38-71. https://doi.org/10.1002/047011276X

[2] Taqieddin, E., Awad, F., & Ahmad, H. (2017). Location-Aware and Mobility-Based Performance Optimization for WSN. *Journal of Wireless Mobile Networks, Ubiquitous Computing, & Dependable Applications, 8*(3), 38-59.

[3] Sinde, R., Begum, F., Njau, K., & Kaijage, S. (2020). Refining Network Lifetime of WSN using Energy-Efficient Clustering and DRL-based Sleep Scheduling. *Sensors, 20*(5), 1540. https://doi.org/10.3390/s20051540

[4] Li, Z. & Xin, P. (2017). Evidence-Efficient Multihop Clustering Routing Scheme for Large-Scale Wireless Sensor Networks. *Wireless Communications and Mobile Computing*, Article-ID: 1914956. https://doi.org/10.1155/2017/1914956

[5] Ebadi, S. (2012). A Multihop Clustering Algorithm for Energy Saving in Wireless Sensor Networks. *International Scholarly Research Network - ISRN Sensor Networks*. https://doi.org/10.5402/2012/817895

[6] Izadi, D., Abawajy, J., & Ghanavati, S. (2015). An Alternative Clustering Scheme in WSN. *IEEE Sens J, 15*(7), 4148-4155. https://doi.org/10.1109/JSEN.2015.2411598

[7] Amer, O., Salem, A., & Noor, S. (2019). Enhanced LEACH Protocol for Increasing a Lifetime of WSNs. *Personal and Ubiquitous Computing, 23*, 901-907. https://doi.org/10.1007/s00779-019-01205-4

[8] Arora, V., Sharma, V., & Sachdeva, M. (2016). A Survey on LEACH and other Routing Protocols in Wireless Sensor Network. *Optik, 127*(16), 6590-6600. https://doi.org/10.1016/j.ijleo.2016.04.041

[9] Comeau, F. & Aslam, N. (2011). Analysis of LEACH Energy Parameters. Workshop on Emerging Topics in Sensor Networks (EmSeNs). *Procedia Computer Science, 5*, 933-938. https://doi.org/10.1016/j.procs.2011.07.131

[10] Tamilarasi, N. & Parvathi, K. (2016). Extension of Network Lifetime using Fuzzy C-Means, Model and Cluster Hierarchy Concept. *International Journal of Recent Trends in Engineering & Research, 2*(8), 302-308. https://doi.org/10.5121/ijdps.2012.3505

[11] Rathi, N., Saraswat, J., & Bhattacharya, P. (2012). A Review on Routing Protocols for Application in Wireless Sensor Networks. *International Journal of Distributed and Parallel Systems, 3*(5). https://doi.org/10.5121/ijdps.2012.3505

[12] Xu, L., Collier, R., & O'Hare, G. (2017). A Survey of Clustering Techniques in WSNs and Consideration of the Challenges of Applying such to 5G IoT Scenarios. *IEEE Internet of Things Journal, 4*(5), 1229-1249. https://doi.org/10.1109/JIOT.2017.2726014

[13] Mahmood, D., Javaid, N., Mahmood, S., Qureshi, S., Memon, A., & Zaman, T. (2013). MODLEACH: A Variant of LEACH for WSNs. *Eighth International Conference on Broadband and Wireless Computing, Communication and Applications*, Compiegne, 158-163. https://doi.org/10.1109/BWCCA.2013.34

[14] Bhavana, T. & Murthy, J. (2014). Spatial Correlation based Clustering-Algorithm for Random & Uniform Topology in WSNs. *IJRET: International Journal of Research in Engineering and Technology, 3*(6), 83-87. https://doi.org/10.15623/ijret.2014.0306015

[15] Zhao, Z., Xu, K., Hui, G., & Hu, L. (2018). An Energy-Efficient Clustering Routing Protocol for Wireless Sensor Networks based on AGNES with Balanced Energy Consumption Optimization. *Sensors, 18*(11), 3938. https://doi.org/10.3390/s18113938

[16] Vuran, M., Akan, Ö., & Akyildiz, I. (2004). Spatiotemporal Correlation: Theory & Applications for WSNs. *Comput Networks, 45*(3), 246-258. https://doi.org/10.1016/j.comnet.2004.03.007

[17] Tsai, M. & Huang, T. (2014). A Sub-Clustering Algorithm based on Spatial Data Correlation for Energy Conservation in Wireless Sensor Networks. *Sensors, 14*(11), 21858-21871. https://doi.org/10.3390/s141121858

[18] Liu, Z., Xing, W., Zeng, B., Wang, Y., & Lu, D. (2013). Distributed Spatial Correlation-based Clustering for Approximate Data Collection in WSNs. *IEEE 27th Int Conf on Adv Information Networking and Applications*. https://doi.org/10.1109/AINA.2013.26

[19] Jorio, A., Fkihi, S., Elbhiri, B., & Aboutajdine, D. (2015). An Energy-Efficient Clustering Routing Algorithm based on Geographic Position and Residual Energy for WSN. *J Comput Networks Commun*. https://doi.org/10.1155/2015/170138

[20] Panchikattil, S. S. & Pete, D. J. (2020). Spatial Clustering with Sequential CH Selection for Energy-Efficient WSN. In: Vasudevan H., Gajic Z., Deshmukh A. (eds) *Proceedings of International Conference on Wireless Communication. Lecture Notes on Data Engineering and Communications Technologies, 36*. Springer, Singapore. https://doi.org/10.1007/978-981-15-1002-1_30

[21] Wang, X., Liu, X., Wang, M., Nie, Y., & Bian, Y. (2019). Spatial Query-Centric Geographic Routing Protocol in Wireless Sensor Network. *Sensors, 19*(10), 2363. https://doi.org/10.3390/s19102363

[22] Zhou, Y., Yang, L., & Ni, M. (2019). Novel Energy-Efficient Data Gathering Scheme Exploiting Spatial-Temporal Correlation for Wireless Sensor Networks. *Wireless Communications and Mobile Computing*. https://doi.org/10.1155/2019/4182563

[23] Pourpeighambar, S. & Sabaei M. (2013). Spatial Correlation Aware Protocols for Efficient Data Aggregation of Moving Object in Wireless Sensor Networks. Scientia Iranica, 20(3), 695-709. https://doi.org/10.1016/j.scient.2012.11.010

[24] Heinzelman, W., Chandrakasan, A., & Balakrishnan, H. (2000). Energy-Efficient Communication Protocol for Wireless Microsensor Networks. *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, 2*, Maui, HI, USA, pp. 10. https://doi.org/10.1109/HICSS.2000.926982

[25] Khalid, Z. & Durrani, S. (2013). Distance Distributions in Regular Polygons. *IEEE Transactions on Vehicular Technology, 62*(5), 2363-2368. https://doi.org/10.1109/TVT.2013.2241092

[26] Kole, S., Vhatkar, K., & Bag, V. (2014). Distance based Cluster Formation Technique for LEACH Protocol in Wireless Sensor Network. *International Journal of Application or Innovation in Engineering & Management (IJAIEM), 3*(3), 334-338.
[27] Heinzelman, W., Chandrakasan, A., & Balakrishnan, H. (2002). An Application-Specific Protocol Architecture for Wireless Microsensor Networks. *IEEE Transactions on Wireless Communications, 1*(4), 660-670. https://doi.org/10.1109/TWC.2002.804190
[28] Sreedharan, P. & Pete, D. (2020). Dynamic Multi Hop Spatial Clustering Scheme for Data Transfer in Wireless Sensor Networks with Enhanced QoS M. *2020 International Conference for Emerging Technology (INCET)*, Belgaum, India, 1-8. https://doi.org/10.1109/INCET49848.2020.9154118
[29] Sreedharan, P. S. & Pete, D. J. (2020). Spatial Clustering Algorithm with Dynamic Multi Hop Routing for Wireless Sensor Networks. *International Journal on Emerging Technologies, 11*(3), 1169-1178.

Authors' contacts:

Panchikattil Susheelkumar Sreedharan, Research Scholar
Dept. of Electronics Engg., Datta Meghe College of Engineering,
Sector-3, Airoli, Navi Mumbai-400 708, Maharashtra, India
E-mail: susheelkumarsredh34@gmail.com

Dnyandeo Jageshwar Pete, Professor & Head
Dept. of Electronics Engg., Datta Meghe College of Engineering,
Sector-3, Airoli, Navi Mumbai-400 708, Maharashtra, India
E-mail: pethedj@rediffmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210204162414

Original scientific paper

Bidirectional ConvLSTMXNet for Brain Tumor Segmentation of MR Images

M. Ravikumar*, B. J. Shivaprasad

Abstract: In recent years, deep learning based networks have achieved good performance in brain tumour segmentation of MR Image. Among the existing networks, U-Net has been successfully applied. In this paper, it is propose deep-learning based Bidirectional Convolutional LSTM XNet (BConvLSTMXNet) for segmentation of brain tumor and using GoogLeNet classify tumor & non-tumor. Evaluated on BRATS-2019 data-set and the results are obtained for classification of tumor and non-tumor with Accuracy: 0.91, Precision: 0.95, Recall: 1.00 & F1-Score: 0.92. Similarly for segmentation of brain tumor obtained Accuracy: 0.99, Specificity: 0.98, Sensitivity: 0.91, Precision: 0.91 & F1-Score: 0.88.

Keywords: ConvLSTM; GoogLeNet; Linear Transformation (LT); Notch Filter; X-Net

1 INTRODUCTION

The brain controls and co-ordinate many important body functions. Normal cells generate, grow and die, abnormal cells grow when the body doesn't require them is known as cancer. A brain tumor occurs when abnormal cells produce within any part of the brain. There are two main types of tumors namely, malignant and benign tumors. Benign brain tumors are non-cancerous, malignant tumors are cancerous. Metastatic brain tumors occur when cancer located in another organ of the body spreads to the brain, 40% of all cancers spread to the brain and central nervous system, up to half of metastatic brain tumors are from lung cancer. Among 10,000 populations 5 to 10 people affected Central Nervous System (CNS) tumors in India [1].

Basically, the brain regions diagnosed/scanned by CT, X-ray, Ultrasound, PET and MRI. MRI is preferred over other imaging modalities because not harm and malaco tissue contrast in the brain [2, 3]. MRI produces different types of sequenced contrast images, which allow MRI extraction of valuable information of tumor and sub-regions, the deferent pulse sequences like, T1, T2, T1C and FLAIR. These sequenced images are diagnosed slice by slice manually is a laborious and time consuming process for radiologists/doctors. This manual burden process can be replaced by automatic enhancement, segmentation and classification with the use of computer-vision technique. To boost the visual appearance of an image, segment the Region of Interest (ROI) and classify them into the given class. Image processing is widely used.

In the present study, we present a techniques for enhancement, classification and segmentation of tumor from MR images using Notch filter & Linear Transformation (LT), GoogLeNet and Bidirectional Convolutional Long Short Term Memory (LSTM) X-Net (BConvLSTMX-Net). Classified and segmentation results are compared with other methods (AlexNet, VGG-16 & GoogLeNet) and (Seg-Net, UNet & XNet) respectively.

The remaining contents of the paper are arranged as follows: Section 2 gives the brief review of literature. In Section 3 discuss the present study. Section 4 shows comparative analysis, finally, in Section 5 interpret the present and future scope of the work.

2 STATE OF THE ART WORK

A brief review of literature on the topic of enhancement, segmentation and classification of MR brain tumor image is discussed below.

To enhance the contrast of MRI brain images, deferent spatial domain techniques were proposed like Histogram Equalization (HE) [4, 5, 7, 9, 10], Adaptive Histogram Equalization (AHE) [4, 5], Contrast Limited Adaptive Histogram Equalization (CLAHE) [4, 7], LHE [4], BBHE [5, 10], MMBEBHE [5, 6], BPDHE [5, 6, 8], RMSHE [6], BPDHE [6], DSIHE [6], BPDFHE [7], Deferent Techniques like GHE [8], Modified BHE, Brightness preserving BHE (BBHE) [10], Fuzzy logic based Adaptive Histogram Equalization (AHE) [5], Multi Scale Retinex (MSR) [9] and Non-sub sampled Contour-let Transform (NSCT)-FU [9].Different frequency based domain methods were proposed to enhance MRI brain images. Methods are Gabor Filter [13], Gaussian Filter [13, 23, 30, 29], salt and pepper-noise [13, 23], Median Filter [16, 17, 18, 20, 22, 25, 26, 30], An-isotropic Diffusion Filter[15, 17], Linear Filter [29], Wiener Filter [33], Discrete Wavelet Transform (DWT) [14, 15, 18, 21, 23, 27, 30] and Dynamic Stochastic Resonance (DSR) [17, 27, 29].

Before the revolution of deep learning, traditional semantic segmentation and handcrafted feature based classification methods were used. From the last decades deep-learning based approaches outstanding improvement in enhancement, segmentation and classification of images, they are CNN, RNN, FCN and GCN. Different CNN techniques were used for segmenting the brain of tumors like SegNet [34, 35], U-Net [35, 36, 37] and X-Net [38]. Similarly, AlexNet [39, 40], VGG-16 [39, 40] and GoogLeNet [39, 40] techniques are used to perform classification brain tumor.

From the related work, it is observed that most of the work done on enhancement, segmentation and classification of brain tumor from MR Images, still there is much scope for improvement. In this paper, Bidirectional Convolutional Long Short Term Memory (LSTM) X-Net (BConvLSTMX-

Net) is proposed as an extension of X-Net, The proposed method performs better than the existing methods.

3 PROPOSED METHOD

Here, the study focused on classification of tumor & non-tumor and also segmenting the brain tumor. The flow of the present methods is shown in Fig. 1 and deferent stages are described below.

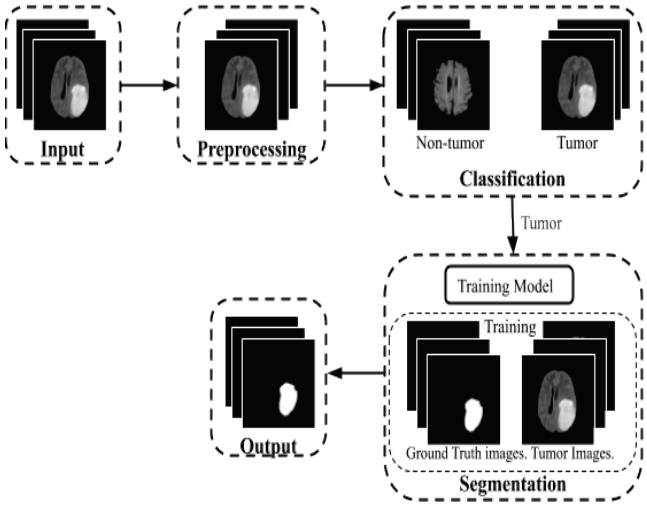


Figure 1 Flow of present method.

3.1 Prepossessing

Initially, we take BRATS-2019 brain images, to improve quality of the image, Notch & LT methods are applied. We tried a different inner & inter class combination of spatial, frequency and fuzzy logic methods, in that Notch & LT method gives good qualitative results.

3.2 Data Augmentation

Since the data-set considered for experimentation is very small i.e., only 284 images, therefore, we artificially augment the training images to create larger data-set to avoid over-fitting. Generally augmented images are obtained by using the geometrical operations like translations, rotation, shear and cropping.

3.3 Classification

For classification of tumor and non-tumor, we used predefined CNN based 22 layered GoogLeNet, The number of variables is small compared to Alex-Net & VGG-Net. The architecture of the Inception layer is given in Fig. 2.

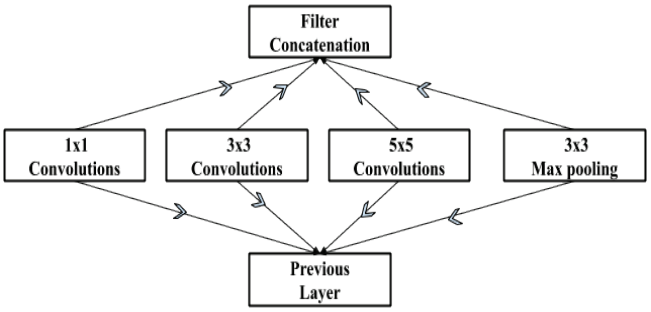


Figure 2 Architecture of the Inception Layer.

3.4 Segmentation

The BConvLSTMXNet method is proposed for segmentation of brain tumor, it is inspired by BConvLSTM [32] and X-Net [33] methods. The different stages of segmentation are discussed below and architecture is shown in Fig. 3.

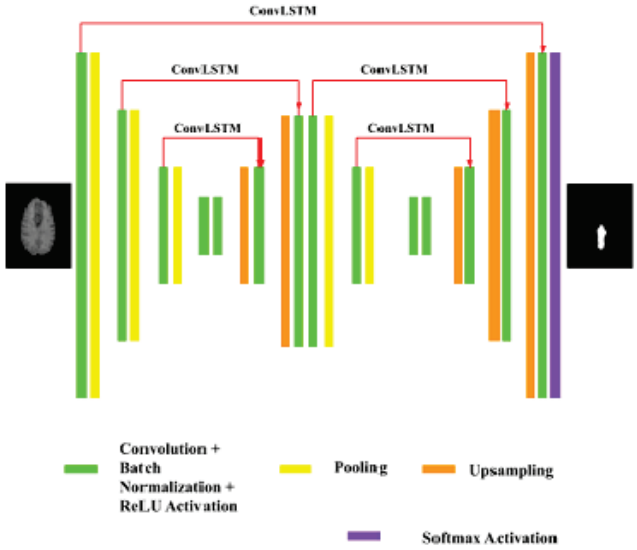


Figure 3 X-Net with BConvLSTM architecture

3.4.1 Encoding Path

The encoding path incorporates a sequence of steps. Each move consists, two convolutional 3×3 filters used for feature extraction along with 2×2 max-pooling function for down-sampling the input image and the activation function i.e, ReLU. Breaking up the down-sampling into multiple stages, features are doubled at each polling stage. The final encoded foot-path makes a big size with information.

3.4.2 Decoding Path

After feature extraction from the encoded path, decoded step to perform up-sample to make segmented mask-of equal size to the input image. Decoded step to perform an up-sample to make a segmented mask of equal size to the input image. In XNet, the encoded steps feature maps are duplicated to decoded steps. The extracted features are mapped to concatenate with BConLSTM, and we used two

encoder-decoder modules in succession. Compared to other networks we avoid larger serial down-sampling of the input, due to the small data-sets. Number of down-sampling in series can determine accurate boundary level on details and also avoid reducing image resolution.

3.4.3 Training and Optimization

An augmented data is trained, so increase the number of samples and lower the over-fitting. Soft dice metric is used as cost function and Adam optimization is used to minimize the cost function. Stochastic gradient based Adam optimization with learning rate 0.0001 [30, 31] is initialized.

The ground truth masks used for training and optimize by using cross-entropy loss.

$$L(N, m) = -\sum_{p=1}^{q} R(m, p) \log t(Q = p \,||\, N) \tag{1}$$

where, N is input pixel, m is the output, $t(Q = p \,||\, N)$ is probability, p given as input and $R(m, p)$ is in Eq. (2).

$$R(m, p) = \left\{ \frac{0 \text{ if } m \neq p}{1 \text{ if } m = p} \right\} \tag{2}$$

Without augmented data testing process is performed. The next section, experimentation and results are described.

4 RESULTS AND DISCUSSIONS

Here we give the detailed experimented discussion.

4.1 Result

For the purpose of experimentation, we have used 284 MRI brain images collected from BRATS-2019 repository to enhance, segment and classify brain images. Notch & LT methods are used to enhance brain image, GoogLeNet & BConvLSTM X-Net based deep convolutional networks are used for classification & segmentation of brain tumor. Result shows in Fig. 4 and Fig. 5.

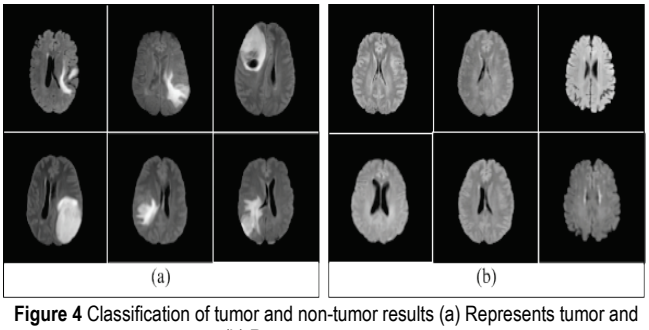


Figure 4 Classification of tumor and non-tumor results (a) Represents tumor and (b) Represents non-tumor.

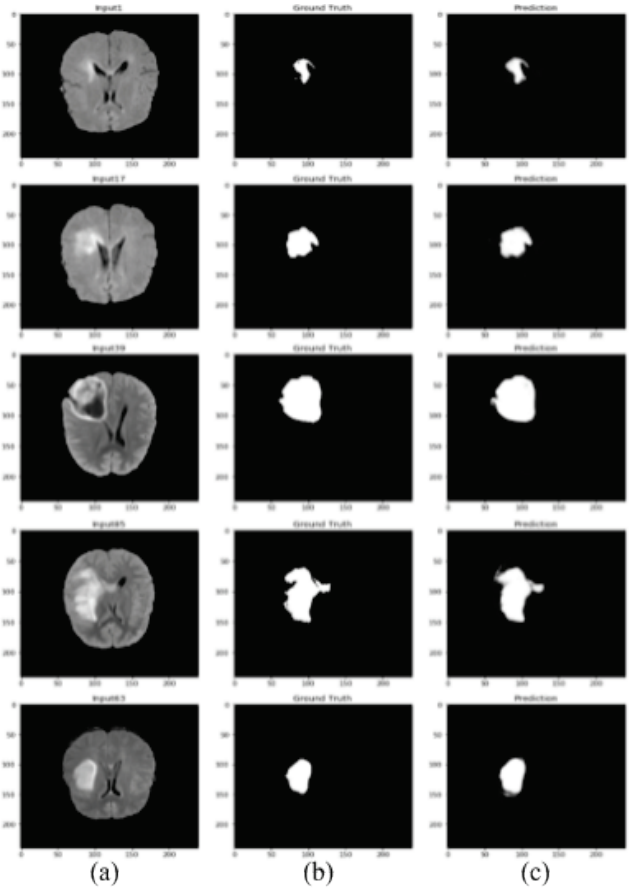


Figure 5 Segmentation of tumor results. Column, (a) Original image, (b) Ground truth and (c) Segmented tumor.

4.2 Discussion

To select the best segmentation and classification method quantitative analysis parameters are used, they are Accuracy, Specificity, Sensitivity, Precision, F1-Score and area under the curve (AUC). Tab. 1 and Tab. 2 gives the different quantitative measure results.

From Tab. 1 and Tab. 2, observed that the presented work obtained good quantitative measure result. In Fig. 4, shows the segmented result. In Fig. 5, the first column is tumor image, the second one is non-tumor images. Classification of tumor & non-tumor training loss and accuracy is shown in Fig. 6, the ROC is shown in Fig. 7 and segmentation Accuracy & Loss is shown in Fig. 8.

Table 1 Performance comparison methods for classification

Methods	Accuracy	Precision	Recall	F1-Score
AlexNet	0.81	0.85	1.00	0.92
VGG-16	0.46	0.86	1.00	0.93
GoogLeNet	0.91	0.95	1.00	0.92

Table 2 Performance comparison methods for Segmentation

Methods	Accuracy	Specificity	Sensitivity	Precision	F1-Score
SegNet	0.92	0.70	0.91	0.86	0.89
U-Net	0.95	0.83	0.95	0.93	0.79
X-Net	0.97	0.94	0.87	0.83	0.88
Proposed	**0.99**	**0.98**	**0.91**	**0.91**	**0.88**

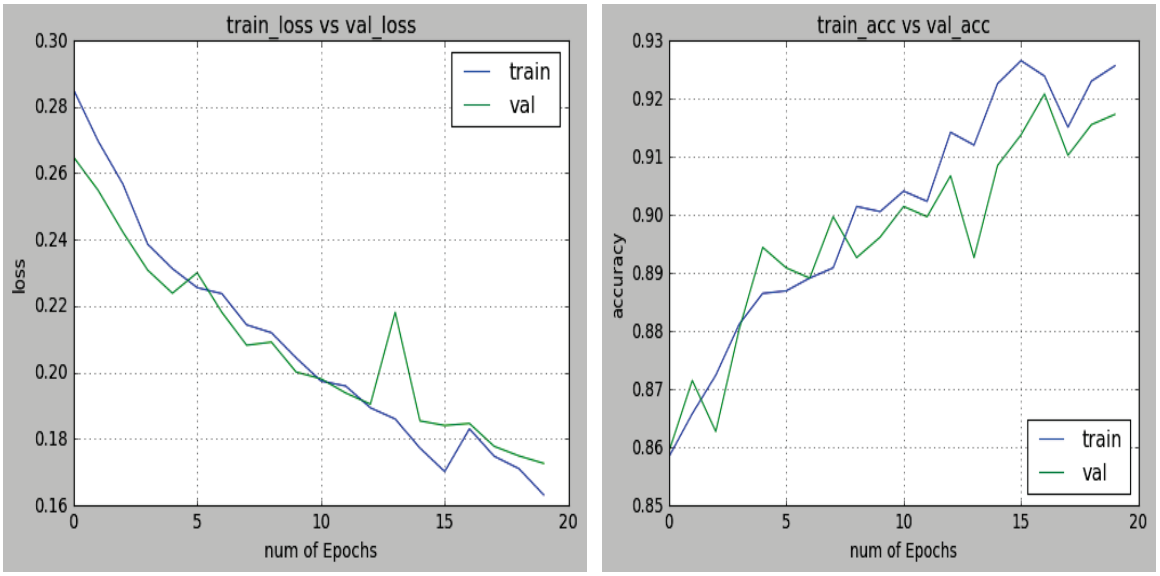


Figure 6 Accuracy and Loss for classification using GoogLeNet

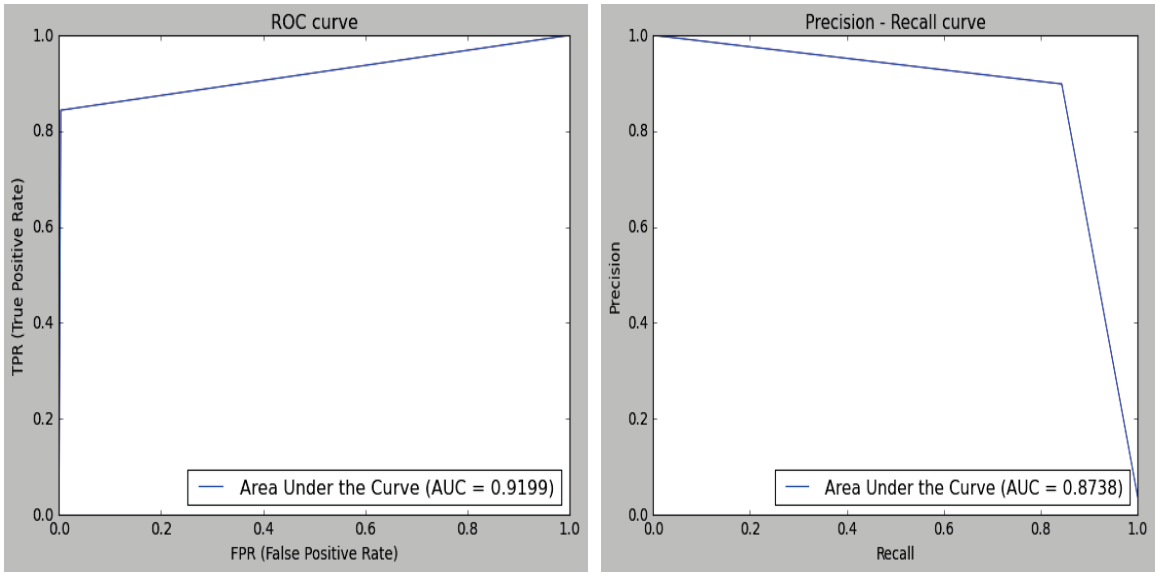


Figure 7 ROC diagrams of the present work for segmentation

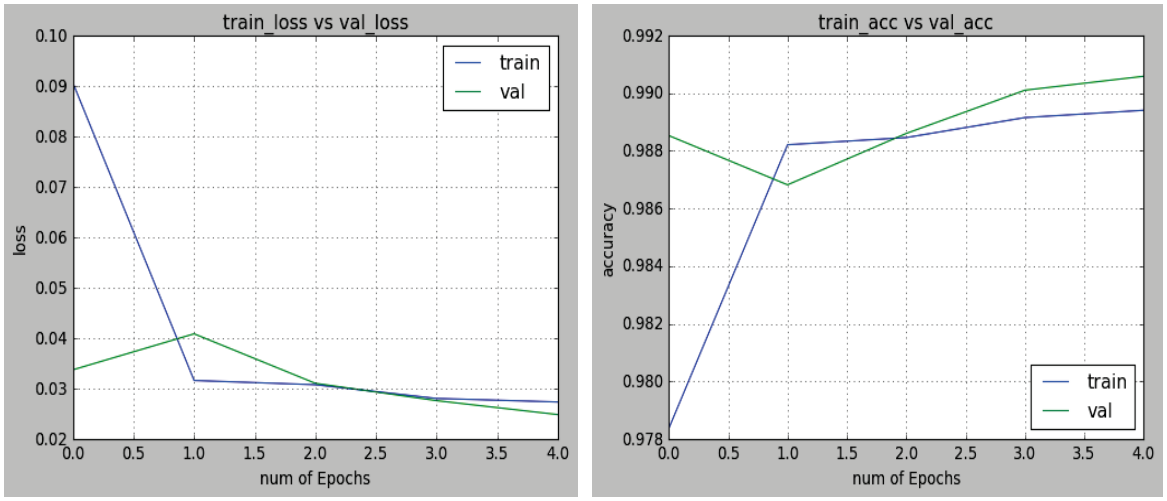


Figure 8 Accuracy and Loss diagrams for the proposed method

5 CONCLUSION

In this paper, proposed a deep-learning based Bidirectional Convolutional LSTM XNet (BConvLSTMXNet) for segmentation of brain tumor and using GoogLeNet classify tumor non-tumor. Evaluated On BRATS-2019 data-set and the results are obtained for classifica-tion of tumor and non tumor with Accuracy: 0.91, Precision: 0.95, Recall: 1.00 & F1-Score: 0.92. Similarly for segmentation of brain tumor obtained Accuracy: 0.99, Specificity: 0.98, Sensitivity: 0.91, Precision: 0.91 & F1Score: 0.88. Further we plan to extend our work towards the segmentation of core (major affected area), enhanced region.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

6 REFERENCES

[1] Knuth: Computers and Typesetting, https://www.nhp.gov.in/world-brain-tumour-daypg.

[2] Bedil, S. S. & Khandelwal, R. (2013). Various Image Enhancement Techniques - A Critical Review. International *Journal of Advanced Research in Computer and Communication Engineering, 2*(3), 267-274.

[3] Agravat, R. R. & Raval, M. S. (2018). Deep Learning for Automated Brain Tumor Segmentation in MRI Images. *Soft Computing Based Medical Image Analysis*, 183-201. https://doi.org/10.1016/B978-0-12-813087-2.00010-5

[4] Kaur, H. & Rani, J. (2016). MRI brain image enhancement using Histogram Equalization techniques. *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, 770-773. https://doi.org/10.1109/WiSPNET.2016.7566237

[5] Subramani, B. & Velucham, M. (2018). MRI brain image enhancement using brightness preserving adaptive fuzzy histogram equalization. *International Journal of Image Systems and Technology, 28*(3), 1-6. https://doi.org/10.1002/ima.22272

[6] Senthilkumaran, N. & Thimmiaraja, J. (2014). A Study on Histogram Equalization for MRI Brain Image Enhancement. *Proc. of Int. Conf. on Recent Trends in Signal Processing, Image Processing and VLSI, ICrtSIV*, 317-325. https://doi.org/10.1109/WCCCT.2014.45

[7] Suryavamsi, R. V., Reddy, L. S. T., Saladi, S., & Karuna, Y. (2018). Comparative Analysis of Various Enhancement Methods for Astrocytoma MRI Images. *International Conference on Communication and Signal Processing (ICCSP)*, Chennai, 0812-0816. https://doi.org/10.1109/ICCSP.2018.8524441

[8] Senthilkumaran, N. & Thimmiaraja, J. (2014). Histogram Equalization for Image Enhancement Using MRI Brain Images. *World Congress on Computing and Communication Technologies*, Trichirappalli, 80-83. https://doi.org/10.1109/WCCCT.2014.45

[9] Zhou, F., Jia, Z.-H., Yang, J., & Kasabov, N. (2017). Method of Improved Fuzzy Contrast Combined Adaptive Threshold in NSCT for Medical Image Enhancement. *Hindawi BioMed Research International, Volume 2017*, 1-10. https://doi.org/10.1155/2017/3969152

[10] Oak, P. V. & Kamathe, R. S. (2013). Contrast Enhancement of brain MRI images using histogram based techniques. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 1*(3), 90-94.

[11] Anand, A. (2017). Brain Tumor Segmentation using Watershed Technique and Self Organizing Maps. *Indian Journal of Science and Technology, 10*, 1-6. https://doi.org/10.17485/ijst/2017/v10i44/120574

[12] Milletari, F., Ahmadi, S.-A., Kroll, C., Plate, A., Rozanski, V., Maiostre, J., Levin, J., Dietrich, O., Ertl-Wagner, B., Bötzel, K., & Navab, N. (2017). Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound. *Computer Vision and Image Understanding, 164*, 92-102. https://doi.org/10.1016/j.cviu.2017.04.002

[13] Kumar, S., Dabas, C., & Godara, S. (2017). Classification of Brain Images: A Hybrid Approach. *Procedia Computer Science, 122*, 510-517. https://doi.org/10.1016/j.procs.2017.11.400

[14] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., & Larochelle, H. (2017). Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis, 35*, 18-31. https://doi.org/10.1016/j.media.2016.05.004

[15] Valente, J., Vieira, P. M., Couto, C., & Lima, C. S. (2018). Brain extraction in partial volumes T2*@7T by using a quasi-anatomic segmentation with bias field correction. *Journal of Neuroscience Methods, 295*, 129-138. https://doi.org/10.1016/j.jneumeth.2017.12.006

[16] Gupta, N., Bhatele, P., & Khanna, P. (2018). Identification of Gliomas from brain MRI through adaptive segmentation and run length of centralized patterns. *Journal of Computational Science, 25*, 213-220. https://doi.org/10.1016/j.jocs.2017.02.009

[17] Srinivas, B. & Rao, G. S. (2018). Unsupervised learning algorithms for MRI brain tumor segmentation. Conference on *Signal Processing and Communication Engineering Systems (SPACES)*, 181-184. https://doi.org/10.1109/SPACES.2018.8316341

[18] Mote, S. R., Baid, U. R., & Talbar, S. N. (2017). Non-negative matrix factorization and self-organizing map for brain tumor segmentation quote. *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 1133-1137. https://doi.org/10.1109/WiSPNET.2017.8299940

[19] Neffati, S. & Taouali, O. (2017). An MR brain images classification technique via the Gaussian radial basis kernel and SVM. *International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, 611-616. https://doi.org/10.1109/STA.2017.8314948

[20] Kushibar, K., Valverde, S., González-Villà, S., Bernal, J., Cabezas, M., Oliver, A., & Lladó, X. (2018). Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features. *Medical Image Analysis, 48*, 177-186. https://doi.org/10.1016/j.media.2018.06.006

[21] Mohseni Salehi, S. S., Erdogmus, D., & Gholipour, A. (2017). Auto-Context Convolutional Neural Network (Auto-Net) for Brain Extraction in Magnetic Resonance Imaging. *IEEE Transactions on Medical Imaging, 36*(11), 2319- 2330. https://doi.org/10.1109/TMI.2017.2721362

[22] Chen, H., Dou, Q., Yu, L., Qin, J., & Heng, P.-A. (2018). VoxResNet: Deep voxelwise residual networks for brain

segmentation from 3D MR images. NeuroImage, 170, 446-455. https://doi.org/10.1016/j.neuroimage.2017.04.041

[23] Abdulraqeb, A. R. A., Al-Haidri, W. A., & Sushkova, L. T. (2018). A novel segmentation algorithm for MRI brain tumor images quote. *Ural Symposium on Biomedical Engineering, Radio electronics and Information Technology (USBEREIT)*, Yekaterinburg, 1-4. https://doi.org/10.1109/USBEREIT.2018.8384535

[24] Islam, M. R., Imteaz, M. R., & Marium-E-Jannat. (2018). Detection and analysis of brain tumor from MRI by Integrated Thresholding and Morphological Process with Histogram based methodquot. *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, Rajshahi, 1-5. https://doi.org/10.1109/IC4ME2.2018.8465663

[25] El-Melegy, M. T. & Mokhtar, H. M. (2014). Tumor segmentation in brain MRI using a fuzzy approach with class center priors. *EURASIP Journal on Image and Video Processing*, 21(2014). https://doi.org/10.1186/1687-5281-2014-21

[26] Velumurugan, T. & Mahalakshmi, S. (2017). Efficiency of Fuzzy C Means algorithm for Brain Tumor segmentation in MR Brain Images. *International Journal of Engineering and Technology (IJET), 8*, 2979-2989. https://doi.org/10.21817/ijet/2016/v8i6/160806261

[27] Velumurugan, T. & Mahalakshmi, S. (2016). A Novel Approach to Find Tumor in MRI Brain Images using Image Segmentation Techniques, 9(40), 43-55.

[28] Sert, E. & Avci, D. (2019). Brain tumor segmentation using neutrosophic expert maximum fuzzy-sure entropy and other approaches. *Biomedical Signal Processing and Control, 47*, 276-287. https://doi.org/10.1016/j.bspc.2018.08.025

[29] Gupta, N., Bhatele, P., & Khanna, P. (2019). Glioma detection on brain MRIs using texture and morphological features with ensemble learning. *Biomedical Signal Processing and Control, 47*, 115-125. https://doi.org/10.1016/j.bspc.2018.06.003

[30] Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. arXiv.1-11. https://doi.org/10.1109/3DV.2016.79

[31] Kingma, D. & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations, Computer Science,* 1-15. https://arxiv.org/abs/1412.6980

[32] Song, H., Wang, W., Zhao, S., Shen, J., & Lam, K.-M. (2018). Pyramid Dilated Deeper ConvLSTM for Video Salient Object Detection. In *Proceedings of the European Conference on Com-puter Vision (ECCV 2018)*, 715-731. https://doi.org/10.1007/978-3-030-01252-6_44

[33] Bullock, J., Cuesta-Lzaro, C., & Quera-Bofarull, A. (2019). XNet: a convolutional neural network (CNN) implementation for medical x-ray image segmentation suitable for small datasets. In: *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 109531Z. https://doi.org/10.1117/12.2512451

[34] Alqazzaz, S., Sun, X., Yang, X., & Nokes, L. (2019). Automated brain tumor segmentation on multi-modal MR image using SegNet. *Comput. Vis. Media, 5*(2), 209-219. https://doi.org/10.1007/s41095-019-0139-y

[35] Daimary, D., Bora, M. B., Amitab, K., & Kandar, D. (2020). Brain Tumor Segmentation from MRI Images using Hybrid Convolutional Neural Networks. *Procedia Comput. Sci., 167*(2019), 2419-2428. https://doi.org/10.1016/j.procs.2020.03.295

[36] Dong, H., Yang, G., Liu, F., Mo, Y., & Guo, Y. (2017). Automatic brain tumor detection and segmentation using U-net based fully convolutional networks. *Commun. Comput. Inf. Sci., 723*, 506-517. https://doi.org/10.1007/978-3-319-60964-5_44

[37] Xu, F., Ma, H., Sun, J., Wu, R., Liu, X., & Kong, Y. (2019). LSTM Multi-modal UNet for Brain Tumor Segmentation. *2019 IEEE 4th Int. Conf. Image, Vis. Comput. ICIVC 2019*, 236-240. https://doi.org/10.1109/ICIVC47709.2019.8981027

[38] Qi, K. et al. (2019) X-Net: Brain Stroke Lesion Segmentation Based on Depthwise Separable Convolution and Long-Range Dependencies. In: Shen D. et al. (eds) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Lecture Notes in Computer Science, 11766*, Springer, Cham. https://doi.org/10.1007/978-3-030-32248-9_28

[39] Bocheva, P. K. B. N. & Nedelcheva, S. (2020). STDP Plasticity in TRN within Hierarchical Spike Timing Model, vol. 1. Springer International Publishing,

[40] Abd-Ellah, M. K., Awad, A. I., Khalaf, A. A. M., & Hamed, H. F. A. (2018). Two-phase multi-model automatic brain tumour diagnosis system from magnetic resonance images using convolutional neural networks. *Eurasip J. Image Video Process., 2018*(1). https://doi.org/10.1186/s13640-018-0332-4

Authors' contacts:

M. Ravikumar, Associate Professor
(Corresponding author)
Computer Science Dept., Kuvempu University,
Shankaraghatta 577451, Shimoga Dist, Karnataka, India
E-mail: ravi2142@yahoo.co.in

B. J. Shivaprasad, Research Scholar
Computer Science Dept., Kuvempu University,
Shankaraghatta 577451, Shimoga Dist, Karnataka, India
E-mail: shivaprasad1607@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210204180217

Original scientific paper

Trust Management Approach for Detection of Malicious Devices in SIoT

Priyanka Hankare*, Sachin Babar, Parikshit Mahalle

Abstract: Internet of Things (IoT) is an innovative era of interrelated devices to provide services to other devices or users. In Social Internet of Thing (SIoT), social networking aspect is used for building relationships between devices. For providing or utilizing services, devices need to trust each other in complex and heterogeneous environments. Separating benign and malicious devices in SIoT is a prime security objective. In literature, several works proposed trust computation models based on trust features. But these models fail to identify malicious devices. This paper focuses on detection of malicious devices. In this paper, basic fundamentals, properties, models and attacks of trust in SIoT are discussed. Up-to-date research distributions on trust management and trust attacks are reviewed and idea of Trust Management using Machine Learning Algorithm (TM-MLA) is proposed for identification of malicious devices.

Keywords: IoT; SIoT; Trust Attacks; Trust Management; Trust properties

1 INTRODUCTION

IoT comprises of large number of devices with ability to sense, gather and produce information from the world around us. The devices interact with one another to deliver wide range of smart services that are utilized by users, manufacturers, and other devices to carry out daily activities [1]. IoT has applicability in many domains like healthcare, smart home and workplaces, intelligent transportation systems, environment monitoring etc. Each device in IoT plays the role of service provider, service requestor, or both. To establish trusted relationships between devices, social networking aspect is used in IoT and this paradigm is called as SIoT. SIoT comprises of various devices/things to gather data, offer services, provide recommendations, make verdicts, and take actions. It has an imperative impact to refresh new advancements of medical services, medical robotics and medical embedded sensor [2]. SIoT also used in crowd-sensing applications [3], coastal management system [4].

In SIoT, the social networking of device owners is used to establish trustworthy social relationship among devices. There are different SIoT relationships among devices. They are:

1. Parental object relationship (POR) exists if devices are owned by the same manufacturer.
2. Co-location object relationship (CLOR) set up among devices if devices are present in a same location.
3. Co-work object relationship (C-WOR) built up between devices working collectively to give a common IoT application.
4. Ownership object relationship (OOR) set up if devices (laptops, smart phones, printers etc.) belong to the same owner.
5. Social object relationship (SOR) exists if devices owners get in touch with one another occasionally or frequently (e.g., devices owned by friends, classmates, colleagues).

Trust performs a key role in IoT. For example, IoT enables real-time alerting, tracking, and monitoring about patient's conditions to doctors. But if the information is not notified on time to doctor by IoT devices, it will be dangerous. So, it is essential to find trustworthiness of device in a network. Security and secrecy are the fundamental encounters in the IoT network. Misbehaving devices may carry out trust attacks based on misuse of trust. In order to fulfil SIoT full deployment the following trust management criterion must be discussed:

1. *Identity*: Identity management handles authentication as well asauthorization. Each device in SIoT has a unique identity. The device hiding its real identity must be detected. Access to SIoT devices, routing information must be authorized.
2. *Availability:* States that SIoT characteristics, SIoT entities, networks and services should be always up to date and work accurately even with failure or malicious attacks on system.
3. *Confidentiality:* It avoids the illegal access to the data and preserves the authorized control on system.
4. *Integrity:* Ensures that data and routing information have not been altered while transferring in a network. The trustee sticks to a bunch of ethics that enables the trustor to accept that the trustee is not malevolent.
5. *Data and Privacy:* The large amount of data is exchanged, shared, processed in SIoT network. In this context, unauthorized access to information is possible. Privacy requirement ensures that identities of SIoT devices must be highly protected from illegal access.
6. *Trust:* Trust could be well estimated in order to find appropriate trustee which can provide the best service for given task of a trustor. Trust management systems have to detect non-trustworthy behavior of device and separate untrusted devices from trusted one.

To date there is a little work on SIoT. Existing methods fails to identify trustworthy and untrustworthy devices. Detecting untrustworthy device is tricky task. In this paper, *Trust Management using Machine Learning Algorithm (TM-MLA)* is proposed to detect malicious device. Paper explores the evolutionary history of trust management for SIoT, examines the SIoT studies and come up with the challenges and idea of TM-MLA.

The paper is organized as follows: Section 2 and 3 discussed research distributions which offer solutions to the trust management and trust attacks in SIoT respectively. Section 4 provides the holistic view on trust management in SIoT, trust model and trust attacks. Section 5 presents the challenges in SIoT. An idea of TM-MLA for detection of malicious devices is presented in Section 6. Evaluation and experimental setup are discussed in section 7. Lastly, Section 8 concludes the paper.

2 EXISTING RESEARCH IN IOT TRUST MANAGEMENT

Chen et al. [6] proposed adaptive trust management protocol in a view of social relationships like using honesty, cooperativeness and community of interest. The protocol defends misbehaving attacks. This protocol is not tested against multitude of dynamically changing atmosphere situations. Trust update depends on recommender node.

Chen et al. [7] proposed an adaptive IoT trust protocol for SoA based IoT systems with adaptive filtering technique. This protocol includes SIoT constraints like scalability, storage and computational costs of devices. For assessing social similarity and filtering trust feedback based on social similarity, three social relationships, i.e., friendship, social contact, and community of interest are considered. This trust protocol is resilient to attacks such as SPA, BMA, BSA and OSA. However, this approach doesn't consider QoS trust factor for trust composition.

Truong N. B. et al. [8] proposed a trust prototype with three aspects that is Reputation, Recommendation, and Knowledge. This prototype finds the trustworthy devices by setting a trust channel between devices and improves the network performance. Only trustor's preferences are taken into account for the calculation of trust score. But trustee's factors like opinion, willingness, and capability are also important for trust calculation. This approach doesn't put forward clarification to confirm the adaptability of the SIoT system.

Ikarm Ud Din et al. [9] did investigation of trust managing practices for IoT. Contributions and limitations of these techniques are presented in a different perspective. This paper provides an overview of how different systems fit together without examining different standards to bring preferred functionalities.

Juan Chen et al. [10] developed a trust architecture by taking into account the technique of Soft Defined Network (SDN) in IoT, and a cross-layer authorization protocol based on IoTrust. Behavior-based Reputation Evaluation Scheme for the device (BES) and an Organization Reputation Evaluation Scheme (ORES) are used for trust establishment. Hypothetical analysis signifies that the developed trust architecture can resilient to modification attack, replay attack, and message dropping attack. This architecture does not work well on heterogeneous devices.

Xiao H et al. [11] proposed a trust model for SIoTon the basis of guarantor and reputation. Credit and reputation are the two parameters used by the model. Every device has its own reputation stored in it. If device provides accurate results then he is rewarded. If device is defective then he has to provide some rewards to other devices. This approach provides same trust value for all devices owned by same user. Storage, computing capacity of objects and energy consumption are the limitations of this model.

Zhiting Lin et al. [12] built a trust model on 5 aspects: 1) mutuality of trustor and trustee; 2) inferential transfer of trust; 3) transitivity of trust; 4) trustworthiness update; and 5) trustworthiness affected by dynamic environment. Behavioral changes in devices, membership changes and the changes in working patterns are considered in this model.

Upul Jayasinghe et al. [13] built a trust model classifier using SVM algorithm into two classes, trustworthy and untrustworthy. For calculation of trust scores, knowledge, experience and reputation trust metrics are used. Event based trust update scheme is used.

Anuoluwapo A. Adewuyi et al. [14] built a trust model, CTRUST for collaborative applications. Trust decay and belief functions are used in model for decaying the past trust values with time and guiding the acceptance of trust recommendations from another node respectively. The model assigns weights to the trust metrics as per their importance. However, the privacy aspects are not considered.

Abdelghani et al. [15] presented a trust management system to detect the malicious devices, block and isolate them using supervised approach of machine learning algorithm. Subjective trust features like Reputation, honesty, quality of provider, similarity, direct experience, rating frequency and rating trend etc. are used to calculate trust score.

Hui et al. [16, 17] implemented a contextualsystemto find out trusted device in SIoT. To calculate the trust between IoT objects and their owners, system considers the concepts from social and physiological science.

Muhammad Ajmal Azad [18] implemented the trust model for preserving privacy of IoT devices as well as user is. The trust score is updated in self-enforcing manner without help of third party. Social relationship between users as well as devices is considered by this model.

Table 1 Categorization of existing research according to SIoT relationship type

Reference paper	Relationship Type	
	Device to device relationship	User to user relationship
[6]		✓
[7]		✓
[8]		✓
[11]		✓
[18]	✓	✓

Tab. 1 summarizes the existing research from relationship type viewpoints. It shows that mostly user to user relationship is considered for trust management. But SIoT network has two important components, user and device. Hence, the social relationship between user-user, device-device and user-device must be taken into account for calculation of trust.

3 EXISTING RESEARCH IN IOT TRUST ATTACKS

Jean Caminha et al., [19] initiated a SIoT method on the basis of machine learning and an elastic slide window method that enabled to detect OO attacks (RA) in IoT. This method differentiates attacker devices from broken devices.

Truong et al., [20] presented a trust composition technique integrated with social trust metrics of the SIoT components such as common interest, cooperativeness and honesty similarity. To calculate weighted sum direct views, global judgements, and personal experiences are used through Bayesian technique. The scheme prevents attacks such as BMA, BSA and SPA. Reputation of device is not considered while trust computation.

Chen et al., [21] presented an access service recommendation scheme for effective service composition in SIoT environment. For trustworthiness analysis of SIoT devices a coherent recommendation metric is introduced. This approach defends attacks such as BMA, BSA and SPA. In this scheme, an energy aware mechanism is taken into account for SIoT privacy and load management. However, SIoT limitations such as device space, scalability and processing capacity have not been considered.

Abderrahim et al. [22] proposed a trust management system that integrates direct-indirect trust, transaction factors and social modelling of trust. This model is resilient to OOA attacks. Kalman filter technique is used to measure trust value and defend probable attacks.

Mariam Masmoudi [23] proposed a trust evaluation model to find out malicious devices using deep learning technique. Subjective trust features like Reputation, honesty, quality of provider, similarity, direct experience, and rating frequency etc. are used to calculate trust score. This approach defends attacks such as BMA, BSA and SPA, DA. However, Specific set of features are used for detection of malicious devices.

Tab. 2 summarizes the existing research from trust attack viewpoints. The features used in literature are not able to identify all types attacks. Specific set of trust features are used for detection of all kinds of attacks. But some features are more related with one type of attack and less with another. For e.g. the similarity has more relation with DA and less with SPA attack.

Table 2 Categorization of existing research according to trust attacks

Ref paper	Trust attacks						
	SPA	BMA	OSA	BSA	WA	OOA	DA
[19]						✓	
[7]	✓	✓	✓	✓			
[20]	✓	✓		✓			
[21]	✓	✓		✓			
[22]						✓	
[23]	✓	✓		✓			✓

Tab. 3 summarizes the existing research on trust features and trust aggregation technique used for trust model implementation. It seems from presented work that there is still lot of work needs to be done in the area of trust management. As shown in Tab. 3, trust aggregation is done with dynamic or static weighted sum approach, fuzzy logic and machine learning algorithms. Though the popular choice for trust aggregation is weighted sum approach, where weights can be assigned to trust features as per their importance in transaction, which trust feature makes more influence on transaction is very difficult to identify.

Table 3 Categorization of existing research according to trust features and aggregation techniques

Ref paper	Trust Features	Trust Aggregation Technique Used
[6]	Honesty, Cooperativeness and Community of Interest	Weighted Sum
[7]	Friendship, Social Contact, and Community of Interest	Bayesian Model
[8]	Reputation, Recommendation, and Knowledge	Fuzzy and Multi-Criteria Utility Theory
[11]	Social Cooperativeness	Probability
[20]	Cooperativeness, Community-Interest, Honesty and Similarity	Weighted Sum
[21]	Coherent Recommendation	Weighted Sum
[22]	Community of Interest	Weighted Sum
[13]	Knowledge, Experience and Reputation	Machine Learning Algorithms

4 TRUST MANAGEMENT IN SIoT

The concept of trust has been studied in numerous fields like psychology, sociology, computer science etc. Each of these fields gives different aspect of trust. Trust is estimate of various qualities like honesty, cooperativeness, willingness, expectation, faith, confidence. Trust has two important entities: trustor and trustee. The trustor has a goal, its own need. It entrusts the trustee by evaluating trustee's competence and willingness.

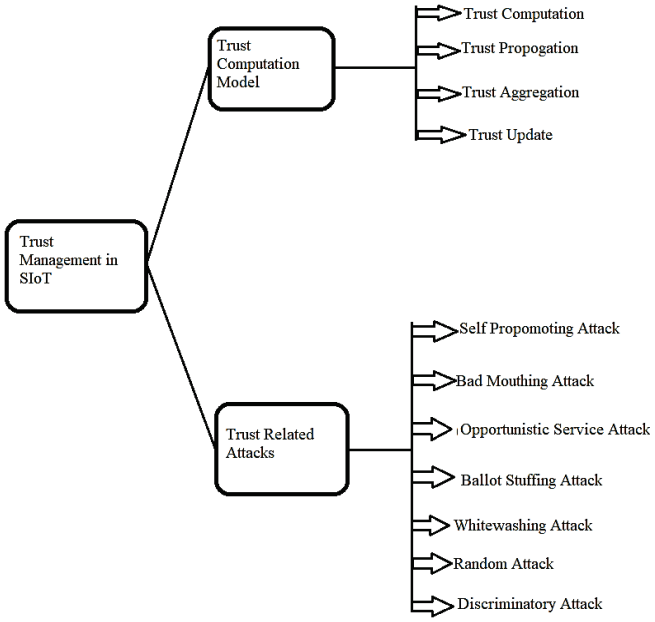


Figure 1 Trust management in SIoT

In today's world so much data is shared among community using devices like apps, computers, sensors, cameras etc. If data is shared with non-trusted clients/devices, it may be used for malicious purpose. For example, Charlie (the trustor/evaluator) wants to use Carlos (the trustee/performer) images placed at Carlos device. Charlie trusts Carlos so he is confident that he will get the images form Carlos device. But at the same time Carlos needs to ensure that Charlie will not misuse the images. It is important that the receiver as well as sender must trust each

other for healthy exchange of data. Trust management becomes a supreme question in SIoT for assurance of reliable trust model and improved object's security [5].

In our study, the trust management is considered from two perspective: trust computation model and trust attacks. Fig. 1 depicts the proposed framework of trust management in SIoT.

4.1 Trust Computation Model

The design dimensions of trust model are: trust composition, trust propagation, trust aggregation and trust update.

4.1.1 Trust Composition

QoS (quality of service) and social trust are the two main ways to determine the trust value. QoS is usually measured by packet delivery ratio, load balance, energy consumption, delay, bandwidth etc. Social trust is estimated by factors like social contact, friendship, community of interest, intimacy, honesty, privacy, centrality, and connectivity etc. In previous research [6-8, 11], trust was computed by considering following properties:

a) Direct: Trust established on direct experiences, interactions.
b) Indirect: Trust constructed on recommendations, feedbacks from other devices or peers. The recommendation relies on surrounding suggestions and global opinions.
c) History: Past interactions or experiences may have impact on present trust level.
d) Context: Trust is context dependent [5]. Trust changes depending on (i) target of task, (ii) time span, and (iii) environment. Trust varies if context is changed.
e) Dynamic: Trust changes non-monotonically with varying situations of environment.

4.1.2 Trust Propagation

It gathers the direct observations and indirect feedbacks for the trust evaluation. Centralized and distributed approaches are used for trust propagation.

a) Centralized approach: For restoring trust value, all devices are connected to centralized entity (e.g physical cloud).
b) Distributed approach: IoT devices store trust observations towards their peer devices. Centralized server is not used by this approach.

4.1.3 Trust Aggregation

Trust is aggregated using methods like static and dynamic weighted sum, Belief Theory, Bayesian Model (BM) and Fuzzy Logic (FL), Regression Analysis.

4.1.4 Trust Update

In general, there are two approaches involving the trust model: time-driven approaches and event-driven approaches. In the time-driven approach, trust reports are collected occasionally. Usually, the latest trustworthiness assessment gets bigger weights. Event-driven approach refers to a device trustworthiness that restructured after an event or transaction is made.

4.2 Trust Attack

Misbehaving or Malicious devices attack SIoT system to disrupt the functionality of SIoT network operations. Different trust related attacks performed by malicious devices are as follows:

1. Self-promoting attack (SPA): can boost its significance by bragging itself in order to be chosen as a service point.
2. Bad mouthing attack (BMA): reduces the likelihood of good devices to be chosen as service points as prominence of these devices are ruined by providing bad trust evaluation against them.
3. Opportunistic service attacks (OSA): perform good services when device reputation falls.
4. Ballot stuffing attack (BSA): increases the chance of malicious devices to be chosen as a service point as good recommendations are provided by other defective devices to them.
5. Whitewashing attack (WA): fades out malicious devices bad image by exiting from the application and then returning again.
6. Random attacks (RA): also called as On-Off Attacks (OOA). A malicious device can provide better and poor services randomly to avoid being rated as low trust device. This attack is hardest to detect.
7. Discriminatory attacks (DA): perform by malicious device on other devices having fewer common friends.

5 CHALLENGES OF SIOT

SIoT faces following number of challenges of trust management.

1. Device capability

Previous trust management solutions can't be applied directly to all SIoT applications as devices are having different computational power, storage capacity, standard, communication stacks, operating system, I/O channels. Trust management algorithm should take into consideration all such device requirements.

2. Handling large network

Communication between devices produces large number of transactions. Existing systems does not scale well to handle such large number of transaction information. The trust management algorithm should be powerful to control the giant number of devices as well as communication between them.

3. Existing device leaving and new device joining

SIoT system evolves with existing device leaving and new device joining. So, trust management algorithm should

consider dynamicity of device like changeable behaviour of device, their membership changes, interaction pattern changes, network topology changes and location changes.

4. Finding trustworthy device

With rising number of devices, it's very difficult to find out trustworthy devices. SIoT makes human's life more comfortable. In today's world so much data is shared among community using devices. If data is shared with non-trusted clients/devices, it may be used for malicious purpose. So, there is a need to design algorithm specifying rules which identify trusted and malicious behaviour of a device and hence enable sharing in controlled manner to avoid malicious attacks.

5. Selection of trust features

Trust is an important challenge in SIoT where device needs to find correct trustee for healthy exchange of data between them. Selecting appropriate trust features is necessary in trust management as accuracy, performance of trust systems depends on this. As shown in table III, specific set of trust features like reputation, honesty, community of interest, similarity, rating frequency are considered for the calculation of overall trust value. The literature work stated in section 2 and 3 rates the best device in SIoT network but fail to detect attacks performed by malicious device. Lastly, in the earlier systems [6, 7] the dynamic change in trust feature criteria is not considered while trust computations. For more accuracy of trust computation there is need to change the features of trust dynamically based on importance of transaction.

6. Trust aggregation

As shown in Tab. 3, most of the previous approaches used weighted sum approach for aggregation of trust values. However, there are numerous shortcomings in this practice. There are several likelihoods when it comes to assess a weighting factor. Systems fails to recognize which feature makes the most impact on trust in specific setting as weights assigned to trust features may vary from one to another. This approach cannot identify malicious and benign behaviour of node. Hence machine learning approach is used for combining the trust scores and detection of malicious devices in this research.

7. Trust update

In [6, 24, 25], the trust update depends on recommendation of other node i.e. the trust update score is computed using value provided by another node or recommender. But what if the recommender node is malicious? In [6, 18, 26, 27] the trust is updated based on previous experience or trust score and ability of node. The ability of device is calculated from his performance in previous task i.e. gain/damage after performing task or good conduct or bad conduct of device or successful/unsuccessful communications, packet received and differentiation etc. What if for longer time there is no interaction among trustor and trustee? It is necessary to take into account the time elapsed for previous interaction while updating the trust. In [14], the trust is decayed if there no interaction between nodes. The trust decay is applied on trust features like recommendation, previous trust value. When new session of interaction is made, previous trust value is decayed. The overall trust is updated based on previous trust effectiveness, direct assessment, and recommendation. Previous trust effectiveness is calculated based on number of interactions in the time interval. After every new interaction with j, previous trust is updated. But what if the node *j* becomes unavailable in the network, and node *i* gets no next chance of interaction with node *j* or *i* provides recommendation to *k* about *j* before next interaction with *j*. In such cases, *i* will provide old trust value. Hence time driven trust approach is used in our research

6 PROPOSED MODEL

The TM-MLAwill focus mainly on fifth, sixth and seventh challenge. TM-MLA will be implemented using trust features as per attack context and dynamically varying surrounding situations to detect the malicious devices. Five attacks will be considered in the proposed model: 1. SPA; 2. BSA; 3. BMA; 4. DA; 5. OSA. Fig. 2 depicts the system architecture.

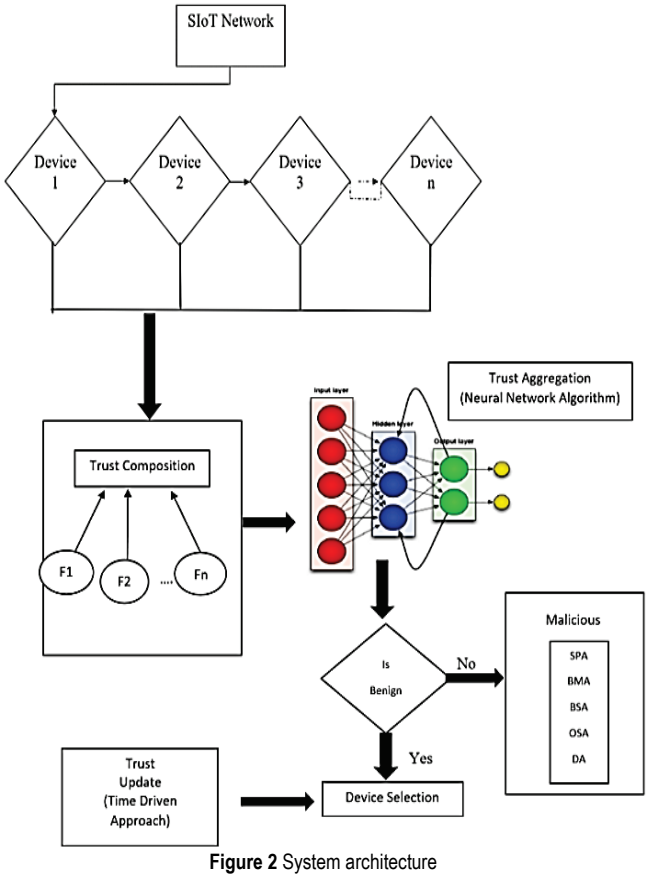


Figure 2 System architecture

TM-MLA will consist of mainly three phases - trust composition, trust aggregation and trust update phase.

- In trust composition phase different features will be chosen as per the context of attack. For trust computation process, the trustee node will be selected from the set of nodes based on trust features.
- In trust aggregation phase machine learning based approach, Artificial Neural Network (ANN) algorithm

will be used to get the trust score to defeat the drawback of past trust aggregation techniques. At the output layer of ANN, a probability is derived which decides whether the trustee for given task of trustor is malicious or benign. After selection of trustee node, the trustor will assign a task to trustee.

- In trust update phase, time driven approach will be used for updating the trust score. In this approach, previous or stored trust values of a device will decay with time and more weightage will be given to latest trust values. After certain number of times, the previous trust values will no longer relevant. The Eq. (1) will be used to decay the trust over time. Depending on result given by trustee for the assigned task, the trust will decay.

$$T_{kl}(\text{current}) = \left(T_{kl}(\text{initial value}) \times (1 - Decay\_rate)\right)^{x} \quad (1)$$

Where: T_{kl}(current) - current trust estimation T_{kl} of trustee node l by trustor node k at time t; T_{kl}(initial value) - trust estimation value at initial time; *Decay_rate* - rate at which trust decays; x - duration required to decay trust value as per decay rate.

After training stage, the model will be utilized to assess the performance of algorithm to detect the trust related attacks.

Table 4 Dataset

Trustor Device	Type of Request	Trustee Device	Trust Features						Trust Score	Malicious (M) or Benign (B)
A	Image	B	-	-	-	-	-	-	0.85	B
B	Video	C	-	-	-	-	-	-	0.33	M
-	-	-	-	-	-	-	-	-	-	B
C	Location request	A	-	-	-	-	-	-	0.68	B

7 RESULT AND ANALYSIS

There are two main parts in SIoT: 1. Devices; 2. Users. $D = D1, D2, \ldots, Dn$ are the set of devices owned by users in network and $U = U1, U2, \ldots, Un$ are the set of users of SIoT network. The Fig. 3 depicts the idea of SIoT network. The communication between users and devices is shown by using the edges between them. Each device will provide services to the users or other devices. For evaluating the performance of proposed model, the data of Facebook, Quora, and Twitter social network will be considered. The request-response patterns, task sharing, interactions among devices will be analyzed for the creation of dataset. This information will be stored in table format as shown in Tab. 4.

The real-world network will be formed between devices like mobiles and laptops. 10,000 records will be used for implementation. Out of 10,000, 80% of the data will be used for training and 20% will be used for testing. To train the model collaborative filtering approach will be used. Theproposed model will find potential trustees for given task of trustor, detect malicious or benign devices in SIoT network.

After implementation of model, the correctness will be analyzed in two ways: 1. By comparing the model with previous approaches and 2. By calculating the accuracy of model using precision, recall, and *F-measure* methods. *Precision* (or positive predictive value) is the ratio of count of accurately-detected-matching records to the count of pair of records that were detected as matching. It is shown in Eq. (2).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall (or sensitivity) is the ratio of count of accurately detected matching records to the total count of matching records in the test set. It is shown in Eq. (3).

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Where: *TP* (true positive) refers to matching instances that are correctly identified as matching by algorithm. *FP* (false positive) refers to non-matching instances that are error-neously-detected as matching. *FN* (false negative) refers to matching instances that mislabeled as non-matching.

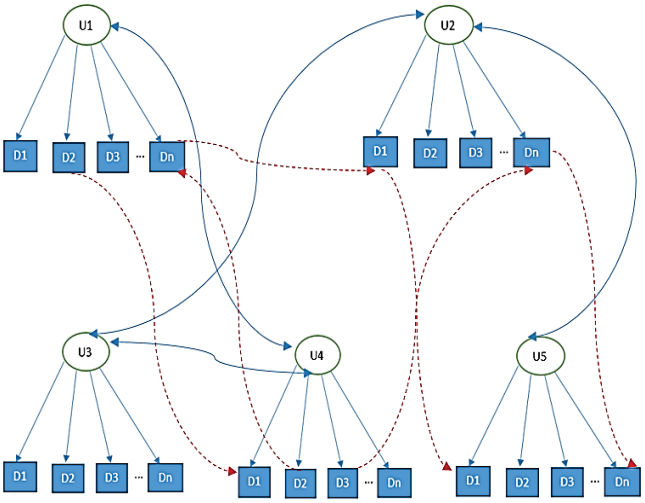


Figure 3 SIoT network

F-measure as shown in Eq. (4), is a combination of precision and recall. It is calculated by taking harmonic mean of precision and recall.

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

The average accuracy of the trust models with weighted mean approach [6, 26, 29] is shown in Fig. 4. Our method of trust update does not depend on recommender node. Thus, there is no chance that untrustworthy node will provide fake recommendations to benign node and good recommendations to malicious node. So, TM-MLA defends against BSA, BMA and SPA. Time driven trust decay

method declines the trust if there is no interaction or less frequent interaction between trustor and trustee. Therefore, the TM-MLA prevents OSA and DA attack. Hence proposed algorithm surely maximizes the accuracy by giving better recall i.e., a TPR, lower FPR and higher TNR as ML based approach is used for trust aggregation. Confusion matrix will be used to exhibit the efficiency of our model against weighted mean methods.

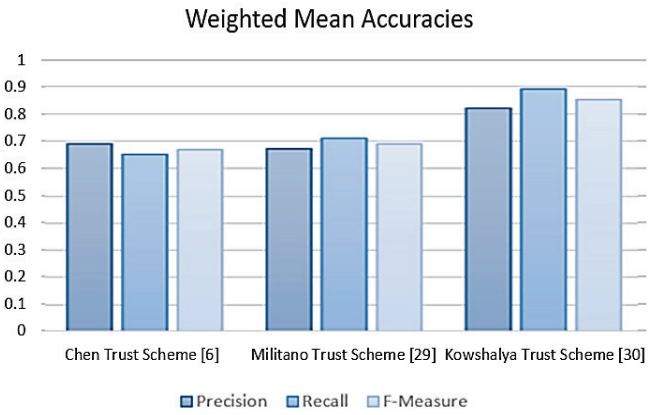


Figure 4 Comparison of Trust Estimation accuracies

8 CONCLUSION

SIoT assures to provide scalable services with trillions of interrelated devices. Trust management in SIoT is an important research issue in previously proposed mechanisms. In this paper, the overview of the SIoT paradigm, basic fundamentals of trust, its properties and trust computation model has been presented. The latest research studies on SIoT trust management and trust attacks have been reviewed. The challenges and trust management model are presented. TM-MLA will detect the malicious devices performing attacks on a system. As trust features will be chosen according to attack context, the better and strong results will be achieved. Machine learning based trust aggregation structure used in the TM-MLA model eliminates the traditional shortcomings of weighted sum. TM-MLA removes the drawbacks of previous trust update. So, it's a more dependable method.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

9 REFERENCES

[1] Ortiz, M., Hussein, D., Park, S., Han, S. N., & Crespi, N. (2014). The cluster between Internet of Things and social networks: Review and research challenges. *IEEE Internet Things J., 1*(3), 206-215. https://doi.org/10.1109/JIOT.2014.2318835

Hassanien, E., Bhatnagar, R., Eldeen, N., Khalifa, M., & Taha, M. H. N. (2020). Toward Social Internet of Things (SIoT): Enabling Technologies, Architectures and Applications. *Springer: Studies in Computational Intelligence book series SCI, vol. 846*. https://doi.org/10.1007/978-3-030-24513-9

[2] Atzori, L., Girau, R., Pilloni, V., & Uras, M. (2019). R2: Assignment of sensing tasks to IoT devices: Exploitation of a social network of objects. *IEEE Internet of Things Journal, 6*(2), 2679-2692. https://doi.org/10.1109/JIOT.2018.2873501

[3] Girau, R., Anedda, M., Fadda, M., Farina, M., Floris, A., Sole, M., & Giusto, D. (2020). Coastal monitoring system based on Social Internet of Things platform. *IEEE Internet of Things Journal, 7*(2). https://doi.org/10.1109/JIOT.2019.2954202

[4] Roopa, M. S., Pattar, S., Buyya, R., Venugopal, K. R., Iyengar, S. S., & Patnaik, L. M. (2019). Social Internet of Things (SIoT): Foundations, thrust areas, systematic review. *Computer Communications, 139*, 32-57. https://doi.org/10.1016/j.comcom.2019.03.009

[5] Chen, I. R., Bao, F., & Guo, J. (2016). Trust-based service management for social Internet of Things systems. *IEEE Transactions on Dependable and Secure Computing, 13*(99), 1-1. https://doi.org/10.1109/TDSC.2015.2420552

[6] Chen, I. R., Guo, J., & Bao, F. (2016). Trust management for SOA-based IoT and its application to service composition. *IEEE Transactions on Services Computing, 9*(3), 482-495. https://doi.org/10.1109/TSC.2014.2365797

[7] Truong, N. B., Um, T. W., & Lee, G. M. (2016). A reputation and knowledge-based trust service platform for trustworthy social internet of things. *Innovations in Clouds, Internet and Networks (ICIN)*, Paris, France.

[8] Kim, B.-S., Hassan, S., & Khan, M. K. (2018). Trust management techniques for the Internet of Things: A survey. *IEEE Access, 7*, 29763-29787. https://doi.org/10.1109/ACCESS.2018.2880838

[9] Chen, J., Tian, Z., Cui, X., Yin, L., & Wang, X. (2019). Trust architecture and reputation evaluation for internet of things. *Journal of Ambient Intelligence and Humanized Computing, 10*(2), 3099-3107. https://doi.org/10.1007/s12652-018-0887-z

[10] Xiao, H., Sidhu, N., & Christianson, B. (2015). Guarantor and reputation-based trust model for social internet of things. *International Wireless Communications and Mobile Computing Conference (IWCMC 2015)*, 600-605. https://doi.org/10.1109/IWCMC.2015.7289151

[11] Lin, Z. & Dong, L. (2018). Clarifying trust in social Internet of Things. *IEEE Transactions on Knowledge and Data Engineering, 30*(2). https://doi.org/10.1109/TKDE.2017.2762678

[12] Jayasinghe, U., Lee, G. M., Um, T. W., & Shi, Q. (2019). Machine learning based trust computational model for IoT services. *IEEE Transactions on Sustainable Computing, 4*(1). https://doi.org/10.1109/TSUSC.2018.2839623

[13] Adewuyi, A. A., Cheng, H., Shi, Q., Cao, J., MacDermott, Á., & Wang, X. (2019). CTRUST: A Dynamic Trust Model for Collaborative Applications in the Internet of Things. IEEE Internet of Things Journal, 6(3), 5432-5445. https://doi.org/10.1109/JIOT.2019.2902022

[14] Abdelghani, W., Zayani, C. A., Amous, I., & Sèdes, F. (2019). Trust Evaluation Model for Attack Detection in Social Internet of Things. In: Zemmari A., Mosbah M., Cuppens-Boulahia N., Cuppens F. (eds) *Risks and Security of Internet and Systems, CRiSIS 2018. Lecture Notes in Computer Science, vol 11391*. Springer, Cham, 48-64. https://doi.org/10.1007/978-3-030-12143-3_5

[15] Xia, H., Xiao, F., Zhang, S., Hu, C., & Cheng, X. (2019). Trustworthiness inference framework in the social internet of

things: A context aware approach. *IEEE Infocom 2019 - IEEE Conference on Computer Communications*, 838-846. https://doi.org/10.1109/INFOCOM.2019.8737491

[16] Xia, H., Hu, C.-Q., Xiao, F., Cheng, X.-G., & Pan, Z.-K. (2019). An efficient social-like semantic-aware service discovery mechanism for large-scale Internet of Things. *Computer Networks, 152*, 210-220. https://doi.org/10.1016/j.comnet.2019.02.006

[17] Azad, M. A., Bag, S., Hao, F., & Shalaginov, A. (2020). Decentralized self-enforcing trust management system for social internet of things. *Internet of Things Journal IEEE, 7*(4), 2690-2703. https://doi.org/10.1109/JIOT.2019.2962282

[18] Caminha, J., Perknnusich, A., & Perkusich, M. (2018). A smart trust management method to detect on-off attacks. *Hindawi Security and Communication Networks, Vol. 2018*, Article ID-6063456. https://doi.org/10.1155/2018/6063456

[19] Truong, N. B., Lee, H., Askwith, B., & Lee, G. M. (2017). Toward a trust evaluation mechanism in the Social Internet of Things. *Sensors 17*(6). https://doi.org/10.3390/s17061346

[20] Chen, Z., Ling, R., Huang, C.-M., & Zhu, X. (2016). A scheme of access service recommendation for the Social Internet of Things. *Int. J. Commun. Syst. 29*(4). https://doi.org/10.1002/dac.2930

[21] Abderrahim, O. B., Elhdhili, M. H., & Saidane, L. (2017). TMCoI-SIOT: A trust management system based on communities of interest for the Social Internet of Things. *Wireless Communications and Mobile Computing Conference, IWCMC 2017*, IEEE, 747-752. https://doi.org/10.1109/IWCMC.2017.7986378

[22] Masmoudi, M., Abdelghani, W., Amous, I., & Sèdes, F. (2019). Deep learning for trust-related attacks detection in social internet of things. *International Conference on e-Business Engineering*, Springer. https://doi.org/10.1007/978-3-030-34986-8_28

[23] Abbas, A. H. & Iqbal, F. (2019). Context based trust formation using direct user-experience in the Internet of Things (IoT). *2019 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, Sydney, Australia, 424-430.

[24] Oualhaj, O. A., Mohamed, A., Guizani, M., and Erbad, A. (2020). Blockchain based decentralized trust management framework. *International Wireless Communications and Mobile Computing (IWCMC 2020)*, Limassol, Cyprus, 2210-2215. https://doi.org/10.1109/IWCMC48107.2020.9148247

[25] Kowshalya, A. M. & Valarmathi, M. L. (2017). Trust management for reliable decision making among social objects in the Social Internet of Things. *IET Networks, 6*(4), 75-80. https://doi.org/10.1049/iet-net.2017.0021

[26] He, Y., Han, G., Jiang, J., Wang, H., & Martinez-Garcia, M. (2020). A trust update mechanism based on reinforcement learning in underwater acoustic sensor networks. *IEEE Transactions on Mobile Computing*. https://doi.org/10.1109/TMC.2020.3020313

[27] Sagar, S., Mahmood, A., Sheng, Q. Z., & Zhang, W. E. (2020). Trust computational heuristic for social internet of things: A machine learning-based approach. *IEEE International Conference on Communication*. https://doi.org/10.1109/ICC40277.2020.9148767

[28] Militano, L., Orsino, A., Araniti, G., Nitti, M., Atzori, L., & Iera, A. (2016). Trusted D2D-based data uploading in in-band narrowband-IoT with social awareness. *IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 1-6. https://doi.org/10.1109/PIMRC.2016.7794568

Authors' contacts:

Ms **Priyanka Hankare**
(Corresponding author)
AISSMS IOIT, Pune
RTO road, Sangamwadi, Pune-411001, India
E-mail: priyankahankare92@gmail.com

Mr **Sachin Babar**
STES, Lonavala
Kusgaon (BK) off. Mumbai-Pune Expressway, Lonavala-410401, India
E-mail: sdbabar@gmail.com

Mr **Parikshit Mahalle**
SKNCOE, Pune
Wadgaon (BK), Pune-411041, India
E-mail: aalborg.pnm@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210204181812

Original scientific paper

Churn Prediction of Employees Using Machine Learning Techniques

Nilasha Bandyopadhyay*, Anil Jadhav

Abstract: Employees are considered as the most valuable assets of any organization. Various policies have been introduced by the HR professionals to create a good working environment for them, but still, the rate of employees quitting the Technology Industry is quite high. Often the reason behind their early attrition could be due to company-related or personal issues, such as No satisfaction at the workplace, Fewer opportunities for learning, Undue Workload, Less Encouragement, and many others. This paper aims in discussing a structured way for predicting the churn rate of the employees by implementing various Classification techniques like SVM, Random Forest classifier, and Naives Bayes classifier. The performance of the classifiers was compared using metrics like Confusion Matrix, Recall, False Positive Rate, and Accuracy to determine the best model for the churn prediction. We found that among the models, the Random Forest classifier proved to be the best model for IT employee churn prediction. A Correlation Matrix was generated in the form of a heatmap to identify the important features that might impact the attrition rate.

Keywords: attrition; churn rate; classification techniques; confusion matrix; feature selection

1 INTRODUCTION

"Attrition" is not a new term for us anymore, as it has become an unavoidable situation in any business or organization, where staff and employees tend to leave due to their personal and professional circumstances. Further, this can cause a huge impact on any organization's growth curve if it is not given any attention, soon [1]. The major battle of employee attrition is right now being fought by the Technology Industries in India. Analysis from LinkedIn shows us that the software industry suffers from the highest turnover rates, which is about 13.2% compared to retail, entertainment, and professional industries. As per Maren Hogan, a talent acquisition expert, following points needs attention:

1. One-third of the new joiners quit, after six months in an organization.
2. After a week of working in a company, few decide on whether they want to continue staying there for the long term or not.
3. Also, a third of heads in companies having more than 100 employees are searching for new job opportunities [2].

Today's Millennial crowd in organizations is often identified as "job-hoppers", as they frequently change or quit their jobs to get to the next step of their career, as compared to the past generations. Rather than staying loyal to one company they often tend to search for better opportunities so that they can keep up in the era of digital progression. If we dig deep then we can find a distinct set of challenges faced by them like industry, proper recognition, communication, ethnicity, age, gender, etc. that drive the employees to leave a particular organization. Challenges faced by talent-hiring consultants are, sorting out the appropriate candidates through resumes and conversation, who will become the asset of the organization, and then if a person quits they need to repeat the entire hiring process. Every time hiring new talent and training them in current technologies involves a great amount of cost to the organization. Apart from this tangible expense, a fair amount of time we need to give the newly employed person to become a productive member of the project [3].

The Human Resource department of any organization generates a plethora of data related to employee's leave, promotion cycle, rewards, wages, various evaluations, conflicts, policies, and benefits. As a researcher, our work is to identify the correct parameters or areas where the employees face issues regularly at their workplace.

In this data-driven study, we will try to analyze the employee's data using some classification techniques and will provide quality insights and suggestions, so that the organization can retain them as well as develop them before it's too late. As HR professionals or managers our main focus should always be on an individual or certain groups of employees, especially towards their specific needs or their situation, then only it can further help an organization to grow more without losing good employees.

1.1 Research Objective

In our study, we will analyze the data of the Technology Professionals, especially their challenges that they face directly or indirectly at their workplace.

Main objectives of this study are:

1. To identify those challenges or input variables that have a huge impact on the employee's intention to leave the organization.
2. To accurately predict which employee will leave the organization in the next few years, using classification models.

2 LITERATURE REVIEW

An evidence-based study by Janet et al. (2017) has combined the already published scholarly reviewed literature on HR Analytics and has concentrated on answering major questions on HR Analytics, how it works, its outcome, and why there is a need for HR Analytics to flourish? They have stated that the interest of people in analytics in the HR domain for the past few years has gradually increased [4].

Later, the authors concluded that the inclusion of HR Analytics in various organizations is very low and proofs on this topic are scattered, hence suggested areas for future research. Many firms or departments say Marketing, Finance, Supply Chain Management organizations today draw insights from the huge data collected from the employees so that they can stay in this competition. The Human Resource department generates massive amounts of data on employee turnover, Return on Investments, and Cost per hire, but somewhere they still face a harder time relating these data with the organization's performance. They should create reports on past performances, administrative tasks, and generate compliance reports to understand the employee's contribution to the organization [5].

HR applications followed by today's organization can act as a mediator between planned HR practices in an organization and the positive outcomes of employees. Hence, Innocenti et al. (2012) have proposed a model that uses survey data that has been collected from over 6000 employees working in almost 37 Italian organizations, and the outcome variables are employee commitment and their job satisfaction. By using the maximum likelihood estimation method and calculating the correlations between different variables it was reported that, there is always a positive effect of experienced HR practices on both affective engagement towards organization and job satisfaction factors [6].

Line managers are considered the assets of that particular organization, so it's necessary to keep them engaged so that they can add value to any organization. Few semi-structured interviews were performed by Sana et al. (2016), to understand the experience and perceptions of the line managers on the level of support and help provided by the HR professionals of their organization [7]. Further, they have stated that the line managers have raised concerns and have suggested ideas for improving few areas like perceptions regarding policies, workload, inadequate training, and HR practices, which we need to pay attention to during any research on the factors related to employee attrition or turnover.

There have been several studies on identifying the parameters that play a role in job satisfaction of the employees and predicting the attrition rate. Many Data Analytics techniques and classification models have been used to predict turnover. In any organization, innovation can be seldom duplicated but once a group of productive employees leaves, that place cannot be replicated easily. So, to retain these employees and predict the turnover rate, a Neural Network, with a 10-fold Cross-Validation was designed for a small Midwest manufacturing company to a greater accuracy [8]. Among Layoffs, Discharges, Unavoidable separation, it was identified that voluntary separation from an organization always proves as the most difficult area because the particular organization loses its investment on talent to its competitors out there. On this same note, Fan et al. (2012) in their study, focused on why technology enterprises in Taiwan are unable to retain their talented employees and they have discussed ways so that the organizations can increase the competitiveness among themselves. Techniques like clustering analysis, hybrid artificial neural networks and other machine learning techniques were applied to forecast the patterns of employee's turnover rate [9]. Again, many Classification models have been used for prediction purposes, on a HR analytics dataset from Kaggle, an online community data site. Correlations between different attributes were evaluated by Sisodia et al. (2017) in their paper [10]. A comparison between different classifiers was drawn using parameters like Accuracy, Precision, True Positive Rate, F-Measure, and few others. Weighted TPR-TNR has been proposed as another performance metric to evaluate the performance of various classifiers, as it especially focuses on the imbalance ratio of any dataset and assigns different weights to TPR (Sensitivity) and TNR (Specificity), which are majorly considered while comparing ROC curve of any model. A mix of balanced and imbalanced datasets was used to evaluate the performance of 12 classifiers using the above metric [11].

To build and maintain a strong relationship between an organization and its employees, Hebbar et al. (2018), in their study initially implemented Logistic Regression on an IBM Employee Attrition dataset available in Kaggle just to get a basic idea, on which outcome group every individual falls [12]. Later on, a comparative study was done with SVM and Random Forest models, and determined the major characteristics of the dataset performing Exploratory Data Analysis and represented the data using different visualization.

With the same dataset (that has been used above), Synthetic Minority Oversampling Technique (SMOTE) was performed by Bhartiya et al. (2019) in their paper, to balance the imbalance dataset, because the count of the "Attrition" parameter with value 0 was greater than "Attrition" with a value of 1. The above technique is often used to generate synthetic data records for that class whose count is very less. Attributes like Gender, Education Field, and Performance Rate were visualized for Attrition parameters thus giving an idea on the relevant features. A comparison between the performance metrics of the classification models provided new insights on improving the work ethics [13].

With redundant data, predicting the correct features becomes a little challenging. So, a superior machine learning model or algorithm called XGBoost gives high accuracy in predicting the attrition rate with fewer running times. Jain et al. (2018) recommend XGBoost as a highly robust model, which easily handles noisy data in a huge dataset, and in their study, it gives an accuracy of about 90% on an online HR dataset [14]. Further, it suggests IT organizations to use this as a top priority, predictive model to identify those employees who are willing to leave in near future and their reasons behind that.

A very common issue that today's IT professionals face is stress disorders. Though organizations do offer a nice workplace environment and different activities or workshops to relieve this stress, still the risk increases among the employees. Various machine learning techniques like Boosting and Decision trees were implemented by Reddy et al. (2018) in their study, and have determined that data on family history of illness, gender and health benefits provided

by employers plays an important role in evaluating this type of risks [15]. Ensemble method gave the highest degree of accuracy and precision compared to Random Forest. General characteristics like having peers to work with and the financial needs of the employees become critical factors for those who are working for a longer tenure in any business or organization. So, for the hospitality industry in the USA, Self et al. (2011) attempted a qualitative study on identifying various factors that might impact an employee's decision to stay back in a company. By analyzing the interview transcripts that were obtained after an in-depth process, four factors were identified: Strong Responsibility towards the company, Financial Requirement, Proper Job Description, and Peers at the workplace has a positive effect on employees [16].

One of the challenges that the big organizations are facing is, motivating their employees and investing in them for their further development. Understanding the importance of investing in employee development and its final results, is very much needed by the organization. A model proposed by Lee et al. (2003) gives us an interrelationship between perceived investments and other job attitudes and the employee's plan to quit an organization. Factor analysis and Exploratory analysis were conducted for assessing the dimensionality and their insights, respectively. Results suggest that the more the employer spends resources on the development of their employees, the more they will be satisfied at their workplace, hence reducing the possibility of an employee quitting his or her job in that organization [17].

Burnett et al. (2019) propose a few topics on which one can use modern technology or tools to measure both employee engagement and the other HRM practices which can improve the same [18]. Different emotional states of employees affect their engagement at the workplace, either directly or indirectly. Further, they have pointed out that to improve on engagements we need to concentrate on three different levels: individual, team and organizational level and have suggested that with the real-time feedback from employees and rigorous research and analysis on the data will help the HRM department to understand the importance of employee engagement in their respective organization.

So, to stay in this competitive market, these technology industries need to continuously evolve in terms of skills and should be ready to embrace the ever-changing products and services. Even employees make themselves proficient in the new skills or technologies and try to search for better job opportunities outside. An analytics-driven approach can help organizations to overcome the situation. Combining the historic record of skills of each employee present in the HR database with the predictive models, Ramamurthy et al. (2015) have proposed an approach that evaluates a set of skills [19]. The algorithm in their study will provide a list of skills to some individuals, where they will fill in their target skills, helping business leaders to find potential candidates and will provide re-skilling offers to them.

One can go for Sentiment Analysis to determine the factors affecting employee retention, and organizations can use these models to understand the concepts of People Analytics. A conceptual study was done to identify key indicators to assess the human factors. Six important areas, like performance leadership, employee engagement, learning, workplace dynamics, and overall organizational development have used sentiment analysis to evaluate various insights. The Enron email corpus test case was incorporated to explain how we can predict the digital footprints. Further encourages implementing various data mining techniques or models to analyze the real-time data for predicting more accurate human factor patterns [20]. In addition to this, often interpersonal environment factors provide insights about employee development in any organization. Liu et al. (2019) in their study have concentrated on a state-owned enterprise in China, extracted the related features, and statistically analyzed the correlation between employee development in organizations and their interpersonal environment. The results of the predictive model prove that colleagues and classmates have a great impact on the growth of employees in their respective workplaces [21].

2.1 Research Gap

After reviewing the existing work, it was observed that many of the studies were following secondary data which is a HR analytics dataset available in an open-source dataset site, to predict employee turnover using Data Mining Techniques. The attributes that they have considered in their study are the generic parameters related to any employee who has already left the organization. Today, if we discuss with the IT professionals, we will get to know that they still face a set of challenges, both at their workplace and in their personal life which results in early attrition. This set of challenges often goes overlooked in this industry by the HR executives.

Every new employee who gets recruited might face a different set of challenges while working. So, analyzing the data of those employees who have already left the organization might not give us the features that apply to the new joiners. Rather, we need to interact with them frequently or take their feedback on a real-time basis, just to get the actual data related to their challenges, like Recognition, Challenging work, Scope of Development, Satisfaction Level, Unhealthy work ethics and Impact on them of their peers leaving an organization. For this reason, we are using primary data in our study that has been collected from employees working in various IT industries.

We need to concentrate more on discussing what they want for their betterment in this organization. Then start predicting who might leave within a couple of years, post this we can offer them proper opportunities. This will not only encourage the employees but will help the organization in retaining its talent.

3 RESEARCH METHODOLOGY

This study is focused on employees from a specific age group that is from 20 to 39 years old, who are considered to be the major contributors to the highest turnover of any organization. In this research, surveys were conducted to get

the raw data from the employees, which is first pre-processed, and then analysis was done to derive meaningful insights.

A questionnaire consisting of 35 questions was circulated among 200 employees and the response rate was around 79%. Among these responses, 83 were male and 75 were female employees. 80% of these employees had working experience of 4 years or less, the remaining 20% had an experience that varies from more than 4 years to 13 years. This survey had combinations of few open and close-ended questions, which includes a Likert scale and few dichotomous answer types. This will help us understand the actual perception of the employees regarding the organization or employer.

The entire questionnaire was designed based on our detailed review of the previous work that has been done by other researchers in this topic and our discussions with a few experts who are involved in the technology industries. Further, these questions have been divided into 5 sections, like Individual Beliefs, Management and Team, Engagement and Encouragement, Talent Development, Organisation and Leadership, to get an overall idea of the employees towards different verticals of an organization.

We are implementing and analyzing a few classification models in R studio.

3.1 Input Data Set

The data collected includes 11 attributes for each employee. The target variable "Quit_in_2years" consists of three classes, they are: "Maybe", "No" and "Yes", thus our study is a multi-class classification. Tab. 1 gives us the details on the attributes that will be used in our study:

Table 1 Dataset Attributes

Sl. No.	Attributes	Data types	Description
1	Age	Numeric	Age of the employees.
2	Gender	Categorical	Gender of the employees
3	Salary_Level	Numeric	Salary window under which the employees fall.
4	Years_of_Experience	Numeric	For how many years that employee is associated with that organisation
5	Satisfaction_Level	Numeric	Degree of satisfaction of the employee at their workplace
6	Discrimination	Numeric	Any discrimination faced based on age, gender or ethnicity
7	Work_Recognition	Numeric	Been given proper recognition of their work in their team or not
8	Challenging_Work	Numeric	To what degree the Employees feel challenged with their daily work
9	Promotion_in_last_year	Numeric	Got promotion in last one year
10	Peers_Leaving	Categorical	To what degree does good colleagues or friends leaving organisation affects them
11	Quit_in_2years	Categorical	Their plan to leave the organisation in 2 years

3.2 Data Pre-Processing

Among the 11 attributes, "Gender", "Peers_Leaving" and the target variable, "Quit_in_2years" are categorical data types. So, to determine the impact of the above predictors on the target variable and evaluate the correlations among the attributes, the categorical fields were converted to numeric values. For example, "Female" was denoted by 1 and "Male" as 2. Under "Peers_Leaving" there were three categories, where "Yes" and "No" were given 1 and 0, respectively, while "Maybe" was denoted as 0.5. Similarly, the values of "Maybe", "Yes" and "No" for the target variable were denoted as 1, 2 and 3 respectively.

Though the null values in the dataset were really less, it was chosen to be replaced by the mean of the whole column rather than dropping the whole entry. To summarize the whole data, and to determine how close these variables have a linear relationship among themselves, we plotted a correlation matrix. This gives us an idea of identifying the features which have weak and strong dependencies.

For example, in Fig. 1, the darkest blue on the scale means there is a positive correlation among the attributes, whereas the dark red means a negative correlation. In the above figure, it can be observed that there is a stronger relationship between "Age" and "Years_of_Experience", again "Satisfaction_Level" and "Work Recognition has a positive correlation, with a coefficient of 0.53. The rest of the variables do not have a strong consistent relationship with each other. We observe that there are Negative Coefficients in the above matrix, this indicates that if the value of one attribute increases then the value of the other attribute will tend to decrease.

Figure 1 Correlation Matrix of the attributes

3.3 Feature Selection and Ranking

This approach helps in recognizing the correct features in any dataset, where we can easily differentiate the features that play a significant role in predicting employee's intention

to leave in the next 2 years, from the other features. Further, it will help in building a reliable model, with greater accuracy. Here, an R package known as "caret", is being used which will automatically give us a report on the importance and relevance of the attributes in our dataset and will help in ranking those features.

So for the feature selection process, RFE (Recursive Feature Elimination) is chosen, which is majorly used with SVM to continuously build a model and simultaneously remove those features that have low weights and discover the optimal number of features. The algorithm is configured to explore all possible subsets of the attributes. Next, to specify ranks to the feature by importance, a method known as LVQ (Linear Vector Quantization) was used, which is a form of ANN (Artificial Neural Network) algorithm and allows us to choose the training instances and learn what those instances should look like.

In Fig. 2, we have ranked all the features as per the target classes. So, it can be inferred that among all the 11 features, "Satisfaction_Level", "Salary_Level", "Work_Recognition", "Gender" and "Challenging_Work" are the top 5 challenges that have a huge effect on the target variable, that is, "Quit_in_2 years". Whereas, "Promotion_in_last_year" and "Peers_Leaving" have the least impact on the employee's decision on leaving the organization in the future.

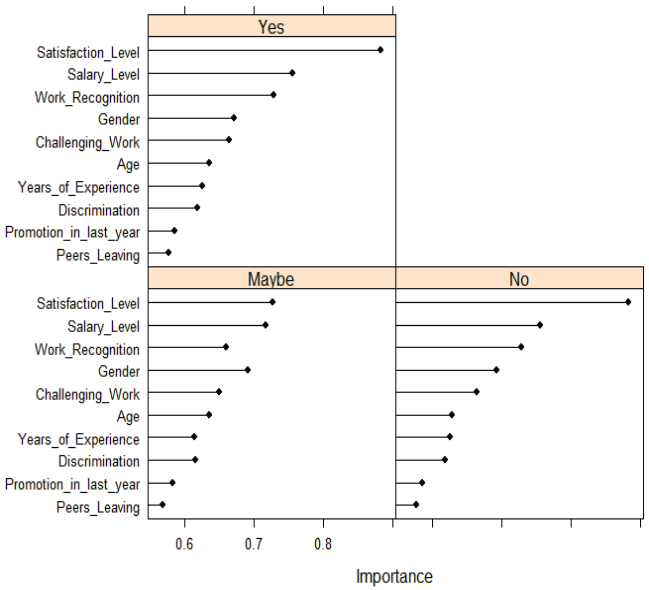


Figure 2 Ranks of the attributes

4 MODELS AND IMPLEMENTATION

In this research, three classification models were used to predict, whether a particular employee will leave the organization or not, based on the challenges he or she is facing currently at the workplace. Here, the classifiers that we are going to implement are the Random Forest classifier, SVM (Support Vector Machine), and Naive Bayes.

As per our research framework, after preprocessing the data, it was split into two parts, that is, train and test dataset in the 70:30 ratio. Trained our classification models by passing the training dataset and then evaluated the most efficient model by predicting the target value using the test dataset.

Our study on analyzing the performance metrics of the models has been bifurcated into two cases.
Case 1: Includes all the three classes of the Target Variable.
Case 2: Here we are including only two classes, that is, "Yes" and "No" of the Target Variable.

56 Employees who are still in the dilemma of whether they will leave their organization or not might affect the accuracy of the model. Hence, we removed them in Case 2 and analyzed the performance metrics.

4.1 Support Vector Machine

It comes under supervised learning techniques, majorly used for classification of data but is often implemented for regression problem statements. In this technique, the data points are separated from each other by a line or a hyperplane, and this division between the two sides categorizes the whole data sets into two or more classes. The space between the two classes is also known as margin, and this should be as large as possible so that we can reduce the error while classification. Package "e1071" is used for the implementation of the said model.

Tab. 2 gives us the Confusion Matrix of SVM, which includes all the classes of the Target variable whereas, Tab. 3 represents the Confusion Matrix for only two classes.

4.1.1 Case 1: Including all the three classes of the Target Variable

Table 2 Confusion Matrix of Test Dataset

		Actual		
		Maybe	No	Yes
Predicted	Maybe	9	2	8
	No	1	6	0
	Yes	5	3	14

4.1.2 Case 2: Including only two classes of the Target Variable (without "Maybe")

Table 3 Confusion Matrix of Test Dataset

		Actual	
		Yes	No
Predicted	Yes	14	5
	No	4	8

4.2 Naive Bayes Classifier

The crux of this classification method is based on the famous Bayes Theorem. It assumes that a particular feature or attribute in a class is independent of the existence of any other feature. The model is easy to build and is particularly useful if we have a huge dataset. With its simplicity in the model, Naive Bayes can outperform other sophisticated classification models for multi-class prediction. Below Tab. 4 and Tab. 5 are the Confusion Matrices for the above model for two different cases that we are considering in our study.

4.2.1 Case 1: Including all the three classes of the Target Variable

Table 4 Confusion Matrix of Test Dataset

		Actual		
		Maybe	No	Yes
Predicted	Maybe	10	2	6
	No	0	7	0
	Yes	5	2	16

4.2.2 Case 2: Including only two classes of the Target Variable (without "Maybe")

Table 5 Confusion Matrix of Test Dataset

		Actual	
		Yes	No
Predicted	Yes	16	5
	No	2	8

4.3 Random Forest Classifier

This model is an ensemble tree-based learning technique. Rather than using a single decision tree for classification of the data, it uses a set of decision trees that randomly selects subsets of data and train the model. Voting will be performed on the predictions from each of these trees and finally, the best solution will be selected. This method helps reduce the overfitting by averaging the results, as compared to traditional decision trees.

For the implementation of the classifier, a package called "randomForest" is used in our study. We can observe the values of predicted and actual instances from Tab. 6 and Tab. 7.

4.3.1 Case 1: Including all the three classes of the Target Variable

Table 6 Confusion Matrix of Test Dataset

		Actual		
		Maybe	No	Yes
Predicted	Maybe	12	1	5
	No	0	7	2
	Yes	3	3	15

4.3.2 Case 2: Including only two classes of the Target Variable (without "Maybe")

Table 7 Confusion Matrix of Test Dataset

		Actual	
		Yes	No
Predicted	Yes	15	4
	No	3	9

5 RESULTS AND DISCUSSION

So, to choose the best classifier for this study, we are comparing the existing performance metrics, say Model's Accuracy, Recall, Specificity, Precision, F-Measure, Area Under Curve (AUC) and another metrics that we are considering is Weighted TPR-TNR. For comparing the results of multi-class classification we are using the Macro Average Method for parameters like Recall (Sensitivity), Specificity, Precision, and F-Measure. This method helps in determining the performance of the overall system. As our data is a balanced dataset, we are using this method to calculate the average of the values that we obtained for each class.

As per our problem statement, we are mainly concerned with the people leaving the organization, thus to acquire complete knowledge to overcome this, parameters like Recall and AUC play a huge role along with the accuracy of the models.

5.1 Case 1: Comparing the performance metrics for all the Target Variable classes

Table 8 Final Results of the Classifiers

Metrics	SVM	Naive Bayes	Random Forest
Accuracy	60.42%	68.75%	70.83%
Recall or TPR or Sensitivity	0.5939	0.6768	0.7061
FNR	0.4061	0.3232	0.2939
TNR or Specificity	0.7874	0.8294	0.8445
FPR	0.2125	0.1705	0.1555
Precision	0.6558	0.7504	0.7196
F-Measure	0.6108	0.6983	0.70833
Weighted TPR-TNR	0.6584	0.7277	0.7522
AUC	61.64%	67.31%	73.48%

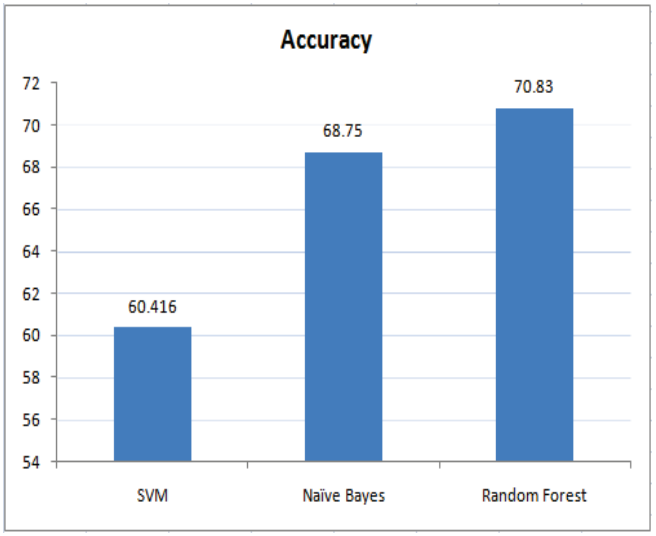


Figure 3 Accuracy of all the Models

From the above graph, we observe that the Random Forest classifier has achieved a far better prediction accuracy of 70.83% when compared to other classifiers.

Simultaneously, one must look for the Recall and Precision value apart from the model's accuracy. From Tab. 8 we can see that for Random Forest classifier the Recall value has increased but the Precision value is slightly less than Naive Bayes. Values of weighted TPR-TNR are the highest in the case of Random Forest than the other two models.

ROC curve is a trade-off between sensitivity and specificity, where the curve of a perfect classifier should have the highest Recall (True Positive Rate) with the lowest False Positive Rate. So, to summarize the performance of the classifiers we take the calculated area under the ROC curve

into consideration, which is also known as AUC. So, the higher the AUC, the greater will be the accuracy of the model.

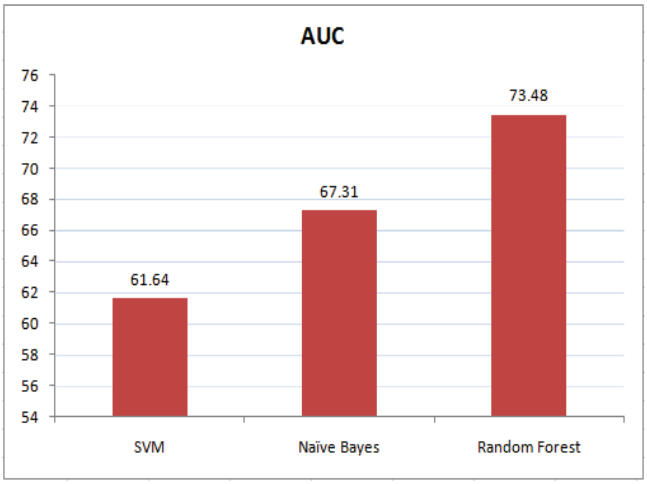


Figure 4 Comparing the AUC of the Models

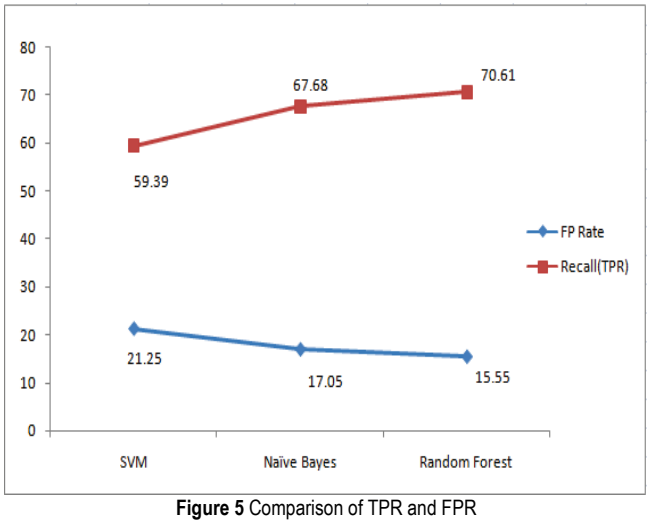


Figure 5 Comparison of TPR and FPR

With the highest AUC and lowest False positive rate, the Random Forest classifier stands out from the rest of the models.

5.2 Case 2: Comparing the performance metrics for only two Target Variable classes (without "Maybe")

Compared to Case 1, it can be observed that the accuracy of each model has increased by quite a percentage after excluding those employees who still had some difficulty in deciding on leaving their organization in the next two years. Both Naive Bayes classifier and Random Forest classifier have obtained an accuracy of 77.42%. Recall and F-Measure value of Naives Bayes is greater than the other two models, whereas if we observe Tab. 9, we can state that the Precision and weighted TPR-TNR for Random Forest classifier has increased, compared to SVM and Naive Bayes.

Table 9 Final results of the Classifiers

Metrics	SVM	Naive Bayes	Random Forest
Accuracy	70.97%	77.42%	77.42%
Recall or TPR or Sensitivity	0.7778	0.8889	0.8333
FNR	0.2222	0.1111	0.1667
TNR or Specificity	0.6154	0.6154	0.6923
FPR	0.3846	0.3846	0.3077
Precision	0.7368	0.7619	0.7895
F-Measure	0.7568	0.8205	0.8108
Weighted TPR-TNR	0.6834	0.7301	0.7513
AUC	69.66%	75.21%	76.28%

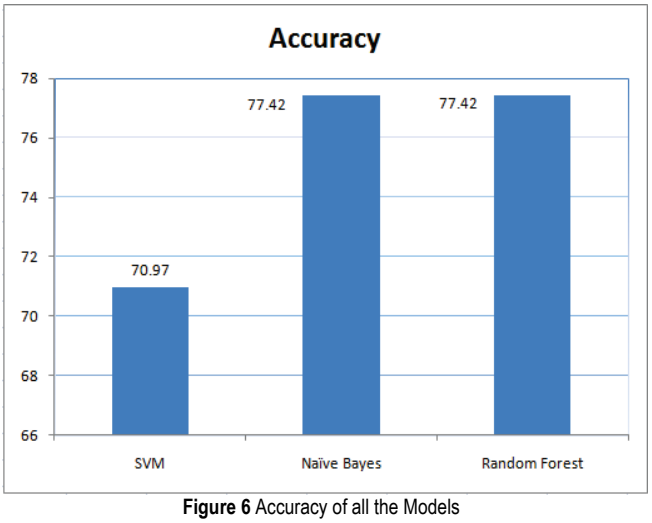


Figure 6 Accuracy of all the Models

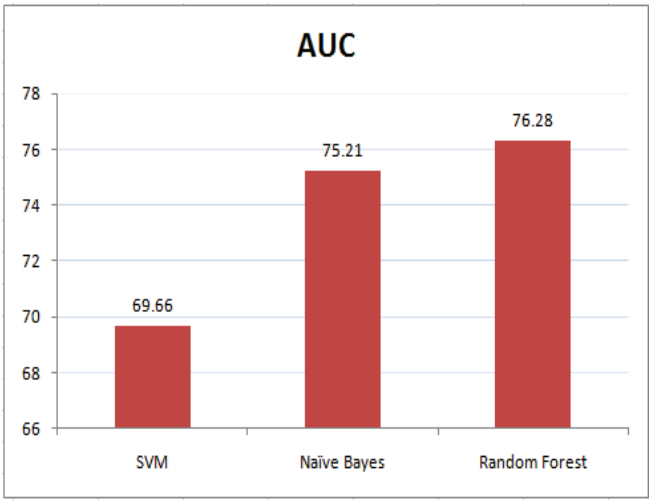


Figure 7 Comparing the AUC of the Models

Considering the area under the ROC curve we can see from the above figure, that the Random Forest classifier still has a lead of 6.62% from SVM and 1.07% from Naive Bayes classifier.

So, it can be stated that with the lowest False Positive Rate and highest AUC, Random Forest Classifier proves to be a good model in this case as well.

Adding to this, as our study focuses more on predicting employees who might leave in near future, we should never forget the False Negatives in this case. That is, those employees who are planning to leave but the model somehow does not predict them correctly. We need to identify these False Negatives and should find ways to reduce this.

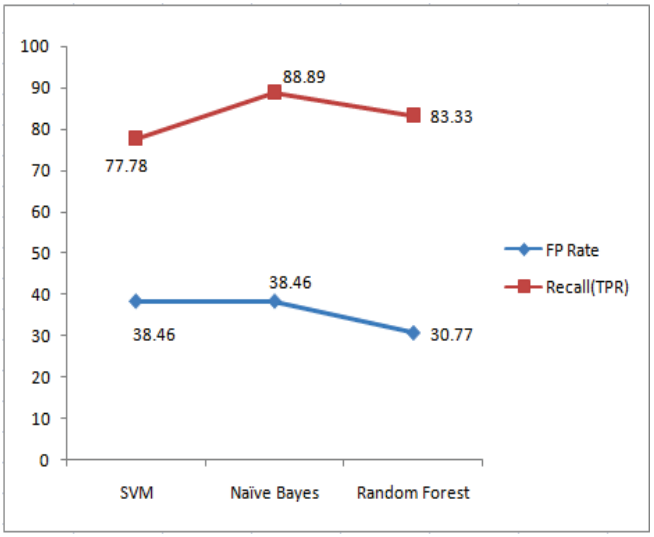


Figure 8 Comparison of TPR and FPR

6 CONCLUSION

As per our discussion above, employees leaving organization has a major impact on the development of these technology organizations. Often the challenges or issues faced by the employees at the workplace or in their personal life have a great impact on their early attrition from the organization.

In our study we have identified the important factors that affect the employees, resulting in future attrition. To help with the analysis, data were collected from professionals working in IT industries. Majority of the attributes we considered did not have a significant correlation with each other. Further to get the top features that have a positive impact on employees, a method called RFE (Recursive Feature Elimination) was chosen for variable selection. This method helped in removing redundant and less important variables and highlighted features which has more impact on the target attribute. In addition to this, LVQ (Linear Vector Quantization) was introduced to rank all the attributes as per their importance.

Secondly, our goal was to accurately predict those employees who are planning to leave the organization in the next 2 years, using a few classification models. Techniques like SVM, Naive Bayes, and Random Forest classifiers have been implemented in this study. So, to analyze the pattern we bifurcated our analysis into two cases. In the first case, we considered all the Target Variable classes but for the second case, we removed those employees who still had their doubts about leaving a particular organization in near future. Observing the results, we conclude that the models implemented in Case 2 gave good accuracy as compared to Case 1. The most efficient model in our study was the Random Forest classifier giving us the highest accuracy and Recall value when compared to the other models.

Apart from getting a good raise and promotion, there have been other kinds of challenges faced by today's talent, which the HR executives or managers of the project need to take care. In the future direction, this study can be further extended, by including attributes, like Scope of Development, Views on workload distribution, Career goal discussion and Issues on unhealthy work ethics.

Organizing frequent feedback or a one to one interview on the organization policies can help HR understand the expectations. In our study we had a limited data size of 158 entries, it is suggested that with more data points and features we can achieve higher accuracy from these models.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

7 REFERENCES

[1] Basu Mallick, C. (May 2020). *What is Employee Attrition? Definition, Attrition Rate, Factors and Reduction Best Practices*. https://hr.toolbox.com/articles/what-is-attrition-complete-guide

[2] (June 2019). *Tech industry battles highest attrition rate in the world - and it's costly.* https://www.viglobal.com/2018/06/13/tech-industry-battles-highest-attrition-rate-in-the-world-and-its-costly/

[3] Yadav, S., Jain, A., & Singh, D. (2018). Early Prediction of Employee Attrition using Data Mining Techniques. *2018 IEEE 8th International Advance Computing Conference (IACC)*, 349-354. https://doi.org/10.1109/IADCC.2018.8692137

[4] Marler, J. H. & Boudreau, J. (2017). An evidence-based review of HR Analytics. *The International Journal of Human Resource Management, 28*, 3-26. https://doi.org/10.1080/09585192.2016.1244699

[5] Harris, J., Craig, E., & Light, D. (2011). Talent and analytics: new approaches, higher ROI. *Journal of Business Strategy, 32*, 4-13. https://doi.org/10.1108/02756661111180087

[6] Innocenti, L. & Peluso, A., M. & Pilati, M. (2012). The Interplay between HR Practices and Perceived Behavioural Integrity in Determining Positive Employee Outcomes. *Journal of Change Management*, 12. https://doi.org/10.1080/14697017.2012.728763

[7] Anwaar, S., Nadeem, A., & Hassan, M. (2016). Critical assessment of the impact of HR strategies on employees' performance. *Cogent Business & Management, 3*. https://doi.org/10.1080/23311975.2016.1245939

[8] Sexton, R., McMurtrey, S., Michalopoulos, J., & Smith, A.M. (2005). Employee turnover: a neural network solution. *Comput. Oper. Res., 32*, 2635-2651. https://doi.org/10.1016/j.cor.2004.06.022

[9] Fan, C., Fan, P., Chan, T., & Chang, S. (2012). Using hybrid data mining and machine learning clustering analysis to predict the turnover rate for technology professionals. *Expert Syst. Appl., 39*, 8844-8851. https://doi.org/10.1016/j.eswa.2012.02.005

[10] Sisodia, D. S., Vishwakarma, S., & Pujahari, A. (2017). Evaluation of machine learning models for employee churn prediction. *2017 International Conference on Inventive Computing and Informatics (ICICI)*, 1016-1020. https://doi.org/10.1109/ICICI.2017.8365293

[11] Jadhav, A. S. (2020). A novel weighted TPR-TNR measure to assess performance of the classifiers. *Expert Syst. Appl., 152*, 113391. https://doi.org/10.1016/j.eswa.2020.113391

[12] Hebbar, A., Sanath, P., Rajeshwari, S., & Saqquaf, S. (2018). Comparison of Machine Learning Techniques to Predict the Attrition Rate of the Employees, 934-938.
[13] Bhartiya, N., Jannu, S., Shukla, P., & Chapaneri, R. (2019). Employee Attrition Prediction Using Classification Models. *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, 1-6. https://doi.org/10.1109/I2CT45611.2019.9033784
[14] Jain, R. & Nayyar, A. (2018). Predicting Employee Attrition using XGBoost Machine Learning Approach. *2018 International Conference on System Modeling & Advancement in Research Trends (SMART)*, 113-120. https://doi.org/10.1109/SYSMART.2018.8746940
[15] Reddy, U. S., Thota, A., & Dharun, A. (2018). Machine Learning Techniques for Stress Prediction in Working Employees, 1-4. https://doi.org/10.1109/ICCIC.2018.8782395
[16] Self, J. & Dewald, B. (2011). Why Do Employees Stay? A Qualitative Exploration of Employee Tenure. *International Journal of Hospitality & Tourism Administration, 12*, 60-72. https://doi.org/10.1080/15256480.2011.540982
[17] Lee, C. H. & Bruvold, N. T. (2003). Creating value for employees: investment in employee development. *The International Journal of Human Resource Management, 14*(6), 981-1000. https://doi.org/10.1080/0958519032000106173
[18] Burnett, J. & Lisk, T. C. (2019). The Future of Employee Engagement: Real-Time Monitoring and Digital Tools for Engaging a Workforce. *International Studies of Management & Organization, 49*, 108-119. https://doi.org/10.1080/00208825.2019.1565097
[19] Ramamurthy, K., Singh, M., Davis, M., Kevern, J.A., Klein, U., & Peran, M. (2015). Identifying Employees for Re-skilling using an Analytics-Based Approach. *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, 345-354. https://doi.org/10.1109/ICDMW.2015.206
[20] Gelbard, R., Ramon-Gonen, R., Carmeli, A., Bittmann, R., & Talyansky, R. (2018). Sentiment analysis in organizational work: Towards an ontology of people analytics. *Expert Syst. J. Knowl. Eng., 35*. https://doi.org/10.1111/exsy.12289
[21] Liu, J., Li, J., Wang, T., & He, R. (2019). Will Your Classmates and Colleagues Affect Your Development in the Workplace: Predicting Employees' Growth Based on Interpersonal Environment, 71-78. https://doi.org/10.1109/BigDataService.2019.00016
[22] Levenson, A. (2018). Using workforce analytics to improve strategy execution. *Human Resource Management, 57*, 685-700. https://doi.org/10.1002/hrm.21850
[23] Thite, M. (2010). All that Glitters is not Gold: Employee Retention in Offshored Indian Information Technology Enabled Services. *Journal of Organizational Computing and Electronic Commerce, 20*, 7-22. https://doi.org/10.1080/10919390903482390
[24] Srivastava, D. & Tiwari, P. (2020). An analysis report to reduce the employee attrition within organizations. *Journal of Discrete Mathematical Sciences and Cryptography, 23*, 337-348. https://doi.org/10.1080/09720529.2020.1721874
[25] Aliyu, O. & Nyadzayo, M., (2016). Reducing employee turnover intention: a customer relationship management perspective. *Journal of Strategic Marketing*, 1-17. https://doi.org/10.1080/0965254X.2016.1195864
[26] Schiemann, W. A., Seibert, J. H., & Blankenship, M. H. (2018). Putting human capital analytics to work: Predicting and driving business success. *Human Resource Management, 57*, 795-807. https://doi.org/10.1002/hrm.21843
[27] Robinson, M. (2018). Using multi-item psychometric scales for research and practice in human resource management. *Human Resource Management, 57*, 739-750. https://doi.org/10.1002/hrm.21852
[28] Hendrick, R. Z. & Raspiller, E. E. (2011). Predicting Employee Retention through Preemployment Assessment. *Community College Journal of Research and Practice, 35*, 895-908. https://doi.org/10.1080/10668920802421561
[29] Book, L., Gatling, A., & Kim, J. (2019). The effects of leadership satisfaction on employee engagement, loyalty, and retention in the hospitality industry. *Journal of Human Resources in Hospitality & Tourism*, 18, 1-26 https://doi.org/10.1080/15332845.2019.1599787
[30] Kryscynski, D., Reeves, C., Stice-Lusvardi, R., Ulrich, M., & Russell, G. (2018). Analytical abilities and the performance of HR professionals. *Human Resource Management, 57*, 715-738. https://doi.org/10.1002/hrm.21854
[31] Wei, D., Kush, R., & Wagman, M. (2015). Optigrow: People Analytics for Job Transfers, 535-542. https://doi.org/10.1109/BigDataCongress.2015.84

Authors' contacts:

Nilasha Bandyopadhyay, Student
(Corresponding author)
Symbiosis Centre for Information Technology, Pune
Plot No: 15, Rajiv Gandhi Infotech Park, MIDC, Hinjewadi, Phase 1, Pune, Maharashtra 411057, India
8458075463, nilasha.bandyopadhyay@associates.scit.edu

Dr. **Anil Jadhav,** Professor
Symbiosis Centre for Information Technology, Pune
Plot No: 15, Rajiv Gandhi Infotech Park, MIDC, Hinjewadi, Phase 1, Pune, Maharashtra 411057, India
9764294698, anil@scit.edu

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210204183323

Original scientific paper

Bio-Inspired Workflow Scheduling on HPC Platforms

Mandeep Kaur*, Sanjay Kadam

Abstract: Efficient scheduling of tasks in workflows of cloud or grid applications is a key to achieving better utilization of resources as well as timely completion of the user jobs. Many scientific applications comprise several tasks that are dependent in nature and are specified by workflow graphs. The aim of the cloud meta-scheduler is to schedule the user application tasks (and the applications) so as to optimize the resource utilization and to execute the user applications in minimum amount of time. During the past decade, there have been several attempts to use bio-inspired scheduling algorithms to obtain an optimal or near optimal schedule in order to minimize the overall schedule length and to optimize the use of resources. However, as the number of tasks increases, the solution space comprising different tasks-resource mapping sequences increases exponentially. Hence, there is a need to devise mechanisms to improvise the search strategies of the bio-inspired scheduling algorithms for better scheduling solutions in lesser number of iterations/time. The objective of the research work in this paper is to use bio-inspired bacteria foraging optimization algorithm (BFOA) along with other heuristics algorithms for better search of the scheduling solution space for multiple workflows. The idea is to first find a schedule by the heuristic algorithms such as MaxMin, MinMin, and Myopic, and use these as initial solutions (along with other randomly generated solutions) in the search space to get better solutions using BFOA. The performance of our approach with the existing approaches is compared for quality of the scheduling solutions. The results demonstrate that our hybrid approach (MinMin/Myopic with BFOA) outperforms other approaches.

Keywords: BFOA; bio-inspired; cloud computing; HPC; makespan; scheduling; workflow

1 INTRODUCTION

High performance computing is about the use of high-productivity computing resources to solve challenging problems in scientific and engineering domains [1]. The HPC platform could comprise workstations, desktop machines, supercomputers, grid or cloud [2]. Grid computing is a kind of HPC loosely coupled collection of heterogeneous resources that are shared by the grid users for utilizing the ideal and under-utilized capacity of the resources [3, 4]. On the other hand cloud computing is a simplified form of grid computing that provides virtual server instance on shared resources based on user specifications [5].

Scientific computing is becoming more relevant in many research disciplines. A typical application may contain several dependent tasks specified as a workflow, which requires efficient scheduling of the tasks. The scheduling of workflows is a challenging task in HPC environment because of inter-dependency of the tasks that needs to be taken care of while scheduling the workflow tasks [6]. The grid/cloud meta-schedulers are responsible for fetching the matched resources that are capable to run the workflow application and schedule these workflows on the available resources [7]. The most important scheduling criterion for HPC environment is to produce a schedule with minimum schedule length so as to optimize the utilization of the resources [8].

During the past decade there have been several attempts to use bio-inspired scheduling algorithms to obtain an optimal or near optimal schedule of the tasks on a specified set of resources in order to minimize the overall schedule length and to optimize the use of resources. However, as the number of tasks increases, the solution space consisting of mapping of tasks to corresponding resources increases exponentially. Hence, there is a need to devise mechanisms to improvise the search strategies and/or mechanisms of the bio-inspired scheduling algorithms for better scheduling solutions in lesser number of iterations/time.

The objective of the research work in this paper is to use bio-inspired bacteria foraging optimization algorithm (BFOA) along with other heuristics algorithms for better search of the scheduling solution space. The idea is to first find a schedule by the deterministic or heuristic algorithms such as MaxMin, MinMin, and Myopic, and use these as initial solutions in the search space to get better solutions using BFOA. The advantage of using BFOA over other nature inspired evolutionary approaches is that it is computationally efficient and has good global convergence [9]. The performance of our approach with the existing approaches is compared for quality of the scheduling solutions. The results demonstrate that our hybrid approach (MinMin/Myopic with BFOA) outperforms other approaches.

The paper is organized into five sections. The first section provides background details and motivation of the research work. The second section provides overview of the workflow scheduling mechanism and also provides insights into the existing workflow scheduling approaches. The third section describes our proposed workflow scheduling approach and fourth section is about the experimental setup. The fifth section provides detailed discussion on the results and observations. The last section concludes the research work presented in this paper.

2 WORKFLOW SCHEDULING PROBLEM

The dependent-task or workflow application in grid and cloud environment is represented as a standard task graph (STG) or directed acyclic graph (DAG) as shown in Fig. 1. In DAG, graph vertices represent tasks and the edges represent task dependencies [8]. There could be n number of dependent tasks and m number of resources; it is very difficult to predict the best schedule with respect to large

mapping combinations between the tasks and the resources. Therefore, the workflow scheduling problem is a non-deterministic polynomial (NP)-complete problem [10]. The workflow scheduling problem can be solved by heuristic methods but the complexity of producing an appropriate schedule becomes high. Therefore, the metaheuristic methods are adopted to produce the near optimal schedule in heterogeneous distributed environment.

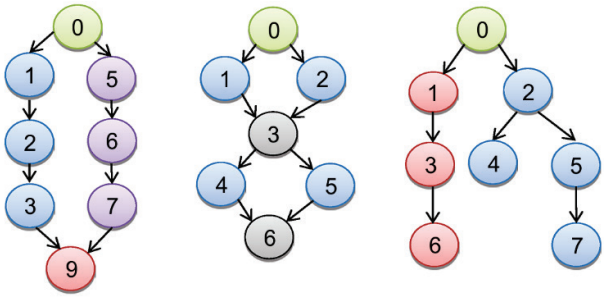


Figure 1 Workflow tasks

Let us assume workflow $W(T, E)$ consists of a set of tasks, $T = \{T_1, T_2, \ldots, T_x, \ldots, T_y, \ldots, T_n\}$, and a set of dependencies among the tasks, $E = \{< T_a, T_b >, \ldots, < T_x, T_y >\}$, where T_x is the parent task of T_y. The set $R = \{R_1, R_2, \ldots, R_m\}$ represents the set of suitable resources in the Cloud. Therefore, the dependent task scheduling problem is the mapping of workflow tasks to Cloud resources ($T \rightarrow R$) so that the makespan M is minimized. The overall timespan of a complete schedule is known as total schedule length or makespan [11].

Generally, a workflow is a set of dependent tasks. The entry task does not have any predecessing task and exit task does not have any successor task [7]. Each dependent task can be executed after the completion of its parent tasks. If a child task is dependent upon more than one parent than it has to wait until all the predecessing tasks complete their execution. The child task becomes a ready task when all the parent tasks complete their execution. If the child task executes upon the same resource where parent task has finished its execution then data transfer time is considered to be zero.

The multiple users can submit multiple workflows to the meta-schedulers. The two important aspects that have been taken care by our meta-scheduler for scheduling the workflow applications are: (1) the parallel handling of multiple workflow applications, and (2) the scheduling of workflow applications on heterogeneous and distributed resources.

3 CURRENT SOLUTIONS IN WORKFLOW SCHEDULING

The current HPC and cloud meta-schedulers use many heuristic algorithms to schedule the workflow applications. The most popular methods are described below.

Myopic-Myopic heuristic is based on the minimum time to compute strategy, where each ready task is assigned to the resource that is capable to complete the ready task at the earliest. Myopic heuristic is one of the simplest scheduling techniques for scheduling dependent tasks in grid environment because it considers a single task while allocating the resource for scheduling. The myopic heuristic is implemented in some real HPC environments such as Condor DAGMan [12]. The Myopic algorithm schedules the ready tasks one after other until all the tasks in ready queue get scheduled. It maps each task to the resource that can process the task at the earliest.

MinMin-This scheduling heuristic prioritizes dependent tasks according the task sizes and schedules the tasks based on the sizes of the tasks [13]. The Min–Min scheduling heuristic maps shortest task on the fastest. The task having minimum expected time for execution over all tasks is selected to be scheduled first on the fastest resource that takes minimum time for execution during each iteration and it keeps scheduling all the tasks until the ready queue is exhausted. MinMin is implemented in real HPC environment such as vGrADS[14] of Rice University USA.

MaxMin-The MaxMin [15] scheduling heuristic prioritizes the dependent tasks according to the expected time to compute, the task that requires the longest execution time is allocated to the fastest resource that is capable to process the task at the earliest. This heuristic arranges the workflow tasks into multiple independent task groups and schedules each group of tasks iteratively. In each iterative step, a task with maximum time to compute is selected to be scheduled on the fastest resource that can process the task at the earliest.

HEFT-Topcuoglu H. et al. proposed a list scheduling method known as Heterogeneous Earliest Finish Time (HEFT) [16], which sets higher priority to the dependent tasks having higher rank value. The rank value is based on the average execution time of each task and average data transfer time between the predecessor and successor task, where the tasks in the critical path have higher rank values. Afterwards, this heuristic sorts the tasks by the descending order of the rank values of the tasks and the task with a higher rank value are set to higher priority. During actual scheduling, tasks in a workflow are scheduled in the order of their priorities, and each task is assigned to the fastest resource that can process the task at the earliest.

PSO-Particle Swarm Optimization (PSO) is a random based scheduling approach that searches the search space to find the near optimal solution. The position of each particle in search space represents a potential solution. The swarm represents the total number of predefined particles. The particle swarm optimization explores the search space by using position vector and terminates when predefined stopping criterion is met. The best particle that provides minimum value for objective function, that is, makespan, is selected as a final solution. In [17], authors have implemented PSO based scheduling approach to minimize the makespan.

GA- Genetic algorithm is also a metaheuristic approach [18] that generates random solutions to achieve near optimal solution. In GA, each individual represents a potential solution. The search space is explored using crossover and mutation operators. The GA terminates after a predetermined stopping criterion is met. The best solution obtained during evaluations is printed as a final solution. The final solution

represents task-resource mapping and scheduling timing. In [19], authors have presented workflow scheduling using GA with the objective of minimization of overall schedule length.

GRASP-The Greedy methods randomized adaptive search procedure (GRASP) [20] is an iterative approach that searches the solution on random basis. In GRASP, predetermined iterations are conducted to search a near optimal solution for scheduling the tasks on available resources. A new solution is generated in each iteration and the best solution among all the iterations is taken as the final solution. This method determines the minimum and maximum time to compute for each task on the available resources. The average time is determined by applying GRASP equation for executing a ready task on the available resources. All the resources that take lesser or equal time to the average time (obtained from GRASP equation) are considered for scheduling decision and any one resource is allocated to the ready task on random basis.

3.1 Related Work

There exists many state-of-the-art works for scheduling workflow tasks in HPC environment but most of the approaches are capable to handle single workflow at a time level-wise.

Rahman et al. [21] have presented a dynamic workflow scheduling approach known as DCP-G that minimizes the workflow execution time dynamically along with reducing the scheduling overhead. Bogdan et al. [22] have introduced an improved critical path using descendant prediction method for workflow scheduling, which is known as ICPDP. This approach performs well for minimizing makespan and for balancing the load of HPC resources. It also minimizes idle time of processing elements to enhance the resource utilization. Wang et al. [23] have presented an extensive approach named look-ahead genetic algorithm (LAGA), which optimizes both makespan and reliability of workflow tasks. LAGA uses an evolution and evaluation method as a two phased methodology. In first phase, the evolution operators of GA decide the task-resource mapping and second phase allows the evaluation steps to govern the task order of solutions using max-min strategy.

Amalarethinam and Selvi [24] have proposed minimum makespan grid workflow scheduling (MMGWS) that minimizes makespan of the workflows in HPC. This approach makes advance reservation of the desired resources and schedules the tasks on the basis of their respective priorities. The results of proposed approach are compared with Min-Min and HEFT scheduling algorithms. Garg et al. [25] have presented an adaptive workflow scheduling (AWS) to optimize makespan considering dynamic availability of the resources. This algorithm also takes care of load balancing by rescheduling the tasks to new resources form overloaded resources.

The existing approaches have not explained that the performance of bio-inspired algorithms deteriorates if the search space is huge. The performance depends upon the quality of scheduling solutions and the computational time to obtain the near optimal schedule. Hence, we are proposing bio-inspired hybrid BFOA approach for better search of the scheduling solution that provides scheduling solutions of better quality within less computational time. The idea is to first find a schedule by the heuristic algorithms such as MaxMin, MinMin, and Myopic, and use these as initial solutions (along with other randomly generated solutions) in the search space to get better solutions using BFOA. Most of the existing workflow scheduling approaches are based on single workflows whereas we have addressed scheduling problem of multiple workflows which are to be scheduled in parallel. The performance of our approach with the existing approaches is compared for quality of the scheduling solutions. The results demonstrate that our hybrid approach (MinMin/Myopic with BFOA) outperforms other approaches.

4 PROPOSED APPROACH

In this paper, a Bacterial foraging optimization algorithm based workflow scheduling mechanism is presented with two different aspects a) Starting with some random solution and searching for the optimal or sub-optimal scheduling solution, 2) Starting with a solution generated by MaxMin, MinMin and Myopic scheduling strategies and then searching for the optimal or sub-optimal schedule from this starting solution.

4.1 Bacterial Foraging Optimization Algorithm (BFOA)

Kevin Passino proposed the Bacterial Foraging Optimization Algorithm (BFOA) in 2002. BFOA provides vigorous search techniques that let a high quality solution to be achieved within a large search space [26].The bacterial foraging algorithm explores the new regions of the search space by chemotaxis and elimination-dispersal process and BFOA exploits the best solutions from the past searches through reproduction process. A bacterium is any solution in the search space, which is represented by a set of parameters. A bacteria foraging optimization algorithm maintains a bacteria population consisting of a set of bacterium that evolves over generations [26].The quality of a bacterium in a bacteria population is determined by an objective function. A typical Bacterial Foraging Optimization Algorithm consists of the following steps:

Algorithm

1. Initialize B, p, N_c, N_{re}, N_s, N_{ed}, P_{ed} and $S(i)$, (i = 1, 2, …, B).
   Choose the initial values randomly for θi where $i=1,2....B$ in the search space. B represents population of bacteria;N_{ch} represents number of chemotactix steps; N_s represents swim length; N_{ed} represents the probability of removal of bacteria; $S(i)$ specifies the size of the step taken in diverse direction during tumbling. The position p of each bacterium in bacteria population B is updated automatically and the iterations stop after meeting the stopping criteria.
2. Elimination loop: $el= el+1$.
3. Reproduction loop: $r=r+1$
4. Chemotactic loop: $c =c +1$

a) For $i = 1, 2, \ldots, B$ take a chemotactic step for i^{th} bacterium.
b) Compute fitness $C(i, c, r, el)$.
c) Let $C(i, c, r, el) = C(i, c, r, el) + C_{cc}(\theta^i(c, r, el), \theta(c, r, el))$
d) Let $C_{previous} = C(i, c, r, el)$ to retain this value until a better cost/fitness is found.
e) Tumble: Create a random vector $\Delta(i) \in \Re p$ with each element $\Delta k(i) \in [-1, 1]$ $(k = 1, 2, \ldots, p)$.
f) Make a movement with a step of size $S(i)$ for i^{th} bacterium in the direction of the tumble.

$$\theta^i(c+1, r, el) = \theta^i(c, r, el) + S(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}}$$

g) Compute $C(i, c + 1, r, el)$
h) Swim.
Let $m = 0$ (Initialize the swin length counter)
While $m < N_s$ Let $m = m + 1$
If $C(i, c + 1, r, el) < C_{previous}$ (if there exists improvement), let $C_{previous} = C(i, c + 1, r, el)$ and let

$$\theta^i(c+1, r, el) = \theta^i(c+1, r, el) + S(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}}$$

Use this $\theta^i(c + 1, r, el)$ to compute the new $C(i, c + 1, r, el)$.
Else, let $m = N_s$. End of while Loop.
i) Move to next bacterium $(i + 1)$, if $i \neq B$
5. If $c < N_c$ go to step 3. In this case, repeat chemotaxis steps, till the end of bacteria life.
6. Go for reproduction.
a. For the given reproduction r and elimination dispersal el, and for each $i = 1, 2, 3, \ldots, B$, let

$$C^i_{health} = \sum_{c=1}^{Nc+1} C(i, c, r, el)$$ be the health of bacterium i.

Sort the bacteria in ascending order of health of bacteria as C_{health}.
b. The B_r bacterium with poor C_{health} values die and the other B_r bacteria with the best values split into two bacteria to keep the population size same.
7. If $r < N_{re}$, move to step 2.
8. Go for Elimination-Dispersal with the pre-determined probability P_{ed}. If $el < N_{ed}$, then go to step 1, otherwise end.

4.2 Problem Definition

Using BFOA algorithm to solve the workflow scheduling problem needs suitable representation of bacterium in the given bacteria population. The problem definition consists of the following:

- The problem consists of a set of resources and the tasks of one or more workflows that need to be scheduled
- The task sequence is fixed, while the resource allocation to the individual tasks varies
- A child task in any workflow can be scheduled only after the completion of its parent tasks.
- The expected execution time (EET) of each task is calculated on the basis of task size (specified in Millon Instructions) and processor's capacity (specified in MIPS (Million instructions per second).

The aim of the proposed work is to minimize the overall schedule length of the task-resource mapping explained in subsequent sections.

4.3 Objective Function

The scheduling of dependent tasks in a workflow focuses on some of the scheduling criteria such as minimizing the total schedule length, minimizing flowtime, minimizing the overall execution cost, executing the tasks within the user specified deadline. The objective function is used to evaluate the current bacteria (population) to produce quality solutions. In our research work, we are taking total schedule length as objective function, which is to be minimized to produce a potential schedule.

Let $T = \{t_1, t_2, ..., t_n\}$ be the n tasks in a given set of workflows that need to be scheduled on a set of m resources $R = \{r_1, r_2, \ldots, r_m\}$. Let B be the start time of the first task, while F be the finish time of the last task in a schedule. The schedule length is defined as the total time span TS_i between B and F for i^{th} schedule. The objective function is to minimize the total schedule length TS_i or makespan over all possible schedules, that is, $\arg_i(\min\{TS_i\}, i \in \text{Schedules})$.

4.4 Initial Bacteria Population in BFOA

The initial population of the bacteria represents random scheduling solutions. That is, the position of each bacterium represents a possible schedule (possible solution in the problem space). The position of each bacteria is an n–dimensional vector, where the i^{th} element of the vector represents the resources ID on which the i^{th} task is executed. The task sequence remains constant, while the resource allocation changes across different schedules. The total number of bacterium is determined by the pre-decided bacteria population size.

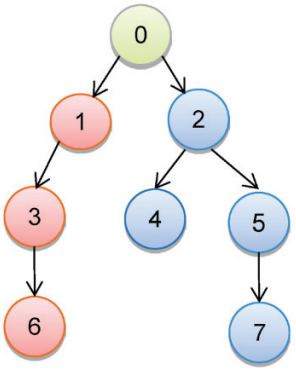


Figure 2 An example workflow

An example of Bacteria position is shown in Tab. 1, where the tasks T_0 to T_7 have fixed order, while the resource string varies across the tasks and also over different schedule solutions. Fig. 2 displays an example of workflow.

Table 1 Bacteria population sample

Position vector (R_1, R_2, …, R_0) of a bacterium							
T_0	T_1	T_3	T_2	T_4	T_5	T_6	T_7
R_1	R_2	R_3	R_1	R_4	R_5	R_6	R_0

The initial bacteria population in the bacterial foraging optimization algorithm could comprise a) Random solutions (positions), or b) Random solutions with a few bacteria positions initialized with solutions obtained by MaxMin, MinMin and Myopic scheduling strategies. The BFOA then searches for the optimal or sub-optimal schedule from this initial population.

The idea behind incorporating these heuristic strategies in BFOA is to enhance the exploration capability of the proposed approach. If there are n jobs and m resources, each job will have n^m combinations. For example, for 100 jobs and 10 resources, there would be 100^{10} combinations. This search space is huge and it may not be computationally feasible to find the near optimal solutions. With initial solutions obtained from some heuristic techniques such as MaxMin, MinMin, and Myopic, the exploration by the evolutionary algorithm would start from these positions or solutions and the algorithms would attempt to improve upon these solutions (find optimal/sub-optimal solutions), if applicable, in relatively lesser amount of time. The following sections describe the steps involved in BFOA in the context of scheduling problem.

4.4.1 Chemotactic Process

Chemotaxis process allows the bacterium to move towards the sources of food. The bacterium swims to change directions during this process and follows the same direction if it finds good fitness over the previously swimming steps. In this process, the bacterium explores search space for better solutions. The tasks are allocated to different combination of resources to achieve better fitness values. The movement of the i^{th} bacterium at the c^{th} chemotactic, r^{th} reproductive, and el^{th} elimination dispersal step can be mathematically expressed as follows:

Where Δ indicates a vector in the random direction whose elements lie in in [−1, 1].

The chemotaxis process allows the bacteria to explore the search space to find better solutions for the given problem. For example, task T1 is allocated to resource R1. The new allocation is 2 (1 (previous resource id) + 0.99987 (value achieved by tumbling process)). Table 1 shows the resource allocation to workflow tasks before and after the chemotaxis process. The task-resource mapping of each bacterium is changed during chemotaxis process if each bacterium finds better fitness; otherwise the previous solution is preserved.

4.4.2 Swarming

Swarming allows the population of bacteria gathers together and moves as concentric patterns of swarms. The cost or fitness of a bacterium position is affected by swarming. The health of bacteria is needed during reproduction step where bacteria with poor health die and bacteria with good health go for reproduction.

The swarming equation of bacteria is affected by the cell-to-cell signaling is given by Eq. (1), where $\|\ \|$ is the Euclidean norm, ωa and ωr are measures of the width of the attractant and repellent signals respectively, M measures the magnitude of the cell-cell signaling effect in the given swarming equation. The swarming equation drives certain weight that is added to cost (fitness value) of each bacterium to determine the health of bacterium.

$$C_{cc}(\varphi^i, \varphi) = \left\{ -M \left(\sum_{k=1}^{B} \mathrm{e}^{-\omega a \left\| \varphi^i - \varphi^k \right\|^2} - \sum_{k=1}^{B} \mathrm{e}^{-\omega r \left\| \varphi^i - \varphi^k \right\|^2} \right) \right\}. \quad (1)$$

For example, if the fitness value (makespan) of a bacterium is 100 seconds, the swarming weight is 44.999, and the health of the bacterium is 100 + 44.999 = 145. The better the health, better would be the bacterium position. The health of bacterium allows it to make decision for the swimming step, that is, whether to swim in the same direction if the health is improving or stay back at the previous position. The health of the bacterium plays a crucial role in reproduction step, where the bacterium with better health survives and bacterium with poor health dies.

4.4.3 Reproduction

The reproduction step allows the bacteria to exploit the search space. The objective here is to search a limited region of the search space with the possibility of improving the local solution. The existing solutions are refined here to improve the fitness value. After chemotactic steps, a reproduction step is followed. In reproduction, the bacteria with bad health die and the bacteria with good health split into two bacteria to keep the population size same [27]. The bacteria are sorted according to their health. The scheduling solutions with higher or poor makespan will be removed from the current bacteria population.

However, this step generates duplicate bacterium (scheduling solutions) in the bacteria population, but it is mandatory to remove the worse population (with poor fitness) from the current solutions.

4.4.4 Elimination-dispersal

This step allows the bacteria to search for the search space to get newer and refined solutions (if any), that is, searching a much larger portion of the search space with the possibility of finding better solutions. In elimination process, a bacterium is stochastically selected for elimination from the population and is replaced by a new bacterium located at a random new location within the search space, according to a predefined probability P_{ed}. For example if the elimination-dispersal probability is 0.25 then 25% of the scheduling solutions or bacteria will selected on random basis and will be replaced by newly generated scheduling solutions or bacteria.

However, this process introduces new population and removes redundant bacteria, which are generated in reproduction step. But it may also kill the best scheduling

solutions found so far. In order to preserve the best population elitism is applied during elimination-dispersal step. If the elitism rate is 10% then the top 10% of bacterium with minimum fitness or make span are preserved, while the rest 90% of bacteria undergo elimination dispersal process.

Tab. 8 is depicting the operational patameter values for BFOA.

4.5 Genetic Algorithm

Each chromose represents a potential solution in the problem space. The individual in population is generated in a similar manner as each bacterium is generated in BFOA. The total chromosomes are decided by the pre-determined population size. The genetic operators are explained below.

4.5.1 Selection

For selecting chromosomes for reproduction from the prevailing population, the tournament selection is applied. Two tournaments are held to select potential parents for mating. Parent 1 with better fitness value has been selected and this method is repeated in the second tournament to obtain second parent 2. The two parent chromosomes are selected for the crossover. Fig. 3 depicts the binary tournament selection procedure.

Table 2 Task resource assignment before chemotaxis

HPC task-resource assignment string							
T_0	T_1	T_3	T_2	T_4	T_5	T_6	T_7
R_2	R_3	R_1	R_2	R_3	R_3	R_0	R_6

Table 3 Task resource assignment after chemotaxis

HPC task-resource assignment string							
T_0	T_1	T_3	T_2	T_4	T_5	T_6	T_7
R_1	R_2	R_3	R_1	R_4	R_5	R_6	R_0

Table 4 Control operational parameters for BFOA

Sr. No	Parameters	Type/Values
1	Bacteria Population	50
2	Maximum number of steps, N_s	3
3	Number of chemo tactic steps, N_c	20
4	Number of reproduction steps, N_{re}	2
5	Number of elimination-dispersal steps, N_{ed}	z2
6	Probability, P_{ed}	0.25
7	Size of the chemotaxis step, $C(i)$	0.8

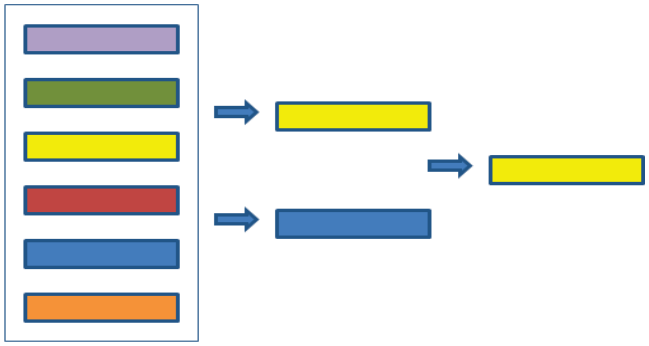


Figure 3 The binary tournament selection

4.5.2 Crossover

The crossover operator allows the individuals to search for new solutions. The parents selected in tournament selection go for reproduction of Offsprings/children during crossover [29]. In The two-point Crossover is used in our proposed approach. Tab. 5 and Tab. 6 are showing before and after crossover points.

Table 5 Before two-point crossover

Parent 1	R_1	R_2	R_3	R_4	R_3	R_4	R_5	R_6
Parent 2	R_0	R_7	R_5	R_9	R_1	R_3	R_2	R_8

Table 6 After two-point crossover

Offspring 1:	R_1	R_2	R_3	R_9	R_1	R_3	R_5	R_6
Offspring 2:	R_0	R_7	R_5	R_4	R_3	R_4	R_2	R_8

4.5.3 Mutation

Next to crossover operation, the mutation operation is applied to one chromosome based on the mutation probability. The mutation operation helps to diversify the population and to obtain new solutions. Replace Mutation is used in the research work presented in this paper. When the mutation probability of is decided by a chromosome with the selection of two random points and the resources of those tasks are interchanged.

Tab. 7 is depicting the controlled parameter values for GA. The fine tuning of controlled parameters is based on the empirical study of the GA based research outcome.

Table 7 Control operational parameters for GA

Sr. No	Parameters	Type/Values
1	Crossover	Two-Point
2	Crossover Probability	0.8
3	Mutation Type	Swap
4	Crossover Probability	0.2
5	Population Size	100

4.6 Particle Swarm Optimization (PSO)

Particle Swarm Optimization is an intelligence technique which is based on a swarm of particles moving in search space and communicating with each other for determining a near optimal solution [30]. Population of particles is known as swarm in PSO. Each particle is represented in a similar manner as a bacterium is represented in bacteria population. The parameters used in the PSO algorithms are:

Velocity (vector): This vector determines the direction in which a particle needs to fly in order to improve its current position in the flock.

pbest (personal best): It is the personal best position (solution) of a given particle found so far.

gbest (global best): Position of the best particle in the entire swarm.

Inertia weight: Denoted by ω, the inertia weight is used to control the impact of the previous history of velocities on the current velocity of a given particle. It can be taken as random value or constant value.

Learning factors: There are two learning factors used in PSO that is, $C1$ and $C2$. The parameter $C1$ represents the attraction of a particle towards its own success, while parameter $C2$ represents the attraction of a particle towards the success of global best position.

Every particle gets updated during each iteration by its personal best value, that is, *pbest* and the global best value, that is, *gbest*. The particle modifies its position with the help of velocity and position vector to explore the search space for better solutions (Doctor Kennedy and Eberhart in 1995).
When the position of a particle is changed, the task-resource mapping is changed. The tasks are assigned to different set of resources as long as the position of each particle is updated. Tab. 8 depicts the values of control parameters used for workflow scheduling in our research work.

Table 8 Control parameters for PSO

Sr. No	Parameters	Values
1	$C1$	1.25
2	$C2$	1
3	ω	0.9
4	$r1$	0.1
5	$r2$	1
6	*Swarm size*	100

4.7 Multiple Workflows Scheduling

In multiple workflows scheduling, a group of tasks are scheduled level-wise in the workflow trees using the heuristic algorithms (MinMin, MaxMin, and Myopic). We explain the multiple workflow scheduling mechanism that we have implemented in our work with one such heuristic algorithm namely, MinMin. We first take all the tasks on level one of all the workflow trees and schedule them using the MinMin algorithm. The task with minimum time to compute is picked first and is scheduled on the fastest resource (the processor which takes shortest time to execute the task). We then schedule the tasks at second level of the workflows. The shortest task is allocated to the fastest resource and this process continues until all the tasks at all the levels are scheduled. MinMin applies its scheduling strategy of shortest job on fastest resource at each workflow level and schedule the all the tasks in the workflows. The other heuristics such as MaxMin and Myopic also perform scheduling according to the levels of workflow. The only difference is their scheduling strategies. MaxMin schedules the longest task on the fastest resource at each workflow level, whereas, myopic chooses the tasks in an arbitrary fashion and schedules them on the fastest resources. The reverse is also true when meta-heuristic algorithms (BFOA, PSO, and GA) allocate a set of tasks to the corresponding resources.

4.8 Experimental Setup

The heterogeneous resources with different processing capacities are simulated and defined in terms of MI/sec (Million Instructions/sec). The size of each task is generated between 8000 MI and 20,000 MI from a uniform distribution. The total number of tasks in the workflow are N. Our workflow generator can generate single workflows with many dependent tasks and multiple workflows (with specific number of tasks in each workflow). In our case study, 10 tasks have been generated in each workflow. If there are 10 workflows then the total number of tasks would be 100. The workflows for evaluation are created using the following parameters:

- Type = Random workflows
- One workflow= 10 tasks each
- $N = \{50, 100, 200, 300, 400\}$
- $M = \{5\}$

Note that for GRASP, 500 iterations are considered and the value of α is set to 0.01.

5 RESULTS AND OBSERVATIONS

The scheduling heuristics are evaluated on the basis of total schedule length. The total schedule length for a set of workflows or dependent tasks is defined by the time span between the start of the first task and the end of the last task. Two sets of scheduling simulations have been carried out. In the first simulation, the search for scheduling solutions was initiated with a random generation of solutions in the search space. The exploitation and exploration of the solution space has been performed in order to refine the existing solutions and to search newer solutions using the metaheuristics algorithms such as BFOA, PSO and GA.

In the second simulation, a hybrid approach is used where heuristic algorithms such as MaxMin, MinMin, and Myopic, have been used to get initial solutions in the search space. These along with randomly generated solutions serve as the starting solutions in the search space to the metaheuristics algorithms for possible improvements in the scheduling solutions. The performance of our hybrid approach is compared with the existing approaches for quality of the scheduling solutions. The results demonstrate that our hybrid approach (MinMin/Myopic with BFOA) outperforms other approaches.

The space complexity increases exponentially with increase in task-resource combinations. For 200 tasks and 5 resources, there exist 5^{200} scheduling solutions. The evolutionary techniques perform well when the search space is small (less number of tasks and lesser number of task-resource combinations). As the search space complexity increases, it is difficult to determine the near optimal solution.

Tab. 9 and Fig. 4 depicts that heuristic algorithm perform well even if the number of workflows with number of tasks go on increasing while the quality of solutions obtained by BFOA and other evolutionary techniques such as GA and PSO starts deteriorating as the search space increases.

Table 9 Metaheuristic Hybrid Approach Vs Heuristic Approaches

Tasks	MaxMin	MinMin	Myopic	GRASP	GA	PSO	BFOA
50	165	145	139	138	134	132	126
100	290	244	226	233	243	242	228
200	602	518	504	501	519	516	509
300	765	633	643	607	677	681	665
400	996	822	811	840	957	925	913

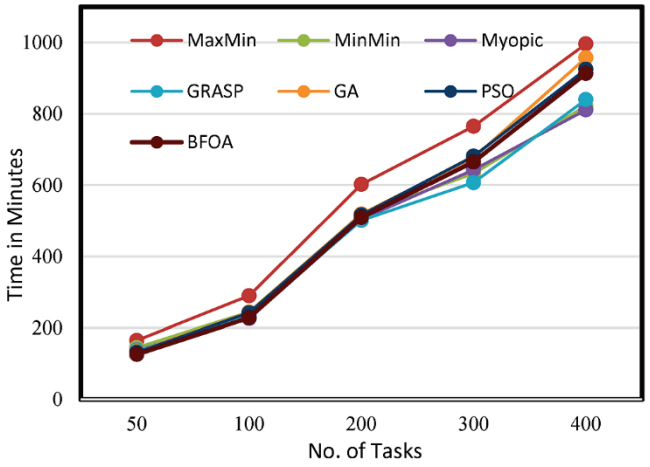


Figure 4 Heuristic approaches vs Metaheuristic approaches

Table 10 Metaheuristic vs Heuristic Approaches

Tasks	MaxMin	MinMin	Myopic	GRASP	GA	PSO	BFOA
50	165	145	139	138	121	120	115
100	290	244	226	233	218	225	202
200	602	518	504	501	476	496	446
300	765	633	643	607	587	581	575
400	996	822	811	840	793	796	774

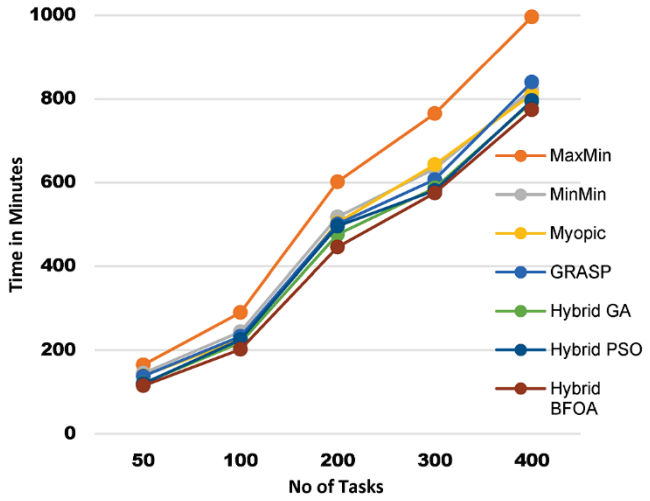


Figure 5 Heuristic approaches Vs Metaheuristic Hybrid Approach

Tab. 10 and Fig. 5 depicts that the hybrid approach improves the quality of scheduling solutions and minimizes the computational time to obtain the near optimal solution. If the metaheuristic algorithms start with solutions obtained by heuristic solution along with some randomly generated solutions, then the quality of solution improves significantly, because the exploration of the search space is guided in a better direction. The solutions generated by hybrid BFOA are improved by 8% for 50 tasks, 11% for 100 tasks, 12% for 200 tasks, 13% for 300 tasks and 15% for 400 tasks, respectively.

6 CONCLUSION

In this paper, we have used the bio-inspired bacteria foraging optimization algorithm (BFOA) together with other heuristic algorithms to find scheduling solutions for multiple workflows. Starting with a random schedule for the multiple workflows, the BFOA attempts to find a scheduling solution. However, this may not always result in optimal or sub-optimal solution and sometimes may take large amount of computation time to get to a desired scheduling solution. Hence, we have used a hybrid approach where heuristic algorithms such as MaxMin, MinMin, and Myopic have been used to get initial solutions in the search space, which then serve as starting points (along with other randomly generated solutions) for getting better solutions using bacteria foraging optimization algorithm. The performance of our hybrid approach presented in this paper is compared with the existing approaches. The results demonstrate that our hybrid approach (MinMin/Myopic with BFOA) outperforms other approaches.

Acknowledgement

The work has been carried out at Centre for Development of Advanced Computing (C-DAC), SPPU Campus using PARAM-HPC setup and we would also like to thank KPIT R & D, Pune to help us in handling of multiple workflows in parallel environment.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

7 REFERENCES

[1] Kaur, M. & Kadam, S. S. (2017). Discovery of resources using MADM approaches for parallel and distributed computing, *Engineering Science and Technology, 20*(3), 1013-1024. https://doi.org/10.1016/j.jestch.2017.04.006

[2] Rankin, S. (2016). An Introduction to HPC. http://www.hpc.cam.ac.uk/getting-help/introtohpc-course

[3] http://www.ianfoster.org/wordpress/wp-content/uploads/2014/01/History-of-the-Grid-numbered.pdf

[4] Kaur, M. (2012)- Semantic Resource Discovery with Resource Usage Policies in Grid Environment, *Int. J. Comput. Sci. Issues, 9*(5), 301-307.

[5] Shawish, A. & Salama, M. (2013). Cloud Computing : Paradigms and Technologies. *Inter-cooperative Collective Intelligence: Techniques and Applications*, 39-68. https://doi.org/10.1007/978-3-642-35016-0_2

[6] http://escience2015.mnm-team.org/wp-content/uploads/2015/09/deelman_escience_2015_keynote.pptx.pdf

[7] Rahman, M., Hassan, R., Ranjan, R., & Buyya, R. (2013). Adaptive workflow scheduling for dynamic grid and cloud computing environment. *Concurrency and Computation: Practice and Experience, 25*, 1816-1842. https://doi.org/10.1002/cpe.3003

[8] Yu, J., Buyya, R., & Ramamohanarao, K. (2008). Workflow Scheduling Algorithms for Grid Computing. *Metaheuristics for Scheduling in Distributed Computing Environments*, 173-214. https://doi.org/10.1007/978-3-540-69277-5_7

[9] Das, S., Biswas, A., Dasgupta, S., & Abraham, A. (2009). Bacterial Foraging Optimization Algorithm : Theoretical

Foundations, Analysis, and Applications. *Foundations of Computational Intelligence, 3*(1), 23-55. https://doi.org/10.1007/978-3-642-01085-9_2

[10] Braun, T. D., Siegel, H. J., Beck, N., Bölöni, L. L., Maheswaran, M., Reuther, A. I., Robertson, J. P. et al. (2001). A Comparison of Eleven Static Heuristics for Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems. *Journal of Parallel and Distributed Computing, 61*(6), 810-837. https://doi.org/10.1006/jpdc.2000.1714

[11] Kaur, M. (2016). FastPGA based scheduling of dependent tasks in grid computing to provide QoS to grid users. *IEEE International Conference on Internet of Things and Applications (IOTA)*, Pune, 418-423. https://doi.org/10.1109/IOTA.2016.7562764

[12] Couvares, P., Kosar, T., Roy, A., Weber, J., & Wenger, K. (2007). Workflow Management in Condor. research.cs.wisc.edu/htcondor/doc/workflow_condor_2007.pdf

[13] Kaur, M. & Kadam, S. (2018). A novel multi-objective bacteria foraging optimization algorithm (MOBFOA) for multi-objective scheduling. *Applied Soft Computing, 66*, 183-195. https://doi.org/10.1016/j.asoc.2018.02.011

[14] Kennedy, K., Cooper, K., Koelbel, C., Tapia, R., & Torczon, L. (2009). The VGrADS Project, http://vgrads.rice.edu/

[15] Kaur, M. & Kadam, S. (2019). Discovery of resources over Cloud using MADM approaches. *International Journal for Engineering Modelling, 32*, 83-92. https://doi.org/10.31534/engmod.2019.2-4.ri.02m

[16] Topcuoglu, H., Hariri, S., & Wu, M. Y. (2002). Performance-effective and low-complexity task scheduling for heterogeneous computing. *IEEE Trans. Parallel Distrib. Syst., 13*(3), 260-274. https://doi.org/10.1109/71.993206

[17] Lei Zhang, B. Y. & Chen, Y. (2006) Task Scheduling Based on PSO Algorithm in Computational Grid, *Sixth Int. Conf. Intell. Syst. Des. Appl., 2*, 696-704. https://doi.org/10.1109/ISDA.2006.253921

[18] Carretero, J., Xhafa, F., & Abraham, A. (2007). Genetic Algorithm Based Schedulers for Grid Computing Systems. *Int. J. Innov. Comput. Inf. Control*, 3(6), 1-19.

[19] Wang, X., Shin, C., Buyya, R., & Su, J. (2011). Optimizing Makespan and Reliability for Workflow Applications with Reputation and Look-ahead Genetic Algorithm. *J. Futur. Gener. Comput. Syst., 27*(8), 1124-1134. https://doi.org/10.1016/j.future.2011.03.008

[20] Resende, M. G. C. & Ribeiro, C. C. (2002). Greedy randomized adaptive search procedures. *DIMACS Series on Discrete Mathematics and Theoretical Computer Science, 50*, 1-29.

[21] Rahman, M., Hassan, R., Ranjan, R., & Buyya, R. (2013). Adaptive workflow scheduling for dynamic grid and cloud computing environment. *Concurr. Comput. Pract. Exp., 25*(3), 1816-1842. https://doi.org/10.1002/cpe.3003

[22] Bogdan, S., Catalin, L., Florin, P., & Valentin, C. (2007). A Hybrid Algorithm for Scheduling Workflow Applications in Grid Environments. *Lect. Notes Comput. Sci., 4804*, 1331-1348. https://doi.org/10.1007/978-3-540-76843-2_15

[23] Wang, X., Shin, C., Buyya, R., & Su, J. (2011). Optimizing Makespan and Reliability for Workflow Applications with Reputation and Look-ahead Genetic Algorithm. *Futur. Gener. Comput. Syst., 27*(8), 1124-1134. https://doi.org/10.1016/j.future.2011.03.008

[24] Amalarethinam, D. & Selvi, F. (2012). A Minimum Makespan Grid Workflow Scheduling algorithm. *2012 International Conference on Computer Communication and Informatics*, Coimbatore, India, 2012, 1-6. https://doi.org/10.1109/ICCCI.2012.6158777

[25] Garg, R. & Singh, A. K. (2015). Adaptive work flow scheduling in grid computing based on dynamic resource availability. *Eng. Sci. Technol. an Int. J., 18*(2), 256-269. https://doi.org/10.1016/j.jestch.2015.01.001

[26] Kaur, M. (2016). Elitist Multi-Objective Bacterial Foraging Evolutionary Algorithm for Multi-Criteria based Grid Scheduling Problem. *IEEE International Conference on Internet of Things and Applications(IOTA' 2016)*, 431-436. https://doi.org/10.1109/IOTA.2016.7562767

[27] Dang, J., Brabazon, A., Neill, M. O., & Edelman, D. (2008). Option Model Calibration Using a Bacterial Foraging Optimization Algorithm. *Springer-Verlag Berlin Heidelb.*, *vol. LNCS 4974*, 113-122. https://doi.org/10.1007/978-3-540-78761-7_12

[28] Bhandari, D., Murthy, C. A., & Pal, S. K. (2012). Variance as a Stopping Criterion for Genetic Algorithms with Elitist Model. *Fundam. Informaticae, 120*, 145-164. https://doi.org/10.3233/FI-2012-754

[29] Maheswaran, M., Ali, S., Siegel, H. J., & Hensgen, D. (1999). Dynamic Mapping of a Class of Independent Tasks onto Heterogeneous Computing Systems. J*ournal of Parallel and Distributed Computing, 59*, 107-131. https://doi.org/10.1006/jpdc.1999.1581

[30] Pandey, S., Wu, L., Guru, S. M., & Buyya, R. (2010). A Particle Swarm Optimization-Based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments. *2010 IEEE International Conference on Advanced Information Networking and Applications*, 400-407. https://doi.org/10.1109/AINA.2010.31

Authors' contacts:

Mandeep Kaur
(Corresponding author)
Dept. of Comp. Sc., Savitribai Phule Pune University,
SPPU Campus, Ganeshkhind Road, Pune, Maharashtra 411007, India
mandeep.gondara@gmail.com

Sanjay Kadam
Centre for Development of Advanced Computing,
SPPU Campus, Ganeshkhind Road, Pune, Maharashtra 411007, India
sskadam@cdac.in

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20200806135648

Original scientific paper

Experimental Optimization of Using Natural Pozzolan in Chloride Ion Exposed Concrete via Taguchi Method

Amir Behshad*, Mostafa Askarian Zardak, Yaser Mohammadi

Abstract: Concrete durability is one of the most important concerns in the field of construction. The environmentally friendly materials that can provide the durability are of great value in the construction of concrete structures. The use of natural pozzolans is one of the cheapest and most efficient methods in this field, which offers a good performance from environmental and economic point of view and satisfies required engineering parameters. In this study, the effect of using natural pozzolan in the manufacture of concrete exposed to sulfate and chloride ion of Oman sea water was investigated. The Taguchi optimization method was used to reduce the number of samples prepared, reduce the cost of experiments and achieve an optimal mix design. The four parameters, namely water to cement ratio, different percentages of natural pozzolan, super-plasticizer and cement grade with different ratios, were considered as problem variables. The Taguchi optimization method proposed 8 mix designs based on the defined levels for the variables. By constructing 96 samples, two parameters of permeability and water absorption from Oman Sea and drinking water were investigated in the samples. By introducing the results of the experiments into the Taguchi method, the final optimal design was presented by this method, and by constructing 12 additional samples and conducting permeability and water absorption experiments, the behavior of this optimal mix design was verified. The appropriate performance of the Taguchi method was demonstrated by obtaining the optimal mix designs from the Taguchi method, constructing this mix design and comparing the results with the regulation limitations. The results showed that it is acceptable to use natural pozzolan under moderate to severe chloride and sulfate ion attacks, but it is not recommended in the extreme environmental conditions.

Keywords: Permeability; Pozzolan; Taguchi Method; Water Absorption

1 INTRODUCTION

Today, the use of natural pozzolans is of great importance due to the improvement of the concrete structure and the increase of concrete performance in aggressive environments considering the economic efficiency and environmental impacts. Replacing the cement with these types of pozzolans (depending on the replacement percentage) has shown significant effects on the compressive strength, initial setting time and durability of concrete. The use of pozzolanic cement containing 55% of a natural pozzolanic material from Turkey demonstrated the ability of this compound to reduce the spread of the alkali-silica reaction [1]. Due to the inconsistency of the structural properties, no certain percentage of natural pozzolans can be determined to improve the concrete properties, and the necessary experiments should be performed for each type of pozzolan. In the study of Najimi et al. (2011) on zeolite as a natural pozzolan, it was found that replacing the cement with 30% of the natural pozzolans in the mix designs can have beneficial effects on water and chloride penetration and corrosion rate [2]. Ezzian et al. (2005) performed some experiments and managed to find an optimal percentage of natural pozzolan between 15 and 20% for cement replacement in the samples at temperatures of 20, 40 and 60 °C [3]. In another study by Nguyen et al. to investigate the size of natural pozzolans for reducing the ettringite production rate, they found that the smaller the particle size, the better the effect on the amount of produced ettringite [4]. In addition to the environmental issues and not requiring much CO_2 production and release process unlike cement, the use of natural pozzolan is economically optimal and results in the reduced heat generation and concrete permeability [5]. In fact, pozzolanic materials are the most important factor in reducing the concrete permeability, which is achieved by reducing the volume of large cavities in concrete [6]. To this end, a study conducted in Algeria by Kaid et al. concluded that the use of Algerian natural pozzolans as a percentage of cement replacement can improve the parameters related to durability, corrosion and resistance to acidic environment [7]. In the present study, the natural pozzolan of Taftan Mountain and the water of the Oman Sea were used. The permeability and water absorption tests were performed using BS EN 12390-8 permeability test standards [8] and BS 1881-part 122 half-hour water absorption test [9].

Merida and Kharchi [10] investigate the concrete properties has been manufactured by blended cement including volcanic origin Algerian natural pozzolan and Portland cement. They included that using such pozzolan improves the pulse velocity and sulphate resistance of concrete.

Omrane et al. [11] investigate the performance and durability of self-compacting concrete (SCC) contain natural pozzolan and recycled concrete aggregates. The results of their research work shows that using pozzolan develop a significant compressive strength to mixtures without it at 120 days. The incorporation of 15% and 20% pozzolan in SCC decreased the penetration depth of chloride ions to 50% and also decreased the mass loss under sulfuric acid attack. A comprehend review regarding the mechanical and durability characteristics as well as the microstructural properties of natural pozzolan based geopolymers and their potential as binding material are reported in Firdous et al. [12] research work. Najimi and Ghafoori [13] find the most proper optimum using hybrid aluminosilicate precursors in alkali-activated concretes made with having different proportions of natural pozzolan (low calcium precursor) and also blast furnace slag (high calcium precursor). Their experimental investigation shows that the optimum sodium hydroxide concentration was dependent on the precursor and

activator combinations as well as the expected fresh, strength, dimensional stability and transport properties.

In this study, by replacing the cement with Taftan natural pozzolans and using the Taguchi method, the optimum percentage of the pozzolan for the minimum permeability and water absorption was obtained in the both Oman Sea water and drinking water. By comparing the optimum and laboratory values with the standards of the "National Concrete Durability Regulation in Oman Sea", the possibility of using this pozzolan in the construction industry was investigated for the coastal cities of the Oman Sea.

2 TAGUCHI METHOD AND TEST MODEL

The Taguchi optimization technique was introduced in the 1950s by the Japanese scientist Taguchi [14]. This method is used to save the number of sample preparation, reduce the cost of experiments and also obtain an optimal design (in this study, pozzolanic concrete mix design). This method is optimized by setting the controllable parameters and reducing the sensitivity of the system performance relative to the uncontrollable parameters of optimal system performance [15-17]. Other advantages of this method include the simultaneous optimization of several factors and evaluation of the impact of controllable factors [16]. The parameters in an experiment are divided into two groups of controllable and uncontrollable parameters.

In the Taguchi method, considering the minimum effects of uncontrollable parameters, it is tried to optimize the system by setting the controllable parameters.

One of the most important parameters in this method is the "signal to noise (S/N) ratio". In terms of quality, Taguchi divided the qualitative characteristics into three categories based on this parameter [18].

1. Larger the better: This parameter is used when the problem objective is to reach the maximum value which is given by Eq. (1).

$$\frac{S}{N} = -10\log_{10}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right) \quad (1)$$

2. Nominal the best: This parameter is used when the problem objective is to produce results around a given value, which is expressed by Eq. (2).

$$\frac{S}{N} = -10\log_{10}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - y_o)^2\right) \quad (2)$$

3. Smaller the better: This parameter is used when the problem objective is to reach the minimum value which is given by Eq. (3).

$$\frac{S}{N} = -10\log_{10}\left(\frac{1}{n}\sum_{i=1}^{n}y_i^2\right) \quad (3)$$

In Eqs. (1)-(3), y_i is the value of the obtained results and n is the number of replications per experiment.

Any level of a factor with the highest S/N ratio is known as the optimal level. In addition, any factor where the difference between its levels (delta) has the highest value is the most effective factor in the test process and, consequently, the most influential factor in the optimization process [19].

The parameters of this study are as follows:

- Taftan natural pozzolan (pozzolan)
- Water to cement ratio (cement with pozzolan) - (w/b)
- Grade of used cement - (c)
- Super-plasticizer - (sp).

The levels (the number of ratios provided for each parameter to prepare the test mix design) and the quantity of each variable parameter related to these parameters are listed in Table (1). For example, for the water to cement ratio, two ratios of 0.4 and 0.45, namely two levels, are considered and for the pozzolan to cement ratio, four ratios of 0, .10, .20 and .25%, namely four levels, are defined.

Table 1 Parameters and related levels

Level	b (kg/m^3)	Sp (%)	W/b (%)	Pozzolan (%)
1	400	0	0.4	0
2	450	0.7	0.45	10
3	-	-	-	20
4	-	-	-	25

In this study, 8 mix designs were used based on the proposed Taguchi method. Each mix design consists of 12 samples, 6 samples for the storage in pond 1 (the pond containing drinking water) and the other 6 samples were used for the storage in pond 2 (the pond containing Oman Sea water). The temperature of the sample storage ponds was kept constant at 28 ± 3 °C. The tests performed on the samples are the BS EN 12390-8 permeability test and BS 1881-part122 half-hour water absorption test, both of which are among the most important tests recommended in the "National Concrete Durability Regulation in Oman Sea" for the evaluation of concrete samples. The experiments were performed on the samples at 28 days. The water permeability test of the relevant standard required more time for the test because of spending more time (72 ± 2 hours for each mix design in each pond) than the other permeability tests.

Taguchi proposed the L_8 orthogonal arrays (8 mix designs) for this experiment, as shown in Tab. 2, which gives the details of the mix designs in Tab. 3 with respect to the proposed levels for each mix design.

Table 2 L8 standard arrays

	Independent Variables			
Test No.	b (kg/m^3)	Sp (%)	w/b (%)	Pozzolan (%)
M1	1	1	1	1
M2	2	2	2	1
M3	2	1	1	2
M4	1	2	2	2
M5	1	2	1	3
M6	2	1	2	3
M7	2	2	1	4
M8	1	1	2	4

Table 3 Details of Taguchi proposed mix designs

Test No.	Independent Variables			
	b (kg/m^3)	Sp (%)	w/b (%)	Pozzolan (%)
M1	400	0	0.4	0
M2	450	1	0.45	0
M3	450	0	0.4	10
M4	400	1	0.45	10
M5	400	1	0.4	20
M6	450	0	0.45	20
M7	450	1	0.4	25
M8	400	0	0.45	25

3 MATERIALS

The cement used in the experiment is the cement type II and the pozzolan used is the Taftan natural pozzolan. The results of chemical analysis of cement and pozzolan used in the experiments are presented in Tab. 4.

Table 4 Physical and chemical properties of Portland cement and mineral admixture

Physical properties	Portland cement	Taftan pozzolan
Specific surface area (g/cm^2)	2850-3000	-
Setting time		
Primary (min)	90-110	-
Secondary (min)	130-180	-
Compressive strength		
3-day (kg/m^2)	180-230	-
7-day (kg/m^2)	290-330	-
28-day (kg/m^2)	410-450	-
Chemical analysis		
SiO_2	21.5-22	60-62
CaO	63.4-64.2	6.5-7.1
Al_2O_3	4.8-5.2	17.8-18.5
Fe_2O_3	3.7-4.1	4.6-5.5
MgO	2.3-2.6	<2.8
SO_3	1.2-1.6	<0.2
Cl	<0.04	<2.5
K_2O	0.6	<2
Na_2O	0.4	-
C_3S	56	-
C_2S	15	-
C_3A	7-7.5	-
Loss of ignition	0.5-0.8	<2.5

The sand and gravel consumed in this study were obtained from a factory in Zahedan which did not meet ASTM C33 standards [20], and the grading was modified, and the grading results after modifying the aggregates are presented in Tabs. 5 and 6. The maximum gravel size, water absorption percentage, and gravel density are 16 mm, 1.1% and 2.75, respectively. Also, the consumed sand has a density of 2.65 and water absorption of 2.1%.

Table 5 Results of gravel grading

Sieve size (mm)	25.4	19	12.5	9.5	#4	#8
Cumulative percentage of grain passed through sieve	100	90.2	40	14	1	0.1
Allowable range ASTM-C33 (%)	100	90-100	20-55	0-15	0-5	–

The grade of cement used (pozzolan + *c*) in this study is 400 kg/m^3 and 450 kg/m^3 (Tab. 3). Due to the necessity for keeping the slump of samples between 70 to 110 mm and the high grade of cement used, the concrete slump of mix designs kept constant with a small percentage of super-plasticizer without increasing the water. Tab. 7 presents the super-plasticizer specifications and the sample mix design. It should be noted that the percentage of super-plasticizer consumption recommended by the manufacturer is 0.6-2.5% by weight of cement. The used super-plasticizer is based on melamine.

Table 6 Results of sand grading

Sieve No.	Allowable range ASTM C 33	Passed percentage (before modification)	Passed percentage (after modification)
200	0	1.2	0
100	2-10	4	7
50	10-30	7	24
30	25-60	20	40
16	50-85	35	69
8	80-100	69	89
4	90-100	84	95
3/8	100	100	100

Table 7 Super-plasticizer specifications

Physical properties	partially concentrated liquid
Color	Light brown
Density	1.2 kg/m^3
pH	6.9
Other specifications	Without chlorine ion and non-flammable

Tab. 8 compares the analysis of the main salts in the waters of the Oman Sea with the salts in the high seas of the world as a measure, considering that the waters of the Oman Sea were used in one of the ponds.

Table 8 Comparison between high seas and Oman Sea water decomposition

Minerals	Minerals existing in sea water (ppm)	Minerals existing in Oman Sea water (ppm)
Calcium	50-480	460
Magnesium	260-1410	1400
Sodium	2190-12200	12500
Potassium	70-550	420
Sulfate	580-2810	2900
Chlorine	3960-20000	21100

4 PREPARATION OF TEST SAMPLES

In this study, ACI-211 [22] was used for the volumetric concrete mix design method. The samples were removed from the mold after 24 h and kept in lime water solution for 7 days. After 7 days, the samples were transferred to two drinking water and Oman Sea water ponds. The permeability test samples are 150 mm cubic samples as per BS EN 12390-8, and 6 samples were prepared for each mix design. Also, the cylindrical samples were 7.5 × 7.5 as per BS 1881-part 122 for water absorption test.

5 EXPERIMENTS

Concrete permeability is usually obtained by calculating the pressurized water that enters a concrete sample. In the research method, according to BS EN 12390-8, the 15 × 15 × 15 cm cubic concrete samples are subjected to constant pressure of 5 atm. After 3 days, the samples are divided into two halves and the amount of permeation is measured [19].

Figs. 1 and 2 show a view of the permeability test device, split sample, and area of water permeation.

The water absorption test is performed according to BS1881 where the 7.5 × 7.5 cylindrical samples are placed in the oven for 72 hours at 110 °C and, after cooling, for 30 minutes inside the water. After drying the surface of the samples, the water absorption percentage of the samples is obtained during this time. In this paper, the water absorption test is performed according to BS1881-part 122.

Figure 1 View of water permeability device in concrete

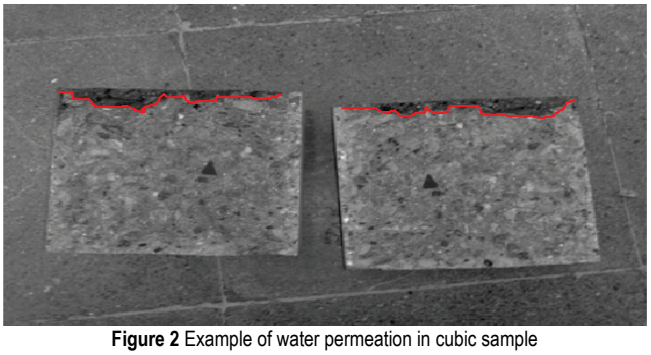


Figure 2 Example of water permeation in cubic sample

The results of the water permeability and absorption tests are presented in Tab. 9. Each value in this table is the average of the results from three concrete samples tested in drinking water and Oman Sea waters.

Table 9 Results of 28-day water permeability and absorption tests

Testing Environment	Sea Water		Drinking Water	
Test / Design No.	Half-hour water absorption (%)	Water permeability (mm)	Half-hour water absorption (%)	Water permeability (mm)
M1	2.71	17.2	2.55	14.6
M2	2.72	17.3	2.56	14.6
M3	2.48	15.5	2.3	12.8
M4	2.79	17	2.58	14.5
M5	2.38	14.4	2.18	11.6
M6	2.49	15	2.26	12.65
M7	2.1	13.7	1.9	11
M8	2.5	16	2.29	13.4

According to Tab. 3 and the results of Tab. 9, it is observed that in all the samples that used pozzolan, the water absorption and permeability values decreased. One reason is that the addition of pozzolan to concrete reduces the large pores to small ones, thereby reducing the concrete permeability. On the other hand, the secondary cement properties of pozzolan cause the pozzolan to consume calcium hydroxide in the cement paste, thereby reducing the amount of lime which is an important factor in the damage caused by sulfate attacks. Pozzolan strengthens the concrete interfacial transition zone (ITZ) (aggregate and paste interface) by converting lime to C-S-H, thereby improving the concrete properties.

Based on the laboratory results (Tab. 9), although sample M7 has the best result among the proposed initial samples, it cannot be stated that the M7 mix design is the most optimal one. Therefore, by presenting the results of the 8 mix designs by the Taguchi optimization method, the most optimal design was obtained (which may not be in the tests).

5.1 Obtaining Optimal Mix Design by Taguchi Method

In this study, the best (most effective) level of each factor was determined for each experiment. Using Minitab software and the plotted graphs and considering the "Smaller the better" for both experiments, the values of the S/N ratios can be used to determine the most effective parameters in each experiment and the optimal mix designs. The graphs in Figs. 3-6 show the results of each experiment.

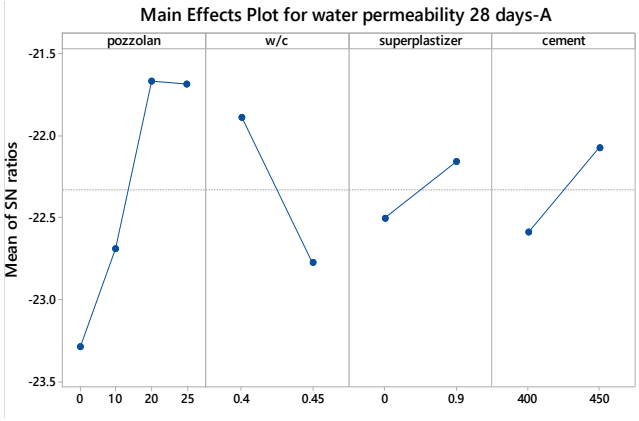


Figure 3 S/N graph for water permeability test of samples in pond 1

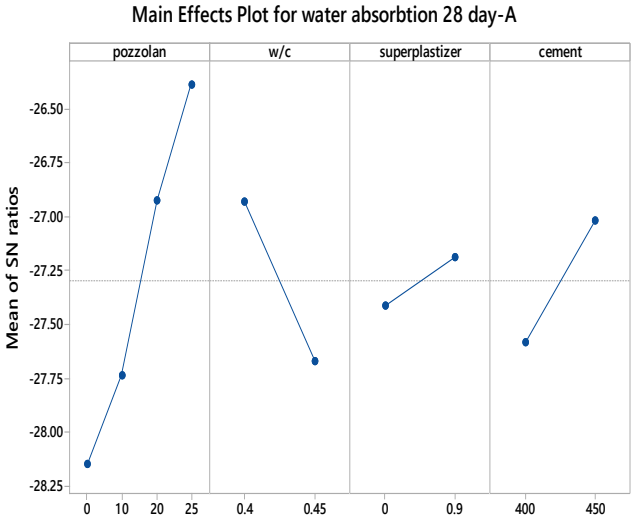


Figure 4 S/N graph for water absorption test of samples in pond 1

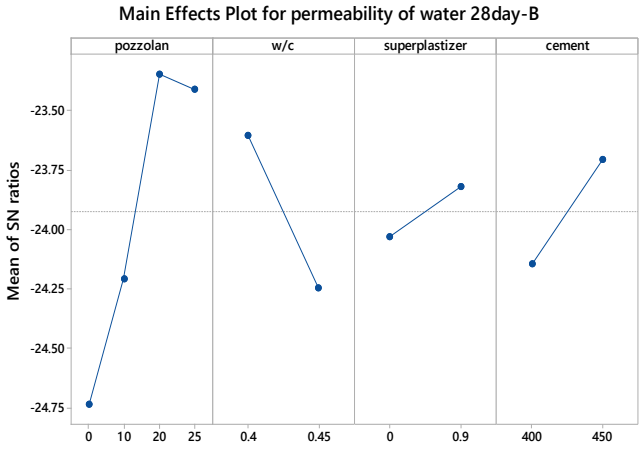


Figure 5 S/N graph for water permeability test of samples in pond 2

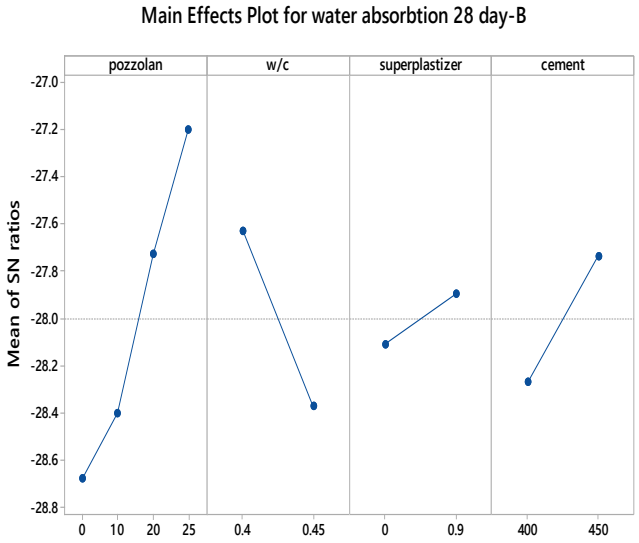


Figure 6 S/N graph for water absorption test of samples in pond 2

Figs. 3-4 are related to pond 1 (drinking water) and Figs. 4-5 are related to pond 2 (sea water). In these figures, in each parameter, the level that has the highest S/N value is specified as the best level of the parameter. The S/N results for Figs. 3-5 are presented in Tab. 10.

According to the graphs of water permeability and absorption test results, the higher the amount of pozzolan consumed, the greater the effect on the results of the water absorption test, but the permeability test graphs show that an increase in used pozzolan up to 20% has a positive effect on our results, but with a 25% substitution, we can see the reduced effect of added pozzolan on the permeability of the samples.

In Tab. 10, the Delta value is the difference between the highest and the lowest parameter levels. The higher this value, the greater the rank parameter that indicates the greater effect of the parameter on the experiment. By observing the rank parameter for the water permeability and absorption tests, it can be stated that the amount of pozzolan consumed has the greatest effect on the results of the experiments.

Also, the next effective parameter is the water to cement ratio (w/b) and ultimately, the amount of super-plasticizer has the least effect on the results of the experiments. Table 11 was obtained using the analysis of variance (ANOVA) and considering the parameters of Table 1 as independent variables. The higher the F-Value in this table, the greater the effect of the parameter on the test result. According to Figs. 3 to 6, the optimal mix design for each experiment and each environment can be obtained. Tab. 12 presents the optimal mix designs.

Tab. 12 presents the predicted values from the optimal mix design. For the validation of these values, 2 mix designs each containing 6 samples were prepared (In the laboratory conditions, the initial samples were cured and constructed, and by comparing the estimated results with the results of the mix designs, this method was validated for predicting the optimal mix design).

Tab. 13 compares the results of the M7 mix design (the best behavior in the proposed Taguchi initial samples) with the optimal Taguchi mix design. The results show that although the half-hour water absorption results did not change in the optimal Taguchi design compared to M7, the concrete permeability improved in the concrete made from the Taguchi optimal design.

Table 10 S/N values for samples in ponds 1 and 2

		Parameter	Level 1	Level 2	Level 3	Level 4	Delta	Rank
Pond 1	Permeability test	pozzolan	-23.29	-22.69	-21.67	-21.68	1.62	1
		w/b	-21.89	-22.77			0.89	2
		sp	-22.50	-22.16			0.35	4
		cement	-22.59	-22.08			0.51	3
	Absorption test	pozzolan	-28.15	-27.73	-26.93	-26.39	1.76	1
		w/b	-26.93	-27.67			0.74	2
		sp	-27.41	-27.19			0.23	3
		cement	-27.58	-27.01			0.57	4
Pond 2	Permeability test	pozzolan	-24.74	-24.21	-23.34	-23.41	1.39	1
		w/b	-23.60	-24.24			0.64	2
		sp	-24.03	-23.82			0.21	4
		cement	-24.14	-23.71				3
		pozzolan	-28.68	-28.40	-27.73	-27.20	1.47	1
		w/b	-27.63	-28.37			0.74	2
		sp	-28.11	-27.89			0.21	4
		cement	-28.27	-27.74				3

Table 11 Analysis of variance (ANOVA) for results of experiments in both ponds

			DF	Adj SS	Adj MS	F-Value	P-Value
Pond 1	Permeability test	pozzolan	1	7.61	7.61	22.53	0.018
		w/bs	1	3.31	3.31	9.81	0.052
		sp	1	0.38	0.38	1.13	0.365
		cement	1	1.16	1.16	3.44	0.162
		Error	3	1.01	0.34		
		Total	7	13.48			
	Absorption test	pozzolan	1	25.6	25.6	180.71	0.001
		w/b	1	7.22	7.22	50.96	0.006
		sp	1	0.405	0.405	2.86	0.189
		cement	1	4.205	4.205	29.68	0.012
		Error	3	0.42	0.141		
		Total	7	37.85			
Pond 2	Permeability test	pozzolan	1	7.66	7.66	19.5	0.022
		w/b	1	2.53	2.53	6.4	0.085
		sp	1	0.21	0.21	0.54	0.517
		cement	1	1.20	1.20	3.06	0.179
		Error	3	1.18	0.39		
		Total	7	12.78			
	Absorption test	pozzolan	1	20.88	20.88	126.61	0.002
		w/b	1	8.61	8.61	52.22	0.005
		sp	1	0.45	0.45	2.74	0.197
		cement	1	4.35	4.35	26.38	0.014
		Error	3	0.49	0.16		
		Total	7	34.79			

Table 12 Optimal mix design values

			pozzolan	w/b	sp	cement	Value obtained from test for optimal mix design
Drinking water	28-day permeability	value	20	0.4	0.7	450	10.88
		level	3	1	2	2	
	28-day absorption	value	25	0.4	0.7	450	1.9
		level	4	1	2	2	
Sea water	28-day permeability	value	20	0.4	0.7	450	13.56
		level	3	1	2	2	
	28-day absorption	value	25	0.4	0.7	450	2.1
		level	4	1	2	2	

Table 13 Compares of results of the M7 mix design with the optimal Taguchi mix design

	Sea water		Drinking water	
Parameter / Test	Half-hour water absorption (%)	Water permeability (mm)	Half-hour water absorption (%)	Water permeability (mm)
M7	2.1	13.7	1.9	11
M (optimization)	2.1	13.56	1.9	10.88

Tab. 14 presents the limitations of the National Concrete Durability Regulation for water permeability and absorption in the Oman Sea and the Persian Gulf [23].

Table 14 Limitations of National Concrete Durability Regulation for water permeability and absorption in Oman Sea

Test	Allowable range		
	Moderate conditions	Severe conditions	Extreme conditions
1. Half-hour water absorption (28 days)	Max 4%	Max 3%	Max 2%
2. Water permeability (28 days)	Max 50 mm	Max 30 mm	Max 10 mm

According to the results of the experiments from the Taguchi optimal mix design and the limitations of the National Concrete Durability Regulation in the Oman Sea, it can be concluded that by observing the limitations of this regulation and the positive effects of Taftan pozzolan on improving the water permeability and absorption properties, the pozzolans can be used as an mineral admixture to improve concrete properties under moderate to severe attack conditions. However, for the use in extreme conditions, due to the strict regulation limitations (Tab. 14), it is not sufficient to simply add this pozzolan to the concrete. Using this pozzolan increased the water absorption value of concrete to 2.1% and the permeability of concrete in the optimal mix design to 14.5 mm, which are close to the maximum allowable values. It is possible to meet the criteria considered in the regulation by adding a higher percentage of super-plasticizer or modifying the aggregates and reducing the water-to-cement ratio.

6 CONCLUSION

In this paper, the effect of using Taftan pozzolan on the water permeability and absorption of concrete structures in the Oman Sea and drinking water was investigated. The Taguchi optimization method was used to reduce the time and cost of sample construction. The four parameters of water to cement ratio, different percentages of natural pozzolan, super-plasticizer, and grade of cementitious materials with different ratios were considered as problem variables. The Taguchi method proposed 8 mix designs, and

192 samples were prepared to validate the results and tested in the both Oman Sea and drinking waters. By comparing the results with the National Concrete Durability Regulation in the Persian Gulf and Oman Sea, it was found that the use of Taftan pozzolan was effective in reducing the water permeability (20%) and absorption in the moderate to severe conditions (25%). In the extreme conditions, the results are close to the critical limits of the regulation, but they failed to meet the regulation criteria. Taguchi method can determine the ratio of the use of pozzolans in the designed concrete mixing ratio with an accuracy of about 2% compared to the laboratory results.

7 REFERENCES

[1] Uzal, B., & Turanli, L. U. T. F. U. L. L. A. H. (2003). Studies on blended cements containing a high volume of natural pozzolans. *Cement and Concrete Research*, *33*(11), 1777-1781. https://doi.org/10.1016/S0008-8846(03)00173-X

[2] Najimi, M., Sobhani, J., Ahmadi, B., & Shekarchi, M. (2012). An experimental study on durability properties of concrete containing zeolite as a highly reactive natural pozzolan. *Construction and Building Materials*, *35*, 1023-1033. https://doi.org/10.1016/j.conbuildmat.2012.04.038

[3] Ezziane, K., Bougara, A., Kadri, A., Khelafi, H., & Kadri, E. (2007). Compressive strength of mortar containing natural pozzolan under various curing temperature. *Cement and Concrete Composites*, *29*(8), 587-593. https://doi.org/10.1016/j.cemconcomp.2007.03.002

[4] Nguyen, V. H., Leklou, N., Aubert, J. E., & Mounanga, P. (2013). The effect of natural pozzolan on delayed ettringite formation of the heat-cured mortars. *Construction and Building Materials*, *48*, 479-484. https://doi.org/10.1016/j.conbuildmat.2013.07.016

[5] Ghrici, M., Kenai, S., & Said-Mansour, M. (2007). Mechanical properties and durability of mortar and concrete containing natural pozzolana and limestone blended cements. *Cement and Concrete Composites*, *29*(7), 542-549. https://doi.org/10.1016/j.cemconcomp.2007.04.009

[6] Bondar, D., Lynsdale, C. J., Milestone, N. B., Hassani, N., & Ramezanianpour, A. A. (2011). Effect of adding mineral additives to alkali-activated natural pozzolan paste. *Construction and Building Materials*, *25*(6), 2906-2910. https://doi.org/10.1016/j.conbuildmat.2010.12.031

[7] Kaid, N., Cyr, M., Julien, S., & Khelafi, H. (2009). Durability of concrete containing a natural pozzolan as defined by a performance-based approach. *Construction and Building Materials*, *23*(12), 3457-3467. https://doi.org/10.1016/j.conbuildmat.2009.08.002

[8] EN, B. S. (2000). 12390-8 Depth of penetration of water under pressure. *British Standards Institution*.

[9] Standard, B. & ISO, B. (1999). Plastics—Determination of water absorption. *PN-EN ISO*, *62*.

[10] Merida, A. & Kharchi, F. (2015). Pozzolan concrete durability on sulphate attack. *Procedia Engineering*, *114*, 832-837. https://doi.org/10.1016/j.proeng.2015.08.035

[11] Omrane, M., Kenai, S., Kadri, E. H., & Aït-Mokhtar, A. (2017). Performance and durability of self compacting concrete using recycled concrete aggregates and natural pozzolan. *Journal of Cleaner Production*, *165*, 415-430. https://doi.org/10.1016/j.jclepro.2017.07.139

[12] Firdous, R., Stephan, D., & Djobo, J. N. Y. (2018). Natural pozzolan based geopolymers: a review on mechanical, microstructural and durability characteristics. *Construction and Building Materials*, *190*, 1251-1263. https://doi.org/10.1016/j.conbuildmat.2018.09.191

[13] Najimi, M. & Ghafoori, N. (2019). Engineering properties of natural pozzolan/slag based alkali-activated concrete. *Construction and Building Materials*, *208*, 46-62. https://doi.org/10.1016/j.conbuildmat.2019.02.107

[14] Ranjit, K. (1990). Roy, 1990. A primer on the Taguchi method. *Soc. Manufac. Eng*.

[15] Ibrahim, M. R., Ismail, N., Leman, Z., & Majid, D. L. A. H. A. (2012). Experimental investigation of HSS face milling to AL6061 using Taguchi method. *Procedia Engineering*, *50*, 933-941.

[16] Ozbay, E., Oztas, A., Baykasoglu, A., & Ozbebek, H. (2009). Investigating mix proportions of high strength self compacting concrete by using Taguchi method. *Construction and building materials*, *23*(2), 694-702. https://doi.org/10.1016/j.conbuildmat.2008.02.014

[17] Türkmen, İ., Gül, R., & Çelik, C. (2008). A Taguchi approach for investigation of some physical properties of concrete produced from mineral admixtures. *Building and environment*, *43*(6), 1127-1137. https://doi.org/10.1016/j.buildenv.2007.02.005

[18] Lyu, S. R., Te Wu, W., Hou, C. C., & Hsieh, W. H. (2010). Study of cryopreservation of articular chondrocytes using the Taguchi method. *Cryobiology*, *60*(2), 165-176. https://doi.org/10.1016/j.cryobiol.2009.10.008

[19] Fowlkes, W. Y. & Creveling, C. M. (1995). *Engineering methods for robust product design: using Taguchi methods in technology and product development*. Addison-Wesley.

[20] Phadke, M. S. (1989). *Quality engineering using robust design: Produkt-u. Prozeßoptimierung mit d. Methoden d. statistischen Versuchsplanung nach G. Taguchi*. Gfmt, Ges. für Management und Technologie. https://doi.org/10.1007/978-1-4684-1472-1_3

[21] (2002). American Society for Testing and Materials, ASTM C33, Standard Specification for Concrete Aggregates.

[22] (2001). American Concrete Institute, ACI 211, Guide for Submittal of Concrete Proportions.

[23] Safehian, M. & Ramezanianpour, A. A. (2013). Assessment of service life models for determination of chloride penetration into silica fume concrete in the severe marine environmental condition. *Construction and Building Materials*, *48*, 287-294. https://doi.org/10.1016/j.conbuildmat.2013.07.006

Authors' contacts:

Amir Behshad
(Corresponding Author)
Faculty of Industry and Mining, Yasouj University,
12345 Yasouj, Iran
amir.behshad@hotmail.com

Mostafa Askarian Zardak
Department of Structural Engineering,
Zahedan University, Iran

Yaser Mohammadi
Ferdowsi University of Mashhad,
Mashhad, Iran

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20200815184439

Original scientific paper

The Three-Objective Optimization Model of Flexible Workshop Scheduling Problem for Minimizing Work Completion Time, Work Delay Time, and Energy Consumption

Neda Karim Ahangar, Majid Khalili*, Hamed Tayebi

Abstract: In recent years, the optimal design of the workshop schedule has received much attention with the increased competition in the business environment. As a strategic issue, designing a workshop schedule affects other decisions in the production chain. The purpose of this thesis is to design a three-objective mathematical model, with the objectives of minimizing work completion time, work delay time and energy consumption, considering the importance of businesses attention to reduce energy consumption in recent years. The developed model has been solved using exact solution methods of Weighted Sum (WS) and Epsilon Constraint ($\mathcal{E}$) in small dimensions using GAMS software. These problems were also solved in large-scale problems with NSGA-II and SFLA meta-heuristic algorithms using MATLAB software in single-objective and multi-objective mode due to the NP-Hard nature of this group of large and real dimensional problems. The standard BRdata set of problems were used to investigate the algorithms performance in solving these problems so that it is possible to compare the algorithms performance of this research with the results of the algorithms used by other researchers. The obtained results show the relatively appropriate performance of these algorithms in solving these problems and also the much better and more optimal performance of the NSGA-II algorithm compared to the performance of the SFLA algorithm.

Keywords: energy consumption; flexible workshop scheduling; makespan; multi-objective optimization; NSGA-II Algorithm; SFLA Algorithm

1 INTRODUCTION

In recent years, the optimal design of the workshop schedule has attracted a lot of attention by the increased competition in the business environment. Workshop scheduling is defined as a strategic problem affecting other decisions in the production chain [1]. The workshop scheduling problem can be divided into two static and dynamic categories. In the static state, n work must be done on the m machine while maintaining a certain sequence. Each task in this environment consists of different operations with the known processing time and processing path on machines [4]. The scheduler tries to optimally allocate a set of resources to a set of tasks to be performed over a period of time. The workshop schedule is a form of classical scheduling problem that has been widely considered in various fields of engineering sciences. Considering the importance of scheduling in workshop and production environments, there is a need for extensive research to address various aspects of the workshop scheduling problem. The flexible workshop flow problem is very common in the real world and has received a lot of attention in recent years. The main problem in the present research is to minimize the total work delay and energy consumption in flexible workshop scheduling problems. A metaheuristic or hybrid metaheuristic algorithms was used to solve this problem and the optimal solutions of each algorithm was compare with each other. Finally it is examined that which of the metaheuristic algorithms provides the best answer to solve the problem in this research. Also, in order to evaluate the efficiency of the method and model, a case study from industry is used in this study to examine the efficiency of the results in the real world. The present study solves the problem of flexible workshop flow scheduling by considering some real-world hypotheses that have not yet been explored. Considering the application of the problem in many production environments, this research can take a small step in solving this problem. The present research aimed to improve operations in the flexible workshop flow problem due to the importance of scheduling.

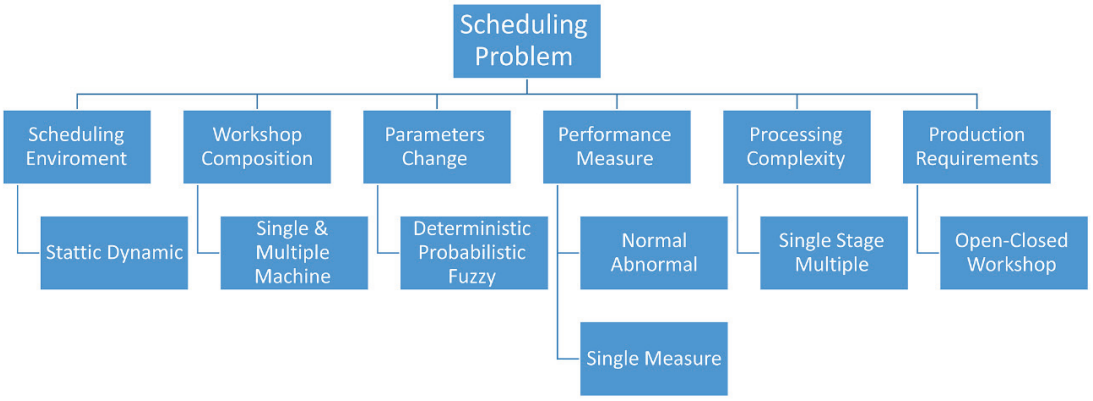


Figure 1 Categorizing scheduling problems [15]

2 LITERATURE REVIEW AND RESEARCH BACKGROUND

The Flexible Workshop Flow Scheduling (FFS) Problem involves sequencing a workshop flow problem where there are at least one or more dissimilar parallel machines at each stage. The objective function of the problem is to minimize the maximum time to complete tasks [2]. Scheduling has been proposed as one of the new research fields since 1954 (Asadi et al., 2015). Scheduling is a decision-making process that plays an important role in production systems so that the performance criteria of any production system can be improved by an effective and efficient scheduling program [1].

Workshop scheduling can be classified into five main categories according to the workshop environment, single machine, parallel machines, workshop flow, closed workshop, and open workshop (Allahverdi and Soroush 2008, Hall 1998). All workshop scheduling issues belong to the NP-Hard class. Tab. 1 shows the types of scheduling issues.

Table 1 Some flexible workshop flow scheduling issues [13]

Authors' name	Model description
Jalalab & Jalab, 2002	Work permutation
Jungwattanakit et al., 2005, 2008, 2009; Yuarima et al., 2009	The problem of flexible workshop flow with the assumption of unrelated parallel machines
Kurz and Askin, 2004; Logendran, Vo and Witt, 2007; Jungwattanakit, 2008; Zandieh and Gholami, 2009; Fattahi et al., 2015	Sequence-dependent preparation on machines
Riane et al., 2001; Alvi and Arbita, 2004	Transport between machines
Naderi et al., 2009	Limited buffer capacity between two consecutive steps
Sawik, 2002; Akrami, 2006	Prioritization of work
Hentous and Benhammadi, 2006	Maintenance constraints
BottaGenoulaz, 2000; Wu et al., 2010	Dynamic uncertainty
Aloe and Arbita, 2004; Aloe and Arbita, 2006	Prerequisite constraint between tasks and wastage times constraint between steps and with the objective function of minimizing the maximum latency
Hong and Wong, 2000; Alisantoso, 2003	The capacity constraint of intermediate warehouses without considering it with the aim of minimizing the maximum completion time, the weighted total time during construction and the weighted total of delay times
Naderi et al., 2009; Janiak et al., 2007; Behnamian and Zandieh, 2011	The problem of flexible multi-objective workshop flow

The problem of a flexible two-stage workshop flow with the same number of parallel machines in the first stage and one machine in the second stage has been investigated in Tran and Ming (2011), aimed at minimizing the maximum completion time. BottaGenoulaz (2000) examined the problem of flexible workshop flow with pre-requisite limit between tasks and waste time constraints between stages and with the objective function of minimizing the maximum latency and proposed six new innovative methods [3]. The flexible workshop flow problem with the objective function of sum of early and late with and waiting time and presented three metaheuristic algorithms and three constructive algorithms for it has been considered by Janiak et al. (2007). The problem of flexible workshop flow with limited waiting time and the objective function of the sum of early and late squares has been investigated by Behnamian and Zandieh (2011) that proposed a discrete colonial competition algorithm to solve it.

The problem of efficient multi-objective energy scheduling, with two objectives: completion time and energy consumption in production systems is developed by Dai et al. (2013) investigated. They used a measurement between completion time and energy consumption [14].

The problem of scheduling tasks in a flexible workshop environment have been investigated by Wang et al. (2018). Their objective function is the total energy consumption and their solution method is a two-step initiative that in the first and second stages have used genetic metaheuristic methods and particle swarm optimization, respectively. Also, a two-objective model is presented by Wong et al. (2018) in which, they have scheduled parallel and identical machines whose goals are the total energy consumption and the time of completion of works. They have used the Epsilon constraint method to accurately solve their model [28].

A study on green planning during a two-machine workshop is presented by Mansouri et al. (2016) presented to examine the relationship between completion time and energy consumption. The developed mathematical model combines the main topics of the workshop: service level and energy consumption. The metaheuristic algorithms have been used to solve large-scale problems [39].

3 RESEARCH METHOD

The present study is an applied research conducted with analytical-descriptive approach. This research presents a multi-objective model of flexible workshop scheduling. The model developed in this research includes three important objectives of minimizing completion time, delay and energy consumption in workshop scheduling issues. The developed model is solved by the multi-objective meta-heuristic algorithm approach. This approach is used to solve multi-objective problems. The present research model includes three objective functions. One objective function seeks to reduce work completion time in workshop flow problems, and the second objective function seeks to minimize work delays; the last objective function seeks to minimize energy consumption levels due to the problems, shortages, and high energy costs in today's world.

The counters, decision variables and problem parameters used in this research are as follows:

Indices and sets	
Machine index, $i = 1,2,\ldots, m$	i, i'
Work index, $j = 1,2,\ldots, n$	j, j'
Operation index of each task, $h = 1,2,\ldots, n_j$	h, h'
Processing speed index	l
A set of machines capable of processing the operations of jh work	E_{jh}
Parameters	

Large positive number	M
Number of work operations j	n_j
Delivery date j	d_j
Processing time of h^{th} operation of j^{th} work on i^{th} machine i	p_{jhi}
Processing speed factor	v_l
Conversion factor for processing speed l	α_l
Conversion factor for machine idle time i	β_i
Decision variables	
Total completion time	$C_{\max}$
Total energy consumption in kWh	TEC
Total delay time	$T_{\max}$
Termination of h^{th} operation of j^{th} work on i^{th} machine	C_{jhi}
The i^{th} machine idle time	θ_i
Work delay time j	T_j
It is a binary variable, it is 1 if the h^{th} operation of j^{th} work on i^{th} machine is processed at speed l, and zero, otherwise	X_{jhil}
It is a binary variable, it is 1 if the h^{th} operation of j^{th} work occurs after the h'^{th} operation of j'^{th} work, and zero, otherwise	$Y_{jhj'h'}$

Accordingly, the problem mathematical model will be as follows:

$$\min z_1 = C_{\max} \tag{1}$$

$$\min z_2 = TEC \tag{2}$$

$$\min z_3 = T_{\max} \tag{3}$$

$$C_{jhi'} - \sum_{i \in E_{j(h+1)}} \left(C_{j(h+1)i} - \sum_{l=1}^{L} \left(\frac{p_{j(h+1)i}}{v_l} \right) \times X_{jhil} \right) \le 0 \quad \forall j, h < n_j, i' \in E_{jh} \tag{4}$$

$$C_{jhi} - \left(\frac{p_{jhi}}{v_l} \right) - C_{j'h'i} \ge \ge -M \times \left(2 - X_{jhil} - \sum_{l'=1}^{L} X_{j'h'il'} \right) - M \times \left(1 - Y_{jhj'h'} \right) \quad \forall j < j', h, h', i, i', l \tag{5}$$

$$C_{j'h'i} - \left(\frac{p_{j'h'i}}{v_l} \right) - C_{jhi} \ge \ge -M \times \left(2 - \sum_{l'=1}^{L} X_{jhil'} - X_{j'h'il} \right) - M \times \left(1 - Y_{jhj'h'} \right) \quad \forall j < j', h, h', i, i', l \tag{6}$$

$$\sum_{i \in E_{jh}} \sum_{l=1}^{L} X_{jhil} = 1 \quad \forall j, h \le n_j \tag{7}$$

$$C_{jhi} \le M \times \sum_{l=1}^{L} X_{jhil} \quad \forall j, h \le n_j, i \in E_{jh} \tag{8}$$

$$\sum_{i \in E_{j1}} \left(C_{j1i} - \sum_{l=1}^{L} \left(\frac{p_{j1i}}{v_l} \right) \times X_{j1il} \right) \ge 0 \quad \forall j \tag{9}$$

$$C_{\max} \ge C_{jhi} \quad \forall j, h \le n_j, i \in E_{jh} \tag{10}$$

$$\theta_i = C_{\max} - \sum_{j=1}^{n} \sum_{h=1}^{n_j} \sum_{l=1}^{L} \left(\frac{p_{jhi}}{v_l} \right) \times X_{jhil} \quad \forall i \in E_{jh} \tag{11}$$

$$TEC = \sum_{i \in E_{jh}} \sum_{j=1}^{n} \sum_{h=1}^{n_j} \sum_{l=1}^{L} \alpha_l \times \left(\frac{p_{jhi}}{v_l} \right) \times X_{jhil} + \sum_{i=1}^{m} \beta_i \times \theta_i \tag{12}$$

$$T_j \ge \sum_{i \in E_{jh}} C_{jhi} - d_j \quad \forall j, h = n_j \tag{13}$$

$$T_{\max} \ge T_j \quad \forall j \tag{14}$$

$$C_{jhi}, T_j, \theta_i, C_{\max}, TEC, T_{\max} \ge 0 \quad X_{jhil}, Y_{jhj'h'} \in \{0, 1\} \tag{15}$$

In the above model, the first objective function (1) maximizes the maximum completion time; the second objective function (2) minimizes the total energy consumption, and the third objective function (3) minimizes the maximum delay time. Constraint (4) causes that the operation of one task does not start until the next operation of that task is completed. Constraints (5) and (6) prevent interference of two operations on a machine. Based on constraint (7), the operation of each task is definitely processed on a machine at a certain speed. Constraint (8) shows the end time of each operation. Based on constraint (9), the completion time of the first operation of each task is a positive value. Constraint (10) indicates the maximum termination time. Constraint (11) indicates the unemployment rate of each machine. Constraint (12) calculates the total energy consumption. Constraint (13) and (14) calculates the amount of latency of each task and the maximum amount of latency, respectively. The constraint (15) shows the problem variables.

Solving the mathematical models and problem analysis:

The meta-heuristic algorithm has been used in this research. In the first step, the main input parameters of this algorithm must be set.

In this section, the input parameter for the NSGA-II algorithm must be adjusted. The experimental design and Taguchi method are used to design the parameter.

The parameters of this algorithm are as follows:

- *nPop*: Initial population size,
- *Pc*: Intersection probability,
- *Pm*: Mutation probability
- *Maxit*: Maximum number of repetitions.

The factors table is as follows:

Table 2 Important factors of NSGA-II algorithm

Parameter	Symbol	Levels		
		1	2	3
nPop	A	30	50	70
Pc	B	0.65	0.75	0.99
Pm	C	0.05	0.25	0.45
Maxit	D	40	60	80

The table is as follows for 4 factors in the three levels of Taguchi:

Table 3 Taguchi L-9 series

No. of experiment	A	B	C	D
1	1	1	1	1
2	2	2	2	1
3	3	3	3	1
4	3	2	1	2
5	1	3	2	2
6	2	1	3	2
7	2	3	1	3
8	3	1	2	3
9	1	2	3	3

3.1 Evaluating the Algorithms Efficiency with Numerical Examples for Large-Scale Problems

In this section, the efficiency of the proposed algorithms to solve the problem of this research will be investigated. The algorithms were coded using MATLAB 2019 and run on a system with 8GB of internal storage and an i7 CPU. In order to evaluate these algorithms in this research, a standard test data set called FJSPLIB, which is available at http://people.idsia.ch, has been used. In this set, standard test problems are used to evaluate the algorithms performance. There is a coded version that has a set of standard problems called Bardata, BRdata, Daudata and Huridata. In this research, Brdata Set has been used, which includes 10 sample problem groups as presented by Brandimart [1]:

Table 4 Specifications of standard problems in the BRdata set

Sample problem number in BRdata	Number of jobs	Number of machines	Total number of operations	Maximum number of machines
MK01	10	6	55	3
MK02	10	6	58	6
MK03	15	8	150	5
MK04	15	8	90	3
MK05	15	4	106	2
MK06	10	15	150	5
MK07	20	5	100	5
MK08	20	10	225	2
MK09	20	10	240	5
MK10	20	15	240	5

In this standard set designed by Brandimart (1993) [19], the parameters of each of the problems in this set are randomly generated between two limits using a uniform distribution. The number of jobs is from 10 to 20, the number of machines is 4 to 15, the number of operations for each job is 5 to 15 and the number of operations for all jobs is 55 to 241. All parameters related to this data set are shown separately in the table above

As these problems are standard and different researchers in different years have used this series of standard problems to evaluate the performance of their chosen algorithm in solving the flexible single-objective workshop scheduling problem by minimizing the completion time, so it is possible to compare the performance of these two algorithms. Research with this series of solutions in single-objective mode is also possible, so first the performance of these algorithms in solving single-objective problem has been examined to determine the performance of these two algorithms in comparison with other algorithms in the research literature and then, three research objectives have been used for solving problem to compare the performance of these two algorithms in relation to each other.

Table 5 The algorithms performance in solving standard problems ([19] Brandimart, 1993)

Sample problem number in BRdata	$n \times m$ number of worker × number	High limit and low limit (LB, UB)	MAPSO method (Nouiri et al., 2015)	Modified Ant Method (IACO) (Wang et al, 2017)	NSGA-II algorithm of the present study	The SFLA algorithm of the present study
MK01	10×6	42,36	41	40	40	41
MK02	10×6	32,24	26	26	26	26
MK03	15×8	211,204	207	204	204	204
MK04	15×8	81,48	65	60	63	65
MK05	15×4	186,168	171	173	168	180
MK06	10×15	86,33	61	60	55	61
MK07	20×5	173,133	173	140	155	145
MK08	20×10	523	523	523	523	523
MK09	20×10	369,299	307	307	307	312
MK10	15×20	296,165	312	208	245	298

Table 6 Time to solution and average results of the proposed algorithms

Sample problem number in BRdata	Best answer BKS	IACO method (Wang et al., 2017)		NSGA-II algorithm of the present study		The SFLA algorithm of the present study	
		CPU	AVG CM	CPU	AVG CM	CPU	AVG CM
MK01	40	40.30	1.09	1.16	40.15	4.01	41.23
MK02	26	26.10	2.16	1.48	26.20	6.09	26.15
MK03	204	204	2.18	9.18	207	10.70	207.10
MK04	62	60	9.02	2.35	63.11	3.87	65.55
MK05	172	173.2	7.10	3.70	168.39	4.88	185
MK06	58	60.30	30.12	10.70	55.5	26.38	61.1
MK07	139	141.5	17.07	3.26	159	26.21	151
MK08	523	523	4.30	11.52	523	189.41	523
MK09	310	315.2	91.99	28.94	308.1	122.87	307.3
MK10	214	213.1	190.11	33.44	254.34	189.41	301.1

In the table above, the upper and lower limit (LB, UB) is the optimal answer if the optimal answer found is time-consuming for the completion time (Makespan), otherwise the upper and lower limit found is set yet. For example, in the MK08 problem series, the optimal solution is 523, which in this row is only 523, which indicates the same case, and both modified ant algorithms (Wang et al., 2017) called IACO and MAPSO method (Nouri et al., 2015) have been able to find the optimal solution; however, heuristic methods have only been able to find the near-optimal 555 solution, and for the rest of the standard problems in the table, the optimal solution has not yet been achieved. The research proposed algorithms are NSGA-II and SFLA, which have obtained acceptable results compared to other algorithms and in the MK08 series problem has also been able to obtain the optimal answer. In the following, the solution time and the average solution results obtained by these algorithms are discussed in order to evaluate the performance of the algorithms both in terms of optimal solution and in terms of solution time.

3.2 The Algorithms Performance in Solving the Proposed Multi-Objective Problem

As mentioned earlier, the proposed mathematical model of the problem has three objectives: minimizing the total completion time (Makespan) C_{max} and the total delay time T_{max}, and finally minimizing the total energy consumption to do the job displayed with *TEC*. NSGA-II and SFLA algorithms have been used to solve this three-objective model and Brdata standard problems have been used to solve this multi-objective model. The status of Pareto answers is as follows, which shows the better performance of NSGA-II algorithm than the SFLA algorithm.

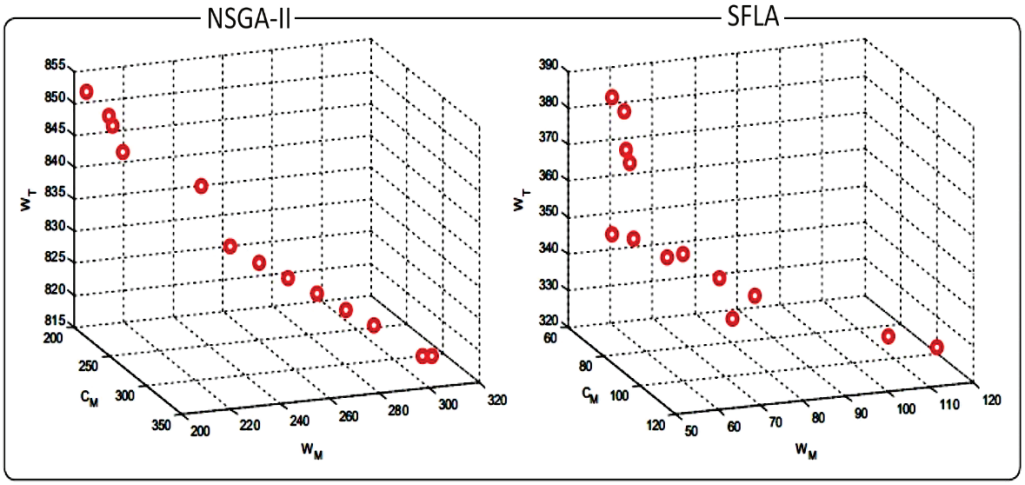


Figure 2 Pareto front set of answers found by the two studied algorithms

The results of the problem objective functions by these two algorithms are provided to solve these standard problems.

Table 7 Results of the algorithms performance in solving the three-objective problem

Name of problem	NSGA-II algorithm	SFLA algorithm
MK01		
C_{max}	40	40
T_{max}	167	169
TEC	36	36
MK02		
C_{max}	26	26
T_{max}	151	151
TEC	26	26
MK03		
C_{max}	204	204
T_{max}	855	852
TEC	199	204
MK04		
C_{max}	61	66
T_{max}	345	366
TEC	63	61
MK05		
C_{max}	173	172
T_{max}	683	687
TEC	173	172
MK06		
C_{max}	62	65
T_{max}	424	398
TEC	55	62
MK07		
C_{max}	139	140
T_{max}	693	695
TEC	139	140
MK08		
C_{max}	523	523
T_{max}	2524	2524
TEC	515	523
MK09		
C_{max}	311	310
T_{max}	2290	2294
TEC	299	301
MK10		
C_{max}	214	214
T_{max}	2053	2082
TEC	204	210

Based on the above table, the results obtained in the objective functions, and as all objectives are minimization, it can be concluded that the relative performance of NSGA-II algorithm in solving this proposed problem than SFLA algorithm is more appropriate, so that in all objectives such as minimizing completion time and energy costs, this algorithm performs much more better than the SFLA algorithm and achieves better results.

3.3 Sensitivity Analysis of NSGA-II Algorithm Parameters

Sensitivity analysis was performed for the NSGA-II multi-objective algorithm as in the single-objective mode. The standard MK02 problem has been solved with different parameters of this algorithm in different modes. Four levels are considered by Taguchi method using each of the main parameters of this algorithm such as *nPop*, *Pc*, *Pm* and *Maxit*.

Table 8 Sensitivity analysis of NSGA-II algorithm parameters

Parameter name	Parameter change interval	Level			
		1	2	3	4
nPop	150-50	50	75	100	150
Pc	0.9-0.6	0.60	0.7	0.75	0.9
Pm	0.5-0.05	0.05	0.1	0.25	0.5
Maxit	100-10	10	30	60	100

The following table also shows the Design of Experiments (DOE) of each orthogonal array.

Table 9 Sensitivity analysis of NSGA-II algorithm parameters

No. of experiment	Factor level number				Average amount of C_{max}
	Pm	*Pc*		*nPop*	*Maxit*
29.4	29.4	3	1	1	28.8
28.2	28.2	4	2	1	29.4
27.8	27.8	2	3	1	28.2
29.2	29.2	1	4	1	27.8
28.8	28.8	2	1	2	29.2
29.0	29.0	3	2	2	28.8
28.8	28.8	4	3	2	29.0
28.4	28.4	1	4	2	28.8
28.8	28.8	1	1	3	28.4
28.8	28.8	2	2	3	28.8
28.6	28.6	3	3	3	28.8
29.4	27.6	4	4	3	28.6
27.6	28.6	3	1	4	27.6
28.6	29.0	1	2	4	28.6
29.0	28.8	4	3	4	28.8
28.8	28.8	2	4	4	29.4

Table 10 The change rate (Delta) of each parameter

Factor	*nPop*	*Maxit*	*Pc*	*Pm*
1	28.55	28.60	28.50	28.40
2	28.95	28.30	28.90	28.75
3	28.65	28.65	28.75	28.50
4	28.50	28.80	28.50	29.00
(Delta) changes	0.45	0.50	0.40	0.60

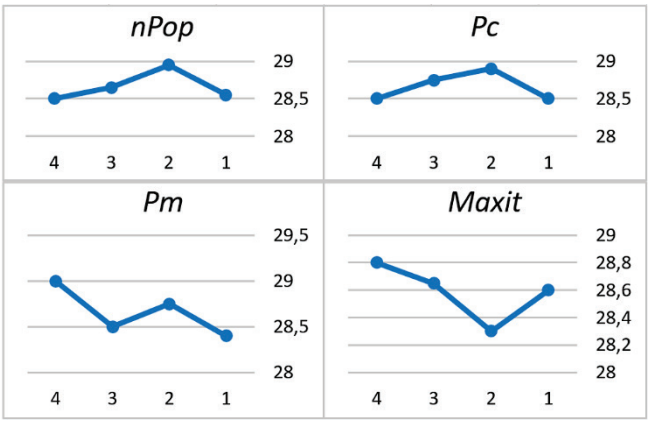


Figure 3 Sensitivity analysis of NSGA-II algorithm parameters

Tab. 10 also indicates the amount of change (Delta) of each parameter of the NSGA-II algorithm. The results of this table shows that the Pm parameter is the most effective parameter and the *Maxit* parameter is the second most effective parameter after *Pm* and PC parameter is the least effective parameter. Therefore, the Pm parameter is the most important and effective parameter in the NSGA-II algorithm in the series of flexible multi-objective workshop scheduling problems.

4 CONCLUSION

The present study aimed to develop a multi-objective mathematical model in the field of flexible workshop flow scheduling. Various optimization techniques have been used in order to achieve the research objectives. In the first step, a set of the problem hypotheses, constraints, and objectives were formulated mathematically. The resulting multi-objective model, as mentioned earlier, is one of the NP-Hard problems that can be solved only in small sizes with exact mathematical methods using GAMS IDE/Cplex software. However, since real-world problems are often larger and more complex, meta-heuristic algorithms were used to solve large problems. Two powerful multi-objective algorithms, namely NSGA-II and SFLA, were also used in this study on a large scale. A set of standard problems in the research literature called BRdata were also used to evaluate the performance of this algorithm, which all researchers around the world use to evaluate their developed methods and solution algorithms. The results show the optimal performance of these algorithms compared to other algorithms used in previous research. The NSGA-II algorithm also performed better than the SFLA algorithm. MATLAB program has been used to code meta-heuristic algorithms. In order to adjust the parameters of the algorithms, the well-known Taguchi method has been used. In addition, according to the sensitivity analysis of algorithms and their performance, the more effective parameter in the performance results of this algorithm in the superior NSGA-II algorithm is the parameter pm. The following issues can be addressed in the future research:

- Developing a mathematical model with items such as adding some constraints such as availability of machinery, possible failure of machinery, start-up time etc.
- Using other new metaheuristic and hybrid algorithms to solve these problems
- Developing new methods such as nonlinear regression and neural network model and neural-fuzzy networks such as FIS and ANFIS, etc. to predict the maximum completion time and new approaches in solving these problems
- Applying other Pareto-based multi-objective methods such as NRGA and SPEA2 for the present research problem
- Coding the problem with other programming languages such as JAVA, etc. and evaluating the performance results
- Using the proposed mathematical model to a real problem in the industry and reviewing the results
- Solving other existing standard models (Bardata, Daudata, and Huridata) in the research literature with the proposed algorithms and comparing their performance in solving single-objective and multi-objective problems.

5 REFERENCES

[1] Alaykyran, K., Engin, O., & Doyen, A. (2007). Using ant colony optimization to solve hybrid flow shop scheduling problems. *Int J Adv Manuf Techno, 35*, 541-550. https://doi.org/10.1007/s00170-007-1048-2

[2] Behnamian, J. & Zandieh, M. (2011). A discrete colonial competitive algorithm for hybrid flowshop scheduling to

minimize earliness and quadratic tardiness penalties. *Expert Systems with Applications, 38*, 14490-14498. https://doi.org/10.1016/j.eswa.2011.04.241

[3] Azar, A. & Sadeghi, A. (2020). *Service Supply Chain and its Performance Evaluation*. Saffar Publications, ISBN: 978-964-388-587-8.

[4] Jenabi, M., Fatemi-Ghomi, S. M. T., Torabi, S. A., & Karimi, B. (2007). Two hybrid metaheuristics for the finite horizon elsp in flexible flow lines with unrelated parallel machines. *Applied Mathematics and Computation, 186*(1), 230-245. https://doi.org/10.1016/j.amc.2006.06.121

[5] Khalouli, S., Ghedjati, F., & Hamzaoui, A. (2010). A meta-heuristic approach to solve a JIT scheduling problem in hybrid flow shop. *Engineering Applications of Artificial Intelligence, 23*, 765-771. https://doi.org/10.1016/j.engappai.2010.01.008

[6] Kaplanoğlu, V. (2016). An object-oriented approach for multiobjective flexible job-shop scheduling problem. *Expert Systems with Applications, 45*(1), 71-84. https://doi.org/10.1016/j.eswa.2015.09.050

[7] Gen, M. & Lin, L. (2014). Multiobjective evolutionary algorithm for manufacturing scheduling problems: state-of-the-art survey. *Journal of Intelligent Manufacturing, 25*(5), 849-866. https://doi.org/10.1007/s10845-013-0804-4

[8] Li, J., Pan, Q., & Xie, S. (2012). An effective shuffled frog-leaping algorithm for multi-objective flexible job shop scheduling problems. *Applied Mathematics and Computation, 218*(18), 9353-9371. https://doi.org/10.1016/j.amc.2012.03.018

[9] Wang, J.-J. & Liu, Y.-J. (2014). Single-machine bicriterion group scheduling with deteriorating setup times and job processing times. *Applied Mathematics and Computation, 242*, 309-314. https://doi.org/10.1016/j.amc.2014.05.048

[10] Wang, L., Wang, S., & Liu, M. (2013). A Pareto-based estimation of distribution algorithm for the multi-objective flexible jobshop scheduling problem. *International Journal of Production Research, 51*(12), 3574-3592. https://doi.org/10.1080/00207543.2012.752588

[11] Wang, L., Zhou, G., Xu, Y., & Liu, M. (2019). An enhanced Paretobased artificial bee colony algorithm for the multi-objective flexible job-shop scheduling. *The International Journal of Advanced Manufacturing Technology, 60*(9-12), 1111-1131. https://doi.org/10.1007/s00170-011-3665-z

[12] Zeballos, L. J. (2010). A constraint programming approach to tool allocation and production scheduling in flexible manufacturing systems. *Robotics and Computer-Integrated Manufacturing, 26*(6), 725-743. https://doi.org/10.1016/j.rcim.2010.04.005

[13] Driss, I., Mouss, K. N., & Laggoun, A. (2015). A new genetic algorithmfor flexible job-shop scheduling problems. *Journal of Mechanical Science and Technology, 29*(3), 1273-1281. https://doi.org/10.1007/s12206-015-0242-7

[14] Wu, X. L. & Wu, S. M. (2015). An elitist quantum-inspired evolutionary algorithm for the flexible job-shop scheduling problem. *Journal of Intelligent Manufacturing, 26*(2), 1-7.

[15] Gao, K. Z., Suganthan, P. N., Chua, T. J., Chong, C. S., Cai, T. X., & Pan, Q. K. (2015). A two-stage artificial bee colony algorithm scheduling flexible job-shop scheduling problem with new job insertion. *Expert Systems with Applications, 42*(21), 7652-7663. https://doi.org/10.1016/j.eswa.2015.06.004

[16] Brandimarte, P. (1993). Routing and scheduling in a flexible job shop by tabu search. *Ann Oper Res, 41*, 157-183. https://doi.org/10.1007/BF02023073

[17] Artigues, C. & Feillet, D. (2008). A branch and bound method for the job-shop problem with sequence-dependent setup times. *Ann Oper Res, 159*(1), 135-159. https://doi.org/10.1007/s10479-007-0283-0

[18] Bagheri, A., Zandieh, M., Mahdavi, I., & Yazdani, M. (2010). An artificial immune algorithm for the flexible job-shop scheduling problem. *Future Gener Comput Syst, 26*(4), 533-541. https://doi.org/10.1016/j.future.2009.10.004

[19] Balaraju, G., Venkatesh, S., & Reddy, B. S. P. (2014). Multi-objective flexible job shop scheduling using hybrid differential evolution algorithm. *Int J Internet Manuf Serv, 3*(3), 226-243. https://doi.org/10.1504/IJIMS.2014.062436

[20] Sadeghi, A., Azar, A., Valmohammadi, Ch., & Alirezaei, A. (2020). Designing a product-service supply chain performance evaluation model in the home appliance industry. *Production and Operations Management, 10*(2), 83-123.

[21] Chiang, T. C. & Lin, H. J. (2013). A simple and effective evolutionary algorithm for multiobjective flexible job shop scheduling. *Int J Prod Econ, 141*(1), 87-98. https://doi.org/10.1016/j.ijpe.2012.03.034

[22] Cwiek, M. & Nalepa, J. (2014). A fast genetic algorithm for the flexible job shop scheduling problem. *Proceedings of the 2014 conference companion on Genetic and evolutionary computation companion, ACM*. 1449-1450. https://doi.org/10.1145/2598394.2602280

[23] Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput, 6*(2), 182-197. https://doi.org/10.1109/4235.996017

[24] Defersha, F. M. & Chen, M. (2010). A parallel genetic algorithm for a flexible job-shop scheduling problem with sequence dependent setups. *Int J Adv Manuf Technol, 49*(1-4), 263-279. https://doi.org/10.1007/s00170-009-2388-x

[25] Della Croce, F., Ghirardi, M., & Tadei, R. (2002). An improved branch-and-bound algorithm for the two machine total completion time flow shop problem. *Eur J Oper Res, 139*(2), 293-301. https://doi.org/10.1016/S0377-2217(01)00374-5

[26] Gao, K. Z., Suganthan, P. N., Pan, Q. K., et al. (2014). Pareto-based grouping discrete harmony search algorithm for multi-objective flexible job shop scheduling. *Inf Sci* (1), 76-90. https://doi.org/10.1016/j.ins.2014.07.039

[27] Geiger, C. D., Uzsoy, R., & Aytuğ, H. (2006). Rapid modeling and discovery of priority dispatching rules: An autonomous learning approach. *J Sched, 9*(1), 7-34. https://doi.org/10.1007/s10951-006-5591-8

[28] Gen, M., Tsujimura, Y., & Kubota, E. (1994). Solving job-shop scheduling problems by genetic algorithm. *IEEE International Conference on Systems, Man, and Cybernetics, Humans, Information and Technology, 2*, 77-82. https://doi.org/10.1109/ICSMC.1994.400072

[29] Heger, J., Branke, J., Hildebrandt, T. et al. (2016). Dynamic adjustment of dispatching rule parameters in flow shops with sequence-dependent set-up times. *Int J Prod Res*, 1-13. https://doi.org/10.1080/00207543.2016.1178406

[30] Jia, S. & Hu, Z. H. (2014). Path-relinking Tabu search for the multi-objective flexible job shop scheduling problem. *Comput Oper Res, 47*, 11-26. https://doi.org/10.1016/j.cor.2014.01.010

[31] Sadeghi, A., Kalanaki, A., Noktehdan, A., Samghabadi, A. S., & Barzinpour, F. (2011). Using Bees Algorithm to Solve the Resource Constrained Project Scheduling Problem in PSPLIB. *Theoretical and Mathematical Foundations of Computer Science*, 486-494. https://doi.org/10.1007/978-3-642-24999-0_67

[32] Kacem, I., Hammadi, S., & Borne, P. (2002). Approach by localization and multiobjective evolutionary optimization for flexiblejob-shop scheduling problems. *IEEE Trans Syst Man Cybern Part C Appl Rev, 32*(1), 1-13. https://doi.org/10.1109/TSMCC.2002.1009117

[33] Li, Y. & Chen, Y. (2014). An improved genetic algorithm of Bi-level coding for flexible job shop scheduling problems. *J Netw, 9*(7), 1783-1789. https://doi.org/10.4304/jnw.9.7.1783-1789
[34] Li, H. & Zhang, Q. (2009). Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II. *IEEE Trans Evol Comput, 13*(2), 284-302. https://doi.org/10.1109/TEVC.2008.925798
[35] Li, J. Q., Pan, Q. K., & Liang, Y. C. (2010). An effective hybrid tabu search algorithm for multi-objective flexible job-shop scheduling problems. *Comput Indust Eng, 59*(4), 647-662. https://doi.org/10.1016/j.cie.2010.07.014
[36] Li, J. Q., Pan, Q. K., Suganthan, P. N. et al (2011). A hybrid tabu search algorithm with an efficient neighborhood structure for the flexible job shop scheduling problem. *Int J Adv Manuf Technol, 52*(5-8), 683-697. https://doi.org/10.1007/s00170-010-2743-y
[37] Moghadam, A. M., Wong, K. Y., & Piroozfard, H. (2014). An efficient genetic algorithm for flexible job-shop scheduling problem. *IEEE International Conference on Industrial Engineering and Engineering Management*, 1409-1413. https://doi.org/10.1109/IEEM.2014.7058870
[38] Mouelhi-Chibani, W. & Pierreval, H. (2010). Training a neural network to select dispatching rules in real time. *Comput Indus Eng, 58*(2), 249-256. https://doi.org/10.1016/j.cie.2009.03.008
[39] Pezzella, F., Morganti, G., & Ciaschetti, G. (2008). A genetic algorithm for the flexible job-shop scheduling problem. *Comput Oper Res, 35*(10), 3202-3212. https://doi.org/10.1016/j.cor.2007.02.014

Authors' contacts:

Neda Karim Ahangar, MSc Student
Department of Industrial Engineering, College of Engineering,
Karaj Branch, Islamic Azad University, Karaj, Iran
negar.karim.ahangar@gmail.com

Majid Khalili, Assistant Professor, Dr.
(Corresponding author)
Department of Industrial Engineering, College of Engineering,
Karaj Branch, Islamic Azad University, Karaj, Iran

Hamed Tayebi, Assistant Professor, Dr.
Department of Industrial Engineering, College of Engineering,
Karaj Branch, Islamic Azad University, Karaj, Iran

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20200922163131

Original scientific paper

The Effect of Corporate Social Responsibility on Brand Performance with the Mediating Role of Corporate Reputation, Resource Commitment and Green Creativity

Ailar Ebrahimi Hesari, Elmira Shadiardehaei, Behzad Shahrabi*

Abstract: Following the strategic decisions of business managers in the community, issues have been raised that lead to changes in the community. One of these issues, which is becoming increasingly important, is consideration of corporate social responsibility. Therefore, the focus of this study was the effect of corporate social responsibility on brand performance with the mediating role of corporate reputation, resource commitment and green creativity. The research method was survey. The statistical population was employees of private banks, 507 of whom participated in the study. Structural equation modelling with SMARTPLS software was used to analyse the data. The results showed the effect of corporate social responsibility is positive and significant on resource commitment, green creativity, corporate reputation and brand performance. The effect of resource commitment, green creativity and corporate reputation is positive and significant on brand performance. Brand performance will be improved if the company supports employee higher education, encourages employees to develop their skills and abilities, implements flexible policies to provide work-life balance for employees, prioritizes employee needs and demands, has plans to reduce the negative effects of the company on the environment, and participates in activities aimed at protecting and improving the quality of the environment.

Keywords: brand performance; corporate reputation; corporate social responsibility; green creativity; resource commitment

1 INTRODUCTION

Following the strategic decisions of business managers in the community, issues have been raised that lead to changes in the community. One of these issues, which is becoming increasingly important, is consideration of corporate social responsibility in customer orientation literature [1]. In order to achieve their goals, businesses need a mechanism that can balance the resources of the company and stakeholders. As a result, the concept of corporate social responsibility is introduced [2, 3]. Based on social responsibility, companies take actions that improve the situation of society and refrain from actions that worsen the status of society, and ultimately have social accountability, attention to capacity to act and act on demands and expectations of society [4]. Corporate social responsibility is generally defined as activities that go beyond the mere economic interests of the company and includes actions to provide social benefits [5-7]. Carroll (1979), who has provided one of the most comprehensive definitions in this regard, considers corporate social responsibility as economic, legal, ethical and cautious expectations that society has of the company during a particular period of time [8]. Homburg et al. [9] claim that corporate social responsibility is a voluntary consideration of corporate concerns in relation to business and non-business processes. In the history of developed economies, corporate social responsibility has been mentioned as a key factor in achieving economic goals and wealth creation [8, 10]. On the other hand, the prerequisite for success of the organization is its proper performance. Indicators such as profitability, survival, growth, achievement of goals and competitive capabilities can be considered as results of good performance. The performance of any company is the result of purposeful activities that are performed in order to earn economic profit. Well-performing companies usually know the channels of profit well and invest in those channels. Although the main goal of organizations is to increase efficiency and profit, in the age of information and globalization, they must respond appropriately to social expectations and ethical norms to achieve success, and best combine such expectations with economic goals of the organization to enable the achievement of higher goals. Studies that have examined the relationship between corporate social responsibility and corporate performance have shown that there is a positive and significant relationship between corporate social responsibility and corporate performance [11-16]. Therefore, it is assumed:

Hypothesis 1: Corporate social responsibility has a positive and significant effect on brand performance.

1.1 Corporate Social Responsibility, Resource Commitment and Performance

Resource commitment refers to the organization's committed effort for business strategies and has always been a key factor in the planning process for success [17]. Resource commitment can also be defined as an organization that uses resources to organize programs so that they result in better performance. When a company integrates environmental policies with its strategies and commits the resources needed to implement and execute them, it will see better performance compared to companies with fewer resources allocated for their green strategies [18]. Resource commitment has a strong impact on success of producers. To ensure the maintenance and continuity of regular export operations, it is essential to understand the attitudes of foreign buyers and conduct advanced marketing activities; in order to build such capacity, resource commitment is very important [19]. When a company is committed to resources, this tendency has a positive effect on the company's performance because it helps to develop the capabilities needed to operate efficiently. Lack of commitment to resources is one of the most important obstacles to performance. Therefore, an increase in resource commitment is needed to increase performance development. Companies

with more resource commitment perform better. Numerous studies have emphasized the importance of green commitment as a key factor in building and maintaining long-term relationships between business partners. Studies that have examined the relationship between corporate social responsibility and commitment have shown that there is a positive and significant relationship between corporate social responsibility and commitment [20-24]. In addition, research has shown that there is a positive and significant relationship between resource commitment and performance [25-27]. Therefore, it is assumed:

Hypothesis 2: Corporate social responsibility has a positive and significant effect on resource commitment.

Hypothesis 3: Resource commitment has a positive and significant effect on brand performance.

1.2 Corporate Social Responsibility, Green Creativity and Performance

Environmental crises such as climate pollution, resource scarcity, declining biodiversity and global warming are the result of high urban growth rates, high population growth, large-scale land development and adverse effects on society, the environment and the economy. One of the issues that can play an important role in greening in organizations is green creativity [28]. Torrance (1995) considers creativity as a process that includes sensitivity to problems, shortcomings, bottlenecks, and inconsistencies. Sensitivity arises after diagnosing a problem or problems, followed by a search for solutions to problems and assumptions for this purpose. The assumptions are then tested and modified and the final results are obtained. Green creativity can help the company take effective environmental action. Green creativity refers to creation of new ideas about green goods, green services, green processes and green practices that are innovative, new, and useful [29, 30]. With their creativity, companies can better respond to environmental changes and acquire new capabilities to achieve better performance. Creativity performance in the literature is one of the main factors in organizational performance that leads to organizational learning, renovation, improvement, learning from failures and adapting to a changing competitive environment [31]. Green creativity can be a new process to produce a product or service, a new management method or policy, or a new solution for business that reduces environmental risks, pollution and the negative effects of energy consumption (such as water, electricity, gas, oil, etc.) [32]. In addition, green creativity is considered as a new perspective, idea, product, service or process that seeks to reduce the negative environmental effects [33, 34].

Studies that have examined the relationship between corporate social responsibility and commitment have shown that there is a positive and significant relationship between corporate social responsibility and commitment [35-39]. In addition, research has shown that there is a positive and significant relationship between creativity and performance [40-44]. Therefore, it is assumed:

Hypothesis 4: Corporate social responsibility has a positive and significant effect on green creativity.

Hypothesis 5: Green creativity has a positive and significant effect on brand performance.

1.3 Corporate Social Responsibility, Corporate Reputation and Performance

Corporate reputation is the overall impact that reflects the perception of stakeholders about the organization's activities and examines the status of the product or service provided [45]. This reputation may be derived from various complementary pathways such as performance in use, advertising, and positive communication [46]. A good reputation prevents stakeholders from understanding negative information. Firms with higher reputation experiences are less likely to be exposed to adverse market reactions or disappointing earnings, and their reputation is less likely to be troubled as they shrink and downsize. In addition, good reputation attracts employees and customers [45]. Organizational researchers examine reputation as a social identity and portray it as an important and intangible resource that may effectively contribute to the organization's performance or even its survival [47]. Marketing authors examine reputation as brand equity [48] and relate it to corporate credibility [49]. Corporate reputation is the result of past activities of the company and may be seen as a mirror of history that can provide information about the company and quality of its products compared to competitors to its target groups [50].

Studies that have examined the relationship between corporate social responsibility and corporate reputation have shown that there is a positive and significant relationship between corporate social responsibility and corporate reputation [51-54]. In addition, research has shown that there is a positive and significant relationship between corporate reputation and performance [50, 54-56]. Therefore, it is assumed:

Hypothesis 6: Corporate social responsibility has a positive and significant effect on corporate reputation.

Hypothesis 7: Corporate reputation has a positive and significant effect on brand performance.

In short, it was once thought that companies were solely accountable to shareholders and employees, or that they had to offer the best product at a low price and high quality, regardless of the consequences. Some developments and changes in the world such as population growth and resource scarcity, environmental pollution, etc. have led to new approaches in the organization and its management; corporate social responsibility is one of these approaches. However, corporate social responsibility has become the dominant paradigm in corporate governance in the last decade, and large and reputable global companies have made social responsibility as part of their strategy. This concept is now strongly pursued in developed countries by stakeholders such as governments, civil societies, international organizations and scientific centers. That is why in the world of competition, corporate social responsibility has become a dominant paradigm. Therefore, the main objective of this study is to investigate the effect of corporate social responsibility on brand performance with the mediating role

of corporate reputation, resource commitment and green creativity. According to the theoretical literature and the theoretical framework obtained from the background, the conceptual model of the study is drawn in Fig. 1.

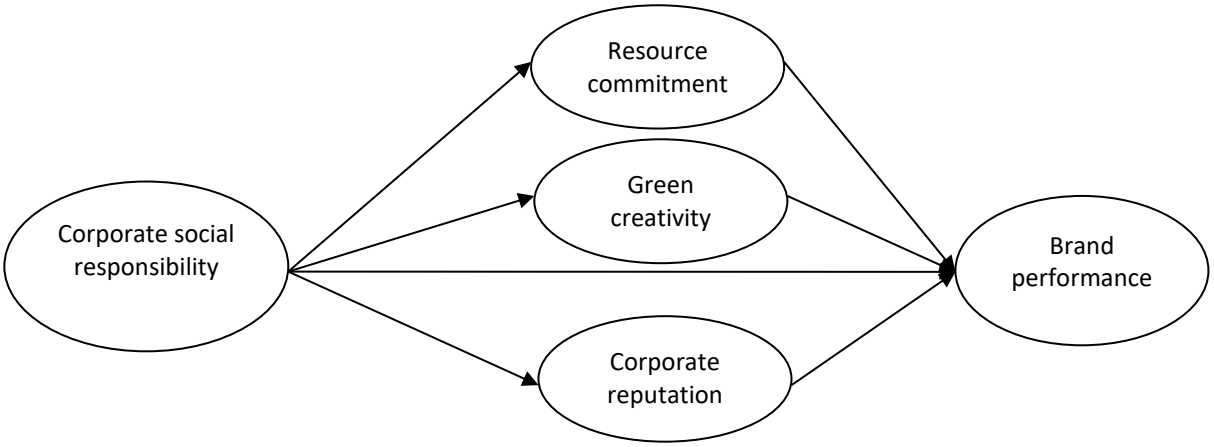


Figure 1 Conceptual model

2 MATERIALS AND METHODS

The research method is descriptive-survey and the research design is correlation using structural equations with partial least squares; because this study examines the relationships between variables through a causal model.

2.1 Statistical Population and Sample

The statistical population was employees of private banks in Iran. For this purpose, 700 questionnaires were sent to employees of private banks in Iran, of which 507 (72.43 %) questionnaires were returned.

2.2 Data Collection Instrument

Corporate social responsibility (CSR): the questionnaire developed by Kim et al. [57] was used to measure corporate social responsibility. This questionnaire has 22 items that measure economic dimension (6 items), legal dimension (6 items), moral dimension (6 items) and humanitarian dimension (4 items). The items were weighed on a 5-point Likert scale from strongly disagree (1) to strongly agree (5).

Resource commitment (RC): the questionnaire developed by Li [58] was used to measure resource commitment. This questionnaire has 4 items. The items were weighed on a 5-point Likert scale from strongly disagree (1) to strongly agree (5).

Green creativity (GC): the questionnaire developed by Chen and Chang et al. [29] was used to measure green commitment. This questionnaire has 6 items. The items were weighed on a 5-point Likert scale from strongly disagree (1) to strongly agree (5).

Corporate Reputation (CR): the questionnaire developed by Park [53] and González-Rodríguez et al. [54] was used to measure corporate reputation. This questionnaire has 4 items. The items were weighed on a 5-point Likert scale from strongly disagree (1) to strongly agree (5).

Brand performance (BP): the questionnaire developed by Lei and Chaoyan [27] was used to measure brand performance. This questionnaire has 3 items. The items were weighed on a 5-point Likert scale from strongly disagree (1) to strongly agree (5).

2.3 Data Analysis

To analyse the data, coefficient of correlation and structural equation modelling with partial least squares (PLS) were used. SPSS and SMARTPLS3 software were used to analyse the data. PLS model is tested and interpreted in two steps: 1) measurement model and 2) structural model. The measurement model or confirmatory factor analysis, to answer questions about validity and reliability of the measurement, determines how latent variables or sub-constructs are measured in the form of more observed variables. The structural model also shows the relationships between constructs (latent variables) and their explanatory power.

3 RESULTS

3.1 Validity and Reliability of Instruments (Measurement Model Testing)

The measurement model test includes checking the reliability (internal consistency) and validity (discriminant validity) of constructs and instruments [59]. To evaluate the reliability of constructs, Fornell and Larcker [60] propose three criteria, 1) reliability of each item, 2) composite reliability of each construct, and 3) average variance extracted. For reliability of each item, the factor load of 0.6 and more of each item in the confirmatory factor analysis indicates the suitability of each item of that construct. Moreover, the factor load of items should be significant at least at the level of 0.01. Bootstrab test (with 700 subsamples) was used to calculate t-value to determine the significance of factor loads. Dillon-Goldstein coefficient (ρ_c) was used to evaluate the composite reliability of each

construct. Acceptable values of ρ_c must be 0.7 or higher. The third criterion of reliability is average variance extracted [60]. Fornell and Larcker [60] recommend AVE values of 0.50 and above, which means that the construct explains about 50 % or more of the variance of its markers. Tab. 1 presents the factor loads, ρ_c, and AVE of the variables. The values of these tables indicate sufficient and appropriate reliability of the constructs.

Table 1 Factor Loads, Composite Reliability and Average Variance Explained of Variables

Variable	Item	Factor	ρ_c	*CR*	*AVE*
Economic social responsibility	1	0.691	0.805	0.860	0.507
	2	0.657			
	3	0.765			
	4	0.722			
	5	0.671			
	6	0.758			
Legal social responsibility	1	0.881	0.887	0.915	0.642
	2	0.782			
	3	0.698			
	4	0.763			
	5	0.829			
	6	0.843			
Moral social responsibility	1	0.653	0.826	0.874	0.537
	2	0.675			
	3	0.779			
	4	0.786			
	5	0.795			
	6	0.698			
Humanitarian social responsibility	1	0.824	0.827	0.885	0.658
	2	0.835			
	3	0.821			
	4	0.764			
Resource commitment	1	0.878	0.907	0.934	0.782
	2	0.918			
	3	0.910			
	4	0.830			
Green creativity	1	0.839	0.901	0.923	0.668
	2	0.861			
	3	0.828			
	4	0.811			
	5	0.833			
	6	0.727			
Corporate reputation	1	0.664	0.798	0.867	0.623
	2	0.803			
	3	0.859			
	4	0.817			
Brand performance	1	0.815	0.793	0.879	0.707
	2	0.857			
	3	0.849			

The Fornell-Larcker (1981) index was used to evaluate validity or discriminant validity of the constructs. This index indicates that AVE of a construct should be greater than correlation of that construct with other constructs. This indicates that the correlation of that construct with its markers is greater than its correlation with other constructs. According to Tab. 2, all dimensions have the highest factor load on their own construct and minimum distance between the factor loads related to their own construct is more than 0.1, suggesting that the constructs have good validity. Tab. 2 reports the results of correlation and Fornell-Larcker index, ie square root of AVE.

Table 2 Matrix of Correlation and Square Root of AVE of Variables

Variable	Mean	SD	CSR	RC	GC	CR	BP
CSR	3.15	0.72	0.86				
RC	2.70	1.05	0.53**	0.88			
GC	3.03	0.92	0.54**	0.63**	0.82		
CR	3.04	0.88	0.55**	0.59**	0.52**	0.79	
BP	2.75	0.91	0.53**	0.57**	0.59**	0.51**	0.84
Note: the numbers on the matrix diagonal are square root of AVE							

According to Tab. 2, the root square of AVE of all variables is greater than their correlation with other variables. Therefore, the second criterion of discriminant validity is established. In addition, the numbers below the matrix diagonal were reported to examine the relationship between the variables. Obviously, the coefficient of correlation between all variables is positive and significant.

3.2 Structural Model Testing

To predict brand performance, the proposed conceptual model was examined through structural equation modeling method; according to the hypotheses, partial least squares method was used to estimate the model. Bootstrap method (with 700 sub-samples) was used to calculate t-values to determine the significance of path coefficients. Figure 2 shows the relationship between variables of the tested model. According to this figure, the effect of corporate social responsibility is positive and significant on resource commitment, green creativity, corporate reputation and brand performance. The effect of resource commitment, green creativity and corporate reputation is positive and significant on brand performance. Numbers in the circles are variance explained of variables.

Table 3 Path coefficients, T-values and variance explained

Variable	β	*t*-value	Variance ex.
On brand performance vie:			
Corporate reputation	0.20**	3.093	0.57
Green creativity	0.26**	4.152	
Resource commitment	0.25**	5.064	
Corporate social responsibility	0.14**	3.448	
On corporate reputation vie:			
Corporate social responsibility	0.56**	15.343	0.32
On green creativity vie:			
Corporate social responsibility	0.54**	13.617	0.29
On resource commitment vie:			
Corporate social responsibility	0.54**	13.529	0.29
Mediating role of corporate reputation in relationship between:			
Corporate social responsibility and brand performance	0.11**	3.032	-
Mediating role of green creativity in relationship between:			
Corporate social responsibility and brand performance	0.14**	3.971	-
Mediating role of resource commitment in relationship between:			
Corporate social responsibility and brand performance	0.13**	4.743	-

*p<0.05; **p<0.01

Tab. 3 reports the estimate of path coefficients and variance explained of the variables. As shown in Tab. 3, 57 % of variance in brand performance, 32 % of variance in corporate reputation, 29 % of green creativity and 29 % of

resource commitment are explained by variables of the model.

The overall fit index in PLS is GOF index and it can be used to check validity or quality of the PLS model in general. This index examines the overall predictive ability of the model and whether the tested model is successful in predicting endogenous latent variables. In the present study, *GOF* = 0.59 indicates the proper fit of the tested model. Values above 0.36 indicate good and acceptable quality of the model.

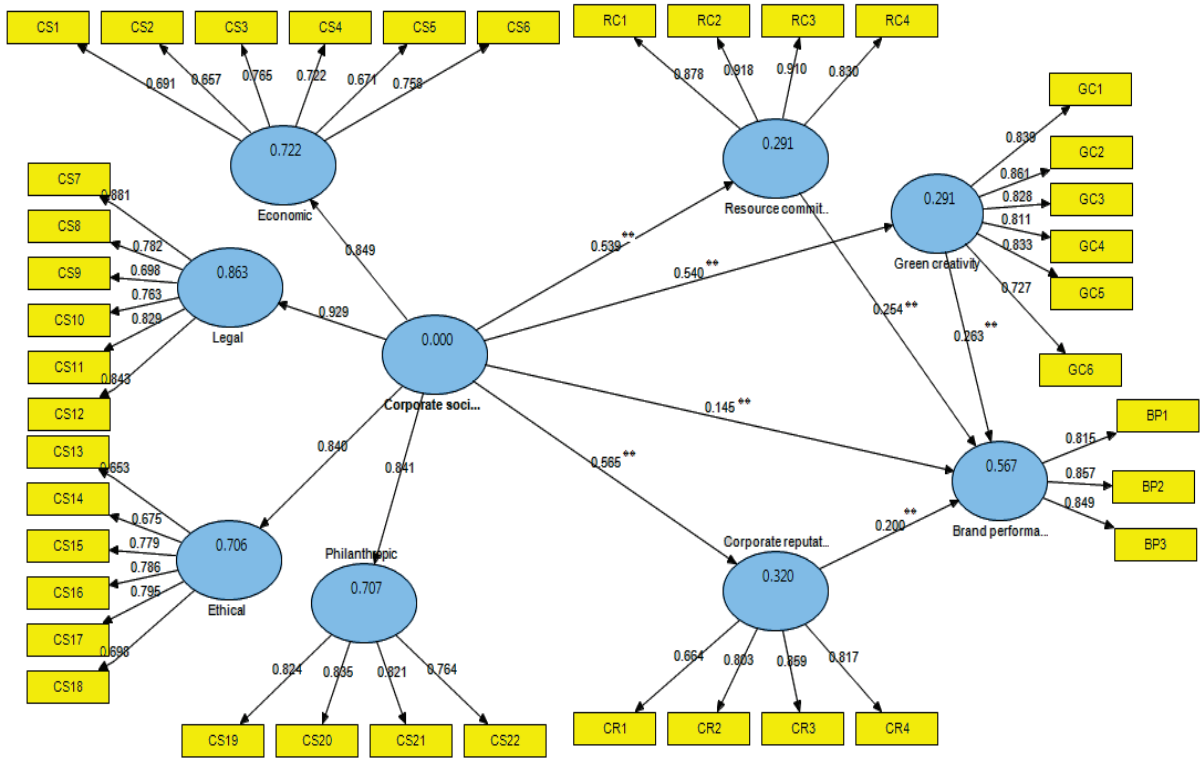


Figure 2 The tested model for predicting brand performance

4 DISCUSSION AND CONCLUSION

The objective of this study was to investigate the effect of corporate social responsibility on brand performance with the mediating role of corporate reputation, resource commitment and green creativity using the structural equation modeling. The results showed that the proposed model fits relatively well with data of this study and can explain 57 % of the variance in brand performance, 32 % of variance in corporate reputation, 29 % of green creativity and 29 % of resource commitment.

The results showed that corporate social responsibility has a positive and significant effect on brand performance. This finding is consistent with Selcuk and Kiymaz [11], Rahman and Yu [12], Lin et al. [13], Oware and Mallikarjunappa [14], Kong et al. [15] and Naseem et al. [16]. To explain this finding, it can be claimed that brand performance will be improved if the company supports employee higher education, encourages employees to develop their skills and abilities, implements flexible policies to provide work-life balance for employees, prioritizes employee needs and demands, provides accurate and comprehensive information to customers, cares about customer rights and satisfaction beyond legal requirements, has plans to reduce the negative effects of the company on the environment, and participates in activities aimed at protecting and improving the quality of the environment.

The results showed that corporate social responsibility has a positive and significant effect on resource commitment, green creativity and corporate reputation. This finding is consistent with Lee and Yoon [22], Chen et al. [23], Chaudhary and Akhouri [37], Ahmad et al. [35], Ko and Choi [39], Park [53], González-Rodríguez et al. [54] and Javed et al. [47]. To explain this finding, it can be claimed that resource commitment to social responsibility activities will increase, employees will have more green creativity and will provide more green ideas and thus corporate reputation will increase if the bank has methods for responding to customer complaints, uses customer satisfaction as one of the indicators of business performance, has been successful in maximizing profits, tries to reduce operating costs, has close monitoring of employee productivity, has long-term strategies, is aware of environmental laws, provides all services with legal standards, considers a duty to fulfill the obligations stated in the contracts and agreements, tries to enforce the rules, tries to comply with all rules related to employment and employee benefits, follows professional standards, monitors the negative effect of activities on the community, is recognized as trustworthy, considers justice to colleagues and clients as one of the main parts of the employee evaluation process, has reliable ways for

employees to report any behavior at work, improves its business-related humanitarian activities, participates in charitable activities, contributes to improvement of society and strives to fulfill its social responsibilities.

The results showed that resource commitment has a positive and significant effect on brand performance. This finding is consistent with Zhang and Walton [25], Wu [26] and Lei and Chaoyan [27]. This finding suggests that brand performance will be improved if the company's managers allocate financial resources in a timely manner, consider the necessary human resources for economic activities, the company's managers have good management skills and the company's managers are committed to their financial obligations.

The results showed that green creativity has a positive and significant effect on brand performance. This finding is consistent with Wu [26], Chen et al. [23], Boso et al. [41], Aeknarajindawat and Jermsittiparsert [40] and Ferreira et al. [43]. To explain this finding, it can be claimed that brand performance will be improved if company employees propose new ways to achieve environmental goals, suggest new green ideas to improve environmental performance, promote new green ideas to others, develop good plans for implementing green ideas, and come up with creative solutions to environmental issues.

The results showed that corporate reputation has a positive and significant effect on brand performance. This finding is consistent with Gonzalez Sanchez and Morales de Vega [56], Kowalczyk and Kucharska [50], González-Rodríguez et al. [54] and Javed et al. [47]. To explain this finding, it can be claimed that corporate performance will be improved if the company is known and reputable, the company has chosen a good name for itself, the company has a good reputation and the company is one of the most important in its industry. In addition, a company with a well-known brand has more power in the competition arena and makes the consumer loyal to the product, and as a result, the corporate performance improves.

Overall, the results showed that corporate reputation, resource commitment and green creativity mediate the effect of corporate social responsibility on brand performance. Therefore, corporate social responsibility leads to increased corporate reputation, resource commitment and green creativity and thus improves brand performance.

5 IMPLICATIONS

It is recommended that bank employees be trained in the skills needed to perform and achieve social responsibilities and to increase environmental awareness, provide regular training on social issues to employees, provide opportunities for employees to engage in social issues, and provide resources to educate employees on social issues. Banks are known as a very professional and successful company by providing quality services to customers and thus gain reputation for the company. It is recommended to allocate financial resources to social issues in a timely manner, consider the necessary human resources for social activities, company managers have good management skills, and company managers adhere to their financial obligations, and the company has sufficient financial resources for investing in innovative environmental methods. Companies should support new green ideas by employees, develop appropriate plans for implementing green ideas, and encourage employees to come up with creative solutions to environmental and social problems.

6 LIMITATIONS

In this study, only a sample of employees of private banks was examined; therefore, generalizations are limited. Moreover, the findings are based on self-report data. It is suggested that qualitative and mixed research methods be used in future studies.

7 REFERENCES

[1] Carroll, A. B. & Brown, J. A. (2018). Corporate social responsibility: A review of current concepts, research, and issues. In *Corporate social responsibility*. Emerald Publishing Limited. https://doi.org/10.1108/S2514-175920180000002002

[2] Bieri, F. (2012). Corporate social responsibility. *The Wiley-Blackwell Encyclopedia of Globalization*.

[3] Crane, A., Matten, D., & Spence, L. (Eds.). (2019). *Corporate social responsibility: Readings and cases in a global context*. Routledge. https://doi.org/10.4324/9780429294273

[4] Banerjee, S. & Wathieu, L. (2017). Corporate social responsibility and product quality: Complements or substitutes? *International Journal of Research in Marketing*, *34*(3), 734-745. https://doi.org/10.1016/j.ijresmar.2017.06.006

[5] Freeman, I. & Hasnaoui, A. (2011). The meaning of corporate social responsibility: The vision of four nations. *Journal of business Ethics*, *100*(3), 419-443. https://doi.org/10.1007/s10551-010-0688-6

[6] Bai, X. & Chang, J. (2015). Corporate social responsibility and firm performance: The mediating role of marketing competence and the moderating role of market environment. *Asia Pacific Journal of Management*, *32*(2), 505-530. https://doi.org/10.1007/s10490-015-9409-0

[7] Ferrell, O. C., Harrison, D. E., Ferrell, L., & Hair, J. F. (2019). Business ethics, corporate social responsibility, and brand attitudes: An exploratory study. *Journal of Business Research*, *95*, 491-501. https://doi.org/10.1016/j.jbusres.2018.07.039

[8] Saeidi, S. P., Sofian, S., Saeidi, P., Saeidi, S. P., & Saaeidi, S. A. (2015). How does corporate social responsibility contribute to firm financial performance? The mediating role of competitive advantage, reputation, and customer satisfaction. *Journal of business research*, *68*(2), 341-350. https://doi.org/10.1016/j.jbusres.2014.06.024

[9] Homburg, C., Stierl, M., & Bornemann, T. (2013). Corporate social responsibility in business-to-business markets: How organizational customers account for supplier corporate social responsibility engagement. *Journal of Marketing*, *77*(6), 54-72. https://doi.org/10.1509/jm.12.0089

[10] Garriga, E. & Melé, D. (2004). Corporate social responsibility theories: Mapping the territory. *Journal of business ethics*, *53*(1-2), 51-71. https://doi.org/10.1023/B:BUSI.0000039399.90587.34

[11] Selcuk, E. A. & Kiymaz, H. (2017). Corporate social responsibility and firm performance: Evidence from an emerging market. *Accounting and Finance Research*, *6*(4), 42. https://doi.org/10.5430/afr.v6n4p42

[12] Rahman, J. M. & Yu, F. (2019). The relationship between corporate social responsibility and firm performance in China. *Risk Governance & Control: Financial Markets & Institutions*. https://doi.org/10.22495/rgcv9i4p4

[13] Lin, W. L., Ho, J. A., Ng, S. I., & Lee, C. (2019). Does corporate social responsibility lead to improved firm performance? The hidden role of financial slack. *Social Responsibility Journal*. https://doi.org/10.1108/SRJ-10-2018-0259

[14] Oware, K. M. & Mallikarjunappa, T. (2019). Corporate social responsibility investment, third-party assurance and firm performance in India. *South Asian Journal of Business Studies*. https://doi.org/10.1108/SAJBS-08-2018-0091

[15] Kong, Y., Antwi-Adjei, A., & Bawuah, J. (2020). A systematic review of the business case for corporate social responsibility and firm performance. *Corporate Social Responsibility and Environmental Management*, *27*(2), 444-454. https://doi.org/10.1002/csr.1838

[16] Naseem, T., Shahzad, F., Asim, G. A., Rehman, I. U., & Nawaz, F. (2020). Corporate social responsibility engagement and firm performance in Asia Pacific: The role of enterprise risk management. *Corporate Social Responsibility and Environmental Management*, *27*(2), 501-513. https://doi.org/10.1002/csr.1815

[17] Lai, F., Li, D., Wang, Q., & Zhao, X. (2008). The information technology capability of third-party logistics providers: a resource-based view and empirical evidence from China. *Journal of supply chain management*, *44*(3), 22-38. https://doi.org/10.1111/j.1745-493X.2008.00064.x

[18] Mittal, S. & Dhar, R. L. (2015). Transformational leadership and employee creativity. *Management Decision*. https://doi.org/10.1108/MD-07-2014-0464

[19] Chugan, P. K. & Singh, S. (2014). Taxonomy for firm-level determinants of export performance. *Universal Journal of Industrial and Business Management*, *2*(1), 6-12.

[20] Mory, L., Wirtz, B. W., & Göttel, V. (2016). Factors of internal corporate social responsibility and the effect on organizational commitment. *The International Journal of Human Resource Management*, *27*(13), 1393-1425. https://doi.org/10.1080/09585192.2015.1072103

[21] Prutina, Ž. (2016). The effect of corporate social responsibility on organizational commitment. *Management: Journal of Contemporary Management, issues 21*(Special issue), 227-248.

[22] Lee, S. & Yoon, J. (2018). Does the authenticity of corporate social responsibility affect employee commitment? *Social Behavior and Personality: an international journal*, *46*(4), 617-632. https://doi.org/10.2224/sbp.6475

[23] Chen, H. L., Hu, Y. C., Lee, M. Y., & Yen, G. F. (2020). Importance of Employee Care in Corporate Social Responsibility: An AHP-Based Study from the Perspective of Corporate Commitment. *Sustainability*, *12*(15), 5885. https://doi.org/10.3390/su12155885

[24] Afsar, B. & Umrani, W. A. (2020). Corporate social responsibility and pro-environmental behavior at workplace: The role of moral reflectiveness, coworker advocacy, and environmental commitment. *Corporate Social Responsibility and Environmental Management*, *27*(1), 109-125. https://doi.org/10.1002/csr.1777

[25] Zhang, J. A. & Walton, S. (2017). Eco-innovation and business performance: the moderating effects of environmental orientation and resource commitment in green-oriented SME s. *R&D Management*, *47*(5), E26-E39. https://doi.org/10.1111/radm.12241

[26] Wu, G. C. (2017). Environmental innovation approaches and business performance: effects of environmental regulations and resource commitment. *Innovation*, *19*(4), 407-427. https://doi.org/10.1080/14479338.2017.1358102

[27] Lei, H. & Chaoyan, W. (2017). The Impact of B2B Brand Orientation on Brand Performance: The Role of Suppliers' Resource Commitment. *Management Review*, (9), 16.

[28] Zameer, H., Wang, Y., & Yasmeen, H. (2020). Reinforcing green competitive advantage through green production, creativity and green brand image: implications for cleaner production in China. *Journal of Cleaner Production*, *247*, 119119. https://doi.org/10.1016/j.jclepro.2019.119119

[29] Chen, Y. S. & Chang, C. H. (2013). The determinants of green product development performance: Green dynamic capabilities, green transformational leadership, and green creativity. *Journal of business ethics*, *116*(1), 107-119. https://doi.org/10.1007/s10551-012-1452-x

[30] Song, W., & Yu, H. (2018). Green innovation strategy and green innovation: The roles of green creativity and green organizational identity. *Corporate Social Responsibility and Environmental Management*, *25*(2), 135-150. https://doi.org/10.1002/csr.1445

[31] Gunday, G., Ulusoy, G., Kilic, K., & Alpkan, L. (2011). Effects of innovation types on firm performance. *International Journal of production economics*, *133*(2), 662-676. https://doi.org/10.1016/j.ijpe.2011.05.014

[32] Van den Bergh, J. C., Truffer, B., & Kallis, G. (2011). Environmental innovation and societal transitions: Introduction and overview. *Environmental innovation and societal transitions*, *1*(1), 1-23. https://doi.org/10.1016/j.eist.2011.04.010

[33] Dangelico, R. M., Pujari, D., & Pontrandolfo, P. (2017). Green product innovation in manufacturing firms: A sustainability-oriented dynamic capability perspective. *Business strategy and the Environment*, *26*(4), 490-506. https://doi.org/10.1002/bse.1932

[34] Seman, N. A. A., Zakuan, N., Jusoh, A., Arif, M. S. M., & Saman, M. Z. M. (2012). The relationship of green supply chain management and green innovation concept. *Procedia-Social and Behavioral Sciences*, *57*, 453-457. https://doi.org/10.1016/j.sbspro.2012.09.1211

[35] Ahmad, I., Donia, M. B., & Shahzad, K. (2019). Impact of corporate social responsibility attributions on employees' creative performance: The mediating role of psychological safety. *Ethics & Behavior*, *29*(6), 490-509. https://doi.org/10.1080/10508422.2018.1501566

[36] Chaudhary, R. & Akhouri, A. (2018). Linking corporate social responsibility attributions and creativity: Modeling work engagement as a mediator. *Journal of cleaner production*, *190*, 809-821. https://doi.org/10.1016/j.jclepro.2018.04.187

[37] Chaudhary, R. & Akhouri, A. (2019). CSR perceptions and employee creativity: Examining serial mediation effects of meaningfulness and work engagement. *Social Responsibility Journal*. https://doi.org/10.1108/SRJ-01-2018-0018

[38] Hur, W. M., Moon, T. W., & Ko, S. H. (2018). How employees' perceptions of CSR increase employee creativity: Mediating mechanisms of compassion at work and intrinsic motivation. *Journal of Business Ethics*, *153*(3), 629-644. https://doi.org/10.1007/s10551-016-3321-5

[39] Ko, S. H. & Choi, Y. (2020). The Effect of CSR Perception on Creativity: The Mediating Effect of Compassion and the

Moderating Effect of CSR Authenticity. *Journal of Digital Convergence*, *18*(3), 59-70.

[40] Aeknarajindawat, N. & Jermsittiparsert, K. (2019). The mediating role of green creativity in the relationship between proactive green innovation, reactive green innovation and the performance of Green product development: A case of Thai sports manufacturing firms. https://doi.org/10.14198/jhse.2019.14.Proc5.45

[41] Boso, N., Donbesuur, F., Bendega, T., Annan, J., & Adeola, O. (2017). Does organizational creativity always drive market performance? *Psychology & Marketing*, *34*(11), 1004-1015. https://doi.org/10.1002/mar.21039

[42] Chen, Y. S., Chang, T. W., Lin, C. Y., Lai, P. Y., & Wang, K. H. (2016). The influence of proactive green innovation and reactive green innovation on green product development performance: The mediation role of green creativity. *Sustainability*, *8*(10), 966. https://doi.org/10.3390/su8100966

[43] Ferreira, J., Coelho, A., & Moutinho, L. (2020). Dynamic capabilities, creativity and innovation capability and their impact on competitive advantage and firm performance: The moderating role of entrepreneurial orientation. *Technovation*, *92*, 102061. https://doi.org/10.1016/j.technovation.2018.11.004

[44] Wu, C. (2016). *Entrepreneurial leadership, organisational creativity, and firm performance in Chinese SMEs context* (Doctoral dissertation, University of Nottingham).

[45] Lai, C. S., Chiu, C. J., Yang, C. F., & Pai, D. C. (2010). The effects of corporate social responsibility on brand performance: The mediating effect of industrial brand equity and corporate reputation. *Journal of business ethics*, *95*(3), 457-469. https://doi.org/10.1007/s10551-010-0433-1

[46] Palmeira, M. (2014). The interplay of products from the same product line: the role of brand reputation. *European Journal of Marketing*. https://doi.org/10.1108/EJM-03-2013-0159

[47] Javed, M., Rashid, M. A., Hussain, G., & Ali, H. Y. (2020). The effects of corporate social responsibility on corporate reputation and firm financial performance: Moderating role of responsible leadership. *Corporate Social Responsibility and Environmental Management*, *27*(3), 1395-1409. https://doi.org/10.1002/csr.1892

[48] Liu, M. & Lu, W. (2019). Corporate social responsibility, firm performance, and firm risk: the role of firm reputation. *Asia-Pacific Journal of Accounting & Economics*, 1-21. https://doi.org/10.1080/16081625.2019.1601022

[49] Gangi, F., Daniele, L. M., & Varrone, N. (2020). How do corporate environmental policy and corporate reputation affect risk-adjusted financial performance? *Business Strategy and the Environment*. https://doi.org/10.1002/bse.2482

[50] Kowalczyk, R. & Kucharska, W. (2020). Corporate social responsibility practices incomes and outcomes: Stakeholders' pressure, culture, employee commitment, corporate reputation, and brand performance. A Polish–German cross-country study. *Corporate Social Responsibility and Environmental Management*, *27*(2), 595-615. https://doi.org/10.1002/csr.1823

[51] Famiyeh, S., Kwarteng, A., & Dadzie, S. A. (2016). Corporate social responsibility and reputation: some empirical perspectives. *Journal of Global Responsibility*. https://doi.org/10.1108/JGR-04-2016-0009

[52] Lu, J., Ren, L., He, Y., Lin, W., & Streimikis, J. (2019). Linking corporate social responsibility with reputation and brand of the firm. *Amfiteatru Economic*, *21*(51), 442-460. https://doi.org/10.24818/EA/2019/51/422

[53] Park, E. (2019). Corporate social responsibility as a determinant of corporate reputation in the airline industry. *Journal of retailing and consumer services*, *47*, 215-221. https://doi.org/10.1016/j.jretconser.2018.11.013

[54] González-Rodríguez, M. R., Martín-Samper, R. C., Köseoglu, M. A., & Okumus, F. (2019). Hotels' corporate social responsibility practices, organizational culture, firm reputation, and performance. *Journal of Sustainable Tourism*, *27*(3), 398-419. https://doi.org/10.1080/09669582.2019.1585441

[55] Weng, P. S. & Chen, W. Y. (2017). Doing good or choosing well? Corporate reputation, CEO reputation, and corporate financial performance. *The North American Journal of Economics and Finance*, *39*, 223-240. https://doi.org/10.1016/j.najef.2016.10.008

[56] Gonzalez Sanchez, M. & Morales de Vega, M. E. (2018). Corporate reputation and firms' performance: Evidence from Spain. *Corporate Social Responsibility and Environmental Management*, *25*(6), 1231-1245. https://doi.org/10.1002/csr.1634

[57] Kim, K. H., Kim, M., & Qian, C. (2018). Effects of corporate social responsibility on corporate financial performance: A competitive-action perspective. *Journal of Management*, *44*(3), 1097-1118. https://doi.org/10.1177/0149206315602530

[58] Li, Y. (2014). Environmental innovation practices and performance: moderating effect of resource commitment. *Journal of Cleaner Production*, *66*, 450-458. https://doi.org/10.1016/j.jclepro.2013.11.044

[59] Mortezaei, A., Sangari, M. S., Nazari-Shirkouhi, S., & Razmi, J. (2018). The Impact of Business Intelligence (BI) Competence on Customer Relationship Management (CRM) Process: An Empirical Investigation of the Banking Industry. *Journal of Information Technology Management, 10*(1), 209-234.

[60] Fornell, C. & Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, 18(3), 382-388. https://doi.org/10.1177/002224378101800313

Authors' contacts:

Ailar Ebrahimi Hesari
Department of Management,
Aliabad Katoul Branch,
Islamic Azad University,
Aliabad Katoul, Iran
E-mail: ailar.ebrahimi@gmail.com

Elmira Shadiardehaei
Department of Management,
Aliabad Katoul Branch,
Islamic Azad University,
Aliabad Katoul, Iran
E-mail: elmira_shadi99@yahoo.com

Behzad Shahrabi
Department of Management,
Aliabad Katoul Branch,
Islamic Azad University,
Aliabad Katoul, Iran
E-mail: b.shahrabi@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210205091148

Preliminary communication

Sound Prohibited Zone for Smart Cities using IoT

Bhawana A. Ahire*, Sachin R. Sakhare

Abstract: One of the dominant causes of noise pollution is the disruptive honking while traveling on a road and is particularly problematic as well as resulting in irritating sound. This results in many effects that need to be addressed taking into consideration stress, depression, mental and physical illnesses. The proposed system comprises of an embedded module that will be of great benefit to citizens of society adjacent to heavily populated roads, students studying in schools near crowded roads, patients admitted to roadside hospitals, etc. Noise Armor's alternative approach needs immense expenditure and personnel for both repair and surveillance as there is the need for expensive noise absorbers. By developing a real-time (smart) honking device that automatically allows vehicles on the road to reduce the volume of horn near conscious areas such as schools , hospitals, old-age homes, baby care centers and other organizations that are adversely damaged, our research aims to develop an effective solution for excessive honking.

Keywords: health hazards; IoT; smart honking; sound absorber; Sound Prohibited Zone; stress

1 INTRODUCTION

Now-a-days noise pollution is the major problem for society. Beeping the horn by cars has been a significant source of community noise pollution, resulting in health risks, frustration or impatience and injuries. As the number of vehicles are extremely increased day by day which results in traffic jams, dangerous and rough drivers and even accidents [1]. Mostly hospitals, schools old ages homes, baby care centers are situated near roadside. The patients, old people, babies and the students are having adverse effect on health due to honking of vehicles [15].

Over the decade, the urban population of India has been increased up to 31.8% during 2001-2011, as per the study [6]. Speedy urbanization has resulted into diverse public health difficulties, including environmental pollution [15, 16]. It is required to initiate the actions to reduce the source of causing pollution in order to fulfil the ever growing population as well as development. Hence, it is essential to minimize factors causing pollution are more factual than their expulsion. Noise is one of the factor causing pollution under the air (Prevention and Control of Pollution) Act, 1981 [7].

Due to increase in vehicle the amount of pollution, that is generated by these vehicles have been increased significantly. In some areas such as in central cities, near hospital, school, baby care centers etc. honking (operating a horn to generate sound) is prohibited by law or regulation [2, 16]. Sound which is measured in decibels can be controlled according to the governance but there is no device to measure and control in real time. To develop an embedded module to reduce a noise intensity for reduce noise pollution. If the surrounding vehicle sound goes beyond the 65 decibels, then it results into annoying effects. Either some sound absorbers should be used to reduce the sound, but it is not economically feasible [17]. So, in order to deal with this problem, some technology should be developed to decrease the noise generated by honking of horn.

An embedded system has been developed to provide the solution for the above problem. This device deals with inter-vehicular contact with proportional (apposite) spectrum, frequency energy requirement) and quantity of material implemented using radio frequency (RF) signals. It is also possible to overlook the accidents caused by the noisy music played within the vehicles in the proposed module [5].

True-time, i.e. the intelligent honking scheme is intended to build a disincentive measure for repetitive honking. A large amount of noise pollution occurs at the time of traffic jam because of the beeping the horns which leads to irritation to the people living in the neighbourhood of the sensitive areas such as (schools, old age home, residential areas as well as hospitals, etc.) [6]. This problem can be easily solved by the currently available device as no horns in the community can be heard. Instead of in the atmosphere inside the receiver car, the source car sends RF signals that produce a beep sound.

In certain instances, the driver is unable to hear the horn, which may lead to an accident, due to the large amount of music [18]. The car's stereo system is shut down by accommodating the stereo system of the car with the horn, which results in reducing the number of incidents. In this system, the stereo system of the car is combined with the horn since the stereo system is automatically paused until the beep of the horn is detected, resulting in road accident avoidance [4]. Alongside the beep, a visual cue emerges that allows the concerned people accompanied by hearing loss to drive safely.

As per emergency, there is also facility of switching to traditional horn to deal with such circumstances such as sudden emergence of humans or animals using dual purpose switch. In the near future, there is a need to improve the existing system with additional design for the automation [14].

Our aim is to find a suitable solution to the problems described, which also leads to a user-friendly culture. In order to reduce noise pollution, IoT framework is used. By using the proposed system, health issues raised for senior citizens due to noise pollution is minimized.

In today’s world, automating horn control of the vehicle is one of most essential demand. Our objective is to control the sound level of horn of vehicles based on location as well as reduce noise pollution using IoT framework. The proposed system consists of an embedded module to reduce noise

intensity of horns automatically when the vehicle enters in the targeted areas.

The performance of this system results in an embedded module that benefits sensitive areas, i.e. residents of communities near busy roads, students studying in schools near busy roads, patients admitted to roadside hospitals.

In the following sections, the remaining paper is arranged: Section 2 focuses on similar work in terms of sound monitoring systems. The suggested approach used for the structure of an embedded module is discussed in Section 3, while Section 4 presents the results obtained, and Section 5 presents the conclusion.

2 RELATED WORK

Up to 55% of aggregate urban noise is caused by traffic produced by vehicles [1]. In [2], the authors researched the essence of horn blowing and examined it in all 107 vehicles. As a result of the research, two-vehicle and car / jeep drivers were found to be 30 percent and 26 percent respectively among all of these heavy vehicle drivers for highest honking, i.e. aggressive horns, as 37 percent of the total population of vehicles.

R. K. Mishra et al. evaluated and analysed traffic volume noise along has rapid transit system corridor [3]. Analysis of traffic volume noise is done with respect to the predicted and actual noise at selected corridors. In order to deal with the situation, the traffic noise pollution is controlled through the design of noise barrier along the road as well as suggested the people to use the public transport [3].

As per survey on the noise level at crowded places, the ambient air quality noise levels (AAQNL) at traffic signals is reported to be 5 dBA whereas for commercial zones found in the spectrum of above/below 80 dBA, at some places 75 dBA is also reported [4, 5]. The effects of noise pollution on human health is very much hazardous which leads to severe headache, migraine, etc. The horns are classified based on their intensity levels by modelling the honk statistics i.e. Jam, Medius, Free. But in order to improve the classification results, pre-honk information is stored and utilised.

As per the Environmental Expert Council, studies show a continuous movement towards an increased cardiovascular hazard if the daytime emission boundary goes beyond 65 Decibels. The dangerous misclassifications have invalidated significant recent research on the extraoral effects of workplace noise. Studies on the health effects of noise are one of the most significant preconditions for a correct exposure assessment in the future [6].

As stated in Health programme of WHO(World Health Organization, 1994) [7], the conflicting outcome of noise is called as a change in the structure as well as physiology, that results in deterioration of practical capability. Temporary exposure to noise results in physiological changes that are easily unexpected. However, noise penetration of sufficient intensity causes duration changes that may not be reversible so readily.

Noise pollution is not regarded as a cause of mental illness, but the enhancement of suppressed mental disorders [10, 11] is intended to intensify and increase. It impairs job performance at school and at work, and reduces motivation [8]. Adults as well as children are also adversely affected. The study suggests that most attention should be paid to the effects of noise induced by excessive vehicle honking. Noise has been postulated to act as the factors that affects the environment [9].

The alternative noise absorber solution needs tremendous resources and manpower for repair and monitoring.

Honk monitoring system is also proposed by Atmadip D. et al. [19] which keeps track of excessive honking produced by the people near traffic signals. The proposed system calculates the number of horns at a certain period of time. If the number of horns goes beyond the limit, the system generates an alarm. The advantage of the system is that it is low-cost. The system if enhanced can be used for analysis of heavy traffic [19].

Acute noise exposure causes the autonomic and hormonal processes, resulting in transient variations such as high blood pressure and heart rate, too. To deal with the noise pollution scenario, Z. Zhao et al. [13] developed an adequate honk controlling system which administers blowing of the horn, locates the position of the sound generating vehicle, and envisions as an image captured by the road mounted cameras. But, the system cannot handle multi-SSL problem as well as localization error is the barrier for the challenging environment scenarios. The vehicle number plate recognition can also be added to improve the visualization results.

3 PROPOSED METHODOLOGY

The proposed system aims to create an efficient solution for excessive beeping the horn by implementing a smart honking system. The proposed system makes it easier for vehicles on the road to minimize horn volume near vulnerable areas such as schools, hospitals, old-age homes, baby care centres, and other adversely affected workplaces. An Admin has accessibility to add various targeted locations like hospitals, old-age homes, baby care centre and other workplaces. The GPS helps to discover the various location data. It enables to find specific places and to find the best path from real-time traffic. The targeted locations such as hospitals, signals, old age homes, baby care centres are extracted from the database server. User that is Drivers has to register and then login to see the Sound Prohibited Areas Zone List.

The proposed system consists of two modules, Software Module (Android App) and other is Hardware Module. The Software module consists of Database Server and an Android App which has GPS tracker and Admin section. The GPS helps to discover the various location data. The targeted locations (sensitive areas) such as hospitals, signals, old age homes, baby care centers are extracted from the database server. The performance of this system would result in an embedded module for the benefit of residents of communities close to sensitive areas such as highways, schools near busy roads, hospitals on the side of the road.

At the Traffic Signal where some drivers starts honking without waiting to turn the signal to green. Many of the

drivers Shows No care or respect about people which required peaceful environment. As the smartphones are easily available with low cost, drivers are familiar with many of the applications. So this technology can be used which will benefited society. Figure 1 shows the structure of System Flow diagram. First the user's location information is extracted with the help of GPS. If the location is present in one of the Sound prohibited zone list then, auto ranging algorithm is applied for optimization to check the shortest distance. The horn of that vehicle is controlled with the help of hardware module installed in the vehicle.

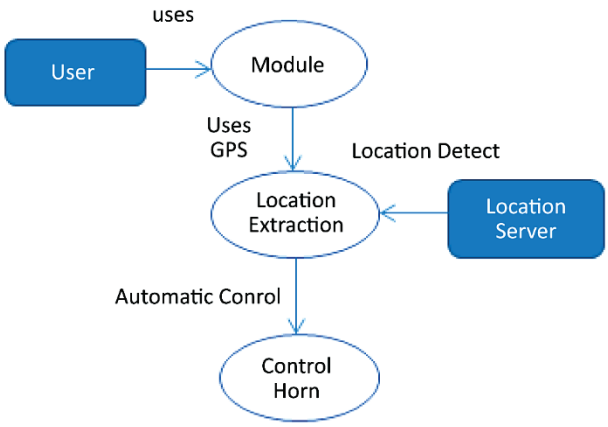


Figure 1 System Flow Diagram

Our proposed system aims at developing the real-time i.e. smart honing system because of which the intensity of the vehicles nearby sensitive areas i.e. schools, hospitals, baby care centers, ola-age homes, etc. is minimized. Sound which is measured in decibels can be controlled according to the governance but there is no device to measure and control in real time. To develop an embedded module to reduce a noise intensity for reduce noise pollution.

Fig. 2 focuses on the Architecture of the proposed system. The architectural diagram consists of e Hardware module consists of Arduino kit, WIFI module and Pizzo buzzer. In the proposed System, first the location is extracted, distance values are read from database then it will check frequency of horn if it is less than greater than minimum decibels then sends alert to driver and trigger the Buzzer and horn will be controlled.

The various targeted locations such as hospitals, signals, old age homes, baby care centres are extracted from the database server. The Auto Ranging Algorithm will be used to check if the vehicle is in the targeted locations or not. If it is, then android will send command to Arduino through WIFI module or Bluetooth. And the intensity of horn will be reduced. As a part of optimization, Auto Ranging algorithm is used which leads to accurate results.

When vehicle entered in the No honking or sound prohibited zone system first checks frequency of horn if Horn intensity greater Then it will sends alert to driver And Automatically Controlled the horn.

The Auto Ranging Algorithm will be applied to check if the vehicle is in the targeted locations or not. If vehicle is in targeted location, android app by using GPS will send command to Arduino through WIFI module. The Pizzobuzzer that is attached to horn will receive the command and intensity of horn that is decibel level of sound of horn will be reduced. When the vehicle passes through targeted locations, the intensity of horn is set to normal decibel as per the previous level.

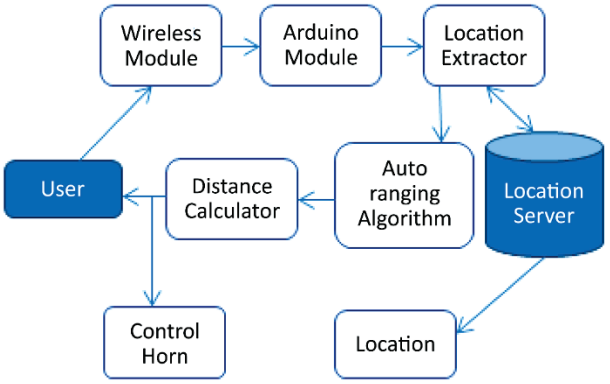


Figure 2 Architecture Diagram of the Proposed System

Big advantage to heart patients, kids and old people as the main target are the areas where these people are mostly suffered. The proposed system i.e. Sound Prohibited Zone for Smart Cities is to ensure that the Piezo buzzer that is attached to horn which will receive the command and intensity of horn that is decibel level of sound of horn will be reduced.

When the vehicle passes through targeted locations, the intensity of horn is set to normal decibel as previous. For the Smart Cities Project, Sound Prohibited Zone can surely become an important part. The system ensures to prevent noise pollution caused due to excessive honking of vehicles. An online monitoring system is to provide these data on the cloud in real-time stored in database.

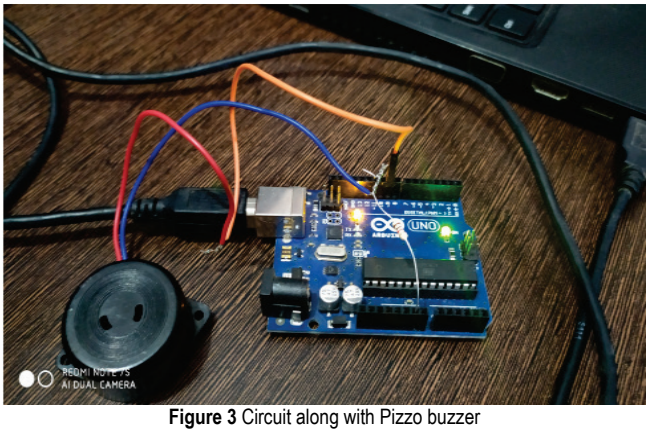


Figure 3 Circuit along with Pizzo buzzer

Sound Prohibited Zone System includes Android Application has Admin and User. In Admin login, It includes various operations like vehicle list which includes the driver list and is only accessible to the admin and not to the user, dustbin list (includes all the entries by the driver) and Add Zone List where admin can add new Sound Prohibited Zones. Admin has to enter location id to add new zone. When Location Id is entered the address of zone is automatically

fetched by using the GPS so that location can easily added. As shown in the below Fig. 3, the circuit is shown with the Pizzo Buzzer.

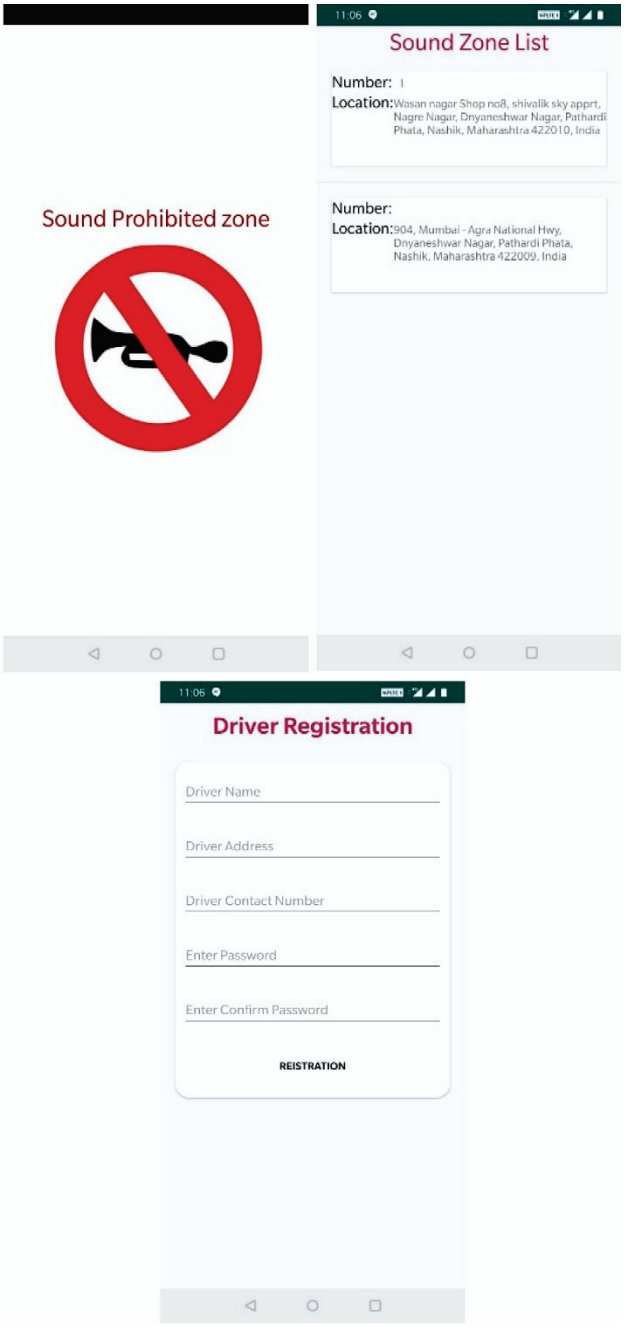


Figure 4 Snapshots (Driver Registration and Sound Prohibited Zone List)

As per snapshots shown in the above Fig. 4, driver registration has to be done. Sound Prohibited zone lists updating can be done by the admin. Admin can add the locations in the Sound Prohibited Zone list with image as well. On clicking that Image you will get to see the Location Id. All the entries by the driver or user are shown to Admin. User don't have access to Driver List. When clicked on the entry by driver, Admin can access its live location.

Admin can access the live location which is secure as only admin has access to the live location of the driver. User has to register to the application. Fig. 5 shows the snapshots after addition of Sound Prohibited Zone in the list by the admin. Registration includes Driver Name, Driver Address, Driver Contact No, and Password. It consist of all the zones entered by Admin. It will direct the driver to the Google Map and will inform if the vehicle is going to enter the zone. Current Location of Driver: The driver can see the its live location.

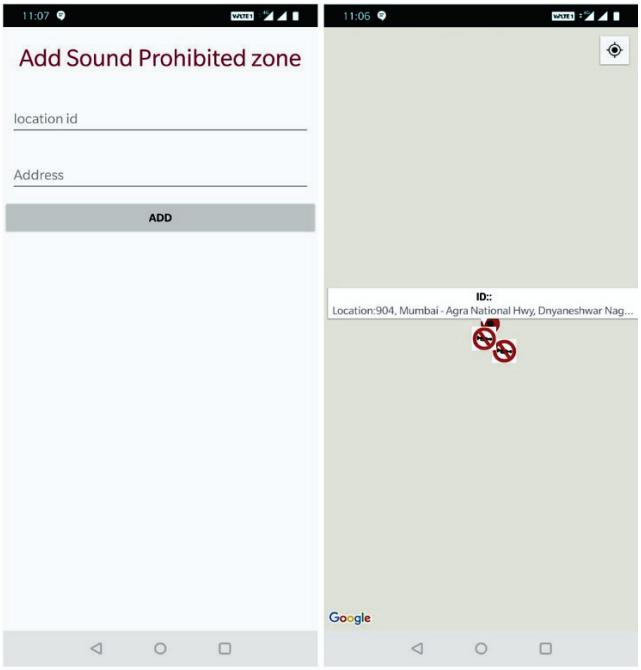


Figure 5 Snapshots (After addition of Sound Prohibited Zone)

4 RESULTS

The proposed system is compared with the existing system with respect to the different parameters like processing speed. The existing systems response time is less whereas, proposed system performs better in terms of processing time. Below graph (Fig. 6) shows the efficiency of the proposed system i.e. system performance.

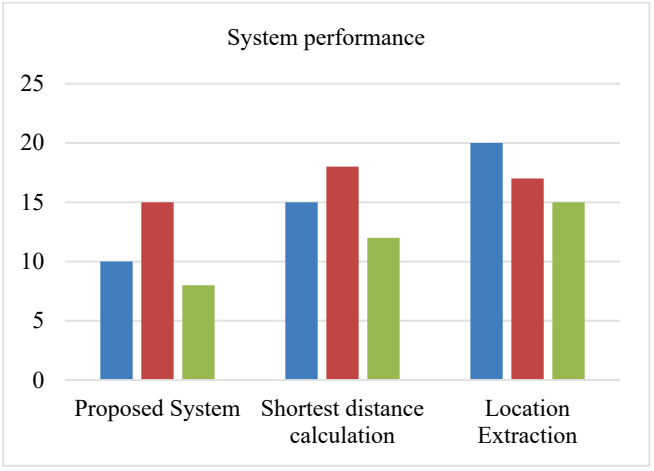


Figure 6 Graph showing System Performance i.e. Comparison with Existing Systems

Accuracy of the proposed system near some of the crowded areas is observed to be around 80%. The pictorial representation of the same is shown in the Fig. 7.

After applying the proposed system, the amount of noise pollution gets reduced. The reduction in the noise pollution is represented using graph in the Fig. 8.

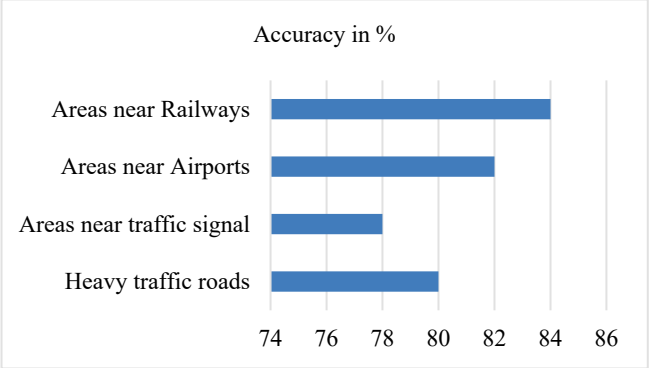


Figure 7 Graph showing System Accuracy near crowded areas.

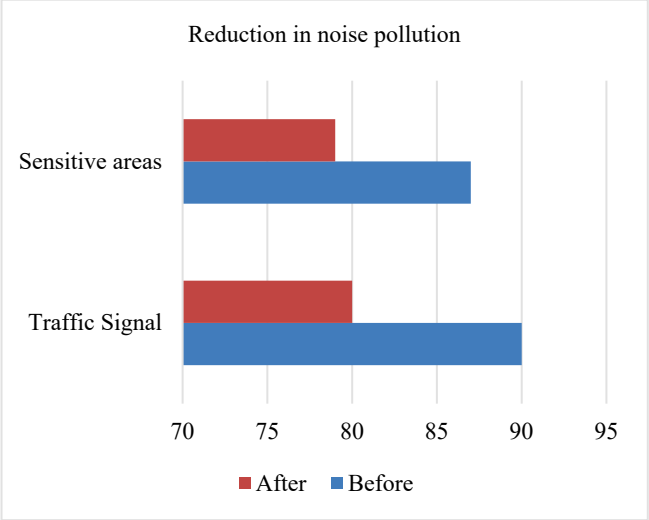


Figure 8 Graph showing Reduction in noise pollution

5 CONCLUSION

For the benefit of residents in communities near high-traffic highways, students studying in schools near busy roads, patients admitted to roadside hospitals, and people from different professions, the proposed framework results in an embedded module. To save many lives caused by their delay in reaching their desired destination browsers. Our proposed system aims at developing the real-time i.e. smart honing system because of which the intensity of the vehicles nearby sensitive areas i.e. schools, hospitals, baby care centers, ola-age homes, etc. is minimized.

The scope for the future work is inability to control the speed of the vehicle. The intensity of horn of vehicle is reduced but the speed of vehicle cannot be controlled. In addition to controlling the intensity of the horn, completely disabling of the horn once the person reaches the Sound Prohibited zone can be also worked on.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

6 REFERENCES

[1] Census of India 2011. Provisional Population Totals. Paper 2, Volume 1 of 2011. Rural-Urban Distribution India series 1. Office of the Registrar General & Census Commissioner, India. 2011

[2] Chakrabarty, N. & Reetesh, R. (2013). Aggressive Driving Case Studies and Mitigations in India. *International Journal of Scientific and Research Publications, 3*(2). www.ijsrp.org

[3] Dey, A., Arka, M., Raktim, P., & Bansari, D. M. (2019). Design of a Smart Real-time Excessive Honking Control System. *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE), 14*(6), 8-12.

[4] Goines, L. & Hagler, L. (2007). Noise Pollution: A Modern Plague. *Southern Medical Journal, 100*(3), 287-294.

[5] Ising, H. & Kruppa, B. (2004). Health effects caused by noise: evidence from the literature from the past 25 years. *Noise Health, 6*(22), 5-13.

[6] Joshi, V., Rajamani, N., Prathapaneni, N., & Subramaniam, L. V. (2013). Traffic density state estimation based on acoustic fusion. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, 478-482. https://doi.org/10.1109/ICASSP.2013.6637693

[7] Shalini, K. & Kumar, B. (2018). Development of traffic noise model (TNM) using regression analysis in Varanasi city, India. *International Journal of Civil Engineering and Technology, 9*(4), 70-76.

[8] Ministry of Environment and Forests, Government of India. No. 14 of 1981, (29/03/1981) - The Air (Prevention and Control of Pollution) Act, 1981, amended 1987. Available from: http://www.moef.nic.in/legis/air/air1.html. (Accessed on 2011 Dec 6).

[9] Mishra, R. K., Parida, M., & Rangnekar, S. (2010). Evaluation and analysis of traffic noise along bus rapid transit system corridor. *Int. J. Environ. Sci. Tech., 7*(4), 737-750.

[10] Niskar, A. S., Kieszak, S. M., Holmes, A. E., Esteban, E., Rubin, C., & Brody, D. J. (2001). Estimated Prevalence of Noise-Induced Hearing Threshold Shifts Among Children 6 to 19 Years of Age: The Third National Health and Nutrition Examination Survey, 1988–1994, United States. *Pediatrics, 108*(1), 40-43. https://doi.org/10.1542/peds.108.1.40

[11] Noise. World Health Organization. 2011. Available from: http://www.euro.who.int/en/what-we-do/health-topics/environmentand-health/noise. (Accessed on 2012 Aug 13).

[12] Vijay, R., Sharma, A., Chakrabarti, T. & Gupta, R. (2015). Assessment of honking impact on traffic noise in urban traffic environment of Nagpur, India. *Journal of Environmental Health Science and Engineering, 13*(10). https://doi.org/10.1186/s40201-015-0164-4

[13] Sakhare, S. R. & Ali, M. S. (2012). Genetic Algorithm Based Adaptive Scheduling Algorithm for Real Time Operating Systems. *International Journal of Embedded Systems and Applications (IJESA), 2*(3), 91-97.

[14] Sakhare, S. R. & Ali, M. S. (2011). An Adaptive Framework for the Selection of Embedded Operating Systems.

International Journal of Scientific & Engineering Research, 2(8). Available from: https://www.ijser.org/researchpaper/An-Adaptive-Framework-for-the-Selection-of-Embedded-Operating-Systems.pdf

[15] Sood, R., Sharma, S., & Yadav, V. K. (2016). Real Time Smart Honking System. *IEEE, International Conference on Information Technology (InCITe)*, 267-270. https://doi.org/ 10.1109/INCITE.2016.7857629

[16] Stansfeld, S. A. & Matheson, M. P. (2003). Noise pollution: non-auditory effects on health. *British Medical Bulletin, 68*(1), 243-257. https://doi.org/10.1093/bmb/ldg033

[17] Tyagi, V., Kalyanaraman, S., & Krishnapuram, R. (2012). Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics. *IEEE Transactions on Intelligent Transportation Systems, 13*, 1156-1166. https://doi.org/10.1109/TITS.2012.2190509

[18] Passchier-Vermeer, W. & Passchier, W. F. (2000). Noise exposure and public health. *Environmental Health Perspectives, 108*(Supplement 1), 123-131. https://doi.org/10.1289/ehp.00108s1123

[19] Zhao, Z., Chen, W., Semprun, K. A., & Chen, P. C. Y. (2019). Design and Evaluation of a Prototype System for Real-Time Monitoring of Vehicle Honking. *IEEE Transactions on Vehicular Technology, 68*(4), 3257-3267. https://doi.org/10.1109/TVT.2019.2893777

Authors' contacts:

Mrs. **Bhawana A. Ahire,** Assistant Professor
(Corresponding author)
Department of Information Technology,
MVP Samaj's K B T College of Engineering,
Gangapur Road, Udoji Maratha Boarding Compound, Nashik, Maharashtra 422013, India
ahire.bhawana@kbtcoe.org

Dr. **Sachin R. Sakhare,** Professor
Vishwakarma Institute of Information Technology,
Survey No. 3/4, Kondhwa (Budruk), Pune – 411048, Maharashtra, India
sachin.sakhare@viit.ac.in
sakharesachin7@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
httpa://doi.org/10.31803/tg-20210205092413

Preliminary communication

Network Performance Measurement through Machine to Machine Communication in Tele-Robotics System

Meghana P. Lokhande*, Dipti Durgesh Patil

Abstract: Machine-to-machine (M2M) communication devices communicate and exchange information with each other in an independent manner to perform necessary tasks. The machine communicates with another machine over a wireless network. Wireless communication opens up the environment to huge vulnerabilities, making it very easy for hackers to gain access to sensitive information and carry out malicious actions. This paper proposes an M2M communication system through the internet in Tele-Robotics and provides network performance security. Tele-robotic systems are designed for surgery, treatment and diagnostics to be conducted across short or long distances while utilizing wireless communication networks. The systems also provide a low delay and secure communication system for the tele-robotics community and data security. The system can perform tasks autonomously and intelligently, minimizing the burden on medical staff and improving the quality and system performance of patient care. In the medical field, surgeons and patients are located at different places and connected through public networks. So the design of a medical sensor node network with LEACH protocol for secure and reliable communication ensures through the attack and without attack performance. Finally, the simulation results show low delay and reliable secure network transmission.

Keywords: Machine-to-Machine (M2M) communication; medical field introduction; network performance security; tele-robotic; wireless sensor network

1 INTRODUCTION

Machine-to-machine communication (M2M) refers to the exchange of information between two devices. It is a key component in the fast-growing Internet of things (IoT), an ever-growing connection of devices in the home, workplace, and industry. M2M communication can be simple, like one-way data transmission, or complex, like multiplexed signals that allow devices to share a decision-making process. In the medical device industry, M2M communication is one of the fastest-growing sectors. According to Global Info Research, the connected medical device market is projected to expand from $939 million in 2018 to $2.7 billion by 2023, with the largest growth forecast for the United States. M2M communication faces various security challenges. Much vulnerability arises from the lack of a central authority and a wireless medium of transmission. Route creation and data transmission are two important functions of the routing algorithm. These two stages must be protected from attackers. The routing Protocol must be strong enough to withstand various attacks. Thus, reliable communication means a secure routing algorithm [1]. Tele-robotics is an integral part of the broader field of telemedicine. The main aim is to provide medical care for long distances and removes the need for both the doctor and the patient to be physically present in the same place. Consultation by far the possibility of diagnosis and treatment can greatly affect the life of patients with limited access to specialist health services [2]. Tele-robotics is practically lead by special doctors where there are no medical institutions or specialists. In addition to medical isolation, Tele-robotic is very important in eliminating medical issues in building countries, disaster places, and war areas where ongoing medical services are not available or do not have time to shift the patient to the hospital [3]. In the mid-1980s, robotic systems were introduced into medicine and today they affect a wide range of medical areas [4, 5]. In a Tele-robotic system, a remote manipulator controls the operator of the object, receives visual and information, and transmits it to the control position. Local and remote systems are called master and slave systems, entire master-slave is a system of remote control of the manipulator program [9] [10]. Figure 1 presents the Tele-robotic system consists of information of melody system for robot-supported telemetry applications.

The underlying system of telerobotic is telepresence. Tele-presence assumes that all information of the remote environment is inherently provided by the operator [11]. The main connection is created in master and slave communication to called machine to machine communication through the internet. The length of nodes from each other is vast; the data transfer delay may disturb the system and ultimately affect the performance of the healthcare professional. The communication quality and performance metrics are important in telerobotic M2M communication. Security and privacy are very important aspects of any communication. Network performance is reduced if a malicious node is present in the network. The incorrect behavior of the node presents malicious packet drop attacks, disturb the routing rules, data transmission is corrupted, packets are dropped and data is lost. So the presented systems provide a low delay and secure communication for the telerobotic community and data security. The system can perform tasks autonomously and intelligently, minimizing the burden on medical staff and improving the quality and system performance of patient care. The LEACH routing protocol is presented to make communication reliable and robust with the medical sensor node network. Once the network is created, network performance of the system is analyzed with or without DoS or Man-in-Middle attacks.

The need for research work provides telerobotics with significant progress to show how the surgery is performed in the operating room. It combines technological and clinical progress in the development of new robotic systems and

surgical techniques to improve the quality and results of surgical intervention. Telerobotic and various imaging techniques act as intermediaries between the surgeon's hand, eye, and surgery site, respectively, but these two elements are part of a larger information system that continues to evolve and affect all aspects of surgery and medicine in general. Providing this technology to surgeons has led to the development of new surgical techniques that would otherwise be impossible. The clinical knowledge gained from these new systems and the understanding of their potential will lead to the development of new and more effective telerobotic systems in the future. The paper is organized as section 1 presents introduction, section 2 shows related work. Section 3 and 4 covers proposed research work and simulation results respectively. Finally last section summarizes points in conclusion.

2 RELATED WORK

M2M communication allows the development of a range of applications, such as smart networks, smart robots, smart transport, and home networks. In [12] analyze the M2M communication and intelligent network environment. The structure of the machine network and its possible applications has been proposed in [13]. Other learning experiments include such as Meteorological Services, Environmental Pollution prevention, integrated video services, etc. [14-17]. In [18] various design issues, such as describing different limitations of sensor assemblies, applications, etc. Was introduces a system for health surveillance in [19]. In [20] proposes an implementation of patient monitoring [21] and [22] presents a patient monitoring WAP system.

Tele-robotic systems should play an important role in the medical field [6]. The first successful remote operation, called Operation Lindbergh, was performed in 2001, using the Zeus Robotic System [7]. Laparoscopic gallbladder surgery was carried on patients in Strasbourg, France, while surgeons were based in New York, USA. Although the first remote operation took place only in 2001 but the tele-operative system was introduced much earlier. The automatic endoscope system uses the vocal arm for minimally invasive surgery to hold the endoscope [8]. The first surgical robot systems build for eye surgery at Northwestern University [23]. It travels through a virtual long-distance traffic center (RCM) at the sclera entry point [24]. A hypodermic needle is attached to the parallel mechanism and inserted through the sclera. A doctor and robot work together with one tool [25]. Tele-manipulator on a big system gives natural feedback. There was a strong sense that there was a large-scale position is not as important as reducing the tremor. In [26, 27] includes virtual devices for continuous hand manipulation, which are movement restrictions applied in software. In [28, 29] ETH developed in Zurich with magnetic control different from the current development. This is a magnetic control micro-robot platform that uses remote communication, using a primer connection. In [30] proposes an alternative parallel robot structure capable of performing virtual RCM motion through the sclera and the entry point. In [31] developed a handheld robotic device called Micron for microsurgery. The Micron concept is to measure the surgeon's hand movements, use advanced filtering technology to separate unwanted movements such as tremors from deliberate movements, and use robotic end effectors to adjust the tip of the instrument for unwanted movements. The Micron detects dynamic motion using an accelerometer and the final effectors of the robot. In [32] advanced robotic system for microsurgical keratoplasty was developed.

In [33] developed more accurate slave telesurgical pathologies of the throat and upper respiratory tract, it has the advantage of distal dexterity and functional self-enhancement. In [34] it has been shown that robotic electrode array inserts significantly reduce input forces when inserting a cochlear implant compared to unmanaged electrode arrays. In [35] build acceleration noninvasive Radiosurgery System that can compensate for the limited target movement, unlike the localization system. The delay could have serious consequences on the performance and control of systems.

In [36] review on the topic of two-way teleoperation, with several approaches, was presented. It covers a range of methodologies, including the passivity-based monitoring proposed to address the above-mentioned problems. In [37] applied the concept of wave variables, an extension of the theory of passivity, to time-delayed teleoperations that are unknown but contain a constant time delay. The work in [38] studied a Force-feedback algorithm, called a-Force, reflection activities, and experimental work; they use the method of minimally invasive surgical applications, communications delays. In [39] several unanswered questions and provides a clear description of IoT security research. Many of the dangers associated with this heterogeneous security infrastructure and privacy policies require a deep explanation. In [40] presents an overview of the access control approach attempts to improve security. The author not only summarizes the approach to access control but also provides an overview of existing limitations and unresolved issues. In [41] provides a model for predicting health risk through a wireless sensor network. The work [42] discussed performance measurement for 4G and 5G enabled architecture in telesurgery.

M2M technology is fast for health monitoring applications without human interference smart devices of WSN and mobile device networks provide remote monitoring of health data. M2M communication addresses several important challenges and the key challenges include security, standardization, privacy, software development, reliability, and communication delay.

3 PROPOSED RESEARCH WORK

M2M is a kind of network in which a large number of smart devices decide to generate information, exchange it, and cooperate without human direct interference. With the different applications and numerous advantages, M2M networks in their design face several technical challenges. One of the key issues holding back the growth of M2M communications is their security.

The presented system will consist of three units: a master station is an expert system, a slave station is a patient system,

and a communication connection that allows data exchange by stations. Block diagram of the proposed system is shown in Fig. 1. The patient on-site, US probe is maintained to operate by the robotic arm system, follow the approach of management master/slave. The thin robot placed on the patient by paramedics. When the robot controller adopts classical force feedback control [43], which allows limiting the force to 20 N for patient comfort and safety, the robot control data update rate is up to 1 kHz without loss or jitter, and teleoperation becomes transparent [44]. The robot is controlled by an open circuit through a communication connection, but by a closed-circuit locally controlled in the patient's position. Eliminate this approach and instability in handling the problems. It has also been shown that the design of this robot allows geometric control [45].

In place of the master, the medical inspector moves the dummy of the ultrasound probe required for echo monitoring. The system also includes an actuator that can control the perception of power and feel the effect of the telerobot.

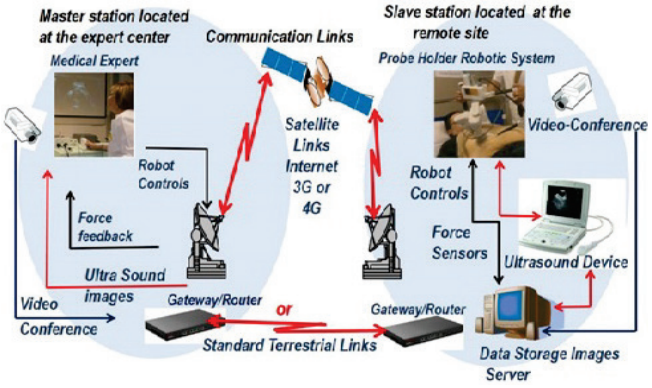


Figure 1 M2M communication through the internet in Tele-Robotics and network performance security [46]

The expert's hand movement is perceived, information is transmitted in real-time to the robot platform via a communication line (ground or satellite), and the operating system copies this movement. An important step in the field of information security is to prevent threats and vulnerabilities at the initial design stage. M2M communication consists of various vulnerabilities and security challenges such as physical security, resource constraints, heterogeneity, delay constraints, and scalability. The problem is to ensure communication security and maintain communication in the presence of adversaries throughout the topology of an unknown, frequently changing multi-hop wireless network. To solve this complex problem and ensure comprehensive security, both stages of communication, route discovery, and data transmission must be protected. However, M2M has faced serious security threats that typically affect end-to-end delay, buffer mechanism, response time, and overall network throughput, packet dropping, less packet delivery ratio, resulting in reduced performance. Therefore, it is economically difficult to manage such networks. Nodes are behaved maliciously by disproportionate network activity and thereby dropping packets to reduce efficiency. For secure or safe communication proposes a LEACH routing protocol with various medical sensor nodes, such as cameras, health checkup machines, computers, printers, etc.

A wireless sensor network uses a hierarchical routing protocol called LEACH, which extends the network life. Using the LEACH Protocol, medical sensor nodes are placed in clusters, and one node of these nodes act as a cluster head (CH) based on their energy and the distance from the medical sensor node to the base station (BS). CH is responsible to collect the data from its cluster and transmit to the BS. In process of data transmission more energy is consumed. LEACH enhances the network life without attack and reduces the power consumption of the WSN. This makes the system more secure and useful in the medical field.

4 SIMULATION RESULT AND DISCUSSION

A square area of 1000 × 1000 m is created by setting up 100 medical sensor nodes in the WSN as the basis for this study. The code is executed in the simulator NS-2 and the initial site settings are given randomly. To model the network, each node is given the initial energy in Joules units, and after the simulation, the total energy consumption is analyzed. First, create a LEACH protocol, and that legitimate traffic is generated. After that introduces a DoS or Man-in-Middle attack to analyze its impact on network performance. Next, measure performance parameters such as throughput, packet delivery speed (PDR), delay, overhead, and energy consumption of node. Experiments have made it possible to increase the service life of the network by about 20-25% using the proposed method. In traditional protocols, CH is created on the probability function, and CHs are selected in each round. The continuous simulation process provided that all nodes are consumed with energy values. The utilized simulation factors shows in Tab. 1.

Table 1 Simulation factors

Parameter	Values
Number of IoT/Machines/Sensor nodes	100, 200, 300, 400, 500, 600
No. of Attackers	10 % of nodes
Pattern of traffic	CBR
Connections number	10
Sink	Base Station (BS)
Area	1000 × 1000 (Long distance communications)
MAC	802.11
Topology	Random deployment
Routing Protocol	LEACH (state-of-art)
Initial Energy	0.5 J
Transmitter energy consumption	16.7 nJ
Receiver energy consumption	36.1 nJ
Simulation Time	200 seconds

Fig. 2 shows the throughput of the network in kbps, in which the *X*-axis represents the medical sensor nodes such as cameras, health checkup machines, computers, printers, etc. and the *Y*-axis represents the throughput values of nodes.

It shows LEACH protocol with and without attack and indicates the performance of the system. LEACH-A indicates the LEACH protocol with the attack. Throughput is the calculation of data successfully transmitted from the sending

node to the receiving node over a specified time. The result shows LEACH without attack gives higher throughput values than LEACH-A with attacks. The Average throughput values of LEACH without attack are 63.88 kbps and LEACH-A with an attack is 58.56 kbps.

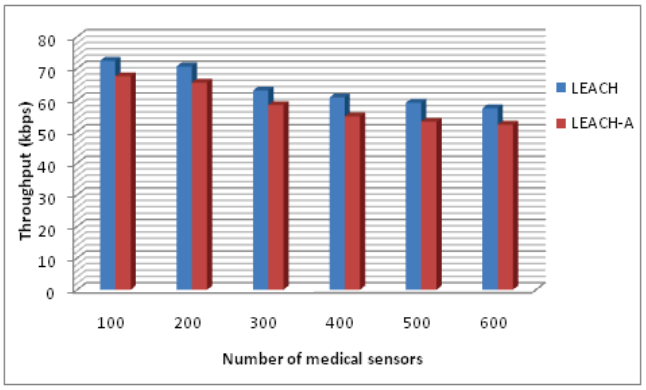


Figure 2 Throughput

Fig. 3 shows the packet delivery ratio (PDR) of the system in percent, in which the *X*-axis represents the medical sensor nodes and the *Y*-axis represents the PDR. PDR calculation is the percentage of the receiving to send packets. The PDR using LEACH protocol without attack is more than LEACH-A with the attack. The PDR values of LEACH without attack are 89.25 % and LEACH-A with an attack is 69.98 % respectively.

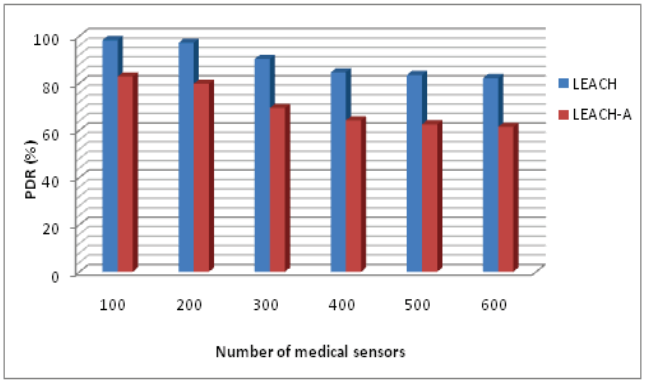


Figure 3 Packet delivery ratio

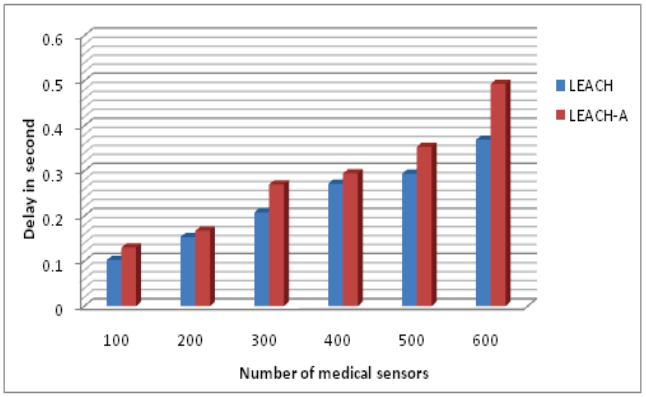


Figure 4 Delay

Fig. 4 shows the Delay in the network where the *X*-axis represents the sensor nodes and the *Y*-axis represents the delay of the system in a second. Delay is the amount of time that the source's physical layer takes to send a packet over a link. The delay of the LEACH protocol without attack is less than LEACH-A with the attack. The delay in second of LEACH and LEACH-A are 0.2327 and 0.2845 respectively and the difference is +0.0518.

Fig. 5 shows the communication overhead of the system in a millisecond, in which the *X*-axis represents the medical sensor nodes such as cameras, health checkup machines, computers, printers, etc. and *Y*-axis represents the communication overhead of the system.

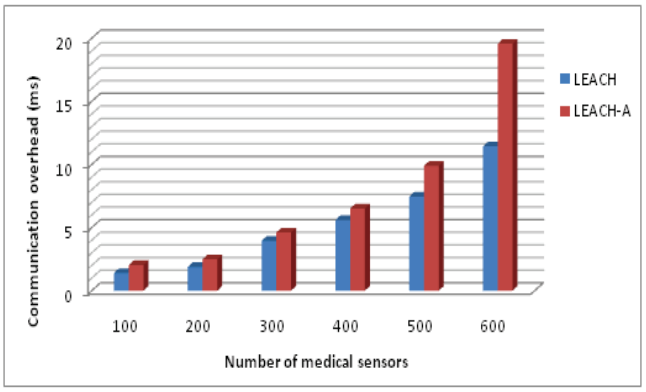


Figure 5 Communication overhead

Communication overhead is the ratio of actual communication time and computed communication time for actual communication time. The communication overhead of the LEACH protocol without attack is less than the LEACH-A protocol with the attack. The communication overhead in milliseconds (ms) of LEACH protocol without attack is 5.28 and LEACH-A with an attack is 7.51 respectively and the difference is +2.23.

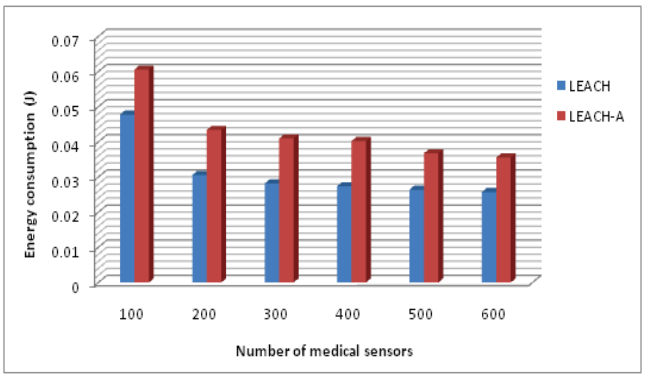


Figure 6 Energy consumption

Energy is the main issue in WSN, in order to enhance energy efficiency, it requires creation of virtual nodes to increase the throughput. Fig. 6 shows the energy consumption of the node. The result shows that the LEACH-A with attack consumes more energy than LEACH without attack. The energy consumption of LEACH protocol without

attack is 0.031 (J) and LEACH-A protocol with an attack is 0.04294 (J) respectively and the difference is +0.01194 (J).

5 CONCLUSION

In healthcare, machine-to-machine communication (M2M) provides a new way of service for a patient's life. M2M utilizes wireless communication technology that enables automatic instructions between devices, such as starting, stopping, sending and receiving data. This paper proposes an M2M communication system through the internet in Tele-Robotics, the main goal is to generate various tasks automatically without human interference and minimizing the burden on the medical staff, and improving the quality and system performance of patient care. Medical telerobotics shows a powerful impact on healthcare. An indicator of this possibility is the fact that Tele-robotics has been considered in many applications and medical specialties, which are revealed in this review. The wireless technology allows the device to receive M2M communication and adjust the implant performance based on biometric feedback from the doctor. With the safety of critical operations in healthcare, securing a network against attacks is highly important. Resources in a sensor network are scarce in terms of energy and the biggest performance parameters in the medical field are delay, throughput, PDR, and communication overhead. Here performance metric parameters measured for LEACH and LEACH-A protocol. Simulation results show LEACH has better performance in terms of energy-saving and low delay over LEACH-A.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

6 REFERENCES

[1] Troccaz, J., Dagnino, G., & Yang, G.-Z. (2019). Frontiers of Medical Robotics: From Concept to Systems to Clinical Translation. *Annual Review of Biomedical Engineering, 21*, 193-218. https://doi.org/10.1146/annurev-bioeng-060418-052502

[2] Yang, G.-Z., Cambias, J., Cleary, K., Daimler, E., Drake, J., Dupont, P. E., Hata, N., Kazanzides, P., Martel, S., Patel, R. V. et al. (2017). Medical roboticsregulatory, ethical, and legal considerations for increasing levels of autonomy. *Science Robotics, 2*(4), p. 8638. https://doi.org/10.1126/scirobotics.aam8638

[3] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60-88. https://doi.org/10.1016/j.media.2017.07.005

[4] Meyer, A., Zverinski, D., Pfahringer, B., Kempfert, J., Kuehne, T., Sündermann, S. H., Stamm, C., Hofmann, T., Falk, V., & Eickhoff, C. (2018). Machine learning for real-time prediction of complications in critical care: a retrospective study. *The Lancet Respiratory Medicine, 6*(12), 905-914. https://doi.org/10.1016/S2213-2600(18)30300-X

[5] Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., Davidson, B., Pereira, S. P., Clarkson, M. J., & Barratt, D. C. (2018). Automatic multi-organ segmentation on abdominal CT with dense vnetworks. *IEEE Transactions on Medical Imaging, 37*(8), 1822-1834. https://doi.org/10.1109/TMI.2018.2806309

[6] Islam, M., Atputharuban, D. A., Ramesh, R., & Ren, H. (2019). Real-time instrument segmentation in robotic surgery using auxiliary supervised deep adversarial learning. *IEEE Robotics and Automation Letters, 4*(2), 2188-2195. https://doi.org/10.1109/LRA.2019.2900854

[7] Laina, I., Rieke, N., Rupprecht, C., Vizcaíno, J. P., Eslami, A., Tombari, F., & Navab, N. (2017). Concurrent segmentation and localization for tracking of surgical instruments. in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2017), Lecture Notes in Computer Science, vol 10434*, Springer, 664-672. https://doi.org/10.1007/978-3-319-66185-8_75

[8] Liao, R., Miao, S., de Tournemire, P., Grbic, S., Kamen, A., Mansi, T., & Comaniciu, D. (2016). An artificial agent for robust image registration. *Proceedings of AAAI Conference on Artificial Intelligence*.

[9] Zhou, X.-Y., Ernst, S., & Lee, S.-L. (2016). Path planning for robot-enhanced cardiac radiofrequency catheter ablation. *2016 IEEE international conference on robotics and automation (ICRA 2016)*, 4172-4177. https://doi.org/10.1109/ICRA.2016.7487610

[10] Bouget, D., Benenson, R., Omran, M., Riffaud, L., Schiele, B., & Jannin, P. (2015). Detecting surgical tools by modelling local appearance and global shape. *IEEE Transactions on Medical Imaging, 34*(12), 2603-2617. https://doi.org/10.1109/TMI.2015.2450831

[11] Shvets, A. A., Rakhlin, A., Kalinin, A. A., & Iglovikov, V. I. (2018). Automatic instrument segmentation in robot-assisted surgery using deep learning. *Proceedings of IEEE International Conference on Machine Learning and Applications (ICMLA 2018)*, 624-628. https://doi.org/10.1109/ICMLA.2018.00100

[12] Toh, S.-H., Lee, S.-C., & Chung, W.-Y. (2008). WSN Based Personal Mobile Physiological Monitoring and Management System for Chronic Disease. *Convergence Information Technology, International Conference on. 1*, 467-472. https://doi.org/10.1109/ICCIT.2008.392

[13] Niyato, D., Xiao, L., & Wang, P. (2011). Machine-to-machine communications for home energy management system in smart grid. *IEEE Communications Magazine, 49*(4), 53-59. https://doi.org/10.1109/MCOM.2011.5741146.

[14] Zhang, Y., Yu, R., Xie, S., Yao, W., Xiao, Y., & Guizani, M. (2011). Home M2M networks: Architectures, standards, and QoS improvement. *IEEE Communications Magazine, 49*(4), 44-52. https://doi.org/10.1109/MCOM.2011.5741145

[15] Jeon, S., Park, K.-W., Ryu, H., & Kim, Y. (2010). A design of M2M-based intelligent operating system for effective pollution control facilities. *2010 International Conference on Information and Communication Technology Convergence, ICTC 2010*. https://doi.org/10.1109/ICTC.2010.5674786

[16] Mitsui, H., Kambe, H., & Koizumi, H. (2010). Student experiments for learning basic M2M technologies by implementing sensor network systems. *2010 9th International Conference on Information Technology Based Higher Education and Training (ITHET)*, Cappadocia, 268-275. https://doi.org/10.1109/ITHET.2010.5480082.

[17] Song, J., Ahn, J. Y., & Kim, S. (2010). A Study on M2M-based System for Hygienic Meteorology Service. *International*

Conference on Security-Enriched Urban Computing and Smart Grid, SUComS 2010. Communications in Computer and Information Science, 78, 39-45. https://doi.org/10.1007/978-3-642-16444-6_6

[18] Wang, S.-C., Chung, T.-C., & Yan, K.-Q. (2008). Machine-to-Machine Technology Applied to Integrated Video Services via Context Transfer. *APSCC '08: Proceedings of the 2008 IEEE Asia-Pacific Services Computing Conference*, 1395-1400. https://doi.org/10.1109/APSCC.2008.67

[19] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y. & Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer Networks, 38*(4), 393-422. https://doi.org/10.1016/S1389-1286(01)00302-4

[20] Boric-Lubecke, O. & Lubecke, V. M. (2002). Wireless house calls: using communications technology for health care and monitoring. *IEEE Microwave Magazine, 3*(3), 43-48. https://doi.org/10.1109/MMW.2002.1028361

[21] Varady, Z. B. P. & Benyo, B. (2002). An open architecture patient monitoring system using standard technologies. *IEEE Transactions on Information Technology in Biomedicine, 6*(1), 95-95. https://doi.org/10.1109/4233.992168

[22] Hung, K. & Zhang, Y. T. (2003). Implementation of a wap-based telemedicine system for patient monitoring. *IEEE Transactions on Information Technology in Biomedicine, 7*(2), 101-107. https://doi.org/10.1109/TITB.2003.811870

[23] Jensen, P. S., Grace, K. W., Attariwala, R., Colgate, J. E., & Glucksberg, M. R. (1997). Toward robot-assisted vascular microsurgery. *Graefe's Arch. Clin. Expl Ophthal., 235*(11), 696-701. https://doi.org/10.1007/BF01880668

[24] Stewart, D. (1965). A platform with six degrees of freedom. *Proc. Instn Mech. Engrs, 180*, 371-386. https://doi.org/10.1243/PIME_PROC_1965_180_029_02

[25] Mitchell, B., Koo, J., Iordachita, I., Kazanzides, P., Kapoor, A., Handa, J., Hager, G., & Taylor, R. (2007). Development and application of a new steadyhand manipulator for retinal surgery. In *Proceedings of the IEEE International Conference on Robotics and automation*, Rome, Italy, 623-629. https://doi.org/10.1109/ROBOT.2007.363056

[26] Abbott, J. J. & Okamura, A. M. (2006). Stable forbidden-region virtual fixtures for bilateral telemanipulation. *Trans. ASME, J. Dynamic Systems, Measmt, Control, 128*(1), 53-64. https://doi.org/10.1115/1.2168163

[27] Abbott, J. J. & Okamura, A. M. (2007). Pseudoadmittance bilateral telemanipulation with guidance virtual fixtures. *Int. J. Robotics Res., 26*(8), 865-884. https://doi.org/10.1177/0278364907080425

[28] Dogangil, G., Ergeneman, O., Abbott, J. J., Pane, S., Hall, H., Muntwyler, S., & Nelson, B. J. (2008). Toward targeted retinal drug delivery with wireless magnetic microrobots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent robots and systems*, Nice, France, 1921-1926. https://doi.org/10.1109/IROS.2008.4650768

[29] Ergeneman, O., Dogangil, G., Kummer, M. P., Abbott, J. J., Nazeeruddin, M. K., & Nelson, B. J. (2008). A magnetically controlled wireless optical oxygen sensor for intraocular measurements. *IEEE Sensors J., 8*(1), 29-37. https://doi.org/10.1109/JSEN.2007.912552

[30] Wei, W., Goldman, R., Simaan, N., Fine, H., & Chang, S. (2007). Design and theoretical evaluation of micro-surgical manipulators for orbital manipulation and intraocular dexterity. In *Proceedings of the IEEE International Conference on Robotics automation*, Rome, Italy, 3389-3395. https://doi.org/10.1109/ROBOT.2007.363996

[31] Ang, W. T., Pradeep, P. K., & Riviere, C. N. (2004). Active tremor compensation in microsurgery. In *Proceedings of the IEEE International Conference of the Engineering in Medicine and Biology Society*, San Francisco, USA, 2738-2741. https://doi.org/10.1109/IEMBS.2004.1403784

[32] Hu, Y., Li, D., Zong, G., & Sun, X. (2005). Robotic system for microsurgical keratoplasty. In *Proceedings of the IEEE International Conference of the Engineering in Medicine and Biology Society*, Shanghai, P. R. China, 5762-5765.

[33] Simaan, N., Taylor, R., & Flint, P. (2004). High dexterity snake-like robotic slaves for minimally invasive telesurgery of the upper airway. In *Lecture notes in computer science, Proceedings of the Medical Image Computing and Computer-Assisted Intervention Conference*, 17-24. https://doi.org/10.1007/978-3-540-30136-3_3

[34] Zhang, J., Wei, W., Manolidis, S., Roland Jr, J. T., & Simaan, N. (2008). Path planning and workspace determination for robot-assisted insertion of steerable electrode arrays for cochlear implant surgery. In *Lecture notes in computer science, Proceedings of the International Conference on Medical image computing and computer-assisted intervention*, 692-700. https://doi.org/10.1007/978-3-540-85990-1_83

[35] The CyberKnife Surgical System, 2008, Accuray Inc. (online). Available from http://www.accuray. com/

[36] Hokayem, P. F. & Spong, M. W. (2006). Bilateral teleoperation: an historical survey. *Automatica, 42*(12), 2035-2057. https://doi.org/10.1016/j.automatica.2006.06.027

[37] Niemeyer, G. (2004). Telemanipulation with Time Delays. *Int J Rob Res., 23*(9), 873-890. https://doi.org/10.1177/0278364904045563

[38] Takhmar, A., Polushin, I. G., Talasaz, A., & Patel, R. V. (2015). Cooperative teleoperation with projection-based force reflection for MIS. *IEEE Trans Control Syst Technol., 23*(4), 1411-1426. https://doi.org/10.1109/TCST.2014.2369344

[39] Lokhande, M. & Patil, D. D. (2020). Security Threats In M2M Framework of IoT. *International Journal of Advanced Science and Technology, 29*(8), 1809-1823.

[40] Lokhande, M. & Patil, D. D. (2019). Access control approaches in Internet of Things. *International Journal of Computer Science and Engineering, 7*(5), E-ISSN: 2347-2693 https://doi.org/10.26438/ijcse/v7i5.11581161

[41] Patil, D. D. & Wadhai, V. M. (2012). Adaptive Real Time Data Mining Methodology for Wireless Body Area Network Based Healthcare Applications. *Advanced Computing: An International Journal (ACIJ 2012), 3*(4). https://doi.org/10.5121/acij.2012.3408

[42] Meshram, D. & Patil, D. (2020). 5G Enabled Tactile Internet for Tele-Robotic surgery. *Third International Conference on Computing and Network Communications, Procedia Computer Science, 171*, 2618-2625. https://doi.org/10.1016/j.procs.2020.04.284

[43] Buss, M. & Schmidt, G. (1999). Control problems in multi-modal telepresence systems. In *Advances in control*. London, Springer, 65-101. https://doi.org/10.1007/978-1-4471-0853-5_3

[44] Guerraz, A. (2002). Etude du télégeste médical non invasif utilisant un transducteur gestuel à retour d'efforts. Grenoble: Université Joseph-Fourier-Grenoble I.

[45] Courreges, F. (2003). Contributions à la conception et commande de robots de télé-chographie. Orléans: Orléans University.

[46] Avgousti, S., Christoforou, E. G., Panayides, A. S. et al. (2016). Medical telerobotic systems: current status and future trends. *BioMed Eng OnLine, 15*(96). https://doi.org/10.1186/s12938-016-0217-7

Authors' contacts:

Meghana P. Lokhande, PhD Scholar
(Corresponding author)
Department of Computer Engineering,
Smt. Kashibai Navale College of Engineering, Pune, India
meghna.ingole1983@gmail.com

Dipti D. Patil, Professor
Department of Information Technology,
MKSSS's Cummins College of Engineering for Women, Pune, India
dipti.patil@cumminscollege.in

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210205094356

Preliminary communication

An Efficient Query Optimizer with Materialized Intermediate Views in Distributed and Cloud Environment

Archana Bachhav*, Vilas Kharat, Madhukar Shelar

Abstract: In cloud computing environment hardware resources required for the execution of query using distributed relational database system are scaled up or scaled down according to the query workload performance. Complex queries require large scale of resources in order to complete their execution efficiently. The large scale of resource requirements can be reduced by minimizing query execution time that maximizes resource utilization and decreases payment overhead of customers. Complex queries or batch queries contain some common subexpressions. If these common subexpressions evaluated once and their results are cached, they can be used for execution of further queries. In this research, we have come up with an algorithm for query optimization, which aims at storing intermediate results of the queries and use these by-products for execution of future queries. Extensive experiments have been carried out with the help of simulation model to test the algorithm efficiency.

Keywords: Conventional SQL; Database Management in Cloud Environment; Distributed Databases; Intermediate Views; Materialized Views; Query Execution Time; Query Optimization

1 INTRODUCTION

In cloud environment, bulk of applications are based on data which are managed by Database Management Systems and that forms a crucial issue on cloud platform. So, service oriented computing in cloud, however is also extended to Database as a Service (DBaaS) [1]. Database management in cloud and efficient data access for cloud users become cumbersome task for cloud service providers. Cloud computing has limit of sharing processing time and storage space for various applications in databases. The proliferation found in diverse applications which impacted exceptional cloud stages bringing about massive growth inside the length of the information created just as devoured through such applications [2]. The most effective method to put together and deal with those large databases in order to get the required information for the clients is found to be the new research area in distributed and cloud environment. The data modelling in cloud platform is the basis of cloud applications and the key issue is the searching algorithms applied [3]. The most effective method to get the information convenient, precisely and dependably; assumes a significant function in the achievement of database model in cloud platform.

In cloud platform, users can hire huge resources for short period of time to execute complex queries more efficiently on large database using group of virtual machines [4]. The hire charges of resources for users can be reduced using better query optimization technique [5]. Thus there is a need of exploring efficient techniques for query execution that would decrease runtime and response time. It will likewise upgrade optimum use of resources in cloud data centers. This paper is in continuation with [6-8], where an extensive survey on various query optimization approaches as well the novel architecture of an intelligent query optimizer for distributed database has been presented.

Contributions - As such, the researchers have designed and developed a technique for query optimization in distributed and cloud environment. The contributions of this research paper are as follows.

- Devise an architecture of query optimizer for distributed databases which is integrated with materialized views that are reused for evaluation of further queries in the system.
- Formulate a query optimization technique that results in a better optimization of database queries.
- Introduce an optimization strategy that may reduce bandwidth requirement in cloud environment by satisfying SLA between customer and cloud service provider.
- Demonstrate the developed model with TPC-H benchmark dataset and series of benchmark queries to test performance of devised query optimization technique.
- Improve resource utilization in cloud by reducing query evaluation time and response time.

The remainder of this research paper is structured as follows. Section 2 presents an outline of related work. The proposed work with cost model is elaborated in section 3. Section 4 describes an experimental setup with workload of benchmark dataset and series of queries for testing performance of query optimizer as well as results for evaluation of the proposed system. Section 5 concludes the paper.

2 RELATED WORK

Optimization of database queries is performed through two phases – search space generation and optimal plan selection from the search space [5]. Researchers have discovered different approaches on query optimization those deals with reduction of communication cost, reduction of execution time and appropriate utilization of system resources.

Execution time of queries can be reduced by eliminating common sub-expressions used. Start-fetch wrapper using request window mechanism is used by Lee R et al. [9] to develop an IGNITE system that eliminates common sub-expressions. However, the communication traffic generated in IGNITE is reduced by Chen G et al. [10] and Dokeroglu et al. [4] by using efficient sets of query execution plans in their research work. The response time of queries is reduced with the help of parallel query processing systems in research works proposed by Garofalakis et al. [11] and Dokeroglu et al. [12]. Giannikis et al. [13] proposed an architecture based on sharing of computation, storage and cache memory by creating batch of queries. Similar concept of resource sharing for query optimization is presented in [14].

Iterative processing method can also be used for query optimization in which actual runtime statistics is collected by continuously monitoring execution of queries [15]. However, based on intermediate results, it would be cumbersome to collect required statistics. Hence Cole and Grafe [16, 17] addressed a multiple plans generation technique at compile time. POP is the progressing query optimization approach invented in [18, 19], where cardinality approximation errors are detected in mid execution of queries. DB2L learning optimizer based on use of misestimates to learn and adjust the statistics to enhance better optimization of further queries [20]. Wang et al. [21] presented adaptive query optimization approach for cloud database system based on execution time as well as monetary costs.

The technique of caching intermediate results is one of the widely used query optimization technique [22], extended by Safaeei A et al. [23] based on multiple sliding windows to improve execution of overlapping queries with common sub-expressions. Laptev et al. [24] presented EARL system and Agarwal et al. [25] proposed the BlinkDB, those iteratively works for collecting larger samples to reach at the desired accuracy. The Shark, presented in [26] caches inter query data with the help of shared memory concept. In distributed cloud environment, CHive [27] and NOVA [28] are the query optimization techniques based on incremental processing of continuous data. Logothetis et al. [29] invented CBP (Continuous Bulk Processing) system in which working state is preserved during query processing to reuse it for future queries. DBaaS can offer an assistance to cloud users that ready to get results with more fragile quality in return of lower cost [30].

Query processing time can be reduced by materializing views generated from select-project join operations [31]. The policies of protecting materialized views cache are introduced by WATCHMAN [32] and DynaMat [33] systems to increase the hit ratio. Ivanova et al. [34] devised an architecture that optimizes query processing by maintaining cache to avoid repetitions of physical operations. The results from MapReduce jobs can be reused by describing them in the form of analytical query languages [35-37]. Perez et al. [38] presented history aware optimizer that archives intermediate results of queries so as to reuse them during future query execution. Intermediate results or views that are to be cached are determined by Cache-on-Demand [39] as well as MQT technique [40] and relations that may be useful in execution of future queries are identified by Kossmann et al. [41].

MapReduce technique results in higher processing cost for queries with more join operations, which can be reduced using pipeline approach where results of a query can be referred by next queries to continue processing. Anyanwu K et al. [42] proposed a data model to minimize the number of MapReduce cycles using pipeline approach. Automatic Query Analyser (AQUA) developed by Wu et al. [43] works for MapReduce in two phases, first phase for minimizing number of MapReduce cycles and second phase for joining intermediate results. MapReduce online system avoids materialization by pipelining midway results from map jobs to reduce jobs [44].

After navigating various approaches on database query optimization, it has been perceived that there is a need of suitable technique to reduce the runtime for database queries that minimizes resource requirement. Conventional SQL cannot predict the future requirements, therefore it reduces the performance than MapReduce technique. MapReduce uses pipeline approach to reduce query processing time, however it require higher processing cost for queries with more join operations. Hence to bridge the gap between MapReduce and Conventional SQL, there is a demand of efficient query optimization technique in distributed and cloud environment.

3 PROPOSED WORK

The primary aim of this research work is to design and develop an intelligent query optimizer that improves resource utilization in distributed and cloud environment by reducing query evaluation time and response time. As such, the proposed query optimizer materializes intermediate views during query evaluation so as to reuse them for execution further queries. Hence, it will reduce I/O operations, communication cost as well as execution time.

3.1 An Architecture

An architecture of the proposed query optimizer presented in Fig. 1 comprises of various components.

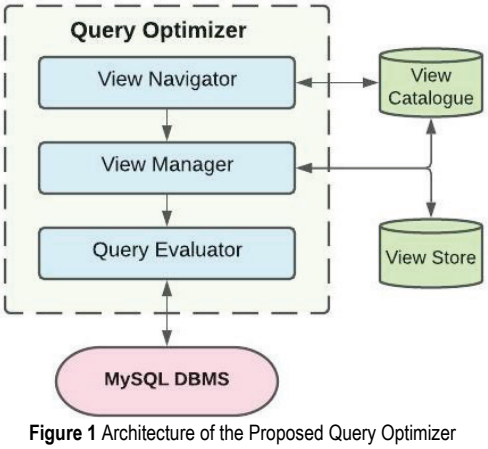


Figure 1 Architecture of the Proposed Query Optimizer

View Navigator searches for appropriate views during query evaluation by navigating View Catalogue with the help of view matching algorithm presented in Tab. 1. It navigates

through each catalogue entry and tries to find the longest match view name stored in view store and returns the location of view store where match is found. Partial intermediate results are also used for evaluation of future queries. This algorithm applies commutative rule for natural join.

View Manager maintains generated views in View Store and also updates View Catalogue. It also keeps track of view reference count as well as read/write timestamp values for every view. These values are used for decision making of view deletions from View Store and corresponding updates in View Catalogue.

Query Evaluator performs the query evaluation task after substituting matching views.

View Store materializes views those are generated from previously evaluated queries.

View Catalogue maintains the catalogue of view names those are materialized in view store. View name matched by View Manager from View Catalogue mapped into View Store.

Table 1 Algorithm for View Matching

	Algorithm ViewMatching
	Input : - Catalog of View names - Viewpattern to be searched **Output**: blocation (Location of matched view name as best location)
1	Begin
2	bvlength ← 0; //initialize Length of longest matched Viewname as best length to 0
3	For each catalogue entry i do
4	Viewname ← catalogue[i].viewname;
5	If (Viewname is substring of Viewpattern) OR (Reverse(Viewname) is substring of Viewpattern)
6	If (length(Viewname) > bvlength) //select the longest //viewname as best viewname
7	bViewname ← Viewname;
8	bViewlength ← length(Viewname);
9	blocation ← catalogue[i].location
10	If length(Viewname)=length(Viewpattern)) //Exact match //found
11	break For; //Terminate searching
12	End If
13	End If
14	End If
15	End For
16	Return blocation;
17	End

3.2 Cost Model

In order to measure the performance of proposed query optimization technique, a cost model has been developed based on various parameters presented in Tab. 2.

Let Q be the query from workload has been divided into n subqueries $\{Q_1, Q_2, \ldots, Q_n\}$ during evaluation and trying to match with k number of materialized views $\{V_1, V_2, \ldots, V_k\}$ in View Store.

Total execution cost of query Q includes execution time, memory requirement, *IO* cost as well as total view matching time from View Store during evaluation of query Q, as shown in Eq. (1). Total view matching time for query Q is calculated as in Eq. (2).

$$CE(Q) \approx \left(\sum_{i=1}^{n} ET(Q_i) + M(Q_i) + IO(Q_i) \right) + VMT(Q) \quad (1)$$

$$VMT(Q) = \sum_{i=1}^{n} VMT(Q_i) \quad (2)$$

where maximum and minimum view matching time from View Store of any subquery Q_i is shown in Eq. (3) and (4) respectively.

Table 2 List of Parameters and symbols used in the cost model

Parameter/ Symbol	Purpose
N	Number of subqueries in query Q
k	Number of materialized views
$CE(Q)$	Total execution cost of query Q
$ET(Q_i)$	Execution time for query Qi, where i = 1,..., n
$M(Q_i)$	Memory requirement for execution of query Q_i
$IO(Q_i)$	Input-output cost require for execution of query Q_i
$VMT(Q)$	Total view matching time require for query Q
$VMT(Q_i)$	View matching time for subquery Q_i, where i = 1,..., n
$VMT(V_j)$	Matching time require for particular View V_j, where j = 1,…, k
$Size(r)$	Size of relation r
$EET(Q_i)$	Execution time require for subquery Q_i
$ETMV(Q_i)$	Execution time of subquery Q_i using materialized view
$ETNMV(Q_i)$	Execution time of subquery Q_i without using materialized view
HR	Hit ratio which is the probability of matching view from View Store
MR	Miss ratio which is the probability of not matching any view from View Store

$$Max\left(VMT(Q_i)\right) = VMT(V_i) \times k \quad (3)$$

$$Min\left(VMT(Q_i)\right) = VMT(V_j) \quad (4)$$

Memory requirement $M(Q_i)$ and IO cost $IO(Q_i)$ for execution of query Q_i are directly proportional to the size of relations r those are involved in query evaluation, shown in Eq. (5) and (6). However processing and *IO* cost will get reduced when view is matched from View Store.

$$M(Q_i) \propto Size(r) \quad (5)$$

$$IO(Q_i) \propto Size(r) \quad (6)$$

Total effective execution time of query Q using materialized view is calculated as in Eq. (7), where effective execution time of subquery Q_i is as shown in Eq. (8).

$$EET(Q) = \sum_{i=1}^{n} EET(Q_i) \quad (7)$$

$$EET(Q_i) = HR \times ETMV(Q_i) + MR \times ETNMV(Q_i) \quad (8)$$

4 EXPERIMENTAL SETUP AND RESULTS

It is crucial to evaluate the performance of proposed algorithm on real infrastructure of large-scale database system. Hence, to ensure the large scale database system infrastructure, the simulator model for query optimization

has been developed in Java which run on the top of MySQL to compare the results of the proposed technique with existing approaches.

4.1 Simulation

The simulation model contains various modules.

Frequent Query Join Holder module maintains a data structure for catalogue where it makes entries of views which are cached. Before executing any query this catalogue is to be searched. The view matching algorithm has been implemented on the top of Zql parser that parses SQL statements for matching views. If any view matches then resulting entry is searched in actual view store.

Frequent Query Join module is responsible for managing view store. It stores results of previously executed queries as views. If any view is not in use for a long period of time then it deletes it and also deletes its entry from catalogue. Hence view searching time will not dominate query execution time.

DBMS Query Runner module is responsible for execution of queries with the help of Zql Parser [45]. Zql parser parses SQL statements and generates Java structures signifying query statements and expressions. Zql parser takes SQL statements with insert, delete, update, select etc. as an input and produces data structure that represent the statements it parsed.

Example:

Select ps_supplycost, pname from part, partsupp, supplier, nation where p_partkey=ps_partkkey and ps_suppkey=s_suppkey and s_nationkey=n_nationkey and n_name='India';

Zql parser extract various parts of this query and generates ZqlQuery structure. The methods getSelect(), getFrom() and getWhere() returns Select, From and Where parts of the query in their respective data structures of Java.

Result Processor module analyses results using various parameters such as total execution time of queries, average runtime per query and number of iterations to be performed.

4.2 Workload

The proposed query optimizer has been tested by generating various workloads using the standard benchmark dataset of TPC-H [46]. The TPC-H database consist of eight individual relations as shown in Tab. 3, which represents data to exercise functionalities of complex analysis application system. The testing has been performed as small scale level, after distributing these benchmark relations randomly over two nodes after the horizontal fragmentation.

TPC-H benchmark comprised of set of 22 original queries to give realistic context that represent the activity of wholesale supplier. With the variant of these 22 base queries, the workload of 50 queries is generated to test the performance of proposed technique. In order to test the effectiveness of proposed system, four variations of workload of queries are generated as shown in Tab. 4. In each query, the selection predicates are generated at random with the help of proper range of probable values as per the benchmark.

Table 3 List of relations in TPC-H database

Sr. No.	Relation/Table Name	Number of records
1	Part	1,89,945
2	Supplier	10,000
3	Partsupp	7,43,870
4	Customer	1,37,542
5	Nation	25
6	Region	5
7	Orders	14,53,561
8	Lineitem	10,59,450

Table 4 Variations in Workload of Queries

Workload	Queries with features
W_1	Queries in which none of them having common sub-expressions in their predicates
W_2	Queries in which some of them having common sub-expressions in their predicates
W_3	Set of repeated queries which are not having common sub-expressions
W_4	Set of repeated queries having common sub-expressions

By using the cost model presented in section 3.2, efficiency of proposed query optimizer has been measured in terms of various performance parameters such as total time for query execution, average runtime per query and number of iterations.

4.3 Results and Discussion

The performance of proposed query optimization technique is analysed and compared with conventional SQL system w.r.t. performance parameters viz. total execution time for workload of queries, average runtime per query and number of iterations required. Extensive experiments have been conducted with the help of all variations of query workloads specified in Tab. 4 and results are compared with conventional SQL system as presented in Tab. 5.

Table 5 Result analysis of Conventional SQL system vs proposed system

Workload	Parameter	Conventional	Proposed
W_1	Average time per query (Sec)	421.79	422.25
	Total time (Minutes)	70.30	70.37
	Number of Iterations	5305870	5306045
W_2	Average time per query (Sec)	336.01	310.25
	Total time (Minutes)	61.60	56.88
	Number of Iterations	4090545	4090765
W_3	Average time per query (Sec)	396.69	111.70
	Total time (Minutes)	132.23	37.23
	Number of Iterations	10611740	5306320
W_4	Average time per query (Sec)	265.09	91.77
	Total time (Minutes)	88.36	30.59
	Number of Iterations	9039020	3078470

As workload W_1 contains queries without any repetition and/or common subexpression, it cannot take benefit of materialization. Hence, proposed system incurs extra overhead of matching views from view store. However, we observed the decrease in running time for queries in workloads W_2 to W_4 that are matched and recycled intermediate views from view store. Fig. 2 shows the

comparative chart on average execution time per query using conventional SQL system and proposed query optimization technique.

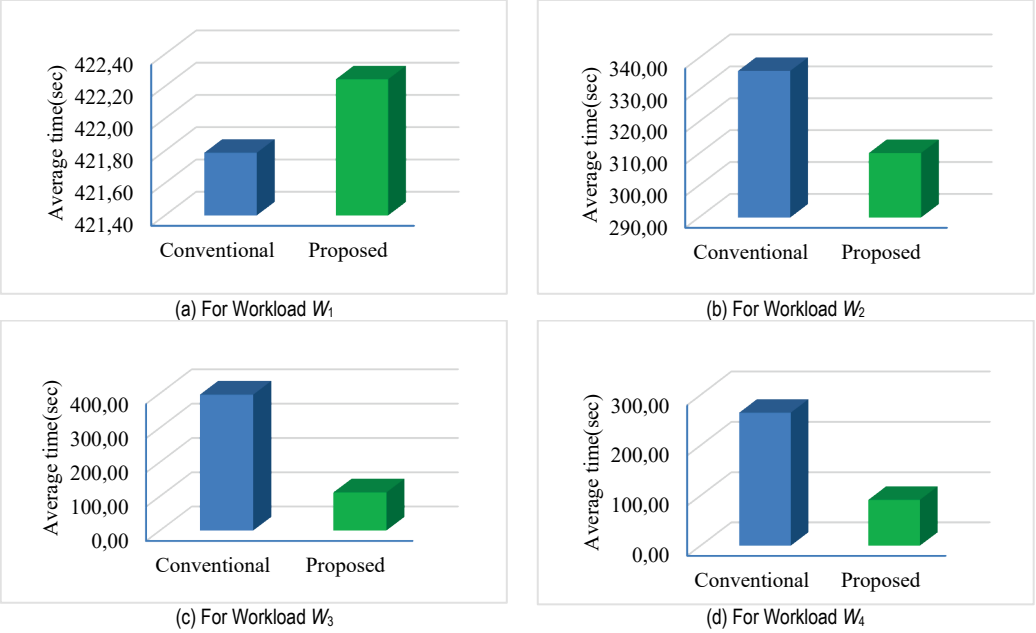


Figure 2 Comparative study on average runtime per query

5 CONCLUSION AND FUTURE WORK

In this paper the novel query optimization technique with architecture of query optimizer has been presented, which aims at recycling intermediate results of previously executed queries for execution of future queries. The queries those contain common subexpressions or repeated, can use these materialized views for their execution so that it results into decrease in the average execution time per query and increase in performance of the system. Hence using this technique resource utilization in cloud environment will also get improved. The proposed system has been tested on small scale infrastructure and it is observed that query execution time is reduced to approximately 30% as compared to the conventional query processing system due to materialization of intermediate views. As a future work, the proposed system will be evaluated on a large scale infrastructure in cloud environment to test its efficiency.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

6 REFERENCES

[1] Hacigümüs, H., Tatemura, J., Hsiung, W. P., Moon, H. J., Po, O., Sawires, A., et al. (2010, July). CloudDB: One size fits all revived. In *2010 6th World Congress on Services* (pp. 148-149). IEEE. https://doi.org/10.1109/SERVICES.2010.96

[2] Abadi, D. J. (2009). Data management in the cloud: Limitations and opportunities. *IEEE Data Eng. Bull.*, 32(1), 3-12.

[3] Zhou, L., He, K., Sheng, X., & Wang, B. (2013). A survey of data management system for cloud computing: models and searching methods. *Research Journal of Applied Sciences, Engineering and Technology*, 6(2), 244-248. https://doi.org/10.19026/rjaset.6.4064

[4] Dokeroglu, T., Sert, S. A., & Cinar, M. S. (2014). Evolutionary multiobjective query workload optimization of Cloud data warehouses. *The Scientific World Journal*, 2014. https://doi.org/10.1155/2014/435254

[5] Doshi, P., & Raisinghani, V. (2011, April). Review of dynamic query optimization strategies in distributed database. In *2011 3rd International Conference on Electronics Computer Technology* (Vol. 6, pp. 145-149). IEEE. https://doi.org/10.1109/ICECTECH.2011.5942069

[6] Bachhav, A., Kharat, V., & Shelar, M. (2017). Query optimization for databases in cloud environment: a survey. *International Journal of Database Theory and Application, 10*(6), 1-12. https://doi.org/10.14257/ijdta.2017.10.6.01

[7] Bachhav, A., Kharat, V., & Shelar, M. (2018). Novel Architecture of an Intelligent Query Optimizer for Distributed Database in Cloud Environment. *Journal of Advanced Database Management & Systems*, 5(2), 28-32.

[8] Bachhav, A., Kharat, V., & Shelar, M. (2018). Processing Distributed Internet of Things Data with Query Optimization in Cloud. *International Journal of Research and Analytical Reviews (IJRAR), 6*(1), 122-124 https://doi.org/10.14257/ijdta.2017.10.6.01

[9] Lee, R., Zhou, M., & Liao, H. (2007, September). Request Window: an approach to improve throughput of RDBMS-based data integration system by utilizing data sharing across concurrent distributed queries. In *Proceedings of the 33rd International Conference on Very Large Data Bases* (pp. 1219-1230).

[10] Chen, G., Wu, Y., Liu, J., Yang, G., & Zheng, W. (2011). Optimization of sub-query processing in distributed data integration systems. *Journal of Network and Computer Applications, 34*(4), 1035-1042. https://doi.org/10.1016/j.jnca.2010.06.007

[11] Garofalakis, M. N. & Ioannidis, Y. E. (1996). Multi-dimensional resource scheduling for parallel queries. *ACM SIGMOD Record, 25*(2), 365-376. https://doi.org/10.1145/235968.233352

[12] Dokeroglu, T., Bayir, M. A., & Cosar, A. (2015). Robust heuristic algorithms for exploiting the common tasks of relational cloud database queries. *Applied Soft Computing*, 30, 72-82. https://doi.org/10.1016/j.asoc.2015.01.026

[13] Giannikis, G., Alonso, G., & Kossmann, D. (2012). SharedDB: killing one thousand queries with one stone. In *Proceedings of the VLDB Endowment, 5*(6), 526-537. https://doi.org/10.14778/2168651.2168654

[14] Agrawal, P., Kifer, D., & Olston, C. (2008). Scheduling shared scans of large data files. *VLDB Endowment, ACM, 2008*. https://doi.org/10.14778/1453856.1453960

[15] Bruno, N., Jain, S., & Zhou, J. (2013). Continuous cloud-scale query optimization and processing. In *Proceedings of the VLDB Endowment, 6*(11), 961-972. https://doi.org/10.14778/2536222.2536223

[16] Cole, R. L. & Graefe, G. (1994, May). Optimization of dynamic query evaluation plans. In *Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data* (pp. 150-160). https://doi.org/10.1145/191843.191872

[17] Graefe, G. & Ward, K. (1989, June). Dynamic query evaluation plans. In *Proceedings of the 1989 ACM SIGMOD international conference on Management of data* (pp. 358-366). https://doi.org/10.1145/66926.66960

[18] Kabra, N. & DeWitt, D. J. (1998, June). Efficient mid-query re-optimization of sub-optimal query execution plans. In *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data* (pp. 106-117). https://doi.org/10.1145/276305.276315

[19] Markl, V., Raman, V., Simmen, D., Lohman, G., Pirahesh, H., & Cilimdzic, M. (2004, June). Robust query processing through progressive optimization. In *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data* (pp. 659-670). https://doi.org/10.1145/1007568.1007642

[20] Stillger, M., Lohman, G. M., Markl, V., & Kandil, M. (2001, September). LEO-DB2's learning optimizer. In *Proceedings of the 27th VLDB Conference* (pp. 19-28).

[21] Wang, C., Arani, Z., Gruenwald, L., & d'Orazio, L. (2018, December). Adaptive Time, Monetary Cost Aware Query Optimization on Cloud Database Systems. In *2018 IEEE International Conference on Big Data* (pp. 3374-3382). IEEE. https://doi.org/10.1109/BigData.2018.8622401

[22] Roy, P., Seshadri, S., Sudarshan, S., & Bhobe, S. (2000, May). Efficient and extensible algorithms for multi query optimization. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data* (pp. 249-260). https://doi.org/10.1145/335191.335419

[23] Safaeei, A. A., Kamali, M., Haghjoo, M. S., & Izadi, K. (2007, May). Caching intermediate results for multiple-query optimization. In *2007 IEEE/ACS International Conference on Computer Systems and Applications* (pp. 412-415). IEEE. https://doi.org/10.1109/AICCSA.2007.370914

[24] Laptev, N., Zeng, K., & Zaniolo, C. (2012). Early accurate results for advanced analytics on mapreduce. In *Proceedings of the VLDB Endowment, 5*(10), 1028-1039. https://doi.org/10.14778/2336664.2336675

[25] Agarwal, S., Iyer, A. P., Panda, A., Madden, S., Mozafari, B., & Stoica, I. (2012). Blink and it's done: interactive queries on very large data. In *Proceedings of the VLDB Endowment, 5*(12), 1902-1905. https://doi.org/10.14778/2367502.2367533

[26] Engle, C., Lupher, A., Xin, R., Zaharia, M., Franklin, M. J., Shenker, S., & Stoica, I. (2012, May). Shark: fast data analysis using coarse-grained distributed memory. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (pp. 689-692). https://doi.org/10.1145/2213836.2213934

[27] Theeten, B. & Janssens, N. (2015). Chive: Bandwidth optimized continuous querying in distributed clouds. *IEEE Transactions on cloud computing*, 3(2), 219-232. https://doi.org/10.1109/TCC.2015.2424868

[28] Olston, C., Chiou, G., Chitnis, L., Liu, F., Han, Y., Larsson, M., et al. (2011, June). Nova: continuous pig/hadoop workflows. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data* (pp. 1081-1090). https://doi.org/10.1145/1989323.1989439

[29] Logothetis, D., Olston, C., Reed, B., Webb, K. C., & Yocum, K. (2010, June). Stateful bulk processing for incremental analytics. In *Proceedings of the 1st ACM symposium on Cloud computing* (pp. 51-62). https://doi.org/10.1145/1807128.1807138

[30] Lang, W., Nehme, R. V., & Rae, I. (2015). Database Optimization in the Cloud: Where Costs, Partial Results, and Consumer Choice Meet. In *7th Biennial Conference on Innovative Data Systems Research (CIDR '15)* (pp. 1-8).

[31] Goldstein, J. & Larson, P. Å. (2001). Optimizing queries using materialized views: a practical, scalable solution. *ACM SIGMOD Record, 30*(2), 331-342. https://doi.org/10.1145/376284.375706

[32] Scheuermann, P., Shim, J., & Vingralek, R. (1996). Watchman: A data warehouse intelligent cache manager. In *Proceedings of the 22nd VLDB Conference* (pp. 1-12).

[33] Kotidis, Y. & Roussopoulos, N. (1999). DynaMat: a dynamic view management system for data warehouses. *ACM Sigmod Record, 28*(2), 371-382. https://doi.org/10.1145/304181.304215

[34] Ivanova, M. G., Kersten, M. L., Nes, N. J., & Gonçalves, R. A. (2010). An architecture for recycling intermediates in a column-store. *ACM Transactions on Database Systems (TODS), 35*(4), 1-43. https://doi.org/10.1145/1862919.1862921

[35] Elghandour, I., & Aboulnaga, A. (2012, May). ReStore: reusing results of MapReduce jobs in pig. *In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (pp. 701-704). https://doi.org/10.1145/2213836.2213937

[36] Olston, C., Reed, B., Srivastava, U., Kumar, R., & Tomkins, A. (2008, June). Pig latin: a not-so-foreign language for data processing. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data* (pp. 1099-1110). https://doi.org/10.1145/1376616.1376726

[37] Lim, H., Herodotou, H., & Babu, S. (2012). Stubby: A transformation-based optimizer for mapreduce workflows. In *Proceedings of the VLDB Endowment, 5*(11), 1196-1207. https://doi.org/10.14778/2350229.2350239

[38] Perez, L. L., et al. (2014, March). History-aware query optimization with materialized intermediate views. In *2014 IEEE 30th International Conference on Data Engineering* (pp. 520-531). IEEE. https://doi.org/10.1109/ICDE.2014.6816678

[39] Tan, K. L., Goh, S. T., & Ooi, B. C. (2001, April). Cache-on-demand: Recycling with certainty. In *Proceedings 17th International Conference on Data Engineering* (pp. 633-640). IEEE.

[40] Phan, T. & Li, W. S. (2008, April). Dynamic materialization of query views for data warehouse workloads. In *2008 IEEE 24th*

International Conference on Data Engineering (pp. 436-445). IEEE. https://doi.org/10.1109/ICDE.2008.4497452

[41] Kossmann, D., Franklin, M. J., Drasch, G., & Ag, W. (2000). Cache investment: integrating query optimization and distributed data placement. *ACM Transactions on Database Systems (TODS), 25*(4), 517-558. https://doi.org/10.1145/377674.377677

[42] Anyanwu, K., Kim, H., & Ravindra, P. (2012). Algebraic optimization for processing graph pattern queries in the cloud. *IEEE Internet Computing, 17*(2), 52-61. https://doi.org/10.1109/MIC.2012.22

[43] Wu, S., Li, F., Mehrotra, S., & Ooi, B. C. (2011, October). Query optimization for massively parallel data processing. In *Proceedings of the 2nd ACM Symposium on Cloud Computing* (pp. 1-13). https://doi.org/10.1145/2038916.2038928

[44] Condie, T., Conway, N., Alvaro, P., Hellerstein, J. M., Elmeleegy, K., & Sears, R. (2010, April). MapReduce online. In *Nsdi, 10*(4).

[45] See http://zql.sourceforge.net/

[46] See http://www.tpc.org/tpch/

Authors' contacts:

Archana Bachhav, Assistant Professor
(Corresponding author)
Department of Computer Science,
KSKW Arts, Science and Commerce College,
Jawahar Road Trimbykeshwar Ta: Tryambakeshwar,
Nashik, Maharashtra 422212, India
Savitribai Phule Pune University,
Ganeshkhind Road, Pune, Maharashtra 411007, India
77.archana@gmail.com

Vilas Kharat, Senior Professor
School of Mathematical and Computing Sciences,
Savitribai Phule Pune University,
Pune, India
laddoo1@yahoo.com

Madhukar Shelar, Associate Professor
Department of Computer Science,
KRT Arts, BH Commerce and AM Science (KTHM) College,
Nashik, India
mnshelar70@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210205101347

Preliminary communication

Ensemble Machine Learning Approaches for Detection of SQL Injection Attack

Umar Farooq

Abstract: In the current era, SQL Injection Attack is a serious threat to the security of the ongoing cyber world particularly for many web applications that reside over the internet. Many webpages accept the sensitive information (e.g. username, passwords, bank details, etc.) from the users and store this information in the database that also resides over the internet. Despite the fact that this online database has much importance for remotely accessing the information by various business purposes but attackers can gain unrestricted access to these online databases or bypass authentication procedures with the help of SQL Injection Attack. This attack results in great damage and variation to database and has been ranked as the topmost security risk by OWASP TOP 10. Considering the trouble of distinguishing unknown attacks by the current principle coordinating technique, a strategy for SQL injection detection dependent on Machine Learning is proposed. Our motive is to detect this attack by splitting the queries into their corresponding tokens with the help of tokenization and then applying our algorithms over the tokenized dataset. We used four Ensemble Machine Learning algorithms: Gradient Boosting Machine (GBM), Adaptive Boosting (AdaBoost), Extended Gradient Boosting Machine (XGBM), and Light Gradient Boosting Machine (LGBM). The results yielded by our models are near to perfection with error rate being almost negligible. The best results are yielded by LGBM with an accuracy of 0.993371, and precision, recall, f1 as 0.993373, 0.993371, and 0.993370, respectively. The LGBM also yielded less error rate with False Positive Rate (FPR) and Root Mean Squared Error (RMSE) to be 0.120761 and 0.007, respectively. The worst results are yielded by AdaBoost with an accuracy of 0.991098, and precision, recall, f1 as 0.990733, 0.989175, and 0.989942, respectively. The AdaBoost also yielded high False Positive Rate (FPR) to be 0.009.

Keywords: Boosting; ensemble learning; Light GBM; SQL injection; web security

1 INTRODUCTION

A Web Application is software that uses internet connected web browsers and has gained high importance for performing different tasks in social, commercial, academic, and other platforms. These web applications are connected to back-end relational databases operated by Structured Query Language (SQL) that hold a huge amount of information like usernames, passwords, bank details, etc., and are used for communication, online transactions, data storage, accessing social networks, etc. Despite all the importance of these web applications it provides a way for hackers and crackers to attack these databases. Securing the web data must be of the utter importance for developers of these web applications.

Almost 98% of web applications are prone to various attacks but the top most one is SQL Injection attack as is listed as number one in the top ten web application security risks by Open Web Application Security Project (OWASP) [1, 2]. This attack has been listed in top ten vulnerabilities by OWASP from last fifteen years [3]. Refined software and other tools are also used nowadays to perform injection attacks controlled by machines [4].

SQL injection is an exploitation technique that compromises the security at database layer of a web application. This vulnerability usually occurs due to insufficient validation of inputs and directly including them in a SQL query. By utilizing these vulnerabilities, an attacker can submit SQL queries legitimately to the database. Generally, any web application is prone to SQL injection attack when any of the following vulnerabilities are present in the web application:

- When filtration, validation, and sanitization of input data from the user is not applied by the web application.
- When the dynamic queries or non-defined calls are given directly to the interpreter.
- When hostile data is used to retrieve sensitive data from the database or dynamic query is concatenated with both hostile data and structure [5].

SQL injection attacks are classified into seven categories: tautologies, illegal/logically incorrect queries, piggy-backed queries, stored queries, inference and alternate encodings [6]. In SQL injection a malicious script is being embedded into a less secure web application through an entry node then bypassed to the back-end database. This script then forces the web application to produce results from the database through queries that shouldn't be executed normally or ever. Using this attack, an attacker can get all the data from the database by bypassing the authentication and authorization of the web application.

SQL injection is a code injection technique that can provide the attacker with an unauthorized access to the sensitive information in the database. It not only gets the unrestricted access but it can also be utilized to disturb data integrity by adding, deleting, or modifying the records in a database. SQL injection attack is primarily focused on exploiting vulnerability in the security of a web application that is when the user input is not correctly validated or filtered, and when user input is not typed strongly and executed unexpectedly. It also occurs when there is weakness in the code, programming language. It is an attack vector for web applications but also can be used to attack any kind of SQL database. Hackers can gain unauthorized access to underlying data, structure, and DBMS. The well understood example of SQL injection attack is tautological one, "SELECT * FROM Users WHERE User-id = 1 or 1=1", where the injection happens due to the true condition using OR. Attackers nowadays use other ways to perform mass SQL injection attacks such as refined tools or botnets for discovering of vulnerable sites [3].

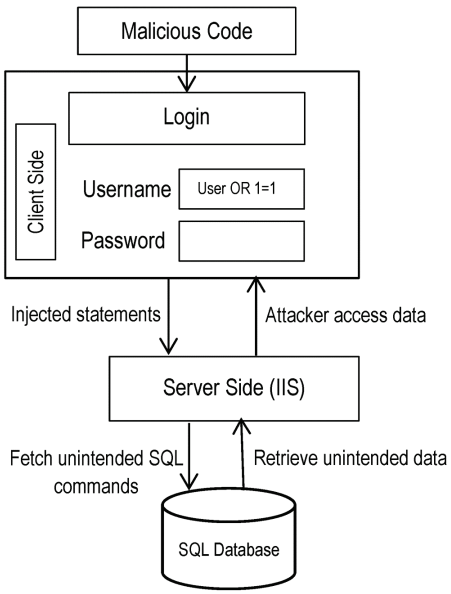


Figure 1 Typical SQL Injection Attack

2 BACKGROUND

In this section, we will briefly mention out all the ten types of SQL injection attack.

2.1 Tautologies

The attacker uses a conditional query wherein the 'WHERE' clause is used to inject and make the condition a tautology that always happens to be true. In example "SELECT * FROM Users WHERE User-id = 1 or 1=1", the query will result all the data in the database the condition of WHERE clause is true. This can be secured by restricting the users to input special characters like single quotes, double quotes, equality, and other symbols that are used to make the malicious queries [7].

Example: SELECT * FROM accountTable WHERE user login= or 1=1

2.2 Piggy-Backed Query

This type is used to retrieve data, modify database, execute commands and perform Denial of Services (DOS) attack. In this attack, attacker tries to inject other malicious queries along with the normal/original query. The original query is true and executed normally while as additional malicious queries are injected without checking. This can be secured by avoiding execution of multiple statements and checking for delimiter in all queries [7].

Example: SELECT * FROM accountTable WHERE user login=umar AND passwd=; drop accountTable user – AND pin=221

2.3 Union Query

This type is used to bypass authentication and extract all data from the database. In this attack, attacker inserts a UNION query into parameter that happens to be weak hence vulnerable. This can be secured by verifying the user inputs strictly and avoid execution of multiple queries on the side of database [7].

Example: SELECT * FROM accountTable WHERE user login= UNION SELECT * FROM accountTable WHERE No=10232 – AND passwd = AND pin=

2.4 Stored Procedures

This type is used to execute remote commands, perform DOS, and for privilege escalation. In this attack, the attacker uses delimiter ";" and stored procedure keywords such as "EXEC", "SHUTDOWN", etc. This can be secured by verifying the user input with a low privileged account for execution and executing stored procedures within a safe interface with appropriate roles [7].

Example: SELECT * FROM accountTable WHERE user login= 'umar' AND passwd = 'farooq'; SHUTDOWN;– ;

2.5 Illegal/Logically Incorrect Queries

This type is used to detect such parameters that are vulnerable to injection and then extract data from the identified database. In this attack, attacker tries to extract all information about database and structure. This can be secured by verifying inputs from user and avoiding the generation of error messages from database [7].

Example: SELECT * FROM accountTable WHERE user login= 'umar"' AND passwd =

2.6 Inference

This type is used to detect such parameters that are vulnerable to injection and then extract data from the database with schema identified. This attack is launched on secured databases and is of two types: Inference blind SQL injection and Inference time SQL injection [7].

Example: 1; IF SYSTEM_USER='sa' SELECT 1/0 ELSE SELECT 5

2.7 Alternate Coding

This type is used to escape from being detected. In this attack, attacker injects encoded text to bypass detection techniques with the help of signatures like EXEC (), Char (), ASCII (), BIN (), HEX (), UNHEX (), BASE64 (), DEC (), ROT13 (), etc. This can be secured by verifying user inputs and prohibition of meta-characters [7].

Example: SELECT * FROM accountTable WHERE user login= 'umar';exec(char(0x59842 352646f776e)) AND passwd ='farooq' AND pin =; SHUTDOWN;–;

2.8 End of Line Comment

SELECT * FROM Accounts WHERE accountName = _admin'--_AND password = _'

This statement logs the hacker as admin user [8].

2.9 Blind Injection

This type is used for asking Boolean (true/false) questions and the information is extracted depending upon the behavior of the web page. The web page functions normally if the injection attack is true, otherwise the web page functions differently [8].

2.10 Timings Attacks

This type is used to derive information with the help of If-Then statements where the attacker notes the timing delays of responses from the database [8].

Generally, SQL injection attack is divided into three types depending upon the mode of transfer of incoming and outgoing data. The three types are in-band, out-of-band, and inferential [9]. In in-band SQL injection attack, the attacker extracts the information from the same channel that is used for sending the query or performing the attack. In out-of-band SQL injection attack, the attacker extracts the information with the help of another channel like email. In inferential SQL injection attack, the attacker does not extract the information using any channels rather launches other attacks to analyze the behavior of the web application.

3 RELATED WORK

Multiple studies and researches have been carried out so far on the field of SQL injection and it's detection by using various approaches like static & dynamic analysis, combined technique, machine learning, Hash technique, Black Box testing, etc. [10].

Static analysis checks whether each stream from a source to a sink is dependent upon an info approval and additionally input purifying routine [11]; though dynamic analysis depends on progressively mining the developer's planned query structure on any information and recognizes assaults by contrasting it against the structure of the real given query [12].

AMNESIA, as a consolidated methodology, is a model-based method that consolidates the static and dynamic analysis for detection and prevention of SQL injection attacks. It uses static analysis in order to make the SQL query models at the time of accessing the database. It then uses dynamic analysis before the queries are sent to database and compares them with the already built statically models [10]. But there are some queries and code snippets generation approaches that make this model less efficient with more error rate [13].

Hidden Markov Model (HMM) has been presented to detect malicious queries with the help of machine learning in two phases: training and running phase. The first phase focuses on collecting known malicious and benign queries and the second phase focuses on detecting injection attacks. Author, by himself, cleared that WHERE clause and piggybacked queries cannot be detected by this model [4].

Detection of SQL injection attack based on Naïve Bayes machine learning algorithm was proposed combined with the mechanism of role-based access [14]. The detection rate with this model is 93%, however future attack cannot be detected with this data and the classifier relies on the labeled data.

4 METHODOLOGY

The main motive of the proposed model is to detect SQL Injection attack. The whole procedure is performed in four stages:

1) The first stage focuses on collecting the dataset that contains proper SQL injection attack queries. For this issue, we created a dataset that contains SQL queries, SQL injection attack queries, and plain text. The labelling of the dataset is done in this stage.
2) The second stage deals with extracting all the features from all the queries and selecting the best of them (a.k.a. Feature extraction and feature selection). Tokenization is used in this stage to divide the queries into tokens.
3) The third stage deals with training the model. The model is trained in this phase with 70% of the dataset (a.k.a. Training part).
4) The fourth stage is focused on using the 30% of dataset that we separated from the collected dataset for testing and evaluating the proposed model with the selected best feature set (a.k.a. Testing part).

4.1 Dataset

The most important part in detecting a SQL injection attack is collecting a meaningful dataset that contains SQL injection attack queries. The main contribution in this paper is a labelled dataset that we manually collected for the said problem. The dataset not only contains SQL injection attack queries but also normal SQL injection queries and plain text queries so that the proposed model will properly comprehend and differentiate between normal and attacking SQL queries. The dataset is collected in three phases: 1) the normal SQL injection queries are collected in first phase, 2) the SQL injection attack queries are collected in the second phase, and 3) the plain text is collected in the third phase. We collected these queries in the text format and applied labelling and preprocessing methods on it and then converted it to a csv file. We applied tokenization on the dataset and formed a new tokenized dataset. The dataset contains a total of 35198 queries with 21 features. The dataset has the following three categories:

4.1.1 Non-Malicious or Normal SQL Queries

These queries, non-malicious in nature, are used to create, maintain, and retrieve database in the form of tables (relational database). The tokens (keywords) used in this type are: (rename, drop, delete, insert, create, exec, update, union, set, Alter, database, and, or, information_schema, load_file, select, shutdown, cmdshell, hex, ascii). Also the dangerous characters used in this type are: --, #, /*, ', ", ||, \\, =, /**/,@@.

4.1.2 SQL Injection Attack Queries/Malicious SQL Queries

These queries are used to execute malicious SQL statements in a web application and bypass the security measures. These queries are also used to add, modify, and delete records in a database in an unrestricted way. The tokens (keywords) used in this type are: , *, ; , _, -, (,), =, {, }, @, ., , &, [,], +, -, ?, %, !, :, \, /. Also the SQL tokens used are: where, table, like, select, update, and, or, set, like, in, having, values, into, alter, as, create, revoke, deny, convert, exec, concat, char, tuncat, ASCII, any, asc, desc, check, group by, order by, delete from, insert into, drop table, union, join.

4.1.3 Plain Text

These are simply in the form of plain text. The tokens (keywords) used in this type are alphabets and digits. The plain text is used in this dataset in order to make sure that the proposed model properly comprehends and differentiated between the SQL query, SQL injection query and the plain text that the user inputs in the login node of any web app.

The detailed description of the collected dataset (features) is given below in Tabs. 1 and 2.

Table 1 Description of features of dataset

S. No.	Feature	Description
1	data	It contains all the full queries
2	no_sngle_quts	Total number of single quotations in a query
3	no_dble_quts	Total number of double quotations in a query
4	no_punctn	Total number of punctuations in a query
5	no_sgle_cmnt	Total number of single line comments in a query
6	no_mlt_cmnt	Total number of multi-line comments in a query
7	no_whte_spce	Total number of white spaces in a query
8	no_nrml_kywrds	Total number of normal keywords in a query
9	no_hmfl_kywrds	Total number of harmful keywords in a query
10	no_prctge	Total number of percentage (%) symbols in a query
11	no_log_oprtr	Total number of logical operators in a query
12	no_oprtr	Total number of operators in a query
13	no_null_valus	Total number of null values in a query
14	no_hexdcml_valus	Total number of hexadecimal values in a query
15	no_db_info_cmnds	Total number of database information commands in a query
16	no_roles	Total number of roles (e.g., Admin, user, etc.) in a query
17	no_ntwr_cmnds	Total number of network commands in a query
18	no_lanage-cmnds	Total number of language commands in a query
19	no_alphabet	Total number of alphabets in a query
20	no_digits	Total number of digits in a query
21	no_spl_chrtr	Total number of special characters in a query

Table 2 Description of labels

S. No.	Label	Description	Count	Ratio
1	0	It represents the normal SQL queries	6888	19.57%
2	1	It represents the SQL injection attack queries	18369	52.19%
3	2	It represents the plain text	9941	28.24%

4.2 Tokenization

The keywords used in SQL injection attack are used to launch operations on the database tables. These keywords play an important role in launching SQL injection attack as the keywords perform the unexpected tasks. So, there is a need to differentiate these keywords form a normal and malicious query. The method of tokenization is used to perform such operation i.e., extract the tokens from the actual queries. In simple terms, tokenization is the process of dividing a query into a list of tokens (keywords). Depending upon these extracted tokens, the proposed model extracts features. Each query is represented by a sequence of numbers where each number represents one of the features represented in Tab. 1.

The suitable determination of these features plays an essential function in detection of SQL injection attack. The reasoning for picking these sorts of features is its capacity to recognize the greater part of SQIA types like redundancies/tautologies, union, piggybacked, illegal/logically incorrect, alternate encodings and stored procedures which are dealt with the same as SQL queries.

Let us take the example of **or 1=1** to understand the concept of tokenization.

By applying the tokenization to the above query, the output is given below and is in accordance with the features listed in Tab. 1:

or 1=1	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	2	2	0	1

4.3 Training Ensemble Models

The main phase is to train the machine learning algorithms for the detection of SQL injection attack with the manually collected dataset. The selected ensemble learning algorithms that we used in our proposed model are Gradient Boosting Machine (GBM), Adaptive Boosting (AdaBoost), Extended Gradient Boosting Machine (XGBM), and Light Gradient Boosting Machine (LGBM). To have a better understanding of how the machine learning models would

perform over the testing data we applied three and five-fold cross-validation where we split the dataset into 3 and 5 parts, respectively. The advantage of cross validation is that all the observations are utilized for both training and testing the models, and each observation is used for testing exactly once.

5 RESULTS AND DISCUSSION

As per the experiments that we conducted, we come to conclusion that our proposed system is enough to detect SQL injection attack queries from normal and plain text queries with 21 features. We focused on making the features as much as possible in order to make the proposed model robust and detect all types of SQL injection attack queries, efficiently. To evaluate the performance of our proposed model we applied the algorithms, ensemble boosting in nature, on the testing data (30% of the original dataset). The classification results that were evolved by the proposed model are near perfection and are depicted in the below tables and figures.

We separated the results in different tables, where in every table represents different classification metrics such as accuracy (Acc.), precision (Pr.), recall (Re.), f1 score (f1), false positive rate (FPR), root mean squared error (RMSE), mean absolute error (MAE), and mean squared error (MSE), to analyze the behavior of our system properly. The results are depicted in below Tabs. 3-14.

5.1 Classification Report

Table 3 Accuracy report of our proposed model

Accuracy			
Classifier	Partition Strategy	3-CV	5-CV
GBM	Training Set = 70% Testing Set = 30%	0.991856	0.990909
AdaBoost		0. 991098	0. 991098
XGBoost		0.992233	0.992233
Light GBM		0.993371	0.993371

Table 4 Precision report of our proposed model

Precision			
Classifier	Partition Strategy	3-CV	5-CV
GBM	Training Set = 70% Testing Set = 30%	0.991791	0.990660
AdaBoost		0.990733	0.990733
XGBoost		0.991400	0.991400
Light GBM		0.993373	0.993373

Table 5 Recall report of our proposed model

Recall			
Classifier	Partition Strategy	3-CV	5-CV
GBM	Training Set = 70% Testing Set = 30%	0.990388	0.989341
AdaBoost		0.989175	0.989175
XGBoost		0.990596	0.990596
Light GBM		0.993371	0.993371

Table 6 F1 score report of our proposed model

F1 Score			
Classifier	Partition Strategy	3-CV	5-CV
GBM	Training Set = 70% Testing Set = 30%	0.991084	0.989997
AdaBoost		0.989942	0.989942
XGBoost		0.992234	0.992234
Light GBM		0.993370	0.993370

Table 7 MAE report of our proposed model

MAE			
Classifier	Partition Strategy	3-CV	5-CV
GBM	Training Set = 70% Testing Set = 30%	0.010321	0.011590
AdaBoost		0.011553	0.011553
XGBoost		0.011742	0.011742
Light GBM		0.009280	0.009280

Table 8 MSE report of our proposed model

MSE			
Classifier	Partition Strategy	3-CV	5-CV
GBM	Training Set = 70% Testing Set = 30%	0.014678	0.016590
AdaBoost		0.016856	0.016856
XGBoost		0.017992	0.017992
Light GBM		0.014583	0.014583

Table 9 RMSE report of our proposed model

RMSE			
Classifier	Partition Strategy	3-CV	5-CV
GBM	Training Set = 70% Testing Set = 30%	0.121152	0.128805
AdaBoost		0.129830	0.129830
XGBoost		0.134135	0.134135
Light GBM		0.120761	0.120761

Table 10 FPR report of our proposed model

False Positives			
Classifier	Partition Strategy	3-CV	5-CV
GBM	Training Set = 70% Testing Set = 30%	0.008	0.009
AdaBoost		0.009	0.010
XGBoost		0.008	0.008
Light GBM		0.007	0.007

5.2 Confusion Matrix

Confusion matrix is a performance measurement for machine learning classifiers with different combinations of actual and predicted values. The above results are calculated with the help of confusion matrix that is used to evaluate the overall performance of our proposed classification system. As the problem we chose is multi-class classification with three classes (normal SQL query, SQL injection attack query, and plain text), hence the confusion matrix is 3×3. The following classification metrics are evaluated:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{(TP)}{(TP + FN)} \tag{3}$$

$$F1\,Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^{n} abs(y_i - \hat{y})}{n} \tag{5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (7)$$

$$FPR = \frac{FP}{FP + TN} \text{ or } 1 - Recall \quad (8)$$

The confusion matrix of our algorithms is given below where 0, 1, and 2 represent normal SQL queries, SQL injection attack queries, and plain text, respectively.

Table 11 Confusion matrix of AdaBoost

AdaBoost				
Predicted	Actual			
		0	1	2
	0	1966	12	21
	1	7	5473	37
	2	6	20	3018

Table 12 Confusion matrix of GBM

GBM				
Predicted	Actual			
		0	1	2
	0	2078	12	15
	1	3	5461	22
	2	8	26	2935

Table 13 Confusion matrix of XGBoost

XGBoost				
Predicted	Actual			
		0	1	2
	0	2060	21	37
	1	7	5388	41
	2	4	28	2974

Table 14 Confusion matrix of LGBM

Light GBM				
Predicted	Actual			
		0	1	2
	0	2095	6	17
	1	1	5418	17
	2	11	18	2977

The classification report of our proposed system is given in Fig. 2 wherein we represented it in graphical form.

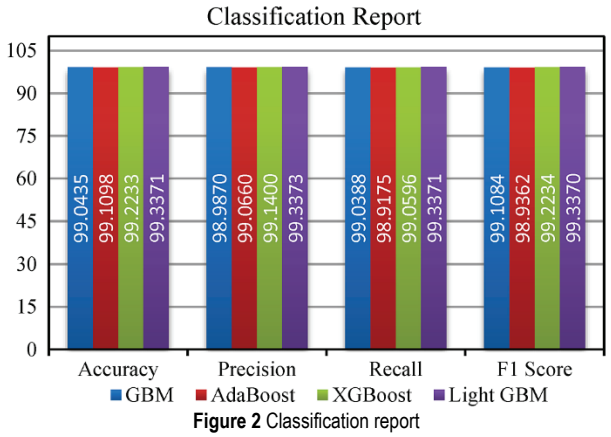


Figure 2 Classification report

The error report, in graphical form, of our proposed system is given in Fig. 3.

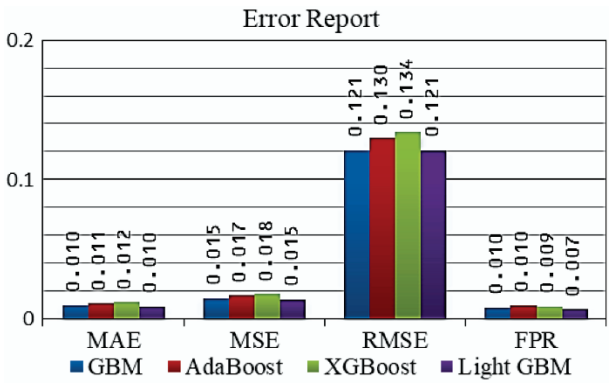


Figure 3 Error report

The classification reports evaluated by our four models are given in Fig. 4.

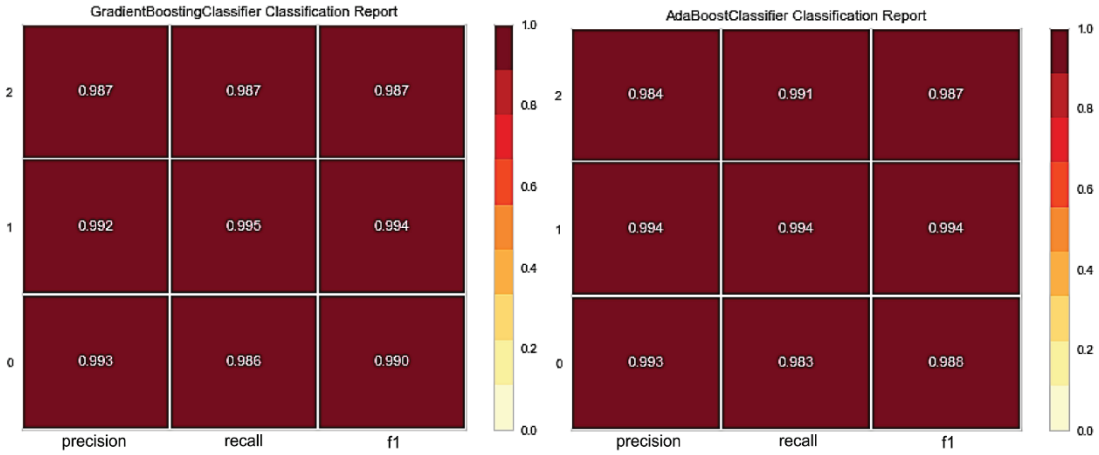


Figure 4 Classification report from GBM, AdaBoost, XGBM, and LGBM, respectively

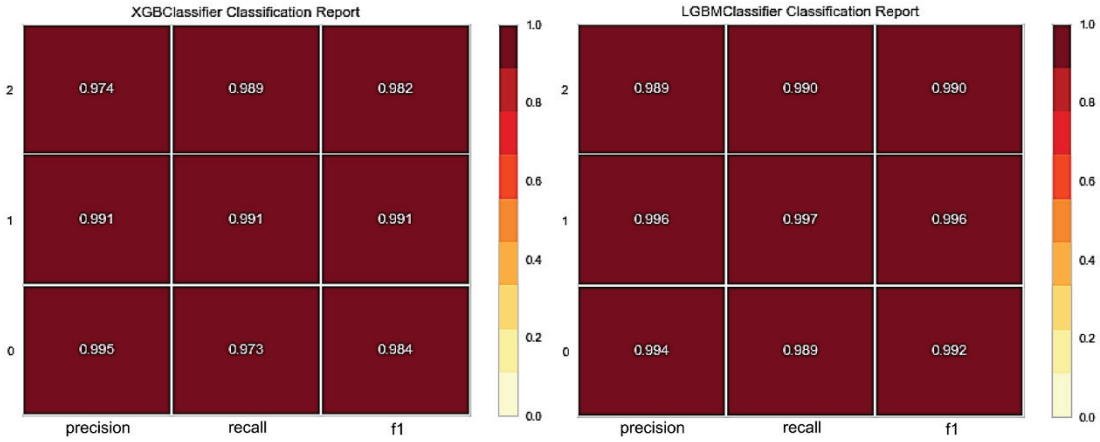


Figure 5 Classification report from GBM, AdaBoost, XGBM, and LGBM, respectively (continuation)

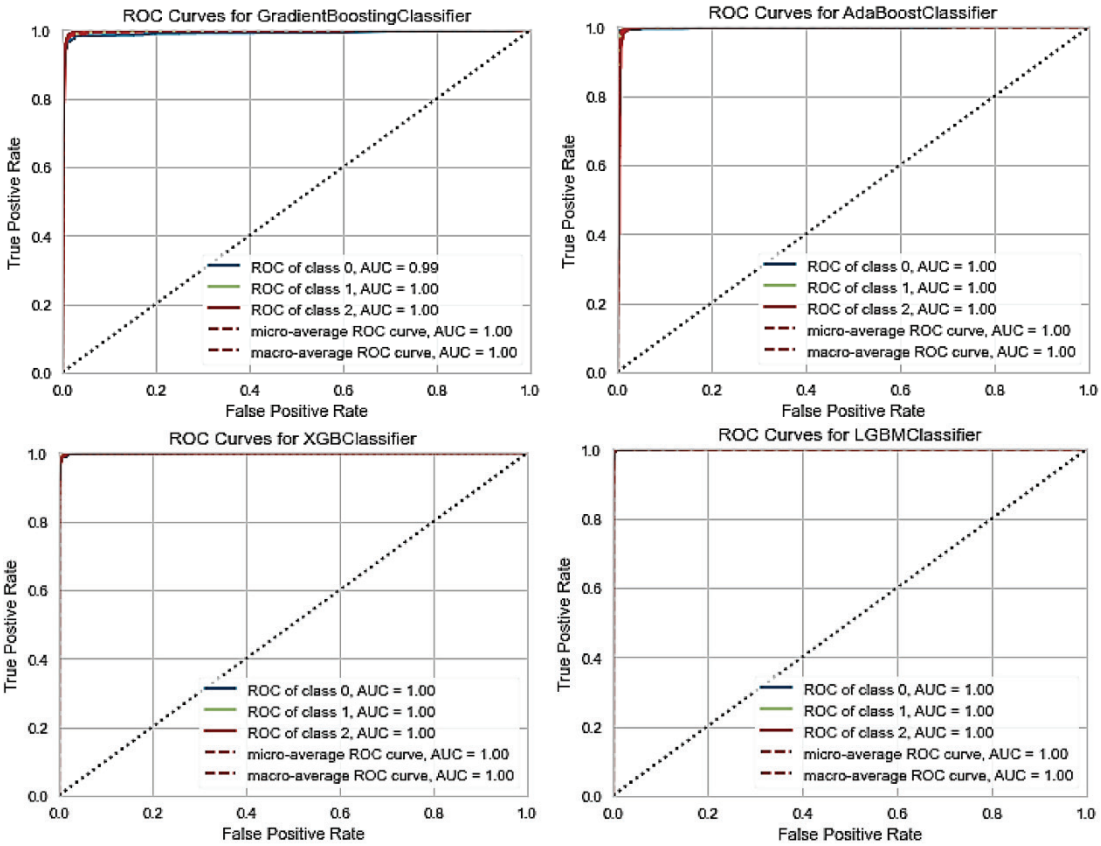


Figure 6 ROC results from GBM, AdaBoost, XGBM, and LGBM, respectively

5.3 Roc Curves

The ROC values evaluated by our algorithms are given in Tab. 15.

Table 15 ROC values of our proposed models

Algorithms	GBM	AdaBoost	XGBoost	Light GBM
ROC Value	0.995449	0.997657	0.999548	0.999845

5.4 Comparative Analysis

The comparative analysis for the research that has been made on SQL injection attack is depicted in the table below (Tab. 16) and we compared them with the proposed model in terms of accuracy. Our proposed model dominates other existing models in terms of accuracy with less error rate.

Table 16 Comparative analysis

Classifiers/Models	Accuracy
SVM, Naïve Bayes, GBM, REGEX [15]	97%
Neural Network system [16]	96.8%
Genetic- fuzzy rule-based system [17]	98.4%
SVM [18]	98%
K-means [19]	98.36%
Our Proposed model (GBM, AdaBoost, XGBM, LGBM)	99.34%

6 CONCLUSION

In this research work, we proposed SQL injection attack detection model based on 21 features in order to increase the efficiency of our classifiers. The main target of our system was particularly SQL injection attack that is increasing day by day while being used with some malicious content to gain unrestricted access to databases and extract sensitive information. These malicious queries can bypass authentication and authorization and can finally alter, modify, and delete the database. Keeping this as our objective, we proposed a robust model for detection of SQL injection attack queries from normal queries and plain text. In this work, the foremost step we carried out was to create a balanced dataset that contains normal and malicious SQL queries. We also introduced plain text to this dataset in order to make the proposed model perform well and differentiate malicious queries from normal and plain text.

The proposed model when applied to the dataset achieves an average accuracy of more than 99% with almost negligible error rate that indicates the selected feature set is quite efficient to discriminate SQL injection attack queries from normal SQL queries and plain text. For real world detection systems, the analysis indicate that our proposed system that is based on ensemble machine learning with the selected features can be applied in such SQL injection attack detection systems. The best test accuracy happens to be 99.34% with 0.007 percent FPR while as the lowest one is 99.11% with 0.009 percent FPR, yielded by LGBM and AdaBoost, respectively. The other two algorithms GBM and XGBM that we used yielded accuracy of 99.19% and 99.22%, respectively.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

7 REFERENCES

[1] OWASP. https://owasp.org/www-project-top-ten/. (Accessed on 18.11.2020).

[2] Farooq, U. (2020). Real Time Password Strength Analysis on a Web Application Using Multiple Machine Learning Approaches. *International Journal of Engineering Research & Technology (IJERT), 9*(12), 359-364.

[3] Moh, M., Pininti, S., Doddapaneni, S., & Moh, T. (2016). Detecting Web Attacks Using Multi-stage Log Analysis. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, Bhimavaram, 733-738. https://doi.org/10.1109/IACC.2016.141

[4] Kar, D., Agarwal, K., Sahoo, A., & Panigrahi, S. (2016). Detection of SQL injection attacks using Hidden Markov Model. *2016 IEEE International Conference on Engineering and Technology (ICETECH)*, Coimbatore, India. https://doi.org/10.1109/ICETECH.2016.7569180

[5] OWASP. https://owasp.org/www-project-top-ten/2017/A1_2017-Injection. (Accessed on 19.11.2020).

[6] Moosa, A. (2010). Artificial Neural Network based Web Application Firewall for SQL Injection. *World Academy of Science, Engineering and Technology, International Journal of Computer and Information Engineering, 4*(4), 610-619. https://panel.waset.org/publications/1001/pdf

[7] Sheykhkanloo, N. M. (2015). SQL-IDS: Evaluation of SQLi Attack Detection and Classification Based on Machine Learning Techniques. *The 8th International Conference on Security of Information and Networks (SIN15)*, Sochi, Russia. https://doi.org/10.1145/2799979.2800011

[8] Kaur, M. & Agrawal, A. P. (2012). Token Sequencing Approach to Prevent SQL Injection Attacks. *IOSR Journal of Computer Engineering (IOSRJCE), 1*(1), 31-37. https://doi.org/10.9790/0661-0113137

[9] Sadeghian, A., Zamani, M., & Ibrahim, S. (2013). SQL injection is still alive: a study on SQL injection signature evasion techniques. In *International Conference on Informatics and Creative Multimedia*, Kuala Lumpur, Malaysia, 265-268. https://doi.org/10.1109/ICICM.2013.52

[10] Halfond, W. G. & Orso, A. (2005). AMNESIA: analysis and monitoring for neutralizing SQL-injection attacks. In *Proceedings of the 20th IEEE/ACM International Conference on Automated Software Engineering*, 174-183. https://doi.org/10.1145/1101908.1101935

[11] Shar, L. K. & Tan, H. B. K. (2013). Defeating SQL injection. *Computer, 46*, 69-77. https://doi.org/10.1109/MC.2012.283

[12] Tajpour, A. & Shooshtar, M. J. Z. (2010). Evaluation of SQL injection detection and prevention techniques. In *Second IEEE International Conference on Computational Intelligence, Communication Systems and Networks*, Liverpool, UK, 216-221. https://doi.org/10.1109/CICSyN.2010.55

[13] Dharam, R. & Shiva, S. G. (2013). Runtime monitors to detect and prevent union query based SQL injection attacks. In *Tenth International Conference on Information Technology: New Generations*, Las Vegas, USA, 357-362. https://doi.org/10.1109/ITNG.2013.57

[14] Joshi, A. & Geetha, V. (2014). SQL Injection detection using machine learning. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, Kanyakumari, IEEE, 1111-1115. https://doi.org/10.1109/ICCICCT.2014.6993127

[15] Kranthikumar, B. & Velusamy, R. L. (2020). SQL injection detection using REGEX classifier. *Journal of Xi'an University of Architecture & Technology, 12*(6), 800-809.

[16] Sheykhkanloo, N. M. (2015). SQL-IDS: evaluation of SQLi attack detection and classification based on machine learning techniques. In *Proceedings of the 8th International Conference on Security of Information and Networks*, USA, 258-266. https://doi.org/10.1145/2799979.2800011

[17] Basta, C., Elfatatry, A., & Darwish, S. (2016). Detection of SQL Injection Using a Genetic Fuzzy Classifier System. *International Journal of Advanced Computer Science and Applications (IJACSA), 7*(6), 129-137. https://doi.org/10.14569/IJACSA.2016.070616

[18] Jagadessan, J., Shrivastava, A., Ansari, A., Kar, L. K., & Kumar, M. (2019). Detection and Prevention Approach to SQLi and Phishing Attack using Machine Learning. I*nternational Journal of Engineering and Advanced Technology (IJEAT), 8*(4), 791-799.

[19] Patel, M. P. & Sivaraman, D. B. (2017). SQL injection Detection for Secure Atomic and Molecular Database node for India. *International Journal of Advance Research and Innovative Ideas in Education (IJARIIE), 3*(2), 3867-3879.

Author's contact:

Umar Farooq,
Department of Computer Science & Technology (Cyber Security),
Central University of Punjab,
City Campus, Mansa Road, Bathinda 151001, Punjab, India
soulatf3@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210205095410

Preliminary communication

An Analytical Study of Code Smells

Lida Bamizadeh*, Binod Kumar, Ajay Kumar, Shailaja Shirwaikar

Abstract: Software development process involves developing, building and enhancing high-quality software for specific tasks and as a consequence generates considerable amount of data. This data can be managed in a systematic manner creating knowledge repositories that can be used to competitive advantage. Lesson's learned as part of the development process can also be part of the knowledge bank and can be used to advantage in subsequent projects by developers and software practitioners. Code smells are a group of symptoms which reveal that code is not good enough and requires some actions to have a cleansed code. Software metrics help to detect code smells while refactoring methods are used for removing them. Furthermore, various tools are applicable for detecting of code smells. A Code smell repository organizes all the available knowledge in the literature about code smells and related concepts. An analytical study of code smells is presented in this paper which extracts useful, actionable and indicative knowledge.

Keywords: code smells; data mining; knowledge repository; refactoring methods; software metrics

1 INTRODUCTION

Today's software development process produces large amount of data. Lesson's learned and best practices in software development process are spread out over literature in various forms such as Code smells, design patterns, idioms etc. Organizing this knowledge into a knowledge repository, extracting insights from this data and making them available to code developers and software practitioners, can assist the software development process. Code smell is a general mechanism to distinguish structural design issues in software projects [1, 2]. Code smell term was formulated by Kent Beck when helping fowler for his refactoring book and has since become an important word in software maintenance vocabulary. Existence of code smell would not interrupt the functionality of system but it would enhance the risk of decay and reduce the software quality of system over time [3, 4]. Many software metrics are available in literature for detection of code smells [5, 6]. Moreover, there are several tools that developers can apply for automatic or semi-automatic detection of code smells in their code. Applying appropriate refactoring actions is the right way to deal with code smells. Refactoring actions can remove Code smells and optimize the quality of software design during maintenance process [7-9].

This paper presents an analytical study of code smells and its related concepts. The significance of this study is to extract some insightful information from inter relation between code smells, software metrics, refactoring actions and detection tools. This paper is organized as follows: background and related work is described in section 2. The design of code smell repository is presented in section 3. Section 4 presents application of different analytical techniques to code smell related information tables and extracting of indicative information that can further enhance the usefulness of a code smell repository. This is followed by conclusion in section 5.

2 BACKGROUND AND RELATED WORKS

In 1999, Beck and Fowler [9] found out that code smells are some indications in source code which don't prevent of its functionality but may reveals lots of problems in future. They presented 22 code smells and some refactoring actions that can be used to develop the design.

Mantyla [10] classified 22 code smells in seven categorized because of their similar features.

Mens and Tourwé [11] presented their survey on refactoring. It includes all aspects of refactoring process such as general ideas, refactoring actions, different formalism and methods, attentions and how refactoring suits the software development process. Walter and Pietrzak [12] pointed out that certain code smells such as divergent change get added as part of the maintenance phase. They proposed that multiple pieces of code need to be analyzed to detect the change. Marinescu [13] promoted the formalization of definition of code smells. He developed the detection to a broader range of code smells and a number of design principle violations. Olbrich et al. [14] demonstrated that in the existence of bad smells, performance of open source projects is degraded. They examined this bad feature for three software projects. God Class and Brain Class were selected by them for their experimental study. They observed that without normalization of size, both smells are harmful for code. In contrast, with normalization of size, outcomes are reversed. Therefore, they evolved that the size of both code smells are major factor for measuring the harmfulness of these smells. An investigation about God Class and Data Class presented by Ferme et al. [15] proves that bad smells are destructive for source code. Different filters were suggested by them to decrease or refine detection rules for code smells. Mahmood et al. [16] investigated several refactoring tools and established their purpose of usage. Also, they examined automation of tools for different code smells. Ganea et al. [17] described that code smells make considerable disadvantages in source code. They presented a tool named "InCode" that is an Eclipse plug-in. This tool is designed for Java programs and has capability of increasing the quality of source code and decreasing the code smells. Yamashita and Moonen [18] presented an empirical study about inter relation of code smells and their effect on occurrence of maintainability issues. They found out that certain inter-smell relations were connected with issues in the maintenance process and some inter-smell relations indicated

through couple artifacts. Yamashita et al. [19] had a survey for detecting a broader range of inter-smell relations. They observed that for various domains some of the code smells have same inter relation and should pay attention to them. Therefore, these inter relations can help practitioners for improving the quality of software systems.

3 BUILDING CODE SMELL REPOSITORY

An extensive literature survey was carried out to gather all the information about Code smells and the related concepts. An initial list of 22 code smells was proposed by Kent Beck and Martin Fowler [20] which has since grown with contributions from several researchers and practitioners into almost 65 code smells. With the increase in number of code smells, Mantyla [10] proposed a classification of code smells into six categories.

Software Metrics use measurable software attributes as indicators of latent software quality attributes [21-23]. Detrition of quality created by presence of code smells can be quickly detected by using one or more related software metrics [24-26]. The literature survey identified that around 49 software metrics are applicative in code smell detection. Software metrics are categorized in many ways and one such classification separates class level metrics from method level metrics.

Tool support is essential, as several code smells can go undetected while programming [27]. Tools are available for automatic or semi-automatic detection of code smells. The detection methods applied by tools are generally established on the calculation of a specific set of composite metrics using the threshold values for these metrics [28]. Numerous tools are accessible but 9 detection tools are popular to use by developers.

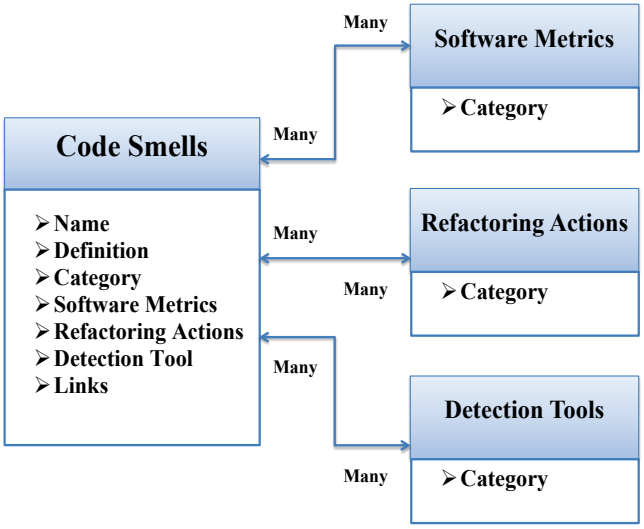


Figure 1 Code smell repository schema

Maintainability is the most important step in software development process [29]. Maintainability can be improved by use of refactoring methods. The term 'Refactoring' was presented by Opdyke [30] in his PhD thesis. Later, Fowler [9] identified that refactoring is a disciplined method for restructuring internal structures of existing source code without changing its external structures [27, 31]. There are around 87 refactoring actions that could be picked from literature, which are classified into six groups.

The schema of Code smell repository presents that code smell operates as the main object of repository which is linked to software metrics, refactoring actions and detection tools. Code smell relations with its corresponding related concepts are many to many. Each of the related concepts has its own details such as name, definition, category and etc. Fig. 1 displays Code smell repository schema. Further 'links' attribute can be used to navigate to different sources of detailed information about code smells.

The code smell repository thus constructed is available at https://serene-tundra-28026.herokuapp.com

3.1 Methodology of Building Code Smell Repository

- Selection of different kind of papers about code smells and related concepts from different journals and internet sources
- Data extraction about code smells and related concepts from the literature
- Organization of code smell knowledge

1) Designing a code smell repository template
2) Designing a code smell repository schema
3) Generating tables between code smells and each related concept for presentation of their relationship
4) Designing a web code smell application using Angular, Material Design, Node JS, Express JS and MongoDB
5) Implementing the code smell repository

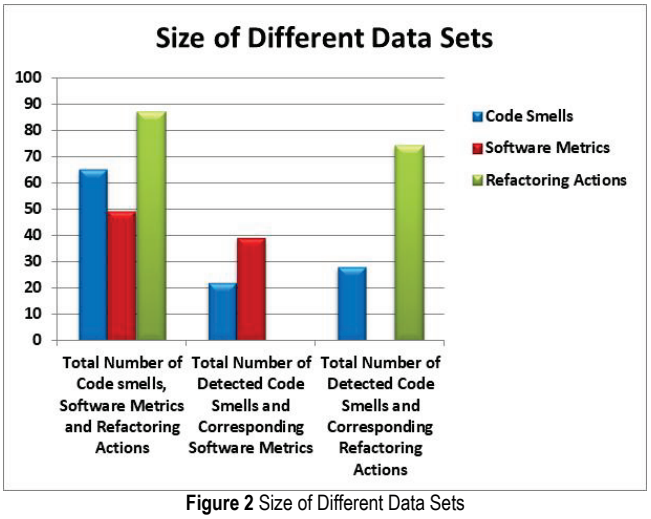


Figure 2 Size of Different Data Sets

4 ANALYTICAL STUDY OF CODE SMELLS AND ITS RELATED CONCEPTS

Data collected in the Code smell repository can be analyzed to gain useful insights into the world of code smells. Analytical study of code smells focuses on inner relation between code smells and its related concepts. Though only 22 code smells are detected by one or more out of 39 software metrics, this table capturing the link among code smells and

software metrics may be subjected to several analytical techniques to extract useful insights. Also, 28 code smells can be removed by 74 refactoring actions. Fig. 2 shows it.

4.1 Identifying Most Significant Set of Software Metrics

Many code smells can be detected by one or more metrics. As a sample, a formula for God Class detection is:

$$WMC \geq VERY\_HIGH \wedge ATFD > FEW \wedge TCC < \frac{1}{3} \qquad (1)$$

where *FEW* is 5 and *VERY_HIGH* is 47.

In contrast, *LOC* alone can detect Large Class [32]. One or more metrics may detect one or more code smells. Therefore, they have a many to many relationship.

Notably, most obvious metric for code smell detection is size metric. *LOC* (number of lines of code) acts as the leader of metrics in detection of code smells as it is used in detection of as many as 8 code smells.

Table 1 Significant Software Metrics and Code Smells Detected by Them

No	Metrics	Detected code smells	Total
1	LOC	God class, Brain class, Long method, Brain method, Duplicated code, Primitive obsession, Large class, Lazy class	8
2	WMC	God class, Brain class, Duplicated code, Large class, Lazy class, Data class, Refused bequest	7
3	VG	God class, Long method, Switch statements, Brain method, Conditional complexity, Duplicated code, Lazy class	7
4	NOM	Message chain, Middle man, Lazy class, Refused bequest, Large class	5
5	LCOM	Feature envy, God class, Large class, Data class	4
6	CBO	Feature envy, God class, Large class, Lazy class	4
7	DIT	Parallel inheritance hierarchy, Duplicated code, Large class, Lazy class	4

Figure 3 Significance of different software metrics in Code smell detection

Both WMC (Weighted Method Count per Class) and VG (McCabe Cyclomatic Complexity per module) are related to Cyclomatic complexity and are used in detection of 7 code smells each. Clearly, 7 metrics (LOC, WMC, VG, NOM, LCOM, CBO and DIT) present the most significant set of software metrics used in detection of a large number of code smells. Tab. 1 shows the significance of detected smells and software metrics. The wordcloud in Fig. 3 gives a visual presentation of Significance of different software metrics in detection of Code smells. Appendix A shows the used metrics abbreviation.

4.2 Identifying Representative Metric for Each Code Smell Category

Code smells are organized into 6 categories [33-36]. All code smells are not categorized in literature. A decision tree classification method is applied on this table where Code smell category acts as a class label. Classification result shows that each code smell category can have one or more representative metric. Result is showed in Tab. 2.

Table 2 Code Smell Category and Its Representative Metric

No	Code Smell Category	Representative Metric
1	Bloaters	LOC
2	Object Orientation Abusers	MNL or WMC
3	Change Preventers	CM
4	Dispensables	DIT
5	Couplers	NOM

This information can be used to advantage in predicting categories of code smell not yet categorized also in designing detection metrics for code smells from a particular category.

4.3 Identifying Association between Software Metrics

Code smell table with related metrics can be also subjected to identifying association between different metrics. The results of apriori algorithm with minimum support: 0.15 (3 instances) and minimum confidence: 0.9 generated 11 one itemsets, 13 two itemsets and 4 three itemsets and corresponding association rules. The three itemsets as given in Tab. 3 bring out most frequently occurring groups of related metrics.

Table 3 Frequency Occurring Metric Groups

No	Most frequently occurring groups
1	Lines of Code, Weighted Method Count per Class, Depth of Inheritance Tree
2	Lines of Code, Weighted Method Count per Class, Coupling Between Objects
3	Lines of Code, Weighted Method Count per Class, Tight Class Cohesion
4	Coupling Between Objects, Lack of Cohesion in Methods, Access To Foreign Data

4.4 Identifying Clustering between Code Smells and Software Metrics

The objective of clustering analysis is to recognize patterns in data and make groups based on those patterns. Thus, if two observations have similar features that mean they have the identical pattern. Consequently, they are included in the same group. Clustering is capable to shows what characteristics frequently appear together. Fig. 4 is result of clustering on detected code smells and corresponding used metrics. After cutting dendrogram at k =

4, one can find out the sets of code smells that appear together. These code smells have similarity as to the detection metrics used by them. Thus there is more chance of them occurring together.

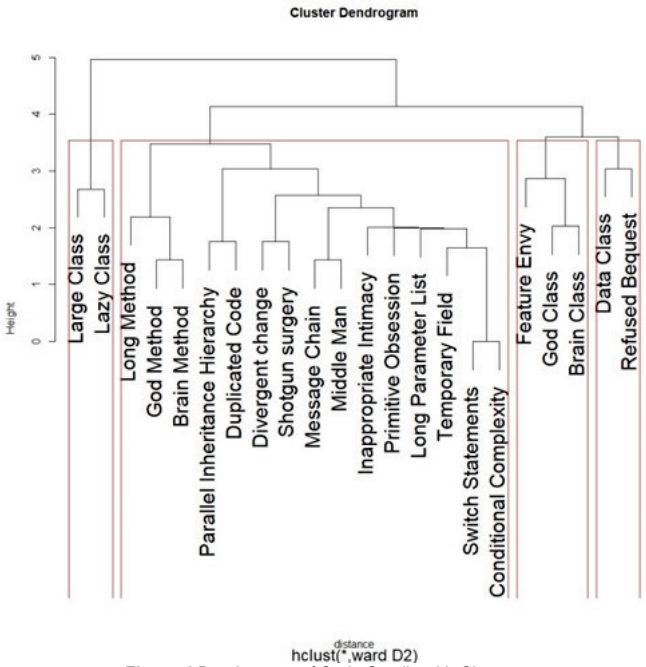


Figure 4 Dendrogram of Code Smells with Clusters

Table 4 Important Refactoring Methods

No	Metrics	Detected code smells	Total
1	Move Method	Switch Statements, Data Class, Feature Envy, Message Chains, Middle Man, Alternative Class with Different Interfaces, Shotgun Surgery, Parallel Inheritance Hierarchies, Inappropriate Intimacy	9
2	Extract Method	Long Method, Switch Statements, Comments, Data Class, Feature Envy, Message Chains, Duplicate Code	7
3	Extract Class	Inappropriate Intimacy, Duplicate Code, Temporary Field, Large Class, Data Clumps, Divergent Change, Primitive Obsession	7
4	Move Field	Feature Envy, Middle Man, Shotgun Surgery, Parallel Inheritance Hierarchies, Inappropriate Intimacy	5
5	Introduce Parameter Object	Long Method, Data Clumps, Long Parameter List, Primitive Obsession	4
6	Preserve Whole Object	Long Method, Long Parameter List, Primitive Obsession, Data Clumps	4
7	Extract Superclass	Alternative Classes with Different Interfaces, Duplicate Code, Refused Bequest, Divergent Change	4
8	Inline Class	Shotgun Surgery, Speculative Generality, Lazy Class, Dead Code	4

4.5 Inter Relation between Code Smells and Refactoring Actions

Refactoring is an important task of maintenance phase that aims at improving latent software quality attributes like understandability, flexibility, and reusability [40]. One or more refactoring actions have been suggested for eliminating of one or more code smells, so the type of correlation between them is many to many. 'Move method' is an important refactoring action that addresses the problem of as many as 9 code smells. Tab. 4 shows the most important refactoring actions that can be used in getting rid of a large set of code smells. Also, Fig. 5 represents a wordcloud of Significance of different refactoring methods in detection of Code smells.

Figure 5 Significance of different refactoring actions in Code smell detection

4.6 Identifying Association between Refactoring Actions

Code smell table with related refactoring actions can be also subjected to identifying association between different refactoring actions. The results of apriori algorithm with minimum support: 0.1 and minimum confidence: 0.9 generated 15 one item sets and 4 two item sets. The best four association rules generated with confidence 1 are as given below in Tab. 5.

Table 5 Association between Refactoring Methods

No	Most frequently occurring groups
1	Movie Field ⇒ Move Method
2	Preserve Whole Object ⇒ Introduce Parameter Object
3	Introduce Parameter Object ⇒ Preserve Whole Object
4	Collapse Hierarchy ⇒ Inline Class

This association indicates the pairs of refactoring actions that is closely linked.

5 CONCLUSION

Organization of knowledge about code smells and related concepts spread out in literature into code smell repository gives rise to tables holding useful information. Applying analytical techniques to these tables can help in improving this knowledge bank further.

Analytical study of code smells and its related concepts gives insightful knowledge about code smells to improve the software development process. Results of this paper are as follows:

- 22 code smells are detected by one or more out of 39 software metrics and 28 code smells can be removed by 74 refactoring actions

- Presenting the table of top 7 software metrics to detect the maximum number of code smells and top 8 refactoring actions to eliminate the maximum number of code smells
- Preparing a wordcloud of Significance of different software metrics and refactoring actions with respect to Code smells
- Presenting of one or more representative software metric for each code smell category by applying a decision tree classification method
- Presenting the most Frequently Occurring Groups of software metrics based on code smells and metrics relationships by applying association apriori algorithm
- Presenting the most Frequently Occurring Groups of Refactoring actions based on code smells and refactoring methods relationships by applying association apriori algorithm
- Applying the hierarchical clustering based on code smells and software metrics relationships to find the similarity of code smells by presenting of a dendrogram. This presents a new way of categorizing code smells

The code smell repository and the extracted insights can assist the developers and software practitioners.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

6 REFERENCES

[1] Vidal, S., Vazquez, H., Diaz-Pace, J. A., Marcos, C., Garcia, A., & Oizumi, W. (2015, November). JSpIRIT: a flexible tool for the analysis of code smells. In *2015 34th International Conference of the Chilean Computer Science Society (SCCC)* (pp. 1-6). IEEE. https://doi.org/10.1109/SCCC.2015.7416572

[2] Mannan, U. A., Ahmed, I., Almurshed, R. A. M., Dig, D., & Jensen, C. (2016, May). Understanding code smells in Android applications. In *2016 IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft)* (pp. 225-236). IEEE. https://doi.org/10.1145/2897073.2897094

[3] Di Nucci, D., Palomba, F., Tamburri, D. A., Serebrenik, A., & De Lucia, A. (2018, March). Detecting code smells using machine learning techniques: are we there yet? In *2018 IEEE 25th International Conference on Software Analysis, Evolution and Reengineering (SANER)* (pp. 612-621). IEEE. https://doi.org/10.1109/SANER.2018.8330266

[4] Firdaus, M. F., Priyambadha, B., & Pradana, F. (2018, November). Refused Bequest Code Smells Detection on Software Design. In *2018 International Conference on Sustainable Information Engineering and Technology (SIET)* (pp. 288-291). IEEE. https://doi.org/10.1109/SIET.2018.8693156

[5] Eisty, N. U., Thiruvathukal, G. K., & Carver, J. C. (2018, October). A survey of software metric use in research software development. In *2018 IEEE 14th International Conference on e-Science (e-Science)* (pp. 212-222). IEEE. https://doi.org/10.1109/eScience.2018.00036

[6] Do Vale, G. A., & Figueiredo, E. M. L. (2015, September). A method to derive metric thresholds for software product lines. In *2015 29th Brazilian Symposium on Software Engineering* (pp. 110-119). IEEE. https://doi.org/10.1109/SBES.2015.9

[7] Khurana, G. & Jindal, S. (2013). A model to compare the degree of refactoring opportunities of three projects using a machine algorithm. *Advanced Computing, 4*(3), 17. https://doi.org/10.5121/acij.2013.4302

[8] Dhaka, G. & Singh, P. (2016, December). An empirical investigation into code smell elimination sequences for energy efficient software. In *2016 23rd Asia-Pacific Software Engineering Conference (APSEC)* (pp. 349-352). IEEE. https://doi.org/10.1109/APSEC.2016.057

[9] Fowler, M. (2018). *Refactoring: improving the design of existing code*. Addison-Wesley Professional.

[10] Mantyla, M. (2003). *Bad smells in software-a taxonomy and an empirical study*. Helsinki University of Technology.

[11] Mens, T. & Tourwé, T. (2004). A survey of software refactoring. *IEEE Transactions on Software Engineering, 30*(2), 126-139. https://doi.org/10.1109/TSE.2004.1265817

[12] Walter, B. & Pietrzak, B. (2005, June). Multi-criteria detection of bad smells in code with UTA method. In *International Conference on Extreme Programming and Agile Processes in Software Engineering* (pp. 154-161). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11499053_18

[13] Marinescu, R. (2005, September). Measurement and quality in object-oriented design. In *21st IEEE International Conference on Software Maintenance (ICSM'05)* (pp. 701-704). IEEE. https://doi.org/10.1109/ICSM.2005.63

[14] Olbrich, S. M., Cruzes, D. S., & Sjøberg, D. I. (2010, September). Are all code smells harmful? A study of God Classes and Brain Classes in the evolution of three open source systems. In *2010 IEEE International Conference on Software Maintenance* (pp. 1-10). IEEE. https://doi.org/10.1109/ICSM.2010.5609564

[15] Ferme, V., Marino, A., & Fontana, F. A. (2013). Is it a real code smell to be removed or not? In *International Workshop on Refactoring & Testing (RefTest), co-located event with XP 2013 Conference*.

[16] Mahmood, J. & Reddy, Y. R. (2014, February). Automated refactorings in Java using IntelliJ IDEA to extract and propogate constants. In *2014 IEEE International Advance Computing Conference (IACC)* (pp. 1406-1414). IEEE. https://doi.org/10.1109/IAdCC.2014.6779532

[17] Ganea, G., Verebi, I., & Marinescu, R. (2017). Continuous quality assessment with inCode. *Science of Computer Programming, 134*, 19-36. https://doi.org/10.1016/j.scico.2015.02.007

[18] Yamashita, A. & Moonen, L. (2013). To what extent can maintenance problems be predicted by code smell detection?–An empirical study. *Information and Software Technology, 55*(12), 2223-2242. https://doi.org/10.1016/j.infsof.2013.08.002

[19] Yamashita, A., Zanoni, M., Fontana, F. A., & Walter, B. (2015, September). Inter-smell relations in industrial and open source systems: A replication and comparative analysis. In *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)* (pp. 121-130). IEEE. https://doi.org/10.1109/ICSM.2015.7332458

[20] Sharma, T. & Spinellis, D. (2018). A survey on software smells. *Journal of Systems and Software, 138*, 158-173. https://doi.org/10.1016/j.jss.2017.12.034

[21] Sharma, M. & Singh, G. (2011). Analysis of Static and Dynamic Metrics for Productivity and Time Complexity. *International Journal of Computer Applications, 30*(1), 7-13. https://doi.org/10.5120/18036-6883

[22] Núñez-Varela, A., Perez-Gonzalez, H. G., Cuevas-Tello, J. C., & Soubervielle-Montalvo, C. (2013). A methodology for obtaining universal software code metrics. *Procedia Technology, 7*, 336-343.

https://doi.org/10.1016/j.protcy.2013.04.042
[23] Tahvildari, L. & Kontogiannis, K. (2003, March). A metric-based approach to enhance design quality through meta-pattern transformations. In *Seventh European Conference on Software Maintenance and Reengineering, 2003. Proceedings*. (pp. 183-192). IEEE.
[24] Sarıman, G. & Kucuksille, E. U. (2016). A novel approach to determine software security level using bayes classifier via static code metrics. *Elektronika ir Elektrotechnika, 22*(2), 73-80. https://doi.org/10.5755/j01.eie.22.2.12177
[25] Srinivasan, K. P. & Devi, T. (2014). A complete and comprehensive metrics suite for object-oriented design quality assessment. *International Journal of Software Engineering and Its Applications, 8*(2), 173-188.
[26] Kaur, S. & Maini, R. (2016). Analysis of various software metrics used to detect bad smells. *Int J Eng Sci (IJES), 5*(6), 14-20.
[27] Hamid, A., Ilyas, M., Hummayun, M., & Nawaz, A. (2013). A comparative study on code smell detection tools. *International Journal of Advanced Science and Technology, 60*, 25-32. https://doi.org/10.14257/ijast.2013.60.03
[28] Fontana, F. A., Mariani, E., Mornioli, A., Sormani, R., & Tonello, A. (2011, March). An experience report on using code smells detection tools. In *2011 IEEE fourth international conference on software testing, verification and validation workshops* (pp. 450-457). IEEE. https://doi.org/10.1109/ICSTW.2011.12
[29] Sharma, T. & Janakiram, D. (2010). Inferring design patterns using the ReP graph. *J. Object Technol., 9*(5), 95-110. https://doi.org/10.5381/jot.2010.9.5.a5
[30] Opdyke, W. F. (1992). Refactoring object-oriented frameworks.
[31] Szőke, G., Antal, G., Nagy, C., Ferenc, R., & Gyimóthy, T. (2017). Empirical study on refactoring large-scale industrial systems and its effects on maintainability. *Journal of Systems and Software, 129*, 107-126. https://doi.org/10.1016/j.jss.2016.08.071
[32] Fontana, F. A., Braione, P., & Zanoni, M. (2012). Automatic detection of bad smells in code: An experimental assessment. *Journal of Object Technology, 11*(2), 5-1. https://doi.org/10.5381/jot.2012.11.2.a5
[33] Roperia, N. (2009). *JSmell: A Bad Smell detection tool for Java systems*. California State University, Long Beach.
[34] Ahmed, I., Ghorashi, S., & Jensen, C. (2014). An exploration of code quality in FOSS projects. In *IFIP International Conference on Open Source Systems*, pp. 181-190. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-55128-4_26
[35] Saranya, G. (2017). *Code smell detection and prioritization of refactoring operations to enhance software maintainability*. Faculty of Science and Humanities Anna University.
[36] Refactoring ·guru· (n.d.c.). Code Smells. https://refactoring.guru/refactoring/smells/ (Accessed on 16 October 2020)

Authors' contacts:

Lida Bamizadeh, research scholar
(Corresponding author)
Department of Computer Science, Savitribai Phule Pune University,
Ganeshkhind Rd, Ganeshkhind, Pune, Maharashtra 411007, India
9503039485, lida_bamizadeh@yahoo.com

Binod Kumar
JSPM's Rajarshi Shahu College of Engineering (MCA Dept.),
Tathawade, Pimpri-Chinchwad, Maharashtra 411033, India
9665548971, binod.istar.1970@gmail.com

Ajay Kumar
JSPM Jayawant, Technical Campus,
Tathawade, Pimpri-Chinchwad, Maharashtra 411033, India
7972095030, ajay19_61@rediffmail.com

Shailaja Shirwaikar
Department of Computer Science, Savitribai Phule Pune University,
Ganeshkhind Rd, Ganeshkhind, Pune, Maharashtra 411007, India
7066046154, scshirwaikar@gmail.com

APPENDIX A

No	Metrics	Abbreviations
1	Number of Lines of Code	LOC
2	McCabe Cyclomatic Complexity per Module	VG
3	Weighted Method Count per Class	WMC
4	Depth of Inheritance Tree	DIT
5	Class Coupling	CC
6	Coupling Between Objects	CBO
7	Lack of Cohesion in Methods	LCOM
8	Number of Parameters per Method	PAR
9	Tight Class Cohesion	TCC
10	Number of Methods	NOM
11	Number of Attributes	NOA
12	Method Lines of Code	MLOC
13	Number of Children	NOC
14	Access To Foreign Data	ATFD
15	Locality of Attribute Accesses	LAA
16	Foreign Data Provider	FDP
17	Number of Accessor Methods	NOAM
18	Halstead Metric	HM
19	Number of Brain Methods	NBM
20	Maximum Nesting Level	MNL
21	Number of Accessed Variables	NOAV
22	Unused Parameters	UP
23	Weight of a Class	WOC
24	Number of Public Attributes	NOPbA
25	Number of Protected Members	NProtM
26	Base-class Usage Ratio	BUR
27	Base-class Overriding Ratio	BOvR
28	Average Method Weight	AMW
29	Number of Lines of Code in a Class	NLOCC
30	Instance Variable per Method in a Class	IVMC
31	Number of Delegate Method	NODM
32	Number of Foreign Fields	NOFF
33	Number of Foreign Methods	NOFM
34	Length of Methods Call Chain	LOMC
35	Number of Variables per Class	NOVC
36	Changing Classes	CHC
37	Changing Methods	CM
38	Dependency-Oriented Complexity Metric	DOCM
39	Number of Concerns per Component	NCC

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20200718142815

Preliminary communication

Experimental Testing of PV Module Performance

Mladen Bošnjaković*, Marinko Stojkov, Boris Zlatunić

Abstract: This study compares the manufacturer's technical data of several PV modules with real measured outdoor technical data. The irradiance effect on several PV modules is examined by the changing a tilt angle and comparing different meteorological situations of sky clearness (clouds) on the modules mounted outdoor and exposed to Sun. Also, the influence of temperature and dust on the performance of a PV panel is under research using measurement methods described in the paper. The measured current and voltage data at the clean surface of the PV module correspond to the declared data of the PV module manufacturer, and in the case of fouling of the module surface with dust, a power drop of 7.39% was measured.

Keywords: dust effect; irradiance; PV generation; solar panel testing; temperature effect

1 INTRODUCTION

The photovoltaic energy is an interesting source of energy as it is renewable, inexhaustible and clean. It can be used in several applications meeting the cost constraints.

The amount of energy produced in the PV module is directly affected by the solar irradiance, which means that in cloudy weather the PV module produces less energy and does not produce electricity at night.

The performance of the PV module is most influenced by the collector tilt angle, which is defined concerning the horizontal position. The optimal tilt angle of a fixed PV module depends on local climatic conditions such as geographical location and season [1]. The total irradiated solar energy is affected by the duration of solar hours and the average monthly cloudiness [2].

The following articles have been selected and studied in the research field. The effect of irradiance and temperature on the performance of a PV panel was investigated by Zuhal ER et al. [3] and Aoun et al. [4]. Damasen Ikwaba Paul [5] analysed the electrical performance of three PV modules with cells connected in different configurations to address the nonuniform illumination effect.

Ramabadran and Badrilal [6] investigated the harmful effects of partial shading of series and parallel connected PV modules and compare their performance.

Abdullateef A. et al. [7] and Mohammed A. F. et al. [8] modelled solar radiation onto a photovoltaic system using MATLAB software. The results obtained by simulation were compared with the experimental results on the photovoltaic module. The influence of dust on panel efficiency has been investigated by Said and Walwil [9], Touati et al. [10], Kale and Singh [11 and Zaihidee [12]. Depending on the amount of deposited dust, the amount of light that penetrates the material of the photovoltaic cell is reduced, and thus the amount of electricity produced. For example, the reduction in PV module conversion efficiency was 10%, 16% and 20% for 12.5 g/m^2, 25 g/m^2 and 37.5 g/m^2 dust density respective (Kumar et al., 2013) [13].

It is clear that the nature of dust particles, such as particle size distribution, and chemical composition also influence the result.

Otsuka et al. (1988) [14] reported that the shape of the dust particle has a strong effect on the adhesion force between the particle and surface, and thus to the amount of dust that accumulates on the panel.

Hegazy, (2001) [15] reported that the overall transmittance reduction increases with decreasing the tilt angle. They reported a 3% difference in transmittance reduction for tilt angles of 28° and 40°.

The experimental study shows what kind of conditions affect the efficiency of a PV panel. PV panels must guarantee cost-effectiveness for investors.

It is a fact that the light spectrum changes when sunlight passes through the atmosphere. To enable an accurate comparison of the characteristics of solar cells tested at different times and in different places, the standardized spectrum and power density (irradiation) for radiation outside the Earth's atmosphere and on the Earth's surface is defined (ISO 9845-1: 1992, IEC EN 60904-3: 1989-02). Air Mass 0 means that light has not passed through the atmosphere. This irradiance on the border of the atmosphere is 1367 W/m^2 and is declared as solar constant.

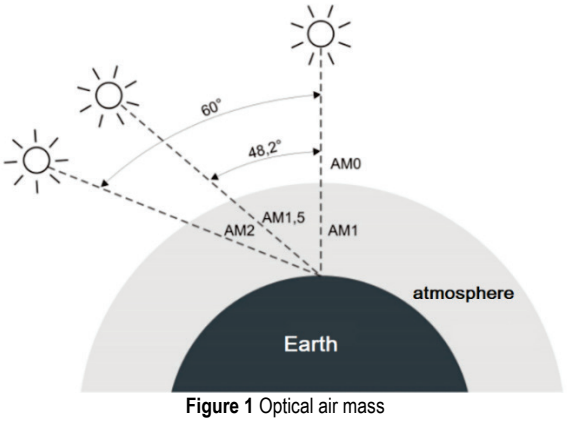


Figure 1 Optical air mass

Ground-level solar radiation with the Sun at its zenith has an optical mass of air AM1, but as a standard ground-level solar

radiation in measurements on solar radiation conversion devices, the distribution of AM1.5 (global) irradiation was adopted. It is the total irradiation (direct and diffuse) that reaches the surface of the sea if the Sun height angle is γ_S = 41.81° (that is, the angle of incidence ζ_S = 48.19°). The normalized AM1.5 G spectrum is set to a value of 1000 W/m^2. The Sun height angle at noon position depending on the time of year and day is illustrated in Fig. 2 and 3.

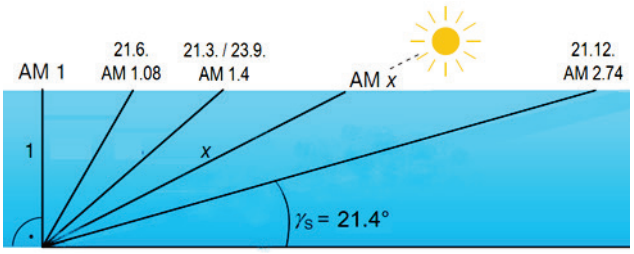


Figure 2 Air mass (at noon position) for Slavonski Brod for different days in the year

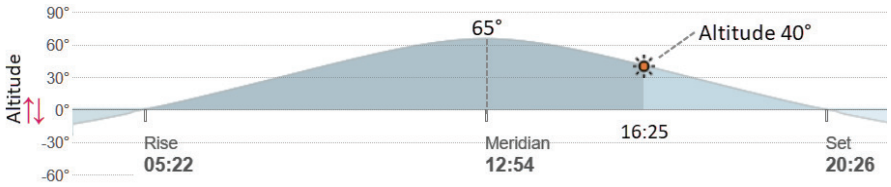


Figure 3 Sun Graph for Slavonski Brod (July 22)

2 MATERIAL AND METHODS

2.1 Description of the Experimental Test

The photovoltaic module is created by connecting photovoltaic cells, where the cells can be connected in series, which increases the voltage or in parallel in a submodule, which increases the output current. The modules produce DC, usually 12 V or 24 V, but there are versions of 6 V and 18 V. During the operation of the PV module, degradations of properties may occur, the causes of which can be very different, for example:

- Degradation or safety issues caused by the failure of cell interconnects, solder bonds, or the bypass diodes that protect in case of shading
- Early degradation in the short-circuit current related to light-induced degradation
- Changes in transmittance associated with changes in anti-reflection coatings, encapsulation discolouration, and delamination
- Corrosion of cells and ribbons (often associated with delamination in the field)
- Junction-box failures, including non-functioning bypass diodes, etc.

In this paper, the aim is not to investigate such phenomena, but for the new module to examine the characteristics in different operating conditions and compare them with the nominal values given by the module manufacturer. Each manufacturer of PV modules takes samples from the production line and tests them following appropriate procedures to ensure the quality of their products. However, it is also interesting to do tests of PV modules that can be bought on the free market. For this purpose, appropriate PV modules were procured and the values of voltage, current, and power were measured and compared with the values stated by the manufacturer.

Equipment by which the tests were performed is (Figs. 4, 5 and 6):

- PV module SOLE U = 18 V, I = 1.11 A, P = 20 W, area 1813 cm^2
- PV module ANBES U = 12 V, I = 0.12 A, P = 1.5 W, area 97.75 cm^2
- PV module Cewaal U = 6 V, I = 0.33 A, P = 2 W, area 81.5 cm^2

Table 1 Declared parameters of photovoltaic modules

	PV module SOLE	PV module ANBES	PV module Cewaal
U (V)	18	12	6
I (A)	1.11	0.12	0.33
P (W)	20	1,5	2

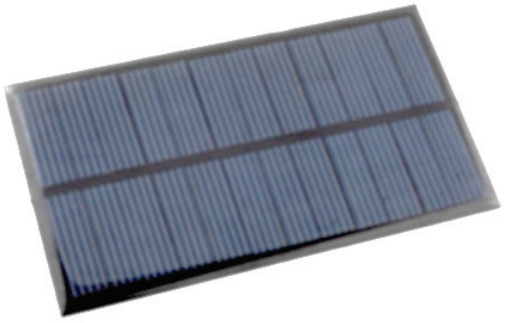


Figure 4 Cewaal PV module 6 V, 0.33 A, 2 W

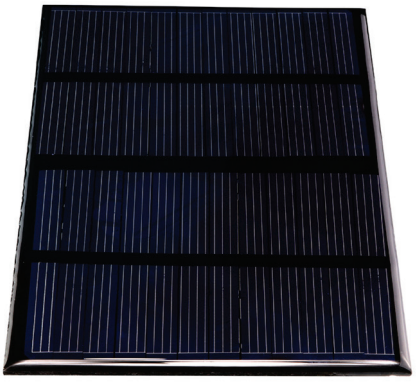


Figure 5 ANBES PV module 12 V, 0.12 A, 1.5 W

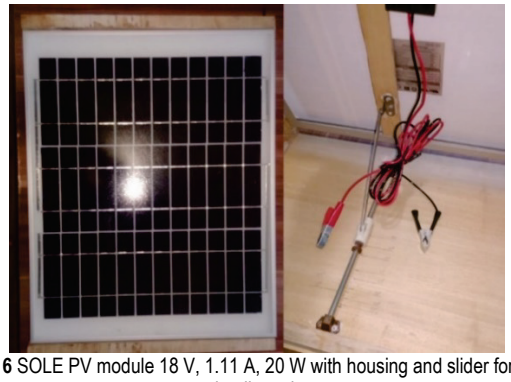


Figure 6 SOLE PV module 18 V, 1.11 A, 20 W with housing and slider for setting the tilt angle

Measuring equipment:

- Measuring Module Inpelanyu C01413 with measuring range 0 - 1000 W, 6.5 - 100 V, current up to 19 A (Fig. 7)
- Lux meter brand Vktech mt30
- Infrared thermometer brand OUTEST GM320

Total (global) solar radiation on a horizontal surface from the spatial angle of 2π steradian consists of radiation that the surface receives directly from the surface of the solar disk and scattered radiation from the sky. The instrument that measures solar radiation from the spatial angle of 2π steradian to the receiving surface is a pyranometer. Because a pyranometer was not available for irradiation measurement, a lux meter was used to measure illuminance (Fig. 7).

Figure 7 Lux meter Vktech mt30, Thermometer OUTEST GM320 and measuring Module Inpelanyu C01413

Illuminance is a measure of photometric flux per unit area or visible flux density. Luminous efficacy in daylight is defined as the ratio of illuminance to global solar irradiation.

$$K_g = E(\mathrm{lx}) \,/ G(\mathrm{W/m^2}) \tag{1}$$

Luminous efficacy models for the clear-sky global and direct beam are mainly related to global solar radiation and solar altitude angle and in some extent of atmospheric conditions. Littlefair [16] gives the expression for luminous efficacy:

$$K_g = 91.2 + 0.702\gamma_S - 0.0063\gamma_S^2 \tag{2}$$

For measuring conditions in Slavonski Brod (season and time of day results in altitude 45°), the value $K_g = 109$ (lm/W) was calculated. So, the amount of irradiation on the surface can be approximately calculated using the expression:

$$G\ (\mathrm{W/m^2}) = E\ (\mathrm{lx}) \cdot 0.00917 \tag{3}$$

2.2 Comparison of Measured Values of Photovoltaic Modules with Values Specified by the Manufacturer

The circuit with components is shown in Fig. 8.

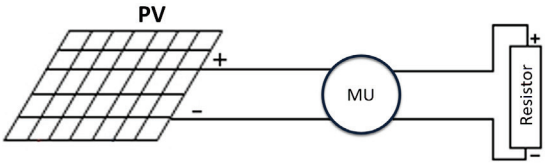


Figure 8 Scheme of the test module

The test is performed outdoors at a brightness of 86,200 lux that applying expression (1) gives 790.5 W/m^2. The measured ambient temperature was 32 °C.

Table 2 Measured values of photovoltaic module parameters

	PV module SOLE	PV module ANBES	PV module Cewaal
U (V)	22.50	12.60	6.80
I (A)	1.23	0.13	0.36
P (W)	27.6	1.6	2.44

The amount of deviation of radiated energy from standard conditions is about 8.3%. For the SOLE module, the voltage deviation is 25%, the current 10.8%, and the power deviation is 38%. From the above, it can be concluded that the module gives slightly better characteristics than the declared ones.

For the ANBES module, the voltage deviation is 5%, the current is 0.5%, and the power is 6.66%. From the above, it can be concluded that the module gives approximately declared characteristics.

For the Cewaal module, the voltage deviation is 13.33%, the current 9.09%, and the power deviation by 22%. From the above, it can be concluded that the module gives slightly better characteristics than the declared ones.

The measured values deviate from those declared by the manufacturer, and the main reason is different test conditions, i.e. a deviation of irradiated energy and higher outdoor temperature which was 32 °C during the measurement. All panels were tested by the manufacturer in standard test conditions (STC - Standard Test Condition) 1000 W / M2; 2 AM1.5; 25 °C.

2.3 Influence of the Tilt Angle on the Performance of the PV Module

The test aims to define the influence of the tilt angle of the module on its efficiency. The SOLE module with the possibility of changing the tilt angle was used for this test (Fig. 9).

The test is performed by connecting the photovoltaic module to the circuit with the measuring module and the consumer and using a slider to set a tilt angle (0-90°). The test is conducted in the open air with measured illumination of 86,200 lumens. Applying expression (1), this illumination corresponds to irradiation of 783.6 W/m^2.

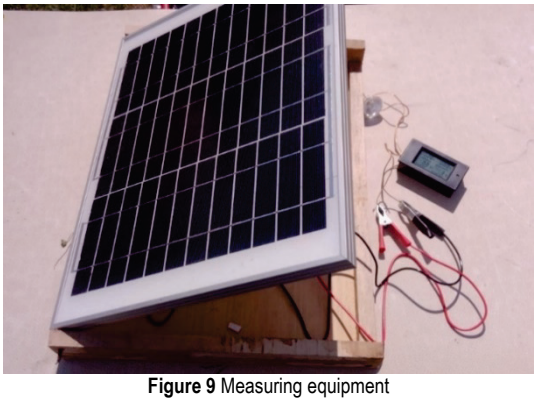


Figure 9 Measuring equipment

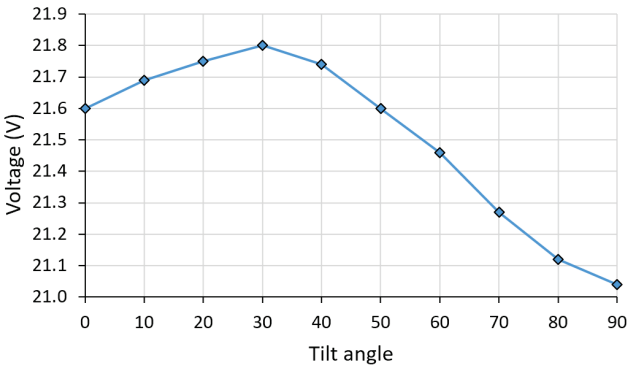


Figure 10 Voltage as a function of tilt angle

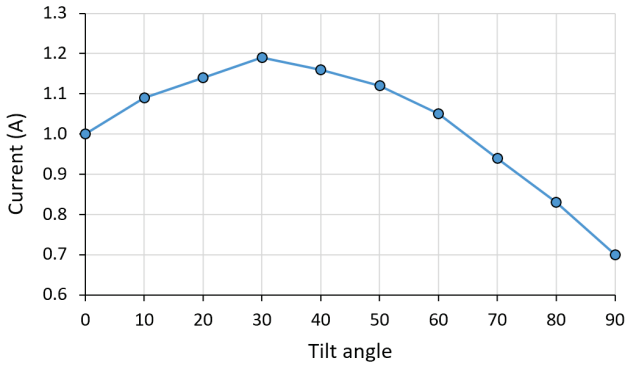


Figure 11 Current as a function of tilt angle

By changing the tilt angle of the photovoltaic module, its performance also changes. The share of direct solar radiation is dominant in the total radiation, and the maximum performance is achieved by placing the surface of the PV module perpendicular to the direction of radiation.

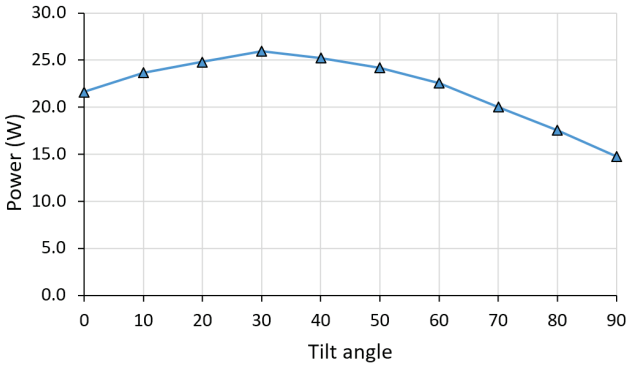


Figure 12 Power as a function of tilt angle

During the test, the module achieved the best performance at the tilt angle of 30°, which is the closest to the optimal tilt angle for the area of Brodsko-Posavska County in Republic of Croatia, which is 35° (Fig. 12).

2.4 Examination of the Influence of Clouds on the Performance of the PV module

The tests were performed using the SOLE module. The two cases were analysed: case when the sky is cloudy and the case for a clear sky. The measured intensity of illumination for the clear sky was 84,500 lumens, and during the cloudy sky at 19,600 lumens. Calculation of irradiated energy for the case of the clear sky gives:

$$G = 84{,}500 \cdot 0.00917 = 775 \text{ W/m}^2 \quad (4)$$

Calculation of irradiated energy for the case of the cloudy sky gives:

$$G\,(\text{W/m}^2) = 19{,}600 \cdot 0.00917 = 180 \text{ W/m}^2 \quad (5)$$

Table 3 Measured values for the clear sky and moderately cloudy sky

	Clear sky	Cloudy
U (V)	21.5	18.7
I (A)	1.13	0.21
P (W)	24.29	3.92

For the cloudy sky, the voltage decreased by 13.02%, the current decreased by 81.41% and the power decreased by 83.74% (Tab. 3). According to the test results, it is noticeable that for the cloudy sky there is a significant drop in current and power while the voltage drops are significantly less.

2.5 Influence of the PV Module Temperature Rise on the Voltage and Current

The test is performed by connecting the SOLE photovoltaic module to the circuit with the measuring module and the consumer. The temperature and voltage of the PV module are then measured every 5 minutes up to 30 minutes. The procedure was performed outdoors at a light intensity of

84,500 lx, which corresponds to irradiation of 775 W/m^2. Module temperature is also affected by the module's regime of operation, wind speed, and ambient temperature [17]. In the observed case, the ambient temperature was 28.9 °C. Although the temperature of the module affects the current, at all measuring points the instrument showed a current of 1.06 A. The reason is the limited accuracy of the instrument of ± 0.01 A.

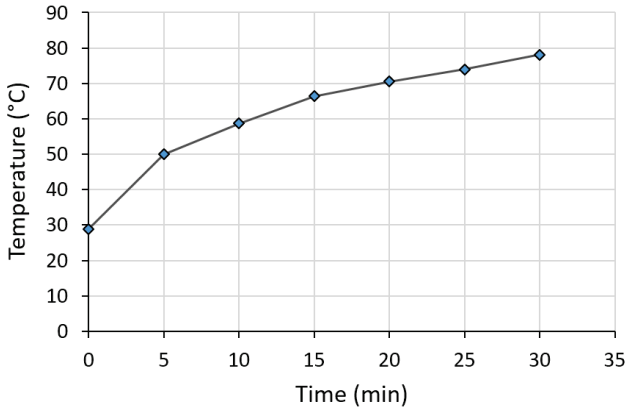


Figure 13 Temperature rise during the measurement

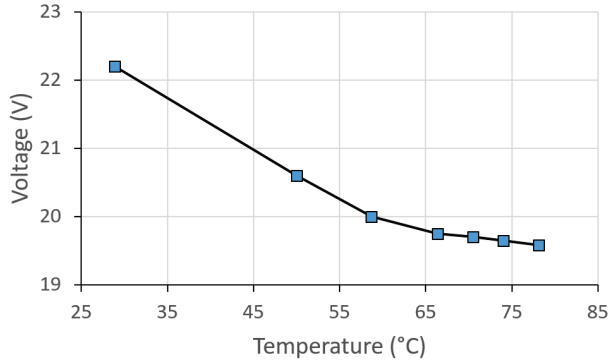


Figure 14 Influence of module temperature on the voltage

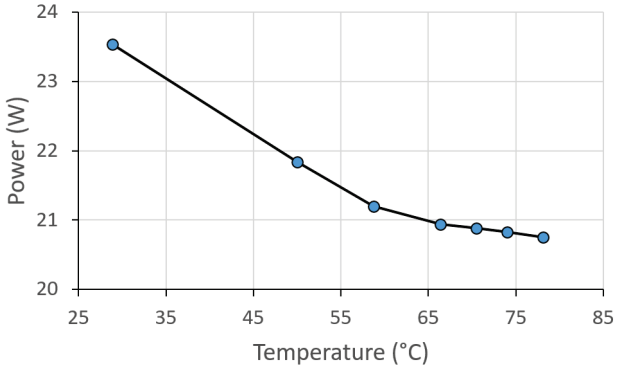


Figure 15 Influence of module temperature on the power

According to Fig. 14, it is noticeable that the voltage slowly decreases with the increasing temperature of the PV module, and the current remains unchanged. This is consistent with the findings of Tofael et al. [18] who concluded that the electrical power produced from a photovoltaic panel depends on the incident solar irradiation, and the temperature of the cells. In the analysed case, when the temperature increases from 28.9 °C to 78.2 °C, the power drops by 11.8% (Fig. 15).

2.6 Examination of the Influence of Dust on the Performance of the PV Module

The test procedure is performed by connecting a photovoltaic module in a circuit with a measuring module and a resistance, then measuring current, voltage, and power under the clean surface (Fig. 16) and surface covered with a layer of dust (Fig. 17). After that, the obtained value is compared. The procedure was performed outdoors, the panel was tilted at approximately 35° facing south. Illumination is measured by a luminometer in the amount of 98,200 lumens. The ambient temperature was 26 °C. Calculation of the amount of irradiated energy:

$$G = 98{,}200 \cdot 0.00917 = 900 \text{ W/m}^2 \quad (6)$$

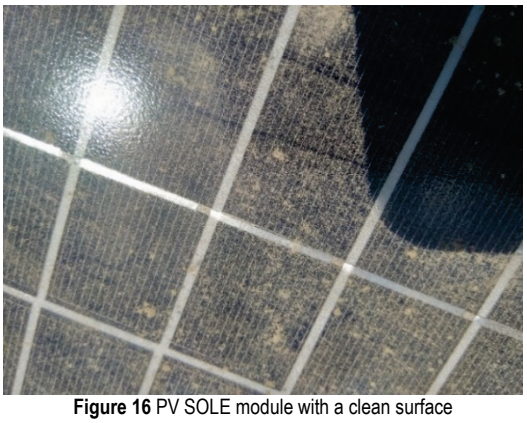


Figure 16 PV SOLE module with a clean surface

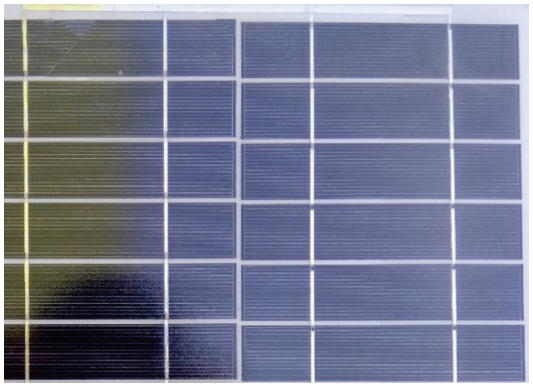


Figure 17 PV module SOLE covered with a layer of dust

Table 4 Measured voltage and current on the clean and fouled PV module

	Clean surface	Fouled surface
U (V)	20.90	20.50
I (A)	1.15	0.91
P (W)	20.04	18.66

According to the test results shown in Tab. 6, it is noticeable that in the case of PV modules with a layer of dust, all measured values are lower than for the case of clean module surface. In the case of the fouled surface, the current is lower 26.37%, and the voltage drop is relatively small and

amounts to 1.95%. A voltage drop and current drop resulted in a power drop of 7.39%.

From the obtained test data, it can be concluded that it is desirable to have a system for cleaning the surface of the PV module to avoid energy losses due to the influence of a layer of dust or other dirt on the surface of the PV module.

3 CONCLUSION

The real-life outdoor experiments support the understanding of parameters that influence the performance of a photovoltaic module in a real application. Measurement methods described and performed analyses in the paper enhance comprehension of the performance of the outdoor mounted PV modules. Influence of temperature effect and dust effect on outdoor mounted PV module surface is presented in the paper. The performed measurements show that the declared values of the PV module are achievable. Due to the increase in the module temperature, a decrease in power by 11.8% occurred, and due to the dust accumulation on the surface of the module, a decrease in the power of 6.9% was registered.

4 REFERENCES

[1] Kumar, A. et al. (2011). Optimization of tilt angle for photovoltaic array, *International Journal of Engineering Science and Technology (IJEST)*, 3(4).

[2] Kimbal, H. H. (1919), Variations in Total and Luminous Solar Radiation with Geographical Position in the United States, *Mon. Weather Rev*., 47, pp. 769. https://doi.org/10.1175/1520-0493(1919)47<769:VITTAL>2.0.CO;2

[3] Er, Z., Rouabah, Z., Kizilkan, G., & Orken, A. T. (2018). Standards and Testing Experiments for a Photovoltaic Module. *European Journal of Science and Technology Special Issue*, 12-15.

[4] Aoun, N., Bouchouicha, K., & Chenni, R. (2017). Performance Evaluation of a Monocrystalline Photovoltaic Module under Different Weather and Sky Conditions. *International Journal of Renewable Energy Research*, 7(1), 292-297.

[5] Damasen Ikwaba, P. (2019). Experimental Characterisation of Photovoltaic Modules with Cells Connected in Different Configurations to Address Nonuniform Illumination Effect. *Journal of Renewable Energy*, Vol. 2019, Article ID 5168259, 15 pages. https://doi.org/10.1155/2019/5168259

[6] Ramaprabha, R. & Mathur, B. (2009). Effect of Shading on Series and Parallel Connected Solar PV Modules. *Modern Applied Science, 3*(10), 32-41. https://doi.org/10.5539/mas.v3n10p32

[7] Jadallah, A. A., Mahmood, D. Y., & Abdulqader, Z. A. (2014). Simulation and Experimental Testing of a PV Panel at Different Operating Conditions. *IJISET - International Journal of Innovative Science, Engineering & Technology, 1*(10), 141-146.

[8] Fares, M. A., Atik, L., Bachir, G., & Aillerie, M. (2017). Photovoltaic panels characterization and experimental testing. *Energy Procedia, 119*, 945-952. https://doi.org/10.1016/j.egypro.2017.07.127

[9] Said, S. A. M. & Walwil, H. M. (2014). Fundamental studies on dust fouling effects on PV module performance, *Solar Energy, 107*, 328-337. https://doi.org/10.1016/j.solener.2014.05.048

[10] Touati, F., Massoud, A., et al. (2013). Effects of environmental and climatic conditions on PV efficiency in Qatar. In: *International Conference on Renewable Energies and Power Quality (ICREPQ'13)*, 20-22 March 2013, Bilbao, Spain, 275-281. https://doi.org/10.24084/repqj11.275

[11] Kale, P. G. & Singh, K. K. (2019). Modeling Effect of Dust Particles on Performance Parameters of the Solar PV Module, *ICEES 2019 Fifth International Conference on Electrical Energy Systems*, SSN College of Engineering, Chennai, India. https://doi.org/10.1109/ICEES.2019.8719298

[12] Zaihidee, F. M., Mekhilef, S., Seyedmahmoudian, M., & Horan, B. (2016). Dust as an unalterable deteriorative factor affecting PV panel's efficiency: Why and how. *Renew. Sustain. Energy Rev., 65*, 1267-1278. https://doi.org/10.1016/j.rser.2016.06.068

[13] Kumar, E. S., Sarkar, B., et al. (2013). Soiling and dust impact on the efficiency and the maximum power point in the photovoltaic modules. *Int. J. Eng., 2*(2).

[14] Otsuka et al. (1988). Measurement of the adhesive force between particles of powdered materials and a glass substrate by means of the impact separation method III: effect of particle shape and surface asperity. *Chem. Pharm. Bull., 36*(2), 741-749. https://doi.org/10.1248/cpb.36.741

[15] Hegazy, A. A. (2001). Effect of dust accumulation on solar transmittance through glass covers of plate-type collectors. *Renew. Energy, 22*(4), 525-540. https://doi.org/10.1016/S0960-1481(00)00093-8

[16] Littlefair, P. J. (1988). Measurements of the luminous efficacy of daylight. *Lighting Research & Technology, 20*(4), 177-188. https://doi.org/10.1177/096032718802000405

[17] Kurnik, J., Jankovec, M., Brecl, K., & Topic, M. (2011). Outdoor testing of PV module temperature and performance under different mounting and operational conditions. *Solar Energy Materials and Solar Cells, 95*(1), 373-376. https://doi.org/10.1016/j.solmat.2010.04.022

[18] Tofael, A., Gonçalves, T., & Tlemcani, M. (2016). Single Diode Model Parameters Analysis of Photovoltaic cell. *5th International Conference on Renewable Energy Research and Applications*, 396-400. https://doi.org/10.1109/ICRERA.2016.7884368

Authors' contacts:

Mladen Bošnjaković, PhD, Assist. Prof.
(Corresponding author)
Technical Department, University of Slavonski Brod,
Trg Stjepana Miletića 12,
35000 Slavonski Brod, Croatia
mladen.bosnjakovic@vusb.hr
https://orcid.org/0000-0001-9882-585X

Marinko Stojkov, PhD, Full Prof.
Mechanical Engineering Faculty, University of Slavonski Brod,
Trg Ivane Brlić Mažuranić 2,
35000 Slavonski Brod, Croatia
mstojkov@sfsb.hr

Boris Zlatunić, BSc, student
University of Slavonski Brod,
Trg Stjepana Miletića 12,
35000 Slavonski Brod, Croatia
bzlatunic@vusb.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210205103310

Subject review

Performance-Based Analysis of Blockchain Scalability Metric

Jyoti Yadav*, Ranjana Shevkar

Abstract: Cryptocurrencies like Bitcoin and Ethereum, are widely known applications of blockchain technology, have drawn much attention and are largely recognized in recent years. Initially Bitcoin and Ethereum processed 7 and 15 Transactions Per Second (TPS) respectively, whereas VISA and Paypal process 1700 and 193 TPS respectively. The biggest challenge to blockchain adoption is scalability, defined as the capacity to change the block size to handle the growing amount of load. This paper attempts to present the existing scalability solutions which are broadly classified into three layers: Layer 0 solutions focus on optimization of propagation protocol for transactions and blocks, Layer 1 solutions are based on the consensus algorithms and data structure, and Layer 2 solutions aims to decrease the load of the primary chain by implementing solutions outside the chain. We present a classification and comparison of existing blockchain scalability solutions based on performance along with their pros and cons.

Keywords: consensus; decentralization; latency; scalability; security; throughput

1 INTRODUCTION

Blockchain is a decentralized, distributed, immutable ledger with a sequence of blocks interlinked and secured using cryptography. Block is a basic unit of blockchain that bundles a set of transactions initiated by participating nodes in the blockchain network. Block is a combination of the block header and block data. Block header generally holds information like current block hash, Merkle root hash: a cryptographic hash of all transactions of the block, timestamp: the time when the block is created, nonce(number used once): 32-bit random number that can be manipulated to get the current block hash within difficulty limit and previous block hash: reference to parent block. Block data portion contains the total number of transactions, transaction details (sender address, the value being transfer, receiver address, transaction fee, etc.). Block header is metadata that is used to verify and validate the block. The first block of the blockchain is called as a *"genesis"* block that does not have any previous block address. The main purpose of blockchain is to develop a network without a central repository and authority. The conceptual framework behind blockchain was first introduced by Haber and Stornetta [1] in 1991, for time-stamping of digital documents to avoid backdating it. An efficient optimization of the hash chain using Merkle tree was first described in the paper. This technology became widely known at the beginning of 2008 when Satoshi Nakamoto introduced Bitcoin: an electronic currency that involves the digital transfer of money [2]. Ethereum [3] was introduced by Vitalik Buterin in 2012 with the addition of Smart Contract as a primary feature. Smart contract was developed in 1997 by Nick Szabo [4] and used for the first time in the Ethereum cryptocurrency (ether) in 2015. Ethereum is not just a platform for the exchange of digital currency, but also a programming language used to build and publish Distributed Applications (DApps) without any downtime and fraud. Various digital cryptocurrencies such as Bitcoin, Ethereum, Ripple, Litecoin and Dagcoin are some examples of this technology. But apart from cryptocurrency, the technology is also instrumental in a variety of domains namely financial sectors such as money transfer, global trade financing, insurance, antimony laundering, KYC and other sectors such as health care, media, logistics, supply chain management, power and utilities, Government, property, E-voting etc. Apart from attractive features and interesting applications, the most challenging task of blockchain is its ***scalability***.

This paper attempts to classify and compare existing scalability solutions of blockchain. These solutions are broadly divided into three layers. Fig. 1 shows the mind map which depicts the taxonomy that classifies the blockchain scalability solutions at a glance. Layer 0 focuses on solutions for data propagation. Layer 2 presents on-chain solutions and Layer 3 focuses on off-chain solutions.

The remainder of the paper is organized as follows: Section 2 defines study methodology with the term Scalability and the related concepts. Section 3 presents scalability solutions in all the three layers. Section 4 compares all the solutions discussed in section 3, based on their performance. Section 5 concludes the paper and section 6 discusses about the future work.

2 STUDY METHODOLOGY

2.1 Scalability

Scalability is defined as the ability to process transactions regardless of volume and the number of participants in the blockchain network. The network is said to be scalable if it is capable to grow along with the demand of user-base [5]. It is also stated as the independence between the speed and number of participants in the network. Scalability is one of cryptocurrencies' primary and urgent concern, especially when it comes to the public blockchain.

The public blockchain should be able to handle millions of users on the network, to become mainstream. It is not a singular property of a system, but it relates several key metrics to teach other. The two most important performance metrics are throughput and latency.

2.2 Throughput

It is the number of transactions confirmed/processed per second (TPS). The most popular and widely used public

blockchains Bitcoin and Ethereum are especially slower than centralized payment processing networks such as VISA and Paypal. Both Bitcoin and Ethereum have extremely low throughput in terms of transaction processing rate as Bitcoin blockchain processes 7 TPS. Ethereum blockchain being faster can process 20 TPS approximately. As opposed to this, PayPal can process 193 TPS and VISA can process 1700 TPS.

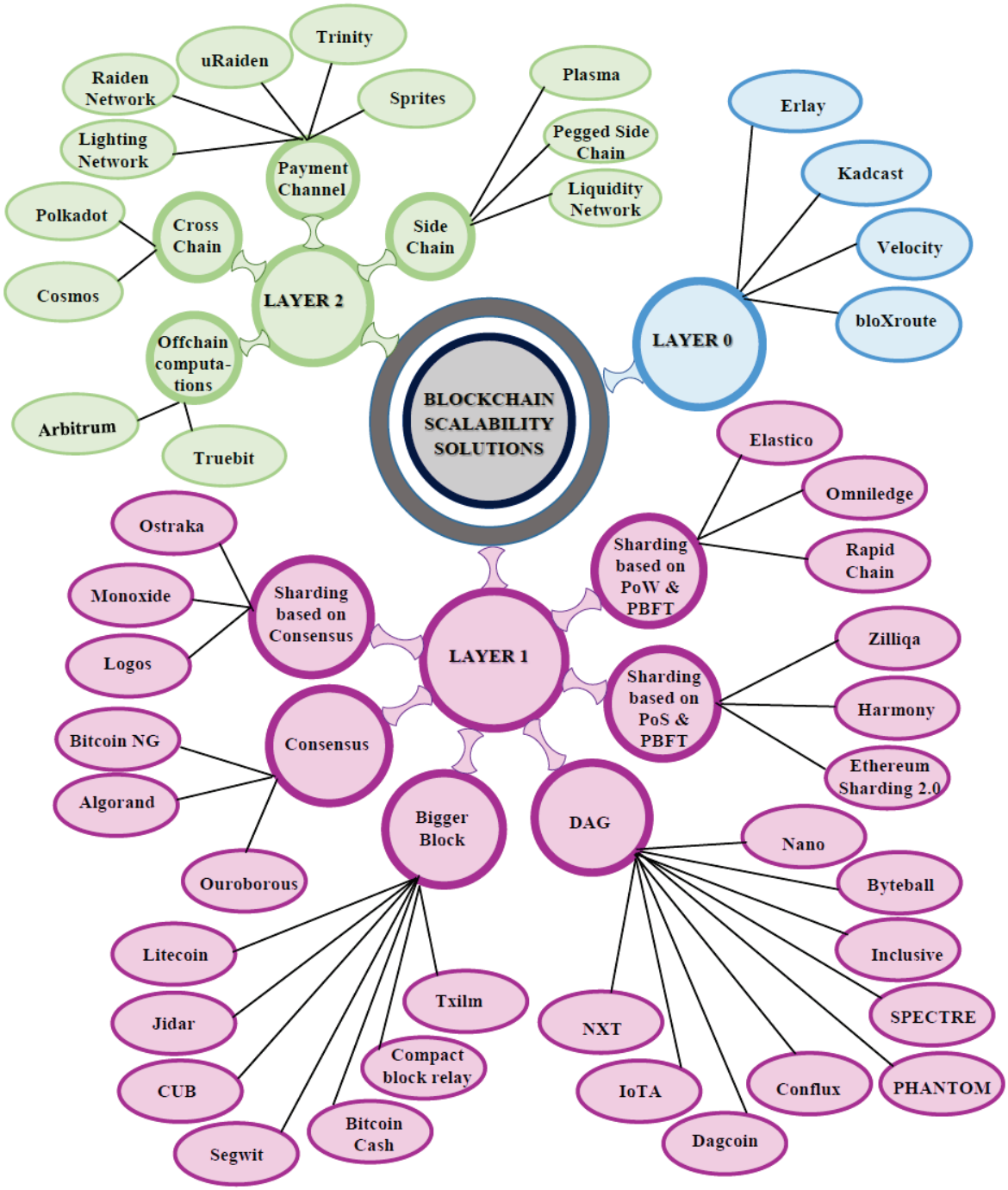


Figure 1 Taxonomy of Existing Blockchain Scalability Solutions

2.3 Latency or Block Time

It is defined as the time between submission and first confirmation of transaction in the blockchain. It is also termed as confirmation time or block time. An increase in number of nodes causes an increase in the number of transactions. Essentially, every single node verifies every transaction, and hence the verification time increases. The confirmation time for Bitcoin is 10 minutes, whereas for Ethereum it is 15 seconds. Thus to cope up with the centralized tech giants of financial sectors such as VISA, MasterCard, and PayPal, some upgradation is needed to scaleup the blockchain technology to increase the user-base like the internet.

2.4 Scalability Trilemma

The corner stones of the scalability trilemma are ***scalability, decentralization,*** and ***security.*** Effectively scaling up the blockchain without compromising on its other two important characteristics namely decentralization and security create alarming challenges to the researchers. Fig. 2 shows the scalability trilemma. The trilemma indicates that decentralization, security, and scalability, cannot co-exist. Blockchain can only possess two of these three properties at a time.

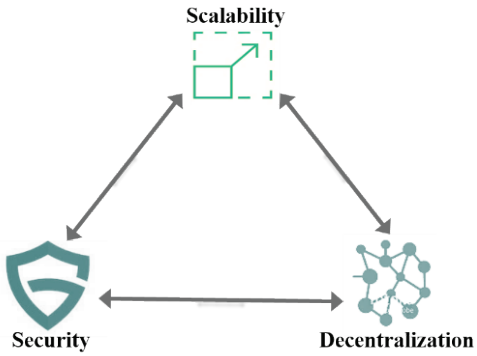


Figure 2 Scalability Trilemma

For instance, to improve scalability, decentralization is compromised by adding a centralized coordinator into the system that reduces the computational power consumed by the proof-of-work algorithm to reach a consensus on a set of transactions. Sacrificing decentralization by creating authority and trust using Hyperledger Fabric greatly improves the scalability and performance of the Blockchain [6]. In case of Bitcoin, reducing the block time improves the transaction throughput by increasing the probability of fork (the new protocols are implemented by splitting the original blockchain into two separate blockchains), which affects the security of the system. Hence, it is essential to balance these three characteristics, particularly for the future development of public blockchain systems. Building a secured system to meet the optimal transaction rate of Bitcoin users remain a formidable challenge.

3 SCALABILITY SOLUTIONS

The following section describes the scalability solutions currently being used in different applications at all three layers.

3.1 Layer 0 Solutions

Layer 0 solutions mainly focus on the propagation protocol for optimizing propagation of information, in the form of transactions and blocks in the blockchain network. The blocks and transactions are transmitted by nodes of the blockchain, but this transmission is not efficient due to high latency and bandwidth. When the block propagation is faster, the larger is the number of blocks added in a shorter block-interval, leading to an increase in transaction throughput. Following are the solutions available in the literature to enhance the propagation protocol:

1) **bloXroute** [7] is the first Blockchain Distribution Network (BDN) that allows faster propagation of blocks and transactions. It allows to increase the block size, reduce the block interval and risk of forks.

2) **Velocity** [8] is an improved block propagation protocol using erasure code (fountain code). This protocol tries to increase the transaction throughput by mining larger blocks.

3) **Kadcast** [9] is a fast, secure and efficient protocol for block propagation. The Kademlia architecture, a well-known structured overlay topology, used for efficient broadcast operation with adjustable redundancy and overhead.

4) **Erlay** [10] is a transaction dissemination protocol that saves 40% bandwidth consumed by a node. Improves the security of the network by allowing more connections to be established at a smaller cost as well as privacy by hardening the network against attack. Effectively, it increases the network connectivity at a very less cost in terms of bandwidth and latency.

Thus using distribution network, erasure code and overlay topology, the block propagation is enhanced along with increasing the throughput and reducing the latency.

3.2 Layer 1 Solutions

Layer 1 solutions concentrate on the consensus algorithms, network and data structure of the blockchain. The execution of these concepts is on-chain, hence these solutions are also referred to as on-chain solutions. There is no need to add anything on top of the existing architecture. The increasing block size of the original blockchain protocol resulted in its modification like bitcoin-cash, bitcoin gold using hard fork. Hard fork is a radical change to a blockchain network's protocol that makes previously invalid blocks/transactions valid (or vice-versa) and all nodes or users need to upgrade to the latest version of the protocol. Data structure like directed acyclic graph has also been studied as a layer 1 scalability solution.

Following are the Layer 1 scalability solutions based on the parameters like block size, sharding using PoS (Proof of Stake) and PoW (Proof of Work) with PBFT (Practical Byzantine Fault Tolerance) and consensus algorithms:

3.2.1 Block Size

The list of transactions are stored in a block and are created periodically. The block size limits the number of transactions stored in a block. Large block size means more number of transactions processed per second. Block interval is the time to generate the next block in the chain. The throughput of the blockchain is directly proportional to the block size and inversely proportional to the block time. As such, the increase in block size leads to slow propagation of blocks in the network and decrease in block time leads to forking of new chains. It is the amount of time between generation of the transaction and adding it to the blockchain. Apart from increasing block size, other solutions are

proposed to compress a number of transactions added to the block [47]. Compact Block Relay and Txilm are some such solutions proposed in BIP152 (Bitcoin Improvement Proposal). The usage of blockchain on mobile phones and low-end PC's is prevented due to the requirement of high storage space. CUB and Jidar are the solutions for it.

Following are the protocols used in the scalability solutions based on block size:

1) **Litecoin** is a peer-to-peer decentralized and open source protocol. Litecoin can handle large volume of transactions than bitcoin. As compared to bitcoins' block interval of 10 minutes, Litecoins' block interval is 2.5 minutes, which is four times faster. Consequently faster block generation supports more number of transactions. The throughput of Litecoin is 56 TPS. It is more resistant to a double spending attack [43]. But it leads to drawbacks such as increased blockchain size and more orphaned blocks (the stale block which is valid and verified, but rejected by the blockchain network due to a time lag in the acceptance of the block). Litecoin uses a new cryptographic algorithm named Scrypt, over a longstanding SHA256 algorithm used by Bitcoin. [44]

2) **SegWit** (Segregated Witness) [11] is the process of separating/segregating the digital signature (witness) of the transaction and is used to increase the block size limit. When certain parts of a transaction are removed, this frees up space/capacity to add more transactions to the chain. The digital signature accounts for 65% of the space in a given transaction. It is defined in BIP141 [45] and designed to solve Bitcoin's malleability and scalability issues. In SegWit block size has been increased from 1 MB to 4 MB and the block is divided into two parts: base transaction block of size 1MB and extended block of size 3MB. Base transaction block contains information about sender and receiver. The digital signature and other data of transactions known as a witness are stored in an extended block. The maximum block size in SegWit is measured in weight, computed as follows:

$$B_{\mathrm{w}} = 3B_{\mathrm{s}} + T_{\mathrm{s}} \tag{1}$$

Where B_{w} is the block weight, B_{s} is the base size: the number of bytes needed to serialize the transaction without witness (3 MB) and T_{s} is the total size: the block size in bytes with transactions serialized including base data and witness data (1 MB).

3) **Bitcoin Cash** (BCC) is a cryptocurrency created from a fork of Bitcoin, in August 2017 [46]. BCC initially increased the block size from 1 MB to 8 MB and later to 32 MB, maintaining the same block interval of 10 minutes. But the large block size leads to centralization, as individual users will not be able to propagate blocks efficiently. Also, it is difficult to verify all transactions within a given time interval.

4) **Compact Block Relay** is a method of reducing the amount of bandwidth used to propagate new blocks to full nodes. Full nodes share almost same mempool contents. Sender sends compact block called sketches to the receiving peer. This compact block includes 80-bytes header of the new block, shortened transaction IDs that are designed to prevent Denial-of-Service (DoS) attacks and some full transactions which the sender predicts the receiver doesn't have yet. Using this information and the transactions already present in mempool, the receiver tries to reconstruct the entire block. The missing transactions are requested from the sender. Once all the transactions are available, the block is generated.

5) **Txilm** on the other hand compresses transactions of each block and saves the bandwidth of the network. Txilm uses a short hash value of TXID to represent a transaction. To avoid hash collisions due to short hash value, the transactions are sorted based on TXID. Thus 80 times data reduction causes an increase in the throughput of the blockchain.

6) **CUB** (Consensus Unit-based Solutions) [12] proposed a new concept called Consensus Unit (CU) that divides different nodes into units. The units of nodes are formed and total blocks of the blockchain are assigned to nodes, to maximize the storage space utilization and reduce the query cost.

7) **Jidar** (Jigsaw-like Data Reduction) [13] in which, each node stores only transactions needed and branches of the Merkle tree from the whole block. This is like selecting pieces from the jigsaw puzzle hence named Jidar. To get complete block data, the fragments are collected from other users and combined into a whole block. But this functionality needs incentive.

The above protocols increases the actual block size, implements enhanced cryptographic algorithm or compresses the transactions to improve the scalability.

Sharding is a widely used solution for scaling distributed databases such as MySQL and MongoDB. It splits the entire blockchain network into multiple smaller groups of nodes called shards or committees. In a blockchain, the shared ledger can be divided into various tasks such as account balances, smart contract code, transaction broadcasting, processing and storage etc. The shards process disjoint transactions in parallel and maintain a disjoint ledger. This results in improved throughput, reduced latency and storage requirements [14]. The sharding solutions are broadly categorized based on consensus as follows:

3.2.2 Sharding Based on PoW and PBFT

PoW consensus is used for committee formation, PBFT consensus for intra-committee communication.

1) **Elastico** [15] provides the first sharding protocol for permissionless blockchains tolerating one-fourth fraction of byzantine faults. It divides the network into multiple committees called shards. Each shard contains a distinct set of transactions. The shard number grows linearly with network size. Throughput is 40 TPS but is only 25% network resilient and 33% committee resilient.

2) **Omniledger** [16] is a distributed ledger based on sharding protocol. It is only 25% resilient to Byzantine faults. To overcome the security issue of Elastico, a bias resistant randomness protocol is used in Omniledger. Apart from PFT for intra shard communication, Byzantine shard atomic

committee is used for cross shard communication. The total and committee resiliency of Omniledger is the same as Elastico.

3) **Rapid Chain** [17] is the first one-third resilient sharding-based blockchain protocol that is highly scalable to large networks. Kademlia routing algorithm is used for inter committee routing. The throughput is greatly increased to 4220 TPS as compared to Elastico. Total resiliency is increased to 33% and committee resiliency to 50%.

4) **Ostraka** [40] architecture scales linearly with the available resources. Ostraka shards are the nodes themselves that runs parallely without affecting the security of the underlying consensus mechanism. The throughput of Ostraka is very high upto 400000 TPS.

Different inter and intra shard communication techniques along with routing algorithms are used to increase the throughput.

3.2.3 Sharding Based on PoS and PBFT

PoS consensus is used for committee formation and PBFT consensus for intra-committee communication.

1) **Zilliqa** [18] allows to process the transaction in parallel and achieve high throughput about thousand times of Ethereum. Zilliqa is susceptible to single shard takeover attacks as it does not support state sharding. Zilliqa's local and global resiliency is the same as Elastico and Omniledger. The throughput of Zilliqa is 2828 TPS.

2) **Harmony** [48] claims to be highly scalable. Along with network communication and transaction sharding, harmony supports state sharding. The distributed randomness process ensures high security. The local and global resiliency of Harmony is the same as Zilliqa, Elastico and Omniledger. In Harmony, one shard contributes to 500 TPS.

3) **Ethereum Sharding 2.0** [49] is the popular sharding based protocol with three phases: Beacon Chain, Shard Chain and State Execution. Beacon chain manages all shards in the network. The consensus rules, rewards and penalties are applied to the validators. Shard chain enables parallel transactions. The operations of the entire system are executed in the State execution phase.

It is observed that there is a sudden growth in the throughput of above solutions due to parallel execution of transactions and state sharding.

3.2.4 Sharding Based on Consensus

Apart from using PoW, PoS or PBFT, other consensus algorithms are used to enhance the performance.

1) **Monoxide** offers linear scaling using asynchronous consensus zones. The blockchain system runs multiple independent and parallel instances called as consensus zones. Each zone is responsible for its own data. It partitions the workload of all key components, without compromising on the decentralization and security of the system [19]. The core and zone-specific data structures, like blocks and transactions are replicated and stored only within their own zones. Mining competition, chain growth, and transaction confirmation are carried out separately and asynchronously in each zone.

2) **Logos** [50] uses Axios, a delegated PBFT consensus algorithm to increase the throughput and minimize latency. Each user on the Logos network has a separate chain to keep track of its transactions and can process in parallel. Sharding adds parallel processing of transactions. The elected delegates validate the transactions.

The asynchronous and delegated PBFT consensus are used to scale-up the performance of the blockchain.

3.2.5 Consensus

Different consensus strategies are used to improve the scalability. Mainly these solutions elects the leader block for the processing of transactions.

1) **Bitcoin NG** (Next Generation) [20] is a protocol that uses Nakamoto consensus, which divides time into epochs. One leader is responsible for transaction serialization in each epoch. Bitcoin-NG introduces key block and micro block. The key block is used only for electing the leader. The PoW mechanism is used by the miners to create the key block. The micro block contains packaged transaction data and is generated by leader. Transactions are processed continuously until new leader is selected. This enhances scalability and reduces transaction confirmation time.

2) **Algorand** [21] is a cryptocurrency built upon a Byzantine Agreement (BA) protocol. Users are selected as committee members using Verifiable Random Function. To reach the next set of transactions, the committee members participate in BA. The participants are replaced by sending a message in BA to avoid targeted attacks. Algorand is highly scalable up to 500,000 users, hence achieves high throughput.

3) **Ouroborous** [22] uses a coin flipping protocol to elect the leader. To determine whether a participant can be elected as a leader, a random number is generated by participants using Verifiable Random Function [19]. In the above listed solutions, Nakamoto consensus, BA protocol and verifiable random functions are used to scale-up the blockchain in terms of the throughput and the users.

3.2.6 DAG (Directed Acyclic Graph)

DAG is a network of nodes that uses topological ordering, where the nodes are connected in order – from earlier to later. The new transaction performed necessitates the validation of two earlier transactions before getting added to the blockchain network. More transactions are validated when new transactions enter the network. This distributed network of double-checked transactions does not need miners and fee for transaction authentication [23]. The scalability is improved by coupling network usage and transaction verification, meaning that a user must handle his/her own transactions in order to use the network [39]. Some DAG based blockchain technologies are as follows:

1) **NXT** [57] is the first crypto-currency to adopt DAG based on blocks instead of using linear linked list structure of blockchain. It is a 100% PoS cryptocurrency, developed in open-source Java. The unique PoS algorithm used in NXT, is independent of implementation of the coin age concept used

by other PoS cryptocurrencies. NXT is also resilient to nothing at stake attacks. The block generation time is 60 seconds and confirmation time is 10 minutes.

2) **Nano** [24] is a trustless, low-latency cryptocurrency that uses novel block-lattice architecture. Each participant has its own blockchain and achieves consensus using delegated PoS voting. Nano offers unlimited scalability, fee-free and instantaneous transaction and runs on low power hardware.

3) **Byteball** [25] is a cryptocurrency platform for smart payments. The transaction itself acts as a unit called ball that connect to each other using DAG. Bytes is the currency for the reward. A DAG is formed by referring one or more parent units. Consensus is achieved by building a main chain which contains most units published by witnesses. Witnesses are trusted and verified addresses which regularly publish sequential units.

4) **Inclusive** [26] protocol proposes to restructure the block chain into a DAG structure that allows transactions from all blocks to be included in the log. The "inclusive" rule is used to select the main chain from within the DAG and to incorporate contents of off-chain blocks into the log. It is verified that there is no conflict with previously included content. An important aspect of the Inclusive protocol is that it rewards fees of accepted transactions to the creator of the block that contains them—even if the block is not part of the main chain. Such payments are granted only if the transaction has not been previously included in the chain, and are decreased for blocks that were published too slowly.

5) **SPECTRE** [27] is specially designed for payments. It is a fast and scalable DAG-based public blockchain. The PoW consensus makes it more secure and resilient to attackers, with only 50% computational power.

6) **PHANTOM** [28] is a protocol for secured transaction confirmation for any throughput that the network supports. It uses a blockDAG that supports faster block generation and higher transaction throughput. PHANTOM uses a greedy algorithm to distinguish between blocks mined properly by honest nodes (a node that behaves as expected) and those that are created by non-cooperating (a node that misbehaves and tries to distribute invalid information) nodes.

7) **Conflux** [29] is a fast, scalable and decentralized system that can process about thousands of TPS and confirms each transaction in minutes. It uses a blockDAG and achieves consensus on the total order of the blocks. The consensus protocol used in Conflux, allows multiple participants to contribute concurrently to the blockchain, preserving the safety. Hence results in faster block generation and higher throughput. The throughput is equivalent to 6400 TPS with latency of about 4.5 to 7.4 minutes, tested on Amazon EC2 clusters.

8) **Dagcoin** [30] was initially built on the top of Byteball. The transactions are stored and ordered using DAG rather than blockchain. Each transaction is treated as a block and accentuates faster and secured confirmations as well as greater throughput. It claims to be faster and securer with the growth of usage.

9) **IoTA** (Internet of Things Application) [51] is the first open-source distributed ledger protocol for the emerging economy of the Internet of Things with feeless micro transactions and data integrity. The key feature of IoTA is Tangle which is the transaction storing and processing mechanism. IoTA is highly scalable, as Tangle can process transactions simultaneously. As more systems are attached to it, the Tangle becomes more secure and efficient at processing transactions.

The DAG solutions generally used for Payments with micro transactions and improves throughput with faster block generation.

3.3 Layer 2 Scalability Solutions

Layer 2 Solutions aims to decrease the load of the main-chain, accomplished by executing some transactions off-chain and shifting computationally intensive tasks on an off-chain platform. The layer 2 solutions are constructed on the top of main blockchain infrastructure. The base level protocols are not altered, instead a smart contract interacts with the blockchain software.

3.3.1 Off-chain Computations

The state of the smart contracts is verified by the validators by imitating the execution of all contracts. But the process of verification is costly, hence decline the scalability of Ethereum. These costly and complex calculations are performed off-chain to enhance the scalability.

1) **Truebit** is an Ethereum smart contract introduced in 2017 by founder and mathematician Jason Teutsch along with the creator of Solidity language Christian Reitwiessner, to facilitate trusted, computationally intensive applications [31]. Computations performed on the main Ethereum blockchain are costly as the transactions are processed by all full nodes on the network simultaneously. The compensation of the computation is given in the form of gas cost. Each block has a maximum gas limit that sets the cap on the total amount of computation performed by all transactions in a block. Hence complex computations are not included in the block. Truebit outsources the complex computations to a verified third party. The third party is trusted as it deposits token into the smart contract and is called as solver. Another third party called challenger, verifies the work done by solver and receives monetary incentives. The challenger identifies exact operation that causes disagreement. Thus the computationally intensive work of Ethereum main blockchain is narrowed down, at the same time true and correct results are recognized.

2) **Arbitrum** [32] protocol performs the verification of smart contract off-chain and improves the scalability. The role of the Verifier is to validate transactions. The fund owned by contract is not consumed for execution of contract. Such contracts are implemented on Virtual Machine (VM). Arbitrum uses mechanism designed to incentivize parties to agree off-chain on what a VM would do, so that the Arbitrum miners verify digital signatures to confirm that parties have agreed on a VM's behavior. A set of VM managers are created by every party, to force to work as per the code. The anonymous assertion is signed by all the managers, only after

agreed upon the new state of VM, otherwise, a disputable assertion is signed to challenge the VM's state change and be engaged in the bisection protocol. The bisection protocol resolves the dispute, identifies and penalizes the dishonest party. Thus, only hashes of contract states are verified and the load of the verifier is minimized allowing contracts to execute privately.

Thus smart contracts and virtual machines are used to reduce the load on the main chain in Truebit and Arbitrum.

3.3.2 Cross Chain

It is an interoperability between independent, heterogeneous blockchains to create a big network of blockchains. Thus the inter communication of independent blockchains can improve the scalability.

1) **Cosmos** [52] is a network of many independent blockchains called zones. The zones are powered by Tendermint BFT. The Tendermint BFT consensus algorithm provides high-performance, consistent, secured and strict fork accountability that controls the behavior of malicious actors. The first zone called as hub uses a governance mechanism enabling the network to adapt and upgrade. The hub and zones can communicate with each other via Inter Blockchain Communication (IBC) protocol to exchange tokens and data.

2) **Polkadot** [33] is a multi-chain protocol that connects heterogeneous blockchains with a relay-chain. Relay-chain enables an independent blockchain called parachain to exchange information and trustless inter-chain translation. Polkadot is a bridge that connects already running blockchains like Ethereum.

Hence secured and trustless intercommunication among the heterogeneous blockchains is achieved using Tendermint and Relay-chain.

3.3.3 Payment Channel

The payment channel is a temporary channel created, on which some transactions are transferred to reduce the load on main chain and to improve the throughput of the entire system.

1) **Lightning Network** [34] is Bitcoin's decentralized scalable solution for faster and high-volume micropayments. It uses a smart contract for instant payments across the network. The key features of Lightning network includes instant payments in milliseconds, high throughput and low cost. But Lightning network has certain drawbacks: (a) scales only transactions but not users, (b) the transactions are less secure than Bitcoin, (c) it works only for Bitcoin's micropayments.

2) **Raiden Network** [53] is an off-chain scaling solution. It is the same as Bitcoin's Lightning Network that facilitates fast, low-fee, scalable, and privacy-preserving payments. The tokens are securely transferred between participants without prerequisite global consensus using balance proofs. The balance proof is digitally signed and hash-locked transfer. The Raiden Network leverages on "off-chain" payment channels to transfer the value. It is not necessary to record each transaction on Ethereum blockchain for completion. Instead of verifying individual transactions, Raiden Network verifies net claims resulted from off-chain transactions. It is interoperable and works with any token that follows Ethereum's standardized token API (Application Programming Interface) (ERC (Ethereum Request for Comment) 20).

3) **µRaiden** [54] is a fast and free off-chain ERC20 token transfer framework, more specialized to a smaller range of applications. Along with all the features of Raiden network, µRaiden allows free off-chain token transfer, fee is incurred for opening and closing of the channel. As it does not support multihop fee transfer, the payments are unidirectional to the predefined receivers.

4) **Trinity** [55] is a universal off-chain scaling solution, with features like real-time payments, low transaction fees, scalability, and privacy protection for main chain assets. The transaction throughput is suddenly increased with the use of a state channel. To enhance privacy Trinity adopts multiple technologies like zero-knowledge proof to protect data security. Trinity works only for payment channels.

5) **Sprites** [35] is a novel payment channel that reduces the collateral cost, which each hob incur along the route. The constant lock times are developed to improve transaction throughput in payment channel networks. The partial deposits and withdrawals are supported without interrupting the payment channel.

The different techniques like state channel, multi hop and constant lock time are used in payment channel solutions.

3.3.4 Side Chain

It is a separate blockchain attached to its parent blockchain using a two-way peg. The two-way peg allows interchanging of assets between the parent blockchain and the sidechain at a predetermined rate. The reverse happens when moving back from a sidechain to the main chain.

1) **Plasma** [36] was proposed by Ethereum co-founder Vitalik Buterin and Joseph Poon in August 2017 as a second-most deployed scalability solution for Ethereum blockchain that aims to increase transaction throughput. Plasma refers to a framework that allows creating unlimited numbers of child chains which are smaller copies of parent blockchain. A tree-like structure is generated by creating more chains on the top of each child chain. The child chain is a customized smart contract designed as per the demand of specific use case. The overall work of main chain will be elevated by each child chain, hence there will be no congestion in the main chain. Plasma is a better solution for decentralized applications for which high transaction fee is obtained from users.

2) **Pegged Side Chain** [37] is a technology that enables transfer of bitcoins and other ledger assets between multiple blockchains. It also prevents the assets from malicious attackers and ensures the atomicity of the transfers. Pegged Side Chains proposed a protocol named Two-way peg, transferring the assets from parent chain to side chain. The coins are sent from parent chain to a special output and are locked until a Simplified Payment Verification (SPV) proof is received on the pegged side chain. After sending the coin,

confirmation period protects the transfer from denial of service attack and deals latency for security. After unlocking, the newly transferred assets cannot be spent on the sidechain to avoid double spending. Same process is used to send coins from the sidechain to the parent chain.

3) **Liquidity Network** [56] proposed Nocust (Non-Custodial) [38], a secured and scalable commit-chain. A new data structure Merkleized Interval Tree: a multi-layered tree is used in Nocust. Every users' balance is stored in exclusive non-crossing interval space. The total balances are verified with the amount recorded in the smart contract, available on the parent chain. There is no limit on funds while transferring, receiving and interacting with parent chain. The real-time transactions are guaranteed by Nocust. The transaction delays are reduced without extra fees and mortgages. With a very low transaction fee, high throughput is achieved while scaling to one billion users.

Two-way peg, child chains and Merkleized Interval tree are implemented in side chain solutions that results in very low transaction fee while increasing the throughput.

Table 1 Comparison of Various Scalability Solutions

Sr. No.	Solutions	Strategy used	Through-put (TPS)	Latency (Seconds)	Block Size (MB)
1	SegWit	Segregate digital sign	7	--	4
2	Byteball	DAG	20-30	60	NA
3	Elastico	Sharding	40	800	1
4	Litecoin	Scrypt	56	150	4
5	Bitcoin-Cash	Increased Block size	61	--	32
6	Bitcoin-NG	Nakamoto consensus	100	NA	NA
7	Ouroboros	Coin-flipping protocol	257.6	120	NA
8	IoTA	DAG and Tangle	500	60	NA
9	Algorand	Byzantine Agreement	875	22	NA
10	LOGOS (Social n/w on blockchain)	Axios	2500	<3	1
11	Zilliqa	Sharding, parallel processing of transactions	2828	--	--
12	Omniledger	Sharding	3500	800	1
13	Conflux	block DAG	6400	270 - 444	NA
14	Nano	block-lattice architecture	7000	1 to 10	NA
15	Rapid Chain	Sharding	7380	8.7	1
16	Monoxide	Asynchron-ous consensus	11694	13-21	1
17	Ostraka	Node sharding	400000	--	1

NA - Not Applicable, "-" - indicates Not available

4 COMPARISON OF SCALABILITY SOLUTIONS

4.1 Comparison of Various Scalability Solutions

As per the literature survey, following scalability solutions are arranged in ascending order based on their performance in terms of throughput. Hafid [41] has categorized scalability solutions only at layer 1 and layer 2. Reference [42] though listed all three layer solutions but has not compared on the basis of pros and cons. From Tab. 1, Ostraka, layer 1 solution of sharding based on consensus, has the highest throughput of 400000 TPS which is invented recently. Among the solutions like Dash, Litecoin, Bitcoin cash and Bitcoin SV, it is observed that increase in throughput is proportional to block size. Block size is not applicable for the DAG scalability solutions, as the transactions are connected to each other. Among all DAG solutions such as Bitcoin NG IoTA, Nano, Ouroborous, Algorand and Conflux, the throughput of Nano is maximum i.e. 7000 TPS. Along with good throughput, solutions like Nano, Logos and Rapid chain have lowest latency.

4.2 Benefits and Limitations of Scalability Solutions

Not only the performance but the other features of existing solutions are compared in terms of their advantages and disadvantages in Tab. 2.

Table 2 Benefits and Limitations of Scalability Solutions

Solutions	Benefits	Limitations
SegWit	• Block size/ capacity increased. • Fixes transaction malleability issue. • Linearly scales the signature-hashing • Reduces UTXO growth	• Causes hard fork on Bitcoin • Needs to be more scalability
Bitcoin Cash	• Increase the throughput	• Lead to Centralization • Difficult to verify large number of transactions within short interval
Txilm	• Saves bandwidth	
Elastico	• Increase in throughput - 40 TPS	• Division of epoch can be influenced by malicious nodes • Total resiliency 25% and committee resiliency 33%
OmniLedger	• Bias resistant randomness protocol used for security • Throughput 500 TPS	• Total resiliency 25% and committee resiliency 33%
Rapid chain	• High throughput 4220 TPS • Total resiliency 33% and Committee resiliency 50%	• Partitioning attack • Responsiveness
Zilliqa	• Throughput 1000 times that of Ethereum	• Susceptible to single shard takeover attack 2) total resiliency 25% and committee resiliency 33%
Harmony	• Highly scalable • Sharding of blockchain state • High security	• Total resiliency 25% and committee resiliency 33%
Spectre	• Specially designed for payments	• Not suitable for smart contracts
Lightning Network	• Low cost • Faster payment • High throughput	• Does not scale users • Less secured • Works only for Bitcoin's micro payment
μRaiden	• Token transfer is free, only fee incurred is for opening and closing of channels.	• Does not support multi hop fee transfer, hence the transfer is unidirectional

Table 2 Benefits and Limitations of Scalability Solutions (continuation)

Solutions	Benefits	Limitations
Trinity	• Real-time payments • Low transaction fees • Scalable • Privacy protection	• Works only for payment channels
Plasma	• Lower transaction cost and faster operations • Secured • Does not need all participants to be online	• Long waiting period to withdraw funds • Complex implementation
Pegged Sidechains	• Communicates among Heterogeneous blockchains	• Transaction fund is saved as a deposit in the trading channel. • The transaction channel depends on complex routing topologies

5 CONCLUSION

Different scalability solutions proposed in the literature are classified and compared based on their performance measures (throughput, latency and strategies used). The solutions are classified into three layers. Layer 0 proposes solutions that uses erasure code and overlay topology, to enhance data propagation and reduce bandwidth usage. Layer 1 describes on-line solutions based on block size, compression of transactions, state and node sharding based on various consensus algorithms, directed acyclic graph etc. Layer 2 focuses on off-line solutions like payment channels, side chain, cross chain and off-chain computations using smart contracts, virtual machines, Tendermint, relay-chain, state channel, two-way peg, child chains and Merkleized interval tree.

This comprehensive study and classification of solutions at different layers can inspire researchers for further enhancement in the scalability of blockchain.

6 FUTURE WORK

The limitations listed in Tab. 2 indicates the areas as recommendations for further work. To enhance the total and committee resilience in sharding solutions. Along with scaling transaction throughput, users also should be scaled-up. There is a scope to improve scalability which is limited to only the payment channels.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

7 REFERENCES

[1] Haber, S. & Stornetta, W. S. (1990, August). How to time-stamp a digital document. In *Conference on the Theory and Application of Cryptography* (pp. 437-455). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-38424-3_32

[2] Nakamoto, S. & Bitcoin, A. (2008). A peer-to-peer electronic cash system. *Bitcoin.* https://bitcoin.org/bitcoin.pdf, 4.

[3] Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, *151*(2014), 1-32.

[4] Szabo, N. (1997). Formalizing and securing relationships on public networks. *First Monday*. https://doi.org/10.5210/fm.v2i9.548

[5] Zhang, R., Xue, R., & Liu, L. (2019). Security and privacy on blockchain. *ACM Computing Surveys (CSUR)*, *52*(3), 1-34. https://doi.org/10.1145/3316481

[6] Scherer, M. (2017). Performance and scalability of blockchain networks and smart contracts.

[7] Klarman, U., Basu, S., Kuzmanovic, A., & Sirer, E. G. (2018). bloxroute: A scalable trustless blockchain distribution network whitepaper. *IEEE Internet Things J.*

[8] Chawla, N., Behrens, H. W., Tapp, D., Boscovic, D., & Candan, K. S. (2019, May). Velocity: Scalability improvements in block propagation through rateless erasure coding. In *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)* (pp. 447-454). IEEE. https://doi.org/10.1109/BLOC.2019.8751427

[9] Rohrer, E. & Tschorsch, F. (2019, October). Kadcast: A structured approach to broadcast in blockchain networks. In *Proceedings of the 1st ACM Conference on Advances in Financial Technologies* (pp. 199-213). https://doi.org/10.1145/3318041.3355469

[10] Naumenko, G., Maxwell, G., Wuille, P., Fedorova, A., & Beschastnikh, I. (2019, November). Erlay: Efficient Transaction Relay for Bitcoin. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (pp. 817-831). https://doi.org/10.1145/3319535.3354237

[11] Lombrozo, E., Lau, J., & Wuille, P. (2015). Segregated witness (consensus layer). *Bitcoin Core Develop. Team, Tech. Rep. BIP*, *141*.

[12] Xu, Z., Han, S., & Chen, L. (2018, April). Cub, a consensus unit-based storage scheme for blockchain system. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)* (pp. 173-184). IEEE. https://doi.org/10.1109/ICDE.2018.00025

[13] Dai, X., Xiao, J., Yang, W., Wang, C., & Jin, H. (2019, July). Jidar: A jigsaw-like data reduction approach without trust assumptions for bitcoin system. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)* (pp. 1317-1326). IEEE. https://doi.org/10.1109/ICDCS.2019.00132

[14] Dang, H., Dinh, T. T. A., Loghin, D., Chang, E. C., Lin, Q., & Ooi, B. C. (2019, June). Towards scaling blockchain systems via sharding. In *Proceedings of the 2019 International Conference on Management of Data* (pp. 123-140). https://doi.org/10.1145/3299869.3319889

[15] Luu, L., Narayanan, V., Zheng, C., Baweja, K., Gilbert, S., & Saxena, P. (2016, October). A secure sharding protocol for open blockchains. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 17-30). https://doi.org/10.1145/2976749.2978389

[16] Kokoris-Kogias, E., Jovanovic, P., Gasser, L., Gailly, N., Syta, E., & Ford, B. (2018, May). Omniledger: A secure, scale-out, decentralized ledger via sharding. In *2018 IEEE Symposium on Security and Privacy (SP)* (pp. 583-598). IEEE. https://doi.org/10.1109/SP.2018.000-5

[17] Zamani, M., Movahedi, M., & Raykova, M. (2018, January). Rapidchain: Scaling blockchain via full sharding. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* (pp. 931-948).

https://doi.org/10.1145/3243734.3243853
[18] Team, Z. (2017). The ZILLIQA technical whitepaper. https://doi.org/10.2139/ssrn.3442330
[19] Wang, J. & Wang, H. (2019). Monoxide: Scale out blockchains with asynchronous consensus zones. In *16th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 19)* (pp. 95-112). https://doi.org/10.1016/j.automatica.2019.108620
[20] Eyal, I., Gencer, A. E., Sirer, E. G., & Van Renesse, R. (2016). Bitcoin-ng: A scalable blockchain protocol. In *13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16)* (pp. 45-59).
[21] Gilad, Y., Hemo, R., Micali, S., Vlachos, G., & Zeldovich, N. (2017, October). Algorand: Scaling byzantine agreements for cryptocurrencies. In *Proceedings of the 26th Symposium on Operating Systems Principles* (pp. 51-68). https://doi.org/10.1145/3132747.3132757
[22] Kiayias, A., Russell, A., David, B., & Oliynykov, R. (2017, August). Ouroboros: A provably secure proof-of-stake blockchain protocol. In *Annual International Cryptology Conference* (pp. 357-388). Springer, Cham. https://doi.org/10.1007/978-3-319-63688-7_12
[23] Pervez, H., Muneeb, M., Irfan, M. U., & Haq, I. U. (2018, December). A comparative analysis of DAG-based blockchain architectures. In *2018 12th International Conference on Open Source Systems and Technologies (ICOSST)* (pp. 27-34). IEEE. https://doi.org/10.1109/ICOSST.2018.8632193
[24] LeMahieu, C. (2018). Nano: A feeless distributed cryptocurrency network. https://nano.org/en/whitepaper (date of access: 24.03. 2018).
[25] Churyumov, A. (2016). Byteball: A decentralized system for storage and transfer of value. https://byteball.org/ Byteball.pdf
[26] Lewenberg, Y., Sompolinsky, Y., & Zohar, A. (2015, January). Inclusive block chain protocols. In *International Conference on Financial Cryptography and Data Security* (pp. 528-547). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-47854-7_33
[27] Sompolinsky, Y., Lewenberg, Y., & Zohar, A. (2016). SPECTRE: A Fast and Scalable Cryptocurrency Protocol. *IACR Cryptol. ePrint Arch.*, *2016*, 1159.
[28] Sompolinsky, Y., & Zohar, A. (2018). PHANTOM: A Scalable BlockDAG Protocol. *IACR Cryptol. ePrint Arch.*, *2018*, 104.
[29] Li, C., Li, P., Zhou, D., Xu, W., Long, F., & Yao, A. (2018). Scaling nakamoto consensus to thousands of transactions per second. *arXiv preprint arXiv:1805.03870*.
[30] Lerner, S. D. (2015). DagCoin: a cryptocurrency without blocks. *White paper*.
[31] Teutsch, J. & Reitwießner, C. (2019). A scalable verification solution for blockchains. *arXiv preprint arXiv:1908.04756*.
[32] Kalodner, H., Goldfeder, S., Chen, X., Weinberg, S. M., & Felten, E. W. (2018). Arbitrum: Scalable, private smart contracts. In *27th {USENIX} Security Symposium ({USENIX} Security 18)* (pp. 1353-1370).
[33] Wood, G. (2016). Polkadot: Vision for a heterogeneous multi-chain framework. *White Paper*.
[34] Poon, J. & Dryja, T. (2016). The bitcoin lightning network: Scalable off-chain instant payments.
[35] Miller, A., Bentov, I., Bakshi, S., Kumaresan, R., & McCorry, P. (2019, February). Sprites and state channels: Payment networks that go faster than lightning. In *International Conference on Financial Cryptography and Data Security* (pp. 508-526). Springer, Cham. https://doi.org/10.1007/978-3-030-32101-7_30
[36] Poon, J., & Buterin, V. (2017). Plasma: Scalable autonomous smart contracts. *White paper*, 1-47.
[37] Back, A., Corallo, M., Dashjr, L., et al. (2014). Enabling blockchain innovations with pegged sidechains. http://www.opensciencereview.com/papers/123/enablingblockchain-innovations-with-pegged-sidechains, 72.
[38] Khalil, R., Gervais, A., & Felley, G. (2018). NOCUST-A Non-Custodial 2nd-Layer Financial Intermediary. *IACR Cryptol. ePrint Arch.*, *2018*, 642.
[39] Pervez, H., Muneeb, M., Irfan, M. U., & Haq, I. U. (2018, December). A comparative analysis of DAG-based blockchain architectures. In *2018 12th International Conference on Open Source Systems and Technologies (ICOSST)* (pp. 27-34). IEEE. https://doi.org/10.1109/ICOSST.2018.8632193
[40] Manuskin, A., Mirkin, M., & Eyal, I. (2020, September). Ostraka: Secure blockchain scaling by node sharding. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS & PW)* (pp. 397-406). IEEE. https://doi.org/10.1109/EuroSPW51379.2020.00060
[41] Hafid, A., Hafid, A. S., & Samih, M. (2020). Scaling blockchains: A comprehensive survey. *IEEE Access*, *8*, 125244-125262. https://doi.org/10.1109/ACCESS.2020.3007251
[42] Zhou, Q., Huang, H., Zheng, Z., & Bian, J. (2020). Solutions to scalability of blockchain: A survey. *IEEE Access*, *8*, 16440-16455. https://doi.org/10.1109/ACCESS.2020.2967218
[43] "Litecoin" https://litecoin.org/
[44] https://www.scmspune.ac.in/journal/pdf/current/Paper%2010%20-%20Jaysing%20Bhosale.pdf
[45] "BIP 141" https://github.com/bitcoin/bips/blob/master/bip-0141.mediawiki#Block_size
[46] "Bitcoin cash" https://news.bitcoin.com/fork-watch-first-bitcoin-cash-block-mined
[47] https://github.com/bitcoin/bips/blob/master/bip-0152.mediawiki#Abstract
[48] "Harmony" https://harmony.one
[49] "Ethereum Sharding 2.0" Buterin. Ethereum Sharding FAQ. Available: https://github.com/ethereum/wiki/wiki/Sharding-FAQ
[50] "Logos" https://logos. network/whitepaper.pdf
[51] "IoTA" https://www. iota.org/get-started/what-is-iot
[52] "Cosmos" https://cosmos.network/resources/whitepaper
[53] "Raiden Network" https://raiden.network/
[54] "μRaiden" https://raiden.network/micro.html
[55] Trinity, Universal off-chain scaling solution, Trinity White Paper. Available: https://trinity.tech/#/whitepaper
[56] "Liquidity Network" https://liquidity.network
[57] "NXT" https://nxtwiki.org/wiki/Whitepaper:Nxt

Authors' contacts:

Jyoti Yadav
(Corresponding author)
Department of Computer Science,
University of Savitribai Phule Pune University,
Ganeshkhind, Pune, Maharashtra 411007, India
yadav.jyo@gmail.com

Ranjana Shevkar
PES's Modern College of Arts, Science and Commerce,
Ganeshkhind, Pune, Maharashtra 411016, India
rsshevkar@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210205104516

Subject review

Evolution of VANETS to IoV: Applications and Challenges

Faisal Rasheed Lone*, Harsh Kumar Verma, Krishna Pal Sharma

Abstract: Advancement in wireless communication technology along with the evolution of low power computational devices, have given rise to the Internet of things paradigm. This paradigm is transforming conventional VANETs into Internet-of- vehicles. This transition has led to a substantial commercial interest; as a result, there has been a significant boost in the field of the Internet of vehicles during the past few years. IoV promises a wide range of applications of commercial interest as well as public entertainment and convenience (collision warning systems, on-demand in-car entertainment, smart parking, traffic information). Applications related to vehicular and passenger safety are particularly of great commercial as well as a research interest as such IoV is going to be a core component in implementing the smart city concept. This paper gives an overview of the transition of conventional VANETs to IoV and highlights the potential applications and challenges faced by the Internet of Vehicles (IoV) paradigm.

Keywords: IoT; IoV; Security; VANET

1 INTRODUCTION

Of the 25-30 billion things envisioned to be connected to the Internet by 2020 [1], vehicles will contribute a compelling number. IoV is considered as a unique application of the Internet of Things (IoT). IoV is a dynamic mobile network establishing communication between heterogeneous networks using Vehicle-to-Vehicle (V2V), Vehicle-to-Roadside (V2R), Vehicle-to-humans (V2H), Vehicle-to-Sensors (V2S) and Vehicle-to-Infrastructure (V2I) [2, 3]. V2X (Vehicle to everything) communication will form the backbone of the Intelligent transportation systems [4, 5]. Conventional VANETs are transforming into the Internet of vehicles. While as VANETs consist of only vehicles connected in an ad-hoc manner exchanging data with each other, IoV spans a bigger network involving entities such as humans, things and other heterogeneous networks. IoV improves upon VANETs by incorporating cellular networks such as LTE, 5G etc., to provide an expansive and reliable communication. IoV uses ad-hoc networks for communication with infrastructure and a cellular network for communication with the backbone network [6, 7]. IoV enables data collection and sharing about surroundings, vehicles, road conditions [8].

Over the years, not much interest has been shown in VANETs as it was not able to provide global sustainable applications and services to users. Nevertheless, with the emergence of IoV, the desired commercial interests have begun to emerge. VANETs can thus be considered as a subset of IoV [9], offering more expansive coverage and application as such IoV is capable of providing services to a much larger area, such as a city or even a country. In comparison to VANETs, IoV treats a vehicle as a smart-entity equipped with various sensors and computational capability. The purpose of the IoV is to assimilate humans, vehicles, things and other heterogeneous networks to boost efficiency, security and safety of transport systems, to provide various services to a smart city (e.g. parking information, traffic information), services to users (e.g. on-demand in-car entertainment, remote diagnostics) and several other useful applications [10, 11].

IoV has brought about an opportunity to improve passenger safety and mobility by utilizing vehicular communication.

The importance of IoV can be highlighted by the Vehicle Infrastructure Integration program (VII), developed by the US department of transportation (DOT), acting as a prime component of the Intelligent Transportation System. The main objective of the program was to develop a wireless communication system to support the Vehicle to Vehicle and Vehicle to Infrastructure communications to enhance safety and traffic mobility [12].

Taking into account the immense benefits offered by IoV and the enormous number of vehicular nodes, it is evident that IoV is going to be one of the largest and most complex networks in the near future. With the decline in prices of electronic components and a neat amalgamation of these onboard devices present immense business opportunities, but at the same time give rise to intimidating research challenges. These issues might look akin to the ones faced in traditional ad-hoc networks, but various inherent characteristics such as the immense size of the network, speed of the vehicular nodes, intermittent connectivity,safety-critical applications, give rise to novel challenges as well as amplify the severity of traditional issues.

The paper is organized in the following way. In Section 2, we present the related work in the field. Section 3 describes the general architecture of IoV, followed by Applications of IoV in Section 4. In Section 5, we give an overview of various challenges faced by the IoV paradigm, and section 6 finally concludes the paper.

2 LITERATURE SURVEY

A common question usually faced is "How is IoV different from IoT and Wireless networks?". The answer lies in the fact that some characteristics of IoV are quite different when compared to IoT and wireless networks. A VANET transforms a vehicle into a mobile router [13]. Both wireless networks and IoV consist of mobile nodes. However, the trajectory followed by the nodes in IoV consists of a defined path which is subject to the road network of an area. In

contrast, the trajectory followed by nodes in a mobile network has no defined path and follow a random walk model. Moreover, IoT lays more stress on *"things"* and *"data awareness"* of these things, while as the main motive of IoV is to integrate humans and vehicles, where vehicles may be considered as an extension of human's capabilities. IoV connects vehicles, humans in and around the Vehicle, systems onboard the Vehicle by incorporating the sensors, vehicles and mobile devices forming a global network enabling it to deliver services and applications to humans around or onboard the Vehicle.

With the growth and advancement of wireless technologies, we can think of integrating vehicles in the network, thus constituting the Internet of Vehicles environment. Communication is made possible by incorporating short as well as long-range communication technologies [14] Many existing wireless technologies can be used to implement the IoV paradigm, including WLAN's, WIMAX, cellular wireless as well as satellite communications. Much research is focused on the wireless aspect of IoV, as a robust wireless scheme can immensely enhance the implementation of IoV by providing relatively better Quality of service (QoS). Each of the current crops of wireless technologies has their share of merits and demerits as far as IoV is concerned.

WLAN has achieved a lot of significance and approval in the market due to robust short-range communication, including high-speed transmissions. WLAN consists of IEEE 802.11 a/b/g/n/ac standards [8] each with a different capacity, modulation technology, bandwidth and coverage area. Performance analysis of IEEE 802.11 a, b, g in-car communication scheme showed that velocity up to 180 km/h has almost negligible impact on the performance [15]. Due to its shorter range, WLAN cannot sustain a VANET for longer duration due to constant movement of vehicles, leading to frequent topology changes. Performance of 802.11p based protocol in Vehicle to vehicle safety applications was analyzed in [16].

Table 1 Comparison of IEEE 802.11 Standards

IEEE 802.11 Standard	Operating Frequency (GHz)	Bandwidth	Beamforming	Coverage	Capacity	Interference	Quality
a	2.4	20	No	Low	Low	High on 2.4/Low on 5	Low
	5						
b	2.4	22	No	Low	Low	High	Low
g	2.4	20	No	Low	Low	High	Low
n	2.4/5	20	Yes	Low	Low	High on 2.4/Low on 5	Low
		40					
ac	5	20	Yes	High	High	Low	High
		40					
		60					
		80					

WiMAX covers IEEE 802.16 a/e/m standards which can cover more than 50 km/h of geographical area and can deliver a theoretical bandwidth up to 72 Mbps [15]. The IEEE 802.16 standard provides support for fixed broadband wireless only. In contrast, as IEEE 802.16e/mobile version supports mobility up to speeds of 160 km/h and various QoS provisions for non-line of sight communications [15]. The leverage WIMAX offers over WLAN is that a user need not frequently compete for entry into the network, instead only once during initial entry. Cellular wireless possesses a considerable potential for IoV due to its widespread penetration and already established robust infrastructure. The cellular network comprises of older technologies such as 3G, 4G, LTE, and more recently 5G. The third-generation networks (3G) are capable of delivering peak data rates of up to 3 Mbps in fixed and 384 Kbps in the mobile environment [17]. Tab. 1 compares the parameters of various IEEE 802.11 standards.

Authors in [18] evaluated the performance of data delivery in VANETs using 3G. They concluded that due to a centralized authority (Mobile Switching Centre), latency might be an issue as most of the VANET applications are delay-sensitive. With the widespread penetration of 4G networks and the advantages, it offers over 3G in terms of bandwidth, latency, coverage, especially LTE networks. It currently is the most feasible wireless technology suited to implement IoV. Nevertheless, with the advent of 5G in coming years, integration of IoV with 5G networks will have to be accomplished due to the benefits 5G is going to offer over 4G. While integrating IoV with 5G, resource management needs to be accomplished in tandem with user requirements [19].

The increased internet connectivity offered by vehicles has led to challenges in securing the vehicular network as the attack surface area increases [20]. Modern vehicles allow devices to be connected via infotainment systems, creating a pathway to inject malware compromising ECU leading to unforeseen consequences. A framework for vehicular malware characterization and protection has been proposed in [20]. The framework uses virtualization to reduce the vehicular attack surface.

A secure authentication protocol for IoV has been proposed in [21] which mitigates attacks such as offline identity guessing attacks, replay attacks, spoofing and also reduced the authentication time. Although secure, the proposed protocol assumes a secure communication channel between RSU and the TA. The protocol also does away with the usage of passwords as it assumes robust physical vehicular security to prevent theft of smart cards.

Impact of attacks on AODV and GPSR was studied in [22]. Authors simulated attacks on both AODV and GPSR in a vehicular network. Scenarios with high traffic as well as

low traffic density, were considered. It was concluded that AODV is affected more than GPSR by the attacks in a vehicular environment and stressed for the need of security mechanisms in a highly dynamic vehicular environment.

A convolutional network-based Intrusion detection system (IDS)was proposed in [23]. The proposed system is able to intercept and detect network attacks by running on a low powered embedded vehicular terminal by real-time data monitoring.

3 IoV ARCHITECTURE

IoV comprises of a variety of heterogeneous networks and devices communicating with each other. IoV requires a robust architecture that can handle heterogeneity, scalability and various other requirements specific to the IoV paradigm [24]. The general architecture IoV comprises of three layers [25] as depicted in Fig. 1.

Perception/Client Layer comprises of sensors responsible for data gathering, environmental information, location, driving patterns and a lot more. The gathered data collected from the sensors is evaluated/transmitted for further course of action.

The network layer enables communication within vehicular nodes (V2V) and other networks and entities. This layer basically supports all communications within IoV.

The application layer hosts actual applications offered by IoV, manages user interaction, storage, decision making based on data analysis and entertainment and convenience applications such as in-car entertainment, traffic information and much more.

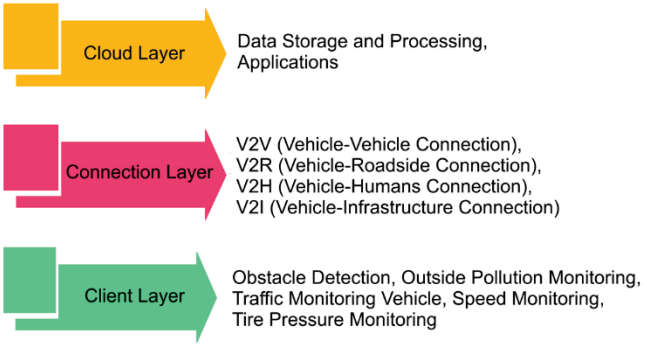


Figure 1 Three-layered IoV architecture.

4 IoV APPLICATIONS

IoV has come up as a very promising paradigm with evolving technology, especially wireless networks and the Internet of Things (IoT). IoV has matured from VANETS to a more significant entity which encapsulates pedestrians, roads, parking lots, city infrastructure, and connects them to provide real-time communication among these entities [26].

IoV provides an extensive and elaborate list of applications [26] which may be broadly grouped as:

- Safety
- Transport Efficiency
- Information/Entertainment/Convenience
- Logistics.

Each of these application categories comprises of multiple applications responsible for enhancing the acceptability of the IoV paradigm; Tab. 2 enlists a few such applications.

Safety in connected vehicles implies that nodes can automatically send real-time crash information, including location information to emergency teams which can expedite the response process and thus help in saving lives.

Cooperative collision warning systems can detect a probable collision and display a warning to the driver to prevent collision [26]. Cooperative forward collision warning systems detect the distance between the cars and alert the driver to prevent rear-end collisions [27]. Driving through an intersection is one of the most challenging tasks a driver faces due to the convergence of multiple traffic streams, thus posing a high possibility of collision [28]. Few approaches have been proposed to avoid collisions at intersections. An abstraction-based algorithm to avoid collisions has been proposed in [29]. An algorithm that can handle a large number of vehicles based on time slots has been proposed in [30]. An intelligent intersection is on the most sought-after safety applications of IoV. Vehicle safety consortium (VSC) has identified some safety-related applications including curve speed warning, pre-crash sensing, cooperative forward collision warning, emergency brake light warning, left turn assistant, traffic signal violation, lane change warning and stop sign movement assistant [26].

Table 2 IoV Applications

Safety	Transport Efficiency	Entertainment /Convenience/ Information
Crash SOS Cooperative Collision Warning Cooperative Forward-Collision Warning Roadside Assistance Left Turn Assistance Lane Change Warning Stop Sign Movement Assistance	Route Guidance and Optimization Green Light Efficiency Traffic Information	Content Streaming Electronic Toll Collection Point of Interest Notification

Some of these applications require Vehicle to Vehicle (V2V) communications while as some require Vehicle to roadside communication (V2R). As such widespread roadside infrastructure for vehicular communication needs to be established in order to implement such safety applications.

Transport efficiency applications such as route guidance and optimization, green light efficiency, can be implemented via IoV. Some of these applications require roadside infrastructure, which some require Vehicle to Vehicle communication only [26]. These applications of IoV

will not only enhance traffic management but will also result in a considerable reduction in journey time, fuel consumption as well as a reduction in pollution levels due to fewer traffic jams. IoV is going to be a key driving force in implementing the smart city concept.

Logistics can benefit from IoV due to its dependence on road transport. Road transportation is essential for timely and efficient delivery of goods. Using IoV in logistics can have far-reaching benefits. IoV can be applied to smart logistics fleet management [31] by a logistic company to improve the delivery of goods. Authors in [31] suggested a smart logistics vehicle management system based on IoV [31]. The primary objective of the proposed system is to equip logistic vehicles with sensors with logistics data control centre as its backbone as visualized in Fig. 2. Working of the proposed system can be explained as follows:

- Logistic vehicles will be equipped with sensors, and these sensors will capture different types of data such as tyre pressure, location, vehicle diagnostics, the status of goods, driving conditions, temperature etc. [31]. A public information centre will provide information such as traffic conditions, weather information and more.
- The goods being transported will be equipped with RFIDs and bar codes that will send data about the goods being carried to the logistics data control centre [31].
- The logistics data control centre will analyze the received data and generate instructions to be sent back based on the result analysis.

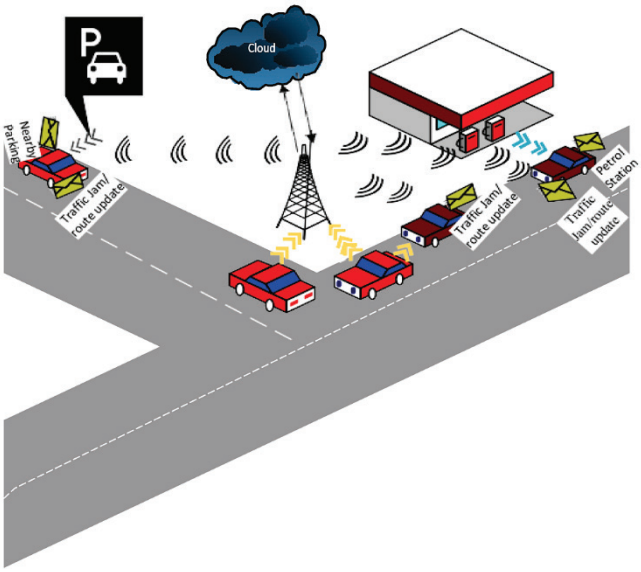


Figure 2 Points of interest and traffic information relayed by nearby roadside units as well as Vehicle-to-Vehicle communication

A prime factor in implementing transport applications in IoV is security and the trust level of the information disseminated. An authentication model for the disseminated information in IoV environment needs to be developed so that bogus and illegitimate information may be stopped from spreading. Bogus information can be detrimental in IoV environment, especially in safety-related applications. Thus, security is a prime concern in IoV and needs further research.

Entertainment/information/convenience include content streaming to in-car entertainment systems, point of interest notifications such as petrol pumps, restaurants, restrooms etc., as represented in Fig. 3. Convenience and entertainment applications range from electronic toll deductions at tolling stations, relay of vehicle diagnostics to the manufacturer for an automated schedule of service, web browsing, streaming to roadside services such as location and price broadcast of fuel stations, parking, restaurants and many more [32].

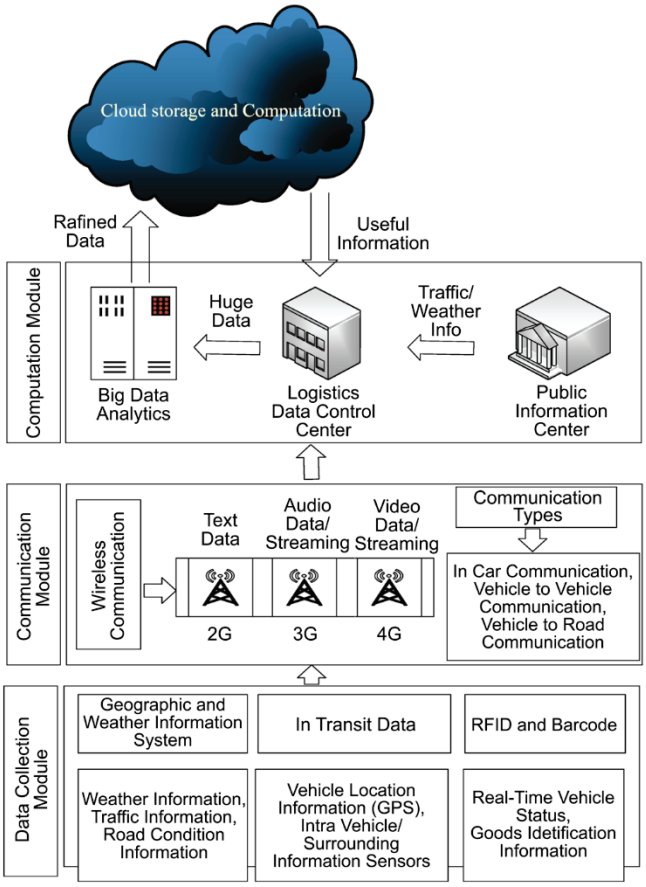


Figure 3 Smart Logistic fleet management system.

5 CHALLENGES

Due to the inherent nature of IoV, it is subject to additional challenges compared to other wireless networks due to its open nature [33]. In IoV no central coordinator can be assumed [26]. Most of the applications are expected to work reliably and efficiently in a decentralized manner [26]. As a result, the need for a single shared channel can be derived. The challenges faced by other wireless networks also inherit themselves into IoV [26], problems faced due to hidden and exposed terminals are evident. Due to the transmission medium being wireless, Medium Access Control (MAC) poses a challenge in the design of VANETs, as many nodes must compete for available channels. The frequency channels assigned to VANET's currently possess bandwidth of 10-20 MHz only. With increasing traffic density and more and more cars being connected, these

channels could choke, causing congestion, posing severe channel management issues [26].

Security is a severe concern in networks, especially wireless networks, and the same applies to VANETs as well [5, 33]. Data integrity, authenticity and trust establishment is a severe issue in vehicular networks [34]. An attacker can fool the network by spreading fake messages [21]. A trust mechanism needs to be established where the receivers are sure enough about the information received. Trust management is a complicated issue in IoV A trade-off must be achieved where users can trust the information without compromising privacy requirements [35]. New paradigms such as Blockchains need to be analyzed for improving user privacy [35, 36, 37]. Achieving a secure communication in IoV is challenging due to a variety of possible attacks [33].

Authors in [38] highlighted attacks that can be carried out in IoV and VANETs. It was concluded that vehicular networks are more susceptible due to their unique characteristics, such as the absence of central authority, mobility, wireless links, cooperativeness, lack of proper lines of defence and scalability [38]. The authors in [38] also proposed defence mechanisms for such attacks but concluded that further research is required With the increasing popularity of autonomous cars, more security issues are creeping [14]. The vulnerability of IoV due to its open nature makes attack detection mandatory [39].

Adoption of VANET equipped vehicles is another issue that needs attention. More and more people must be convinced to buy and use VANET equipped vehicles. The value and benefit provided to a customer in a VANET depend on the total number of customers using VANETs [26] Thus, a key factor is to convince early users to buy VANET equipped vehicles. It can be done by luring customers by offering discounts on VANET equipped vehicles, by installing roadside equipment's for VANETs and enforcement by law. Roadside infrastructure and back end IT connectivity required for VANETs is required to be put in place for smooth and trouble-free deployment.

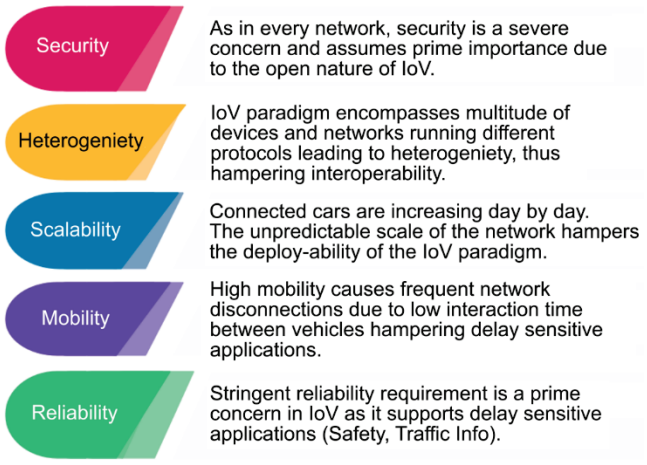


Figure 4 IoV Challenges

The use of cryptographic techniques such as PKI is not particularly suitable for VANETs due issues in key distribution and limited communication time between vehicular nodes [40]. Fig. 4 highlights some of the challenges faced by the IoV paradigm.

6 CONCLUSION

IoV is emerging as a vital component of the smart city concept. It is a unique application of the Internet of Things paradigm. It is a complex network system comprising of heterogeneous devices and networks communicating with each other. As important and beneficial IoV and VANETs are, they have their share of limitations and challenges as discussed. The deploy-ability and rollout of IoV depend on mitigating the issues and challenges faced by these highly dynamic networks. Security and privacy are two factors of utmost importance that need focus. As most of the nodes in a VANET are autonomous, the authenticity and reliability of the disseminated information raise concerns. Trusted communication is the need of the hour as far as VANETs are concerned. A trusted communication can be implemented when a sender is always accepted as a trusted source, and it is made sure that the message in transit has not been tampered with. Thus, security is a prime concern in IoV and VANETs and needs further research. Finally, there are other challenges faced by IoV and VANETs regarding market introduction and demonstration of their capabilities. Efficient methods and policies need to be drafted for the sustainability of services provided as vehicles are introduced as mobile nodes.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

7 REFERENCES

[1] Al-Fuqaha, A. et al. (2015). Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Communications Surveys & Tutorials. 17. Fourthquarter 2015, 17*(4). https://doi.org/10.1109/COMST.2015.2444095

[2] Alouache, L., Nguyen, N., Aliouat, M., & Chelouah, R. (2018). Survey on IoV routing protocols: Security and network architecture. I*nternational Journal of Communication Systems*, 32. https://doi.org/10.1002/dac.3849

[3] Sheikh, M. S., Liang, J., & Wang, W. (2020). Security and Privacy in Vehicular Ad Hoc Network and Vehicle Cloud Computing: A Survey. *Wireless Communications and Mobile Computing, Volume 2020*, Article ID 5129620. https://doi.org/10.1155/2020/5129620

[4] Liang, H., Wu, J., Mumtaz, S., Li, J., Lin, X., & Wen, M. (2019). MBID: Micro-Blockchain-Based Geographical Dynamic Intrusion Detection for V2X. *IEEE Commun. Mag., 57*(10), 77-83. https://doi.org/10.1109/MCOM.001.1900143

[5] Sherazi, H. H. R., Iqbal, R., Ahmad, F., Khan, Z. A., & Chaudary, M. H. (2019). DDoS attack detection: A key enabler for sustainable communication in internet of vehicles. *Sustain. Comput. Informatics Syst., 23*, 13-20. https://doi.org/10.1016/j.suscom.2019.05.002

[6] Marquez-Barja, J. M. et al. (2015). Breaking the Vehicular Wireless Communications Barriers: Vertical Handover Techniques for Heterogeneous Networks. *IEEE Transactions on Vehicular Technology, 64*(12), 5878-5890. https://doi.org/10.1109/TVT.2014.2386911

[7] Tornell, S. M., Patra, S., Calafate, C. T., Cano, J.-C., & Manzoni, P. (2015). GRCBox: Extending Smartphone Connectivity in Vehicular Networks. *Int. J. Distrib. Sens. Networks, 11*(3), p. 478064. https://doi.org/10.1155/2015/478064

[8] Wu, W., Yang, Z., & Li, K. (2016). Internet of Vehicles and applications. *Internet of Things: Principles and Paradigm*s, 299-317. https://doi.org/10.1016/B978-0-12-805395-9.00016-2

[9] Sharma, S. & Kaushik, B. (2019). A survey on internet of vehicles: Applications, security issues & solutions. *Vehicular Communications, 20*, 100182. https://doi.org/10.1016/j.vehcom.2019.100182

[10] Sun, Y. et al. (2017). Attacks and countermeasures in the internet of vehicles. *Ann. des Telecommun. Telecommun., 72*(5-6), 283-295. https://doi.org/10.1007/s12243-016-0551-6

[11] Sheikh, M. S., Liang, J., & Wang, W. (2020). Security and Privacy in Vehicular Ad Hoc Network and Vehicle Cloud Computing: A Survey. *Wirel. Commun. Mob. Comput., vol. 2020*, 1-25. https://doi.org/10.1155/2020/5129620

[12] Nishiyama, H., Ngo, T., Oiyama, S., & Kato, N. (2015). Relay by Smart Device: Innovative Communications for Efficient Information Sharing among Vehicles and Pedestrians. *IEEE Veh. Technol. Mag., 10*(4), 54-62. https://doi.org/10.1109/MVT.2015.2481558

[13] Dua, A., Kumar, N., & Bawa, S. (2014). A systematic review on routing protocols for Vehicular Ad Hoc Networks. *Vehicular Communications, 1*(1), 33.52. https://doi.org/10.1016/j.vehcom.2014.01.001

[14] Abu Talib, M., Abbas, S., Nasir, Q., & Mowakeh, M. F. (2018). Systematic literature review on Internet-of-Vehicles communication security. *Int. J. Distrib. Sens. Networks, 14*(12). https://doi.org/10.1177/1550147718815054

[15] Wellens, M., Westphal, B., & Mähönen, P. (2007). Performance Evaluation of IEEE 802.11-based WLANs in Vehicular Scenarios. *2007 IEEE 65th Vehicular Technology Conference - VTC2007*, Corpus ID: 7404711. https://doi.org/10.1109/VETECS.2007.247

[16] Yao, Y., Rao, L., Liu, X., & Zhou, X. (2013). Delay analysis and study of IEEE 802.11p based DSRC safety communication in a highway environment. *Proceedings - IEEE INFOCOM*, 1591-1599. https://doi.org/10.1109/INFCOM.2013.6566955

[17] Yang, F., Wang, S., Li, J., Liu, Z., & Sun, Q. (2014). An overview of Internet of Vehicles. *China Commun., 11*(10), 1-15. https://doi.org/10.1109/CC.2014.6969789

[18] Zhao, Q., Zhu, Y., Chen, C., Zhu, H., & Li, B. (2013). When 3G meets VANET: 3G-assisted data delivery in VANETs. *IEEE Sens. J., 13*(10), 3575-3584. https://doi.org/10.1109/JSEN.2013.2265304

[19] Aloqaily, M., Balasubramanian, V., Zaman, F., Al Ridhawi, I., & Jararweh, Y. (2018). Congestion mitigation in densely crowded environments for augmenting QoS in vehicular clouds. *DIVANet 2018 - Proceedings of the 8th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications*, 49-56. https://doi.org/10.1145/3272036.3272038

[20] Iqbal, S., Haque, A., & Zulkernine, M. (2019). Towards a security architecture for protecting connected vehicles from malware. *IEEE Vehicular Technology Conference, vol. 2019*. https://doi.org/10.1109/VTCSpring.2019.8746516

[21] Chen, C. M., Xiang, B., Liu, Y., & Wang, K. H. (2019). A secure authentication protocol for internet of vehicles. *IEEE Access, 7*, 12047-12057. https://doi.org/10.1109/ACCESS.2019.2891105

[22] Mintemur, O. & Sen, S. (2017). Attack Analysis in Vehicular Ad Hoc Networks. *Computer Science & Information Technology (CS & IT)*, 35-46. https://doi.org/10.5121/csit.2017.71103

[23] Peng, R., Li, W., Yang, T., & Huafeng, K. (2019). An Internet of Vehicles Intrusion Detection System Based on a Convolutional Neural Network. *2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, Xiamen, China, 1595-1599. https://doi.org/10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00234

[24] Tuyisenge, L., Ayaida, M., Tohme, S., & Afilal, L.-E. (2018). Network Architectures in Internet of Vehicles (IoV): Review, Protocols Analysis, Challenges and Issues. In: Skulimowski A., Sheng Z., Khemiri-Kallel S., Cérin C., Hsu CH. (eds) *Internet of Vehicles. Technologies and Services Towards Smart City. IOV 2018. Lecture Notes in Computer Science, vol 11253*. Springer, Cham. https://doi.org/10.1007/978-3-030-05081-8_1

[25] Huang, J. M. (2013). Research on Internet of Vehicles and its Application in Intelligent Transportation. *Applied Mechanics and Materials*, 321-324, 2818-2821. https://doi.org/10.4028/www.scientific.net/AMM.321-324.2818

[26] Hartenstein, H. & Laberteaux, K. P. (2008). A tutorial survey on vehicular ad hoc networks. *IEEE Communications Magazine, 46*(6), 164-171. https://doi.org/10.1109/MCOM.2008.4539481

[27] Kowshik, H., Caveney, D., & Kumar, P. R. (2011). Provable systemwide safety in intelligent intersections. *IEEE Trans. Veh. Technol., 60*(3), 804-818. https://doi.org/10.1109/TVT.2011.2107584

[28] Toor, Y., Mühlethaler, P., Laouiti, A., & De La Fortelle, A. (2008). Vehicle ad hoc networks: Applications and related technical issues. *IEEE Commun. Surv. Tutorials, 10*(3), 74-88. https://doi.org/10.1109/COMST.2008.4625806

[29] Colombo, A. & Del Vecchio, D. (2011). Supervisory control of differentially flat systems based on abstraction. *Proceedings of the IEEE Conference on Decision and Control*, 6134-6139. https://doi.org/10.1109/CDC.2011.6160759

[30] Colombo, A. & Del Vecchio, D. (2012). Efficient algorithms for collision avoidance at intersections. *HSCC'12 - Proceedings of the 15th ACM International Conference on Hybrid Systems: Computation and Control*, 145-154. https://doi.org/10.1145/2185632.2185656

[31] Sharma, N., Chauhan, N., & Chand, N. (2016). Smart logistics vehicle management system based on internet of vehicles. *The 4th International Conference on Parallel, Distributed and Grid Computing, PDGC 2016*, 495-499. https://doi.org/10.1109/PDGC.2016.7913245

[32] Hossain, E. et al. (2010). Vehicular telematics over heterogeneous wireless networks: A survey. *Comput. Commun., 33*(7), 775-793. https://doi.org/10.1016/j.comcom.2009.12.010

[33] Raya, M. & Hubaux, J. P. (2007). Securing vehicular ad hoc networks. *J. Comput. Secur., 15*(1), 39-68. https://doi.org/10.3233/JCS-2007-15103

[34] Bagga, P., Das, A. K., Wazid, M., Rodrigues, J. J. P. C., & Park, Y. (2020). Authentication protocols in internet of vehicles: Taxonomy, analysis, and challenges. *IEEE Access, 8*, 54314-54344. https://doi.org/10.1109/ACCESS.2020.2981397

[35] Butt, T., Iqbal, R., Salah, K., Aloqaily, M., & Jararweh, Y. (2019). Privacy management in social internet of vehicles: Review, challenges and blockchain based solutions. *IEEE Access, 7*, 79694-79713. https://doi.org/10.1109/ACCESS.2019.2922236

[36] Taiyaba, M., Akbar, M. A., Qureshi, B., Shafiq, M., Hamza, H., & Riaz, T. (2020). Secure V2X Environment using Blockchain Technology. *EASE '20: Proceedings of the Evaluation and Assessment in Software Engineering*, 469-474. https://doi.org/10.1145/3383219.3383287

[37] Ramaguru, R. B., Sindhu, M., & Sethumadhavan, M. (2019). Blockchain for the Internet of Vehicles. In: Singh M., Gupta P., Tyagi V., Flusser J., Ören T., Kashyap R. (eds) Advances in Computing and Data Sciences. *ICACDS 2019. Communications in Computer and Information Science, vol 1045*. Springer, Singapore. https://doi.org/10.1007/978-981-13-9939-8_37

[38] Sakiz, F. & Sen, S. (2017). A Survey of Attacks and Detection Mechanisms on Intelligent Transportation Systems: VANETs and IoV. *Ad Hoc Networks, 2017.* https://doi.org/10.1016/j.adhoc.2017.03.006

[39] Hasan, M., Mohan, S., Shimizu, T., & Lu, H. (2020). Securing Vehicle-to-Everything (V2X) communication platforms. *IEEE Transactions on Intelligent Vehicles, 5*(4), 693-713. https://doi.org/10.1109/TIV.2020.2987430

[40] Malhi, A. K., Batra, S., & Pannu, H. S. (2020). Security of vehicular ad-hoc networks: A comprehensive survey. *Comput. Secur., 89*, p. 101664. https://doi.org/10.1016/j.cose.2019.101664

Authors' contacts:

Faisal Rasheed Lone
(Corresonding author)
BGSB University,
National Highway Road 144A, Dhanore,
Rajouri, Jammu and Kashmir 185234, India
fasulone@gmail.com

Harsh Kumar Verma
NIT Jalandhar,
GT Road Byepass, NH-1,
Jalandhar 144011, Punjab, India
vermah@nitj.ac.in

Krishna Pal Sharma
NIT Jalandhar,
GT Road Byepass, NH-1,
Jalandhar 144011, Punjab, India
sharmakp@nitj.ac.in

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https:doi.org/10.31803/tg-20201017094254

Subject review

Application of Compromise Programming in the Energy Generation Planning

Fisnik Osmani, Atanas Kochov*, Mirjeta Ilazi

Abstract: The need for the right electricity planning is closely related to the analysis that has to be done on its generating capacity. Of course, in order to make such planning, the choice of the method by which the process is carried out must be considered. The inclusion of multiple objectives (goals) enables us to create models that are useful in researching and expanding electricity generation systems. Compromised programming used in this paper aims to find compromise solutions among different conflicting objectives in the electricity generation system. One such case is the goal of minimizing the cost and CO2 emissions. Using this logic, the paper deals with an example of electricity generation planning for a given region.

Keywords: compromising programming; emissions; energy planning; generation costs; MCDM

1 INTRODUCTION

This paper assumes annual electricity which can be generated by the use of available energy sources including: coal, wind energy, biomass, hydropower, and solar energy. With the application of compromise programming, taking into account the minimization of energy generation cost and CO_2 emissions, the optimal solution is achieved by calculating groups of compromise for different weights.

As a concept, planning can be defined as the first and the most basic function of management. When planning needs to be based on results and a lot of goals, then of course the right methods are needed, in order for optimal solutions to be acheived. The linear programming method, as the traditional approach to modeling decisions about power generation, relies on some basic assumptions about situation modeling and in this approach the decision maker intends to choose a well-defined objective.

The transition from the current electricity system to a renewable electricity supply poses immense economic, technological, and policy challenges. Energy system models represent the complexity of interactions in combined processes from extraction of primary energy to the use of the final energy to supply services and goods. [1]

Last year, important changes in the sectors of energy supply have happened, which have had important implications for energy planning. By 2050, 49–67% of primary energy will be supplied by renewable energy sources (RES), by lowering energy use and decarbonizing energy supply in the built-up environment [2]. Electricity generation has some environmental impacts, which however are not taken into account when allocating resources efficiently. According to an estimate, gross demand for energy in Kosovo is expected to increase by 4.6% a year over the 2010-2025 periods. [3]

Defining objectives is an essential part of the planning process, useful to induce creative alternatives and to derive the attributes (criteria) on which the alternatives will be assessed. Practically, the decision maker seeks consistent compromise between several objectives, many of which may be in conflict with one another. In addition to the economic issue, an emission issue should also be considered in the operation of an industrial consumer in order to reduce greenhouse gases like NO_2, SO_2, and CO_2 in the atmosphere. [4, 5]

This paper provides a simplified electricity planning model with some energy planning assumptions in order to reduce operational complexity. The introduction of environmental costs is done through a related variable, such as CO_2 emissions of environmental sustainability.

Based on international analysis and reports for the Western Balkans, it is clear in what indicators / key components the greatest focus is required

The inclusion of multiple objectives (goals) enables us to create models that are useful in researching and expanding electricity generation systems. Multi Criteria Decision Making Method is the method preferred by many authors, referring to the Compromise, Topsis, Pareto, Electre, Promethee, AHP method etc. [6-9].

The conceptual approach applied in the paper where we are limited to only two purposes, namely the cost of generation and emissions, requires an additional discussion because other methods with the inclusion of more factors are currently available.

2 ELECTRICITY GENERATION PLANNING, THROUGH THE APPLICATION OF COMPROMISED PROGRAMMING.

2. 1 Formulation of the Problem within the Conditions of Republic of Kosovo

For our region, annual electricity consumption is 5000 GWh [10], which can be generated by the use of the following energy sources: coal, wind power, biomass, Hydropower, and solar power. The cost of generation, as well as the generated environmental emissions, should be minimal.

Tab. 1 shows the generation costs and CO_2 emissions for the options considered. To refer to electricity output, both options are set in relation to the kWh generated.

Under these assumptions, the formulation of the problem is:

Minimized $z(x) = [z_1(x),\ z_2(x)]$

$z_1(x) = (c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 + c_5x_5)/\sum x_i$, euro/MWh

$z_2(x) = (e_1x_1 + e_3x_3)/\sum x_i$, g/MWh

$z_1(x) = (28x_1 + 85x_2 + 71x_3 + 77.3x_4 + 136.5x_5)/\sum x_i$, euro/kWh

$z_2(x) = (1015x_1 + 101x_3)/\sum x_i$, g/kWh

The above functions are subjected to the following limitations

$\sum x_i \geq 5000$ GWh; $x_i \geq 0,\ \forall i$

$x_2 \leq 200$ GWh; $x_3 \leq 100$ GWh;

$x_4 \leq 500$ GWh; $x_5 \leq 200$ GWh.

Table 1 Consumption, costs and emissions of CO_2 from electricity generation [11, 12]

Nr.	Energy source	Consumption, GWh/year		Costs, euro/MWh		Emissions CO_2, g/kWh	
1	Coal	x_1	4000	c_1	28 [11]	e_1	1015
2	Wind	x_2	200	c_2	85 [12]	e_2	
3	Biomass	x_3	100	c_3	71[12]	e_3	101
4	Hydropower	x_4	500	c_4	77.3 [12]	e_4	
5	Solar	x_5	200	c_5	136.4 [12]	e_5	
Energy Demand			5000				

2.2 Application of Multi-Objective Programming

Multi-criteria decision-making, or MCDM [13], is implemented when there are some contradictory criteria, that is to say, several criteria that matter but cannot be optimized at the same time, especially in energy problems. Multi-objective programming or vectorial optimization [15, 14] techniques address the simultaneous optimization of some of the objectives that are subject to a set of commonly linear constraints, since an optimization solution cannot be determined for some objectives, MOP is used to gain community potential solutions that are efficient solutions (Pareto Optimal) rather than finding a single optimal solution. The elements of this efficient group are the possible solutions so there are no other possible solutions that can achieve the same or better performance for all objectives and in the best way for at least one objective.

Thus, to generate an efficient MOP, a model group can be formulated as follows:

Eff. $z(x) = [z_1(x), ..., z_q(x)]$, based on $x \in F$.

When *Eff.* means seeking efficient solutions (in terms of minimizing and maximizing).

F, on the other hand, represents the possible group, and x indicates the vector of decision variables. Therefore, based on this, in our case, first, it is necessary to calculate the elements of the cost matrix [15]. This matrix is the result of the optimization of each objective, the costs of electricity generation, and CO_2 emissions separately, giving the second goal the appropriate value for optimal solution of the first. Thus, a square matrix is obtained in which the level of conflict between the goals is reflected rather than finding a single optimal solution. The elements of this efficient group are the possible solutions. There are no other possible solutions that can achieve the same or better performance for all objectives, or in the best way for at least one objective. Thus to generate an efficient MOP model group can be formulated as follows: Eff. $z(x) = [z_1(x), ... z_q(x)]$

Optimization is done with the application of the phpsimpleex program [16], from which we get the solutions: Minimizing the cost $z_1(x)$, the following solution is obtained: Cost, $z_1(x)$ = 28.00 euro/kWh and emissions CO_2, $z_2(x)$ = 1015 g/kWh with: x_1 = 5000 GWh of coal and other components $x_2 = 0$; $x_3 = 0$; $x_4 = 0$; $x_5 = 0$.

The interpretation that can be done is to maximize the use of coal, which generates the lowest cost that it has. Minimizing CO_2 emissions $z_2(x)$, we gain:

Cost $z_1(x)$ = 40.41 euro/MWh and CO_2 emissions $z_2(x)$ = 814.02 g/kWh with: x_1 = 4000 GWh, x_2 = 200 GWh, x_3 = 100 GWh, x_4 = 500 GWh, x_5 = 200 GWh.

In this case, they make most of all clean energy sources, making the coal meet the fixed total needs. The gained values of optimism with the application of the simplex method [17] are listed in Tab. 3.

Table 2 Ideal and Anti-Ideal Points used in Compromise Programming

	Cost, euro/MWh	Emission CO_2, g/kWh
Cost	28.00	1015.00
Emission e CO_2	40.41	814.02

The main diagonal elements are called ideal locations, hence the solution in which both goals (objectives) reach their optimum value. In reality, the ideal point is unattainable, but it is useful to determine the most appropriate solution in order to homogenize decision-making units. The most appropriate solution should be chosen from the group of efficient solutions. This group can be approximated by the limitation method.

This method optimizes a goal (objective), including other constraints set as a parametric barrier. For each value of this parameter, a certain point of efficient solutions will be gained. The method in our case, has been applied, in minimizing the cost, for a number of relevant values of CO_2 emissions. This group is determined by ideal and anti-ideal emission values, for an increase of 10 g / kWh variation.

The formulation is as follows:

$z_1(x) = (28x_1 + 85x_2 + 71x_3 + 77.3x_4 + 136.5x_5)/\sum x_i$, euro/kWh

$z_2(x) = (1015x_1 + 101x_3)/\sum x_i$, g/kWh

The subject and the following restrictions

$\sum x_i \geq 5000$ GWh; $x_i \geq 0,\ \forall i$

$x_2 \leq 200$ GWh; $x_3 \leq 100$ GWh; $x_4 \leq 500$ GWh;

$z_2(x) = k$, $814.02 < k < 1015.00$

The gained values are shown in Tab. 3. In this table, the participation of different items in the generation of electricity

for the options being considered can be observed. We can see how many different options are coming up with optimal solutions as we change the cost of CO_2 generation and emission. Thus, for example, wind energy only enters the solution when CO_2 emissions are limited to 904.02 g/kWh, bringing the cost of production to 33.36 euros/kWh. However, if the cost of generation is limited, for example, 30.93 euro/kWh, we only have room for the optimal mix of coal, biomass, and hydropower generation.

Table 3 Efficient set of solutions

ENERGY SOURCE			COAL	WI.	BIOM.	HYD.	SOL
Cost of energy production			28	85	71	77.3	136.4
Nr.	$z_1(x)$	$z_2(x)$	x_1	x_2	x_3	x_4	x_5
1	40.41	814.02	4000	200	100	500	200
2	39.34	824.02	4049	200	100	500	151
3	38.27	834.02	4099	200	100	500	101
4	37.20	844.02	4148	200	100	500	52
5	36.13	854.02	4197	200	100	500	3
6	35.54	864.02	4246	154	100	500	0
7	34.98	874.02	4296	104	100	500	0
8	34.42	884.02	4345	55	100	500	0
9	33.86	894.02	4394	6	100	500	0
10	33.36	904.02	4443	0	100	457	0
11	32.88	914.02	4493	0	100	407	0
12	32.39	924.02	4542	0	100	358	0
13	31.91	934.02	4591	0	100	309	0
14	31.42	944.02	4640	0	100	260	0
15	30.93	954.02	4690	0	100	210	0
16	30.45	964.02	4739	0	100	161	0
17	29.96	974.02	4788	0	100	112	0
18	29.48	984.02	4837	0	100	63	0
19	28.99	994.02	4887	0	100	13	0
20	28.52	1004.02	4940	0	60	0	0
21	28.05	1014.02	4995	0	5	0	0
22	28.00	1015.02	5000	0	0	0	0

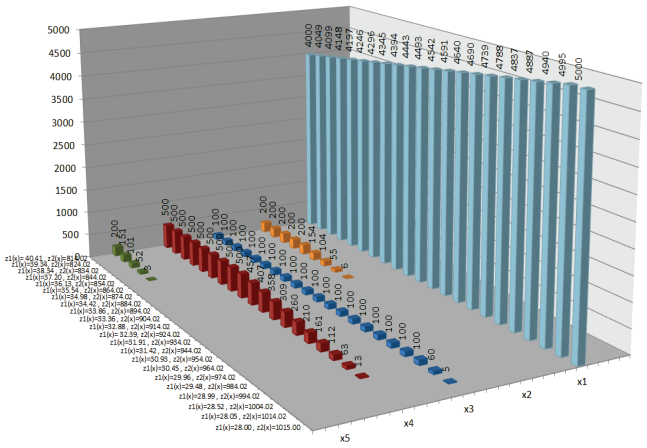


Figure 1 Total efficient solution from Tab. 3

With the obtained values, the graph is developed, which also points to the ideal point. All the solutions represented on the curve graph are viable and efficient solution. But to solve the problem, it is necessary to choose one of the items. In our case, such is the value of the minimum generation cost or the minimum emission of CO_2, where the optimal point will move along the right or left curve.

Therefore, the displayed curve provides decision-making assistance because it shows all of those efficient optimization solutions to the problem in order to find the solution that best suits the interests of decision-makers. So, it seems that in the group of solutions, it is possible to choose an optimal solution, by means of compromise programming, described below.

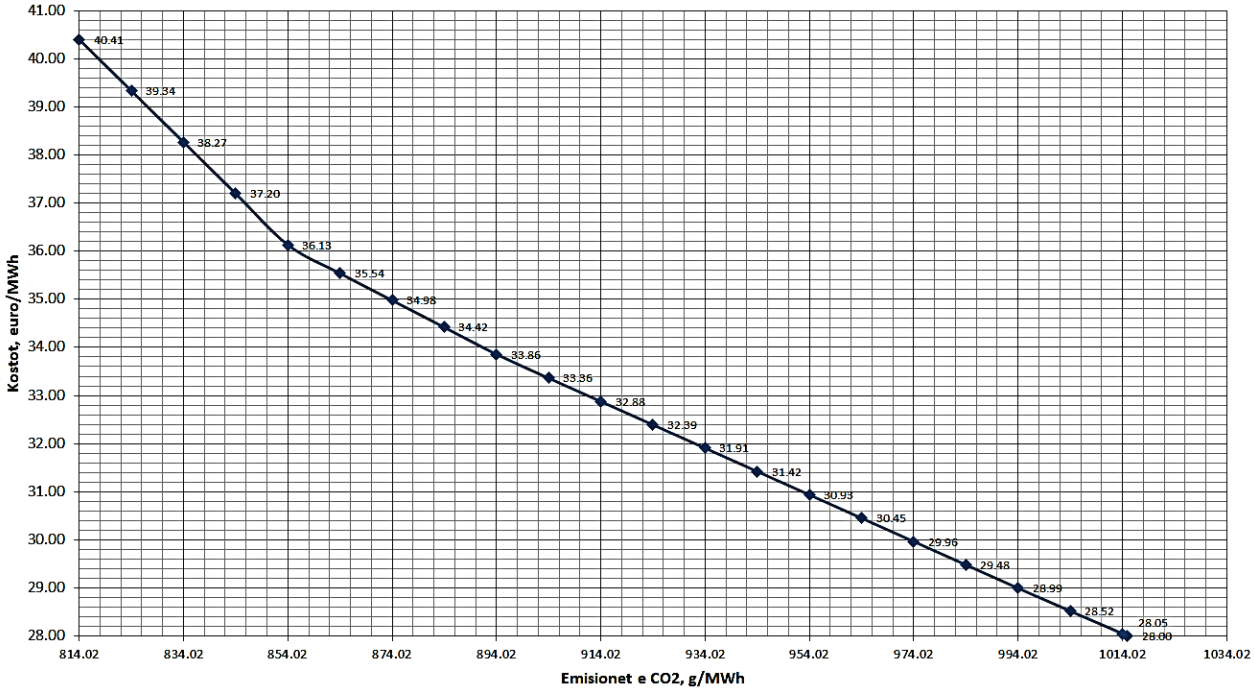


Figure 2 Compromise group for equal weight criteria

3 COMPROMISE PROGRAMMING

The compromise programming was used to select the optimal element from a range of efficient solutions as proposed by Zeleny. This rule is called Zeleny's axiom and is expressed as follows: "Taking into consideration solutions in target space z_1 and z_2, the preferred solution will be the closest to the ideal point" (Zeleny 1973, 74) [18].

If we consider that the decision-making center behaves in a rational way, it will select that effective point or zone of effective point groups that are closer to the ideal point.

Compromising programming begins by setting the ideal point, the coordinates of which are given by the optimum values of the various objectives of the decision-maker. The ideal point is usually inadequate. If possible, then there is no conflict between the objectives. When the ideal point is inadequate, the optimal elements or compromise solutions are provided by an efficient solution that is closer to the ideal point.

The ideal alternative coordinates are given by optimal values usually when targets (goals) are measured in different units, so that the amount of proximity stakes does not make sense, without having dimensional homogeneity. Therefore, it will be necessary to continue with the normalization of objectives (goals). Thus, the degree of proximity as the relative deviation between the j objective and its ideal value is determined by:

$$d_j = \frac{\left[z_j^* - z_j(x)\right]}{\left[z_j^* - z_{*j}\right]}$$

where d_j represents the degree of proximity of the normalized objective j and z^* j is the anti-ideal of this objective - the worst possible value for objective j in an effective set (efficiency). The normalized degree of proximity is limited between 0 and 1. Thus, when a target reaches its ideal value, its proximity is zero; on the contrary, this scale becomes equal to one when the objective in question reaches an equal value with the anti-ideal. If we now represent W_j preferences that the decision-making center relates to the discrepancy between achieving the j and its ideal goal, compromise programming is consistently consistent in seeking more efficient solutions closer to the ideal. So, this programming process is based on whether the optimal solution is to find the closest point to the ideal. This proximity is measured by the mathematical concept of distance.

There are many distances, apart from the Euclidean, the best known, and the question is which ones should be used. In fact, the process simplifies, as it is shown that the range of distance solutions from the ideal point is the one that is minimal, the so-called distance of Manhattan L_1 and Chebysev L_∞ respectively are minimal.

These points are unchanged as they depend on the given weighting (weighting) of each known target and must reflect the preferences of the decision-making center. These weights are placed in the expression of the distance, so that the result is as follows:

$$L_1 = W_1 \frac{z_1(x) - z_1^*}{z_{*1} - z_1^*} + W_2 \frac{z_2(x) - z_2^*}{z_{*2} - z_2^*}$$

$$L_\infty = \max[W_1 z_1(x) - z_1^*,\ W_2 z_2(x) - z_2^*]$$

where: z_j^* - the ideal value, z_{*j} - the anti-ideal value, W_j - the weight of each objective.

For example, we have calculated the tradeoffs for different weights. Here, the same weighting is considered for the two objectives, twice more important for one of them, and also one of them four times more important than the other. The resulting groups of compromise are shown in Tab. 4

Table 4 Compromise solutions L_1 and L_∞

Weight factor	Distance	Cost	Emission CO_2	Coal	Win.	Bio.	Hyd.	Solar
		euro/MWh	g/kWh	GW/year	GW/year	GW/year	GW/year	GW/year
W_1=1	L_1	36.55	850.17	4178	200	100	500	22
W_2=1	L_∞	36.33	852.15	4188	200	100	500	12
W_1=1	L_1	33.30	905.37	4450	0	100	450	0
W_2=0.8	L_∞	34.11	889.58	4372	28	100	500	0
W_1=1	L_1	34.38	884.65	4348	52	100	500	0
W_2=0.9	L_∞	35.00	873.71	4294	106	100	500	0
W_1=08	L_1	36.34	852.05	4187	200	100	500	13
W_2=1	L_∞	38.23	834.39	4100	200	100	500	100
W_1=0.9	L_1	36.44	851.11	4183	200	100	500	17
W_2=1	L_∞	37.83	838.12	4119	200	100	500	81

For example, we calculated the groups of compromises for different weights. Here the division of the same weight is considered for both objectives, ($W_1 = 1$, $W_2 = 1$), ($W_1 = 1$, $W_2 = 0.9$), ($W_1 = 1$, $W_2 = 0.8$) and ($W_1 = 1$, $W_2 = 1$), ($W_1 = 08$, $W_2 = 1$), ($W_1 = 0.9$, $W_2 = 1$). The weight groups, placed in the expression of the distance (L_1, L_{inf}), resulting in compromise, are reflected in Tab. 3.

Any type of energy in optimal solutions can be seen in Tab. 2, introducing the cost of generation and the respective CO_2 emissions. As an illustration, a compromise has been made in detail on the case of the same importance for two objectives that are considered certain.

This compromise set is graphically shown in Fig. 4.

As it can be observed, wind energy appears in this set of compromises, albeit with a slightly lower share. However, solar energy does not appear, because the generation cost is very high, as shown in Tab. 5 that is observed.

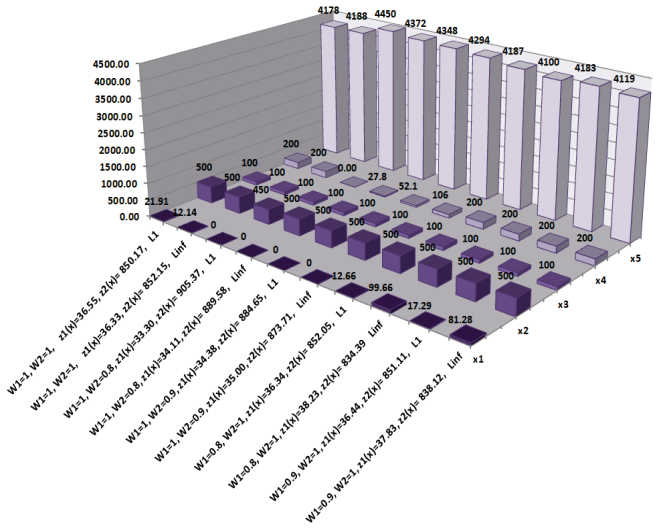


Figure 3 Graphically solutions of L_1 and L_∞

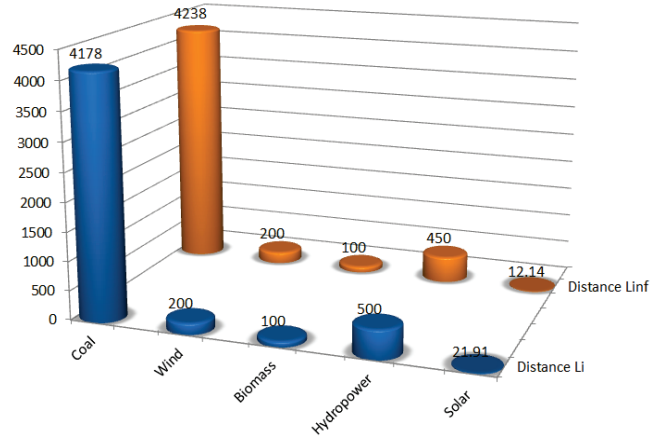


Figure 4 Compromise values of L_{inf} and L_i Distance

Table 5 Compromise solutions for energy sources L_1 and L_∞

Energy Sources	Distance L_i	Distance L_∞
Coal	4178.09	4237.85
Wind	200	200
Biomass	100	100
Hydropower	500	450
Solar	21.91	12.14

4 CONCLUSION

Optimization is an engineering discipline where extreme values of design criteria are required and often there are numerous conflicting criteria to be addressed. Meeting one of these criteria comes at the expense of another. Therefore, multi-purpose optimization deals with conflicting objectives. From various papers presented in the literature, it is noted that no single optimization approach is superior; rather, the selection of a specific optimization method depends on the type of information given in the problem, designer preferences, solution requirements, and the availability of software application.

In the treated example of electricity planning for a given region, the introduction of the CO_2 emission criterion along with generation costs modifies the outcome of the electricity planning. So by reducing the amount of coal, the more expensive options are offered, but with lower CO_2 emissions. The simplifications used for this example are useful in that the results are used not as absolute values but rather as indicators for decision making. In-depth analysis would include better definition of assumptions, taking into account a greater number of criteria for decision-making, such as other socio-economic or environmental variables.

The compromised programming approach gives the decision maker the most efficient system for generating electricity, which is closer to reality than that achieved by optimizing a single objective, as has been done in the past. But besides programming with compromise, actually the analytical hierarchy process method [19] is quite preferred by decision-makers. Although AHP is not without theoretical difficulty, its iterations with a decision-making team makes it a very convenient tool to extract preferential weight within the context of energy planning.

Notice

The paper was presented at MOTSP 2020 – International Conference Management of Technology – Step to Sustainable Production, which took place from 30th September – 2nd October 2020 in Bol, island Brač (Croatia). The paper is not and will not be published anywhere else.

5 REFERENCES

[1] Tobias, W., Marcel, D., & Jutta, F. (2019). Combining scenario planning, energy system analysis, and multicriteria analysis to develop and evaluate energy scenarios. *Journal of Cleaner Production, 242*, 118414. https://doi.org/10.1016/j.jclepro.2019.118414

[2] Lobaccaro, G., Croce, S., Lindkvist, C., Munari Probst, M. C. Scognamiglio, A., Dahlberg, J., Lundgren, M., & Wall, M. (2019). A cross-country perspective on solar energy in urban planning: Lessons learned from international case studies. *Renewable and Sustainable Energy Reviews, 108*, 209-237. https://doi.org/10.1016/j.rser.2019.03.041

[3] Building stock study-Kosovo, World Bank, 2016.

[4] Lee, N. C., Leal, V. M. S., & Dias, L. C. (2018). Identification of objectives for national energy planning in developing countries. *Energy Strategy Reviews, 21*, 218-232. https://doi.org/10.1016/j.esr.2018.05.004

[5] Khodaei, H., Hajiali, M., Darvishan, A., Sepehr, M. & Ghadimi, N. (2018). Fuzzy-based heat and power hub models for cost-emission operation of an industrial consumer using compromise programming. *Applied Thermal Engineering, 137*, 395-405. https://doi.org/10.1016/j.applthermaleng.2018.04.008

[6] Kasprzak, E. M. & Lewis, K. E. (2001) Pareto analysis in multiobjective optimization using the collinearity theorem and scaling method. *Struct Multidisc Optim, 22*, 208-218. https://doi.org/10.1007/s001580100138

[7] Kolios, A., Mytilinou, V., Lozano-Minguez, E., & Salonitis, K. (2016). *A Comparative Study of Multiple-Criteria Decision-Making Methods under Stochastic Inputs*. Energies MDPI. Basel, Switzerland. https://doi.org/10.3390/en9070566

[8] Saaty, T. L. (1980). *Multicriteria Decision Making: The Analytic Hierarchy Process*. McGraw-Hill, New York. https://doi.org/10.21236/ADA214804

[9] Saaty, T. L. & Gholamnezhad, A. H. (1982). High-Level Nuclear Waste Management: Analysis of Options. *Environmental Planning, 9*, 181-196. https://doi.org/10.1068/b090181
[10] Energy Strategy of the Republic of Kosovo 2017 – 2026
[11] Kosovo Energy Corporation - Kosovo B Thermal Power Plant Division B January 2020
[12] Evaluation of Power Supply Option for Kosovo, the World Bank – IBRD –IDA / World Bank Group, August 2018.
[13] Pohekar, S. D. & Ramachandran, M. (2004). Application of multi-criteria decision making to sustainable energy planning—A review. *Renewable and Sustainable Energy Reviews, 8*, 365-381. https://www.elsevier.com/locate/rser https://doi.org/10.1016/j.rser.2003.12.007
[14] Odu, G. O. & Charles-Owaba, O. E. (2013). Review of Multi-criteria Optimization Methods – Theory and Applications. *IOSR Journal of Engineering (IOSRJEN), 3*(10). https://doi.org/10.9790/3021-031020114
[15] Projected Costs of Generating Electricity. 2015 Edition, International Energy Agency Nuclear Energy Agency Organisation for Economic Co-Operation and Development. https://www.solar-united.org/wp-content/uploads/2017/02/ ElecCost2015SUM.pdf
[16] Simplex method software - www.phpsimpleex.com
[17] Luenberger, D. G. & Ye, Y. (2008). *Linear and Nonlinear Programming*, Fourth Edition, Stanford University, p. 72.
[18] Cochrane, J. L. & Zeleny, M. (1973). *Compromise programming in multiple criteria decision making*, University of South Carolina Press, Columbia.
[19] Saaty, T. L. & Kats, J. M. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research, 48*(1), 9-26. https://doi.org/10.1016/0377-2217(90)90057-I

Authors' contacts:

Fisnik Osmani, Prof. ass. dr.
University of Mitrovica,
Faculty of Mechanical Engineering and Computer,
Ukshin Kovaqica, 40000 Mitrovica, Kosovo
fisnik.osmani@umib.net

Atanas Kochov, Prof. dr.
(Corresponding author)
University "St. Kiril & Methodius",
Faculty of Mechanical Engineering,
Karpos II bb 1000 Skopje, North Macedonia
atanas.kochov@mf.edu.mk

Mirjeta Ilazi, MSc, PhD Candidate
University "St. Kiril & Methodius",
Faculty of Mechanical Engineering,
Karpos II bb 1000 Skopje, North Macedonia

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20210205105921

Professional paper

Predictive Maintenance of Cash Dispenser Using a Cognitive Prioritization Model

Archana Dixit*, Amol B. Mahamuni

Abstract: In this technical paper, we address the issue of predicting cash dispenser (addressed as 'Device' henceforth) failure by harnessing the power of humungous data from service history, logs, metrics, transactions, and plausible environmental factors. This study helps increase device availability, enhanced customer experience, manage risk & compliance and revenue growth. It also helps reduce maintenance cost, travel cost, labour cost, downtime, repair duration and increase meantime between failures (MTBF) of individual components. This study uses a cognitive prioritization model which entails the following at its core; a) Machine Learning engineered features with highest influence on machine failure, b) Observation Windows, Transition Windows and Prediction Windows to accommodate various business processes and service planning delivery windows, and c) A forward-looking evaluation of emerging patterns to determine failure prediction score that is prioritized by business impact, for a predefined time window in the future. The model not only predicts failure score for the devices to be serviced, but it also reduces the service miss impact for the prediction windows.

Keywords: Cash Dispenser Failure Prediction; Cognitive Prioritization Model; Feature Engineering; Machine Learning; Predictive Maintenance (PdM)

1 INTRODUCTION

Maintenance is a vital area that controls major cost savings and revenue. According to the International Society of Automation (ISA), more than $647 billion is lost each year due to downtime. All businesses have strived to achieve cost effective maintenance, high availability and customer satisfaction. Retail Banks are no different. They too have grappled with maintenance processes to alleviate downtime for a long time. Lack of a formal commitment and the easiness with which consumers can switch to competitors make the process of building trust among consumers even tougher for retail bankers. The consumers need to be able to get the service they want on a desired time. Otherwise, 66% of them will simply switch banks on account of an unfulfilling service needs.

Cash dispenser devices are getting more and more powerful everyday by supporting capabilities such as Passbook updating, Cash-Cheque Deposits, Money Transfer, Statement printing, Voucher Printing, Electronic Purse Upload, Wire Transfer, Update Personal Information, Cross-Selling, Advertising, Special Interest and so on. Sudden failure and unplanned maintenance of these devices can have substantial impact on not only banking specific services but also revenue generating streams. It also directly impacts customer relationships. Therefore, there is an urgent need to rise from preventative or reactive maintenance to predictive maintenance.

This study introduces predictive maintenance capabilities that can immensely help

- Identify failures in advance leading to the optimization of logistics and associated functions
- Lower downtime by a significant factor
- Spill over benefits from optimized cash in transit activities
- Increase reliability and operational efficiency.

This study uses three rolling time windows 'Observation', 'Transition' and 'Prediction' to determine explanatory features and window of prediction. Features pertaining to inventory and machine components are treated as static while features pertaining to maintenance/service history, errors generated by the machines, environmental changes, and transactions are treated as dynamic. Dynamic features are designed in view of aforementioned time windows. Observation window depicts the slice of data consumed by the model during model training and prediction generation. Transition window signifies the time required to process, prepare and make the data compatible to generate predictions. Predictions are generated for four weeks into the future to allow adequate time for engineering team to act upon the predictions and perform maintenance activities on the identified devices or their components.

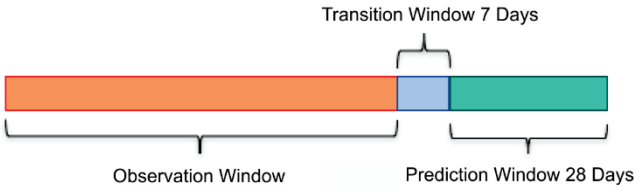


Figure 1 Time Windows for Predictive Maintenance Model

The remaining structure of this paper is organized as follows. Section 2 represents existing and related work on predictive maintenance. Section 3 focuses on the proposed method employed and describes the following; data used, exploratory data analysis, feature engineering, architecture of the proposed solution followed by model training and its evaluation. Section 4 discusses the results obtained in detail. Section 5 reviews the impact, stance and contribution of this solution. Section 6 highlights the most relevant aspects of this study and how it could easily be extended to other sectors. Finally, Section 7 concludes this paper by throwing light on the plausible approaches that could be leveraged in the future to make this Predictive maintenance solution more robust.

2 RELATED WORK

An extensive study of literature, purposes and available approaches is explained by the authors in [1]. Work carried out by other authors [4, 5] in this area also focuses primarily

on generic approaches around predictive maintenance. Some of them focus on challenges and reliability aspect of PdM in their study [6], while others focus only on the cost and economic aspect [2, 15, 19]. While there are studies done for Industrial and IoT devices [10, 11, 12, 14, 16], there are only a handful studies available for cash dispenser devices [21] wherein only event/error logs are considered as variables to perform Time Series Classification. Strong indicators of failures could be gleaned out from service history, metrics, transactions, etc. which is missing from the existing work in this field. The focus of this study is to harness the power of data points from service history, logs, metrics, transactions, and plausible environmental factors to get a holistic view around the device and component failure.

3 PROPOSED METHOD

This study uses a well-defined leading methodology used by industry data miners all over the world called Cross Industry Standard Process for Data Mining (CRISP-DM) (depicted in Fig. 2). It entails framing business questions, understanding data, preparing it for model training, modelling, evaluation and deployment.

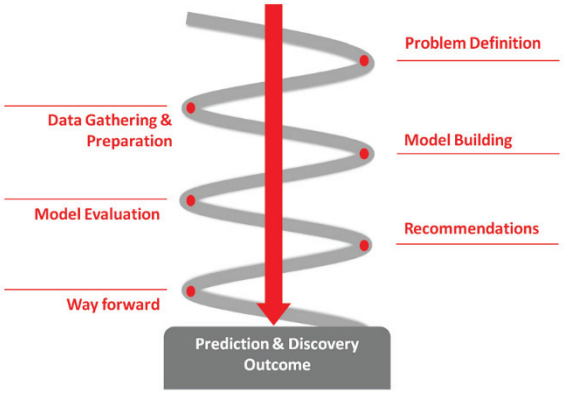


Figure 2 Processes performed in CRISP-DM Methodology

This study leverages two years' worth of data with approx. 10K inventory. Multidimensional data encompasses crucial information from inventory, service history, error logs, transactions, and environment. The grain is at the device component level. Translation Activity Report (TAR Code) from inventory is used to identify modules/components for prediction (TAR Code structure is illustrated in Fig. 3). Preliminary analysis of data reveals that Cheque and Cash Deposit Module (CCDM) is the most profoundly used component of the device. Therefore, it is chosen to illustrate the approach used in this study.

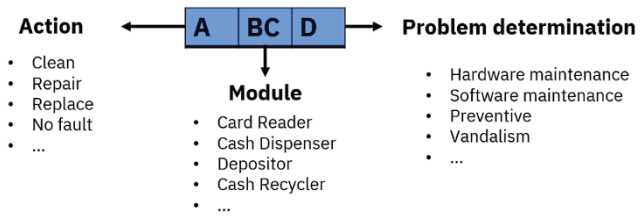


Figure 3 TAR code structure

3.1 Exploratory Data Analysis (EDA)

An extensive EDA is performed during this study to understand data distribution before applying any model. Univariate, Bivariate & multivariate analysis is performed to understand distribution, association & correlation of these variables better. The following Figs. 4, 5, 6 & 7 are some of the visuals (illustrative not exhaustive) depicting the same.

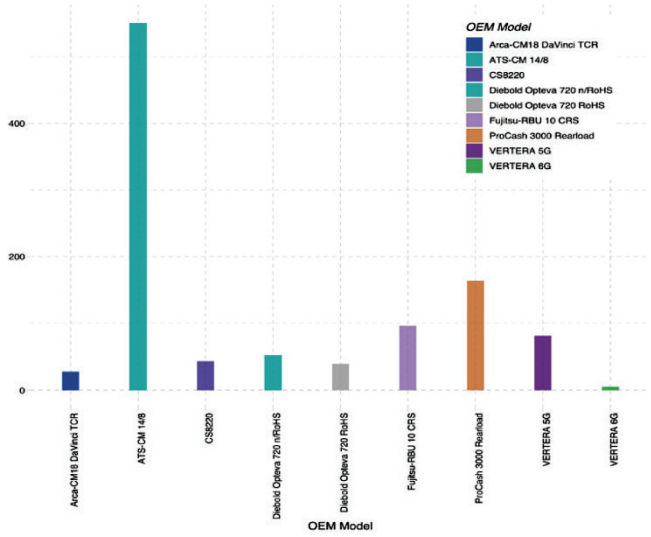


Figure 4 Inventory by Original Equipment Manufacturer (OEM)

Most of the inventory used in this study belongs to one specific OEM model. In addition, CCDM module accounts for most of the service tickets confirming that it is indeed the most often used component during transactions.

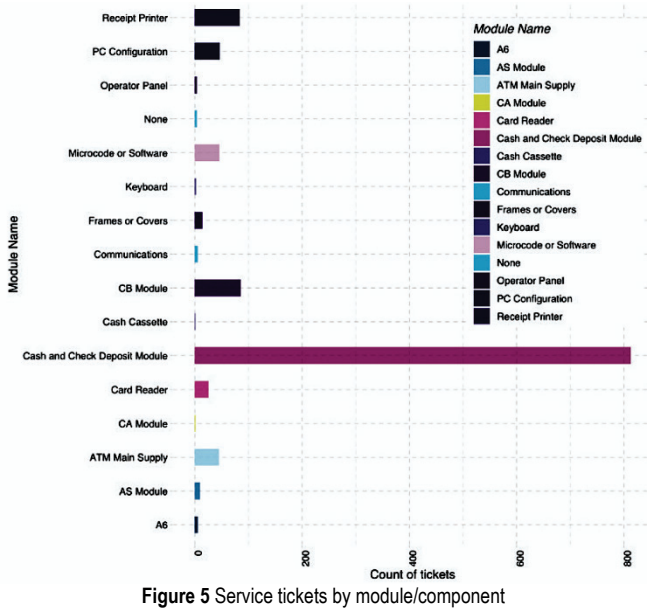


Figure 5 Service tickets by module/component

Service tickets show a downward trend, aptly so as banks strive hard to reduce these numbers (Fig. 6).

Fig. 7 depicts primary problem areas for the CCDM module.

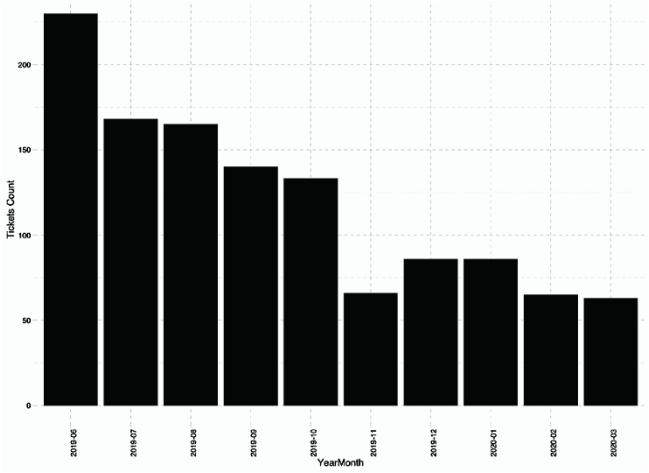


Figure 6 Service tickets Monthly Trend

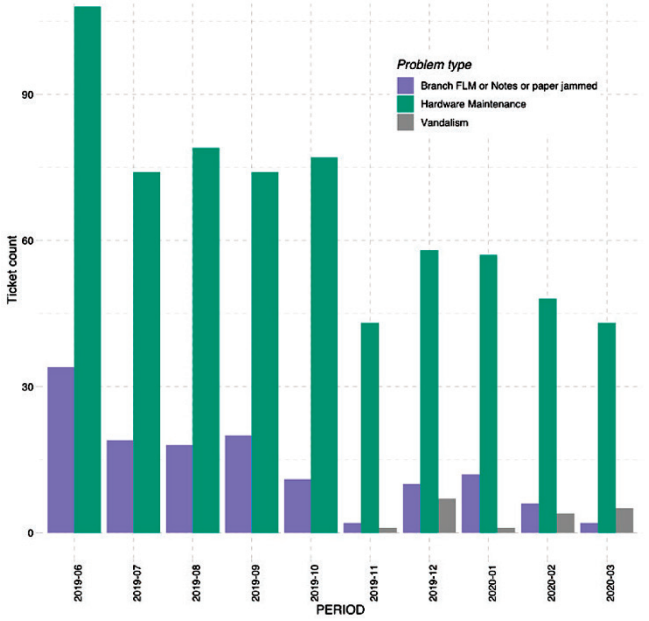


Figure 7 Problem areas for CCDM module

Our EDA does not exhibit any prominent issues with the data with respect to missing values or outliers. Data is mostly clean and tidy.

3.2 Feature Engineering

Predictions are generated at module or component level and therefore, features are also engineered for each module. As described earlier, there are two types of features used in this study; static & dynamic. Static features are time windows independent and dynamic features respect time windows mentioned beforehand.

Inventory details like 'OEM', 'Age', 'TAR Action', 'TAR problem' etc. are static. Only active inventory (Devices that are not yet deinstalled) is considered for this study. 'Last Failure', 'Uptime', 'Last Visit' etc. are computed using Observation Window and most recently closed ticket. Transaction features 'Operation Count', 'Served Count', 'Reversed' etc. are created by performing aggregation. Error features 'Err Type 1 Count', 'Err Type 2 Count' and so on are computed for each error type. Error time elapsed features are also computed for each error type. Mean Time Between Failure (MTBF) features are computed too using error information. Features Avg MTBF, Weighted MTBF, etc. are computed using statistical measures. Ticket features 'Last Preventive', 'Preventive Ticket Count', 'Repair Count', etc. are computed using aggregation and business rules. These dynamic features contribute to significant chunk of explanatory variables used in this study. Target variable has two classes; 'Failure' & 'Operational' making it a binary classification supervised machine-learning problem. All these features are generated for each module.

3.3 Predictive Maintenance Architecture

The following flow represents the predictive maintenance (PdM) architecture employed in this study (Fig. 8). Data from various sources is collected and integrated into a single coherent source. Various data wrangling and munging techniques are leveraged to make it tidy and compatible for machine learning model training. Features are generated with respect to both time windows and business rules.

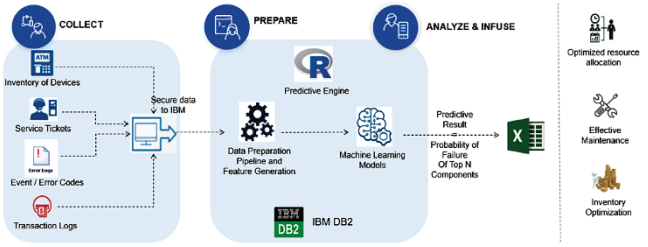


Figure 8 Predictive Maintenance Architecture

This study uses DB2 database to extract and store predicted results for subsequent actions. It leverages open source programming language R to perform data extraction, data preparation, data transformation, statistical modelling, machine learning model building, training & evaluation. Meaningful actionable insights are also extracted using R's visualization libraries. Generated predictive insights are used for optimized resource allocation, effective maintenance and inventory optimization.

3.4 MODEL TRAINING & EVALUATION

Data for CCDM module suffers from imbalance of classes; 'failure' & 'Operational'. Therefore, sampling techniques; upsampling, downsampling, SMOTE are employed to create a dataset with balanced classes. Inbuilt weightage techniques provided in R libraries are also leveraged to overcome class imbalance issue. Cross validation and bootstrapping strategies are used in splitting train and test datasets to tackle overfitting issues. Numerous experiments are carried out for each module and algorithm combination by tuning various hyperparameter nobs to arrive at the optimal parameter values resulting in best performance

of the model. Figs. 9 - 11 depict results for few of such experiments.

The following algorithms are used to perform model training and prediction of failure for each module; RandomForest, XGBoost, Support Vector Machines, Adaboost, and Generalized Linear Model. Multiple models are trained and validated before selecting a best one for scoring and prediction on new unseen data.

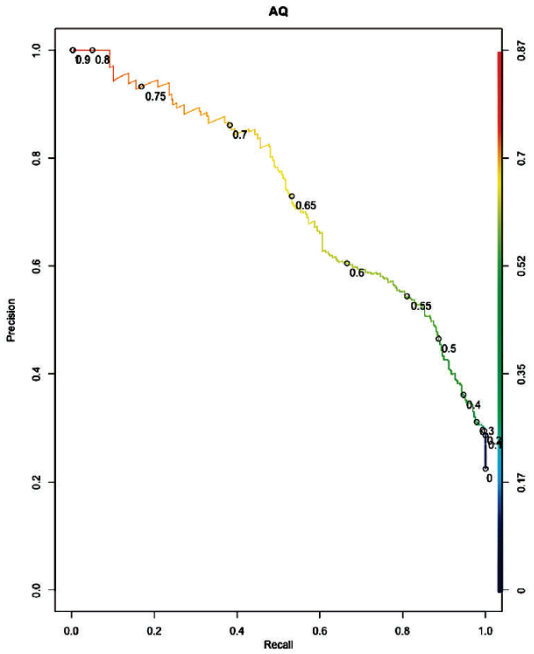


Figure 9 Recall and Precision for AQ (CCDM) module for experiment 1

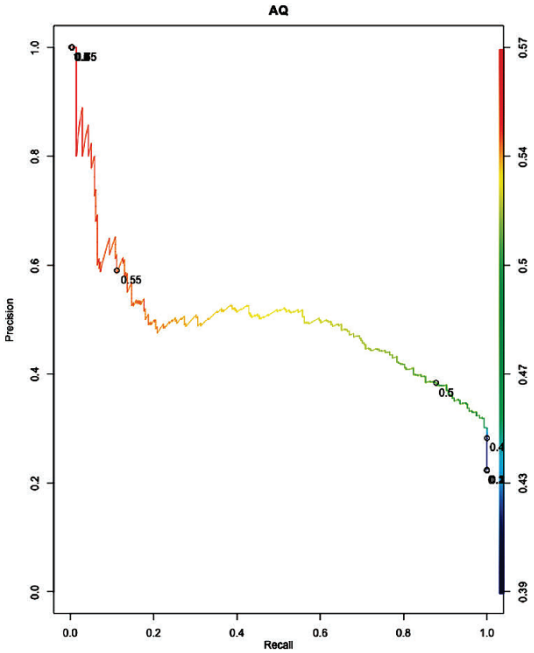


Figure 10 Recall and Precision for AQ (CCDM) module for experiment 2

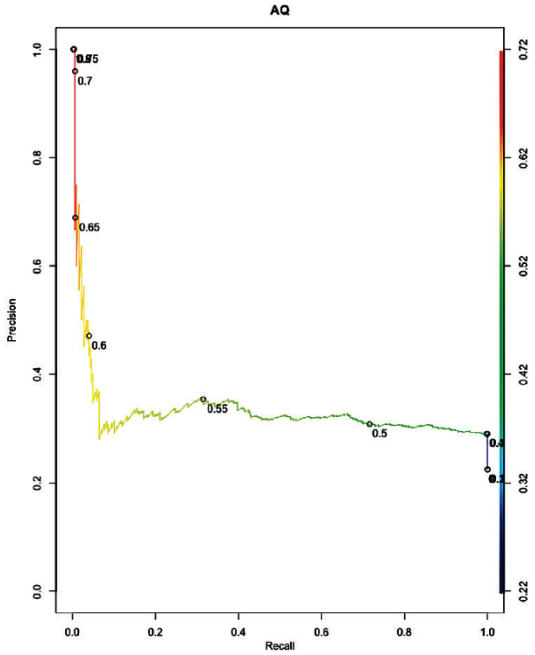


Figure 11 Recall and Precision for AQ (CCDM) module for experiment 3

In this study we are more concerned about how many of the actual failures are predicted as failures by the model, i.e. recall rather than vanilla accuracy or the model. Recall is mathematically defined as the ratio of accurately classified positive classes by model and actual positive classes. Therefore, in our study we focused on reducing Type II error or false negatives to achieve the best possible recall.

4 RESULTS

Subsequently, it is observed that the ensemble algorithm Random Forest performs best for the data at hand. It gives us the highest recall and outperforms other algorithms. Generated insights could also be categorized as high, medium or low based on the prediction probability of each module/component to specify severity of the failure. High severity failures may then be attended to on a priority basis. This novel framework preserves time-driven business process in perspective while generating predictions to facilitate adaption based on relatedness, reliability and efficiency factors.

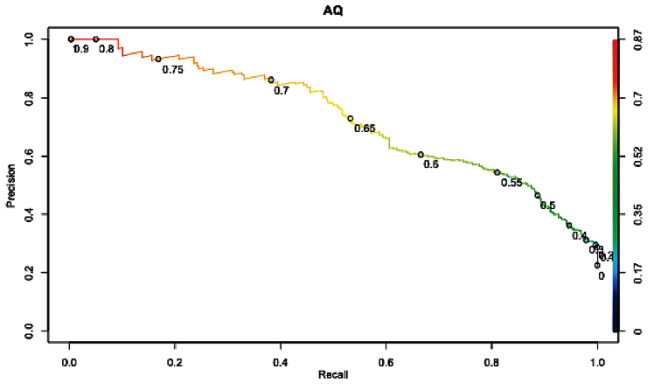


Figure 12 Recall and Precision for AQ (CCDM) module using Random Forest

5 OUTLOOK & IMPACT

There may be a reduction of approx. 40 - 50% in reactive tickets if generated predictive insights are taken into account and acted upon well within the predictive window.

Fig. 13 describes predictions generated for two rounds for illustrative purposes. Prediction impact analysis is performed by visualizing device count with prediction failures vs device count with reactive tickets for failure in future prediction window. There were sixteen out of total thirty devices that were predicted in round 1 would have reactive tickets in future. Similarly, there were ten out of total twenty-five devices that were predicted in round 2 would have reactive tickets in future.

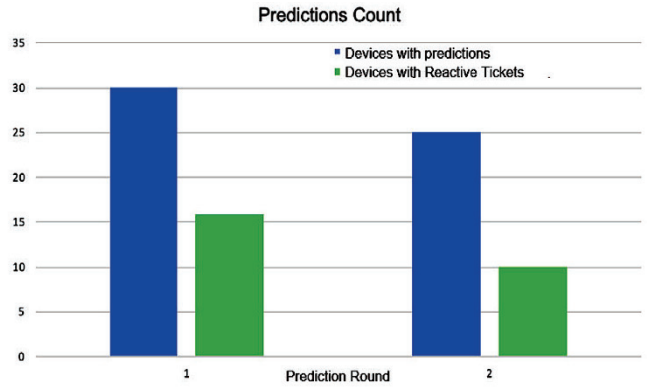


Figure 13 Predictions vs Reactive Tickets for Prediction Rounds

Further, such analysis shows that there is an increase in device availability by 2% if this solution is implemented for predictive maintenance. Meantime between failures of individual components is almost doubled. The solution identifies failures in advance leading to the optimization of logistics and associated functions. Downtime is also lowered by a significant factor. First touch resolution is increased within a week of ticket logging. Repair duration, labor cost and travel costs are lowered noticeably too. Retail banks are able to attain spillover benefits from optimized cash in transit activities, as they relate to second line maintenance.

6 CONCLUSION

In this work, we have studied how machine learning heuristic models augmented with sophisticated feature engineering mechanism could be used to empower retail bankers by predicting component that are mostly likely to fail in the next four weeks. Multidimensional data from disparate touch points of devices is employed to facilitate predictive maintenance in real time.

This study could easily be prolonged to predict failure in other devices such as POS machines, Self-checkout kiosks, etc.

7 WAY FORWARD

This study has several avenues of enhancement to add more capabilities into the overall Predictive maintenance solution.

- Sensor Data could be married with existing multidimensional data to enrich and form 360-degree view of devices and respective components
- Inclusion of data indicating any kind of social unrest leading to vandalism may help boost model's prediction power.
- This could easily be extended to predict failures in Data centers, servers, and other IT infrastructure devices
- The state-of-the-art techniques Deep Learning, LSTM, HTM may also be leveraged to improve model performance and thereby prediction capabilities
- Prescriptive capabilities could further enhance the solution and assist engineers in resolving issues with quick turnaround.
- This approach may be experimented with Python programming language to design an overall offering.

Notice

This paper was presented at IC2ST-2021 – International Conference on Convergence of Smart Technologies. This conference was organized in Pune, India by Aspire Research Foundation, January 9-10, 2021. The paper will not be published anywhere else.

8 REFERENCES

[1] Ran, Y., Zhou, X., Lin, P., Wen, Y., & Deng, R. (2019). A Survey of Predictive Maintenance: Systems, Purposes and Approaches. Retrieved from: https://arxiv.org/abs/1912.07383

[2] S. Spiegel, F. Mueller, D. Weismann, J. Bird (2018). Cost-Sensitive Learning for Predictive Maintenance. Retrieved from: https://arxiv.org/abs/1809.10979

[3] Tuggener, L., Amirian, M., Rombach, K., Lörwald, S., Varlet, A., Westermann, C., & Stadelmann, T. (2019). Automated Machine Learning in Practice: State of the Art and Recent Results. https://doi.org/10.1109/SDS.2019.00-11

[4] Miller, K. & Dubrawski, A. (2020). System-Level Predictive Maintenance: Review of Research Literature and Gap Analysis. Retrieved from: https://arxiv.org/pdf/2005.05239.pdf

[5] Bousdekis, A., Lepenioti, K., Apostolou, D., & Mentzas, G. (2019). Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0. *IFAC-PapersOnLine, 52*(13), 607-612. https://doi.org/10.1016/j.ifacol.2019.11.226

[6] Meyer Zu Wickern, V. - Challenges and Reliability of Predictive Maintenance. (2019). Challenges and Reliability of Predictive Maintenance. https://doi.org/10.13140/RG.2.2.35379.89129

[7] Poór, P. & Basl, J. (2019). Predictive maintenance as an intelligent service in Industry 4.0. *Journal of Systems Integration, 10*(1), 3-10. https://doi.org/10.20470/jsi.v10i1.364

[8] Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018). Machine Learning approach for Predictive Maintenance in Industry 4.0. *The 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, 1-6. https://doi.org/10.1109/MESA.2018.8449150

[9] Schmidt, B., Sandberg, U., & Wang, L. (2014). Next Generation Condition Based Predictive Maintenance. *The 6th International Swedish Production Symposium*,

https://www.researchgate.net/publication/318775487_Next_Generation_Condition_Based_Predictive_Maintenance.
[10] Bonci, A., Longhi, S., & Nabissi, G. (2019). Predictive Maintenance System using motor current signal analysis for Industrial Robot. *The 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1453-1456. https://doi.org/10.1109/ETFA.2019.8869067
[11] Cakir, M., Ali Guvenc, M., & Mistikoglu, S. (2020). The Experimental Application of Popular Machine Learning Algorithms on Predictive Maintenance and the Design of IIoT based Condition Monitoring System. *Computers & Industrial Engineering*, 106948. Available online 30 October 2020. https://doi.org/10.1016/j.cie.2020.106948
[12] Maashi, M., Alwhibi, N., Alamr, F., Alzahrani, R., Alhamid, A., & Altawallah, N. (2020). Industrial Duct Fan Maintenance Predictive Approach Based on Random Forest. *The 9th International Conference on Information Technology Convergence and Services (ITCSE 2020)*, 177-184. https://doi.org/10.5121/csit.2020.100516
[13] Raza, A. & Ulansky, V. (2016). Modelling of predictive maintenance for a periodically inspected system. P*rocedia CIRP 59*, 95-101. https://doi.org/10.1016/j.procir.2016.09.032
[14] Vibha, K., Pal, H., Bendigeri, M. V., & Panna, N. (2020). Development of Predictive Maintenance System of Motors. *International Journal of Advanced Science and Technology, 29*(7), 2344-2351.
[15] Jing Yu, D. T. & Wing-Keung, A. L. (2020). An Economic Perspective on Predictive Maintenance of Filtration Units. Retrieved from: https://arxiv.org/abs/2008.11070
[16] Zheng, H., Paiva, A. R., & Gurciullo, C. S. (2020). Advancing from Predictive Maintenance to Intelligent Maintenance with AI and IIoT. Retrieved from: https://aiotworkshop.github.io/published/AIoT_4_Zheng_TechnicalPaper_KDD2020.pdf
[17] Serradilla, O., Zugasti, E., & Zurutuza, U. (2020). Deep learning models for predictive maintenance: a survey, comparison, challenges and prospect. Retrieved form: https://arxiv.org/pdf/2010.03207.pdf
[18] Satta, R., Cavallari, S., Pomponi, E., Grasselli, D., Picheo, D., & Annis, C. (2017). A dissimilarity-based approach to predictive maintenance with application to HVAC systems. Retrieved from: https://arxiv.org/ftp/arxiv/papers/1701/1701.03633.pdf
[19] Alsyouf, I. (2007). The role of maintenance in improving companies' productivity and profitability. *International Journal of Production Economics, 105*(1), 70-78. https://doi.org/10.1016/j.ijpe.2004.06.057
[20] Guillaume, A., Vrain, C., & Wael, E. (2020). Time series classification for predictive maintenance on event logs. Retrieved from: https://arxiv.org/pdf/1911.10996.pdf

Authors' contacts:

Archana Dixit, Data Scientist
(Corresponding author)
IBM India
archana.dixit@in.ibm.com

Amol B. Mahamuni, Program Director
IBM India
mamol@in.ibm.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20200224122509

Professional paper

Lance Design for Scrap Melting Aggregates

Ladislav Kovar, Pavel Novak*, Tomas Hapla

Abstract: Metallurgical aggregates, which are used for metal waste melting, are equipped with lances for blowing gaseous media. These gaseous media allow especially scrap melting and intense homogenization of the resulting melt. In connection with this, the blowing systems are developed both for blowing of gaseous media itself on the bath surface or into the melt and for blowing of the gas mixture with powdered substances. When designing the blowing systems and the individual lance tips and nozzles, it is necessary to respect certain criteria, the derivation of which is based on long-term experience and acquired knowledge in this field. The submitted paper summarizes the design recommendations, including the determination of the selected design characteristics of the nozzles for blowing gaseous media on the bath surface in the metallurgical aggregate. These design characteristics help designers and engineers to design systems that meet the high demands on quality, efficiency and operational reliability.

Keywords: blowing lance; gas blowing; iron scrap melting; metal scrap processing; nozzle; steel production

1 INTRODUCTION

If gaseous medium is blowed into metallurgical aggregate for waste metal processing, then we want to achieve certain technological effect at melt processing or gaseous products. Within this context the character of mutually interaction between blown gas and melt or generally told by environment inside of reactor, which is substantially affected by lance design, is very much important. Wide spectrum of various lance designs are used in thermal reactors at different industrial branches which are intended to different purposes [1-5]. In the following text we will focuse most of all to refine lances used in metallurgical aggregates for scrap iron processing.

2 FACTORS AFFECTING REFINE LANCES DESING

Metallurgical aggregates for iron waste processing (melting) are equipped with various types of lances for gaseous media blowing. Design of the lances is given by the purpose for which the lances are designed. The main technological process is provided by the so-called refining lances by which refining oxygen is blown to the surface or below the melt level. It depends on the construction of the aggregate [6, 3].

Lance design has a significant influence on the parameters of the output stream of refining oxygen, which significantly influences not only the metallurgical parameters but also the economic indicators of the aggregate [7, 8].

It is therefore necessary that the lance design respects also requirements for lance service life and reliability of the production unit in addition to the above-mentioned indicators. Multinozzle lances (lance tips) with convergent-divergent and cylindrical nozzles (see Fig. 1) are designed for refining oxygen blowing, the number of which depends on the intensity of the O_2 blowing and in case of oxygen converter there are from three to six nozzles. The mutual position of the nozzles depends on the design of the lance tip. When designing and assessing nozzle function, we proceed with the following parameters:

- maximum O_2 mass flow rate through lance and the corresponding O_2 resting pressure in the lance,
- optimum number of nozzles and cross-sections ratio,
- static pressure value in front of the nozzle,
- corresponding value of the flow pulse and the action coefficient of the outlet stream on the melt bath,
- dimensions of the reactor working profile (spraying, angle between nozzles),
- the area on the bath surface which is affected by the outlet stream and the penetration of the O_2 outlet stream into the melt.

Lances for blowing of gaseous media on the melt surface in the working space of high-temperature reactors for metal making are exposed to severe working conditions [9]. In the case of oxygen blowing into the steel bath, there is intensive heat generation, metal and slag spraying on the lance tip [10], and at the same time waste gases are released. It is reported that the surface temperature of the bath (reaction zone) reaches values from 2000 to 2500 °C [11] and the temperature difference between the remaining volume of the bath is in the range of 400 to 700 °C [12]. Operating experiences shows that the following factors have a decisive impact on the service life and perfect function of the lance:

- radiation effect of the bath and reactor lining,
- thermal effect of exhausted gaseous products,
- effect of melt spraying on the lance tip,
- aerodynamic characteristics of the lance,
- position of the nozzle in the workspace - distance of the nozzle from the bath surface,
- quality of the materials used for lance manufacturing,
- manufacturing method,
- blowing intensity,
- number, shape and dimensions of individual nozzles in the lance,
- design of the cooling system (central, peripheral).

All these factors are related to the heat load of the lance and especially its tip. The outer surface of the lance is therefore exposed to a high heat load. To prevent breakage of the lance material, we must keep the material temperature

within the appropriate limits and hence there is the need to cool the lance. Lance inner surface is intensively cooled.

Water is used as the cooling medium. However, the lance material is exposed to high thermal stresses, which may cause cracks especially at the weld points between the copper lance tip and other steel parts. The design of the lances should meet the following requirements:

- simplicity with regard to production and assembly,
- use of suitable materials,
- minimum heat consumption from the reactor working space,
- optimal media supply with respect to pressure losses,
- optimal shape of the cooling system, avoiding incrustation and excluding film boiling,
- low noise intensity,
- as small splash and melt ejection as possible together with maximum oxygen utilization,
- as long service life as possible, service life should not be lower than the duration of the high temperature reactor campaign.

3 FORGED AND CAST REFINING LANCES TIPS

It is said that in the case of the basic oxygen furnaces, the lance service life is about several hundred heats. It may be from 100 to 500 heats and it depends on many individual parameters of steel melting shop [11].

The material used to manufacture lance tips is electrolytic copper [13]. It is a material which has a high thermal conductivity and therefore the thermal stresses do not appear at such intensity as it would be in the case of other materials. Conversely, it was shown that the deoxidation elements such as e.g. phosphorus, which are contained in a limited amount in the copper acts against diffusion of oxygen into the surface of the lance tip and thus prevent cracking. Although these small amounts of chemical elements reduce, to a certain extent, the thermal conductivity, they can, on the other hand, prolong the service life of the lance.

The tips of the upper refining lances of the steelmaking aggregates (LD converters, heard furnaces. EAF) are, at the present time, manufactured in the form of forgings and (or) castings [12]. Forgings have to be machined to the final form, and the individual parts are then soldered and welded together.

Examples of the lance tips designed for so called upper blowing to the oxygen converter and which consist of several forgings are shown in the Fig. 1 and Fig. 2.

While the lance tip in Fig. 1 consists of six forged parts and has six connecting points (nine places for total lance head assembly), in Fig. 2 there is a lance tip consisting of three forged parts (Fig. 3) and it has only three connecting points (six places for total lance tip assembly).

Another advantage of the lance tip in Fig. 2 is the fact that the connecting plane between the inlet and the outlet part is moved further from the front face of the lance tip, thereby making it better protected from the thermal load from the front face of the lance tip.

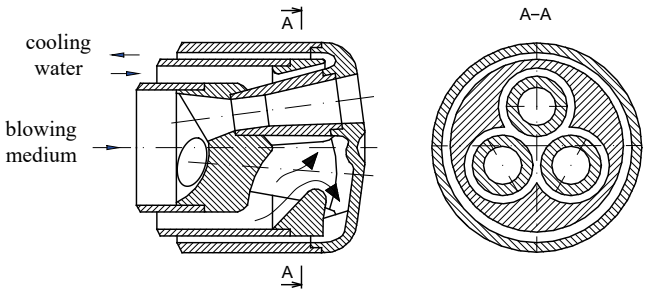


Figure 1 Three-stream lance tip consisting of four parts

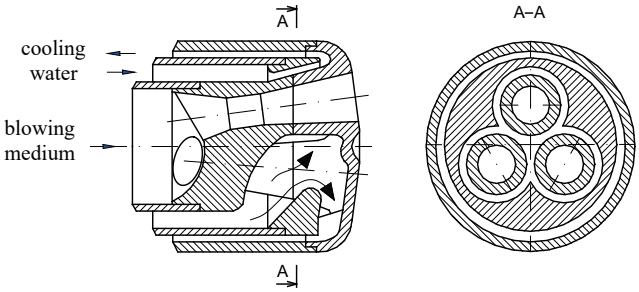


Figure 2 Three-stream lance tip consisting of three parts

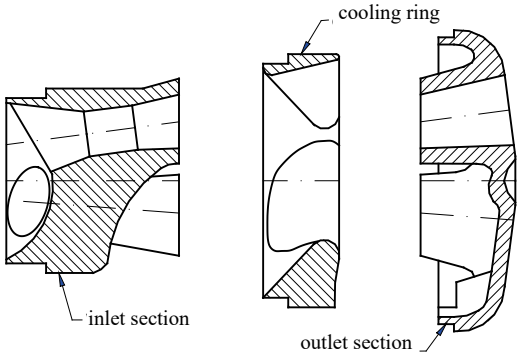


Figure 3 Main parts of three-stream lance tip from Fig. 2

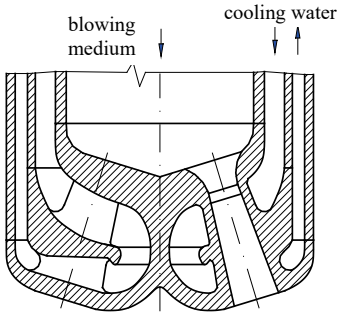


Figure 4 Longitudinal cross-section of the cast lance tip

Cast components, on the other hand, save material, reduce production costs and simplify production and assembly. An example of a cast lance tip is shown in Fig. 4 and Fig. 5.

4 COOLING OF REFINING LANCES

The cooling system design is of great importance for increasing the lance service life. Lances are cooled by technical water at a pressure of 0.5 to 1.0 MPa. Cooling water

velocity inside the lance tip is in the range of 1.8 to 6 m/s. Stagnation areas of the cooling water flow should not be in this system, as the bubble boil then change to the film boil and thus the coefficient of heat transfer from the nozzle wall to the cooling water decreases sharply and the lance tip may be burnt through.

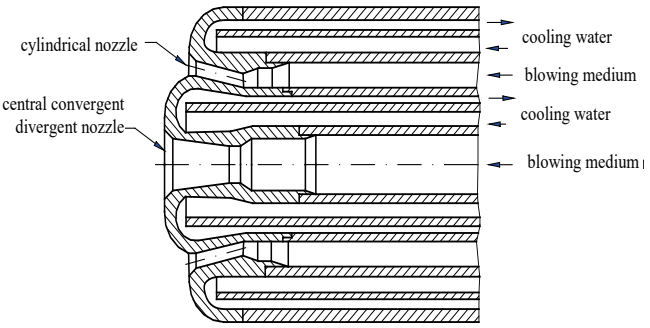


Figure 5 Multinozzle lance tip with central convergent-divergent nozzle

Under operating conditions there are thermodynamic changes in both the working space of the high temperature aggregate and the parameters of the cooling water. These are in particular the heat transfer coefficient from the working space to the lance tip material, the cooling water flow velocity and the purity of the inner walls changes.

Good cooling of the lance and inner nozzles also depends on the thickness of the walls. The thickness of the lance tip wall depends on its dimensions, design, construction, production technology and type of use, and ranges from 6 to 12 mm [12].

Because of lance high heat load the thermal stresses arise during operational period, which may ultimately lead to the deformation of the particular outer tube [14]. Bellows or stuffing compensators are used to compensate these adverse effects of thermal expansion.

If we go back to the layout (design) of the lance tips in the Fig. 1 and Fig. 2, then cooling water is directed through the cooling rings, which can be divided or in whole.

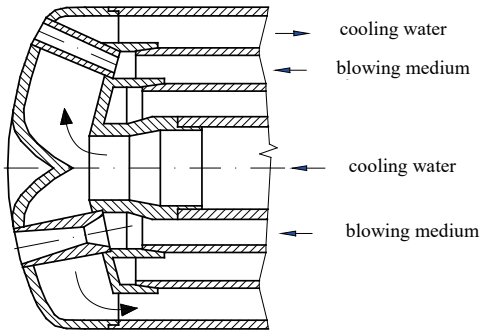


Figure 6 Multinozzle lance tip with central front part cooling

Structural design of cooling channels must ensure continuous velocity changes of cooling water flow in the whole area of cooling ring. Cooling water velocity gradually increases and reaches a maximum when entering into the output annulus formed by the lance outer pipe and the pipe which divides water inlet and outlet.

Central convergent-divergent nozzle with a single or two-layer nozzle arrangement (see Fig. 5) or profiled nozzles with a central cooling system (see Fig. 6) are lance tips designed to increase the cooling effect.

5 RECOMMENDATIONS FOR MASS FLOW AND TOTAL PRESSURE IN FRONT OF NOZZLES SELECTION

Assuming that the stagnation pressure of the blown oxygen in front of the nozzle p_0 will be in the range of 0.6 to 1.3 MPa and the mean value of the oxygen blowing intensity will be 3 $m_N^3/t \cdot min$, then for determination of oxygen mass flow rate for oxygen converter derived regression dependency which is expressed as a function of the maximum value of the oxygen mass flow rate $Q_{m\text{-}max}$ (kg/s) to the weight of the melt G_{tav} (t) is used in following shape:

$$Q_{m\text{-}max} = 0.0731 \cdot G_{tav}^{0.996} \tag{1}$$

Above equation was derived from the parameters of operated oxygen converters.

In the case of hearth furnaces, the following parameters of blown oxygen were considered:

- pressure of blown oxygen in front of the nozzle $p_0 = 0.6$ up to 1 MPa,
- the intensity of blowing will be around 1.2 $m_N^3/t \cdot min$

For hearth furnace relation listed below is recommended:

$$Q_{m\text{-}max} = 0.0086 \cdot G_{tav}^{1.208} \tag{2}$$

Value of the oxygen stagnation pressure before the nozzle p_o (MPa) is another characteristic which is one of the determining parameters of the flow momentum values and also of blown oxygen concentration.

For converter operation following relation was established:

$$p_o = 0.6755 \cdot G_{tav}^{0.104} \tag{3}$$

p_o - oxygen stagnation pressure before the nozzle (MPa).

For hearth furnace an analogous relation is valid:

$$p_o = 0.5041 \cdot G_{tav}^{0.107} \tag{4}$$

6 RECOMMENDATIONS FOR NOZZLE GEOMETRY SELECTION

Convergent-divergent nozzles with the neck of a cylindrical shape are the most commonly used. Diameter of the cylindrical shape neck together with nozzle outlet diameter are other important characteristics.

Important thing is that in design flow mode of oxygen through the nozzle and considered one-dimensional isentropic flow of ideal gas without considering flowing losses so called flow critical state occurs in the throat.

Throat diameter is then nozzle critical diameter d_{kr}. Value of critical diameter d_{kr} (m) is calculated from known values of stagnation pressure p_0 (MPa) and stagnation temperature T_0 (K) at corresponding oxygen mass flow ($Q_{m\,max}$) and at selected number of nozzles in lance tip n (1) according following relation.

$$Q_{m\text{-}max} = n \cdot S_{kr} \cdot \frac{p_o}{\sqrt{T_0}} \cdot \left(\sqrt{\frac{2\kappa}{\kappa+1} \cdot R} \right) \cdot \left(\frac{2}{\kappa+1} \right)^{\frac{1}{\kappa-1}} \quad (5)$$

After substituting for κ (for oxygen $\kappa = 1.4$) and R (for oxygen $R = 259.78$) and editing we can write the following expression for n nozzles

$$Q_{m\text{-}max} = 0.042483 \cdot n \cdot S_{kr} \cdot \frac{p_o}{\sqrt{T_0}} \quad (6)$$

This expression can also be expressed depending on the critical diameter (d_{kr}) of the nozzle (see Fig. 9).

$$Q_{m\text{-}max} = 0.033366 \cdot n \cdot d_{kr} \cdot \frac{p_o}{\sqrt{T_0}} \quad (7)$$

Among the number of nozzles in the lance tip n, the diameter d_{kr} and melt weight G_{tav}, using the previous expressions (1) - (7) and assuming that the total temperature before the nozzle is $T_0 = 25$ °C, then for oxygen converter following relationship is valid

$$d_{kr} = 0.00751 \cdot G_{tav}^{0.446} \cdot \frac{1}{\sqrt{n}} \quad (8)$$

and this expression for hearth furnace

$$d_{kr} = 0.00298 \cdot G_{tav}^{0.5505} \cdot \frac{1}{\sqrt{n}} \quad (9)$$

Graphic representation of critical diameter change, which are expressed by relation (8) is in Fig. 7.

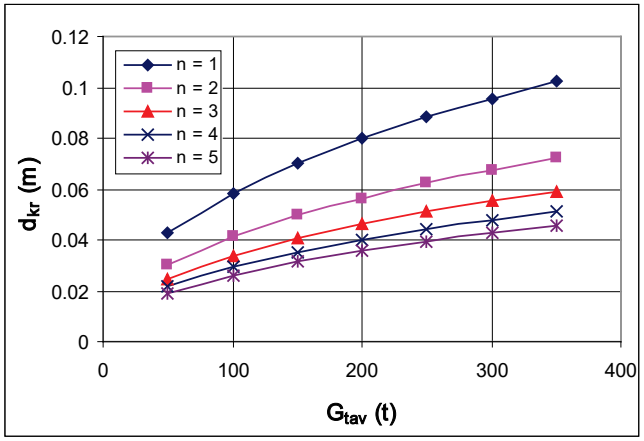


Figure 7 Functional dependence $d_{kr} = f(G_{tav})$

An important characteristic of the oxygen nozzle is the ratio of the output and critical cross-section (S_v and S_{kr}) expressed depending on the ratio of the stagnation pressure in front of output cross-section of the nozzle p_{ok} (surrounding environment pressure) and stagnation pressure in front of the nozzle p_o [15]. For convergent-divergent nozzle in design mode, condition $p_{ok} = p_v$ is valid, where p_v is the static pressure in the output cross-section. Then following relation is valid.

$$\frac{S_v}{S_{kr}} = \frac{0.24943}{\left[\left(\frac{p_{ok}}{p_o} \right)^{1.46} - \left(\frac{p_{ok}}{p_o} \right)^{1.73} \right]^{0.5}} \quad (10)$$

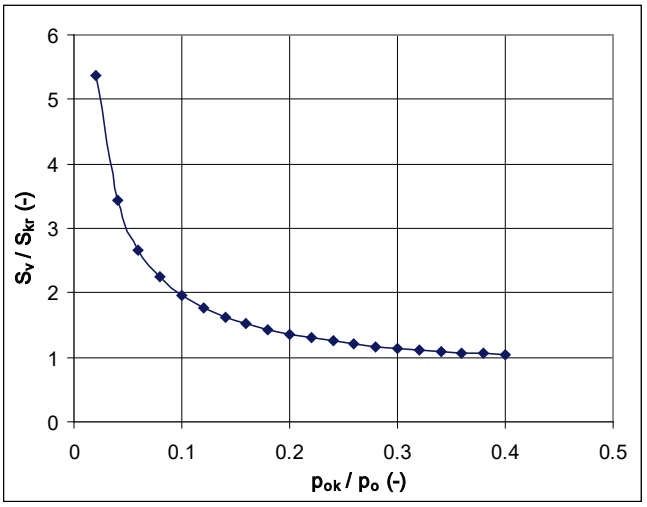


Figure 8 Functional dependence $S_v/S_{kr} = f(p_{ok}/p_o)$

This function (10) is graphically shown in the Fig. 8 and Fig. 9 shows selected nozzle dimensions and parameters.

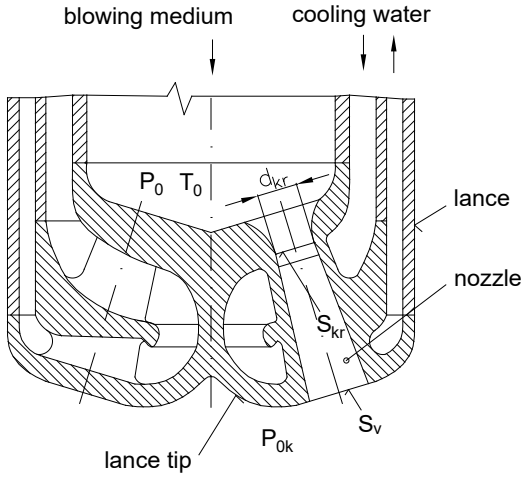


Figure 9 Representation of selected dimensions and areas of the nozzle

7 CONCLUSION

Quality of blowing systems for high-temperature reactors for metal production and processing is a necessary

prerequisite for quality and continuous production. Utility properties of blowing systems depends, among other things, especially on high-quality design of their construction.

The main results of the presented article are the following:

1) operating requirements for lances (nozzles) have been established,
2) requirements for lance (nozzle) design have been established,
3) formulas (1), (2), (3) and (4) may be used to determine the maximum mass flow and total pressure before the nozzle for oxygen converter and hearth furnace, provided that the flow conditions are observed (see Chapter 5),
4) formulas (8), (9) and (10) were derived to calculate the critical nozzle diameter for oxygen converter and hearth furnace,
5) the area (diameter) of the nozzle output cross-section can be determined using formula (11).

Selected main design recommendations and characteristics that were mentioned in the submitted paper help designers and engineers to design systems that meet the high demands on service life, quality, efficiency and operational reliability [16].

Acknowledgments

This paper was created with the support of the SGS project at the Department of Production Machines and Design, Faculty of Mechanical Engineering VSB-TU Ostrava.

List of used symbols

$Q_{m\text{-}max}$	oxygen mass flow rate, $kg \cdot s^{-1}$
G_{tav}	weight of the melt, t
p_o	oxygen stagnation pressure before the nozzle, MPa
d_{kr}	critical diameter, m
T_0	stagnation temperature, K
n	number of nozzles in tuyere, -
S_v	output cross-section, m^2
S_{kr}	critical cross-section, m^2
p_v	static pressure in the output cross-section, MPa
p_{ok}	static pressure in the environment into which the gas stream flows, MPa
$p_{ok} = p_v$	
R	specific gas constant, $J \cdot kg^{-1} \cdot K^{-1}$
κ	adiabatic coefficient, -

8 REFERENCES

[1] Bhatacharya, T. et al. (2016) Design considerations of supersonic oxygen lances for a basic oxygen furnace (BOF). *5th International Conference on Process Development in Iron and Steelmaking including 2nd International Process Integration Forum for the Steel Industry, SCANMET V*, Lulea, 13-15 June, Sweden. p 1-10

[2] Li, M. M., Li, Q., Li, L., He, Y. B., & Zou, Z. S. (2014). Effect of operation parameters on supersonic jet behaviour of BOF six-nozzle oxygen lance. *Ironmaking & Steelmaking, 41*(9), 699-709. https://doi.org/10.1179/1743281213y.0000000154

[3] Kovář, L. (2015) Bottom tuyeres of high temperatures units – division, requirements and selected operating parameters. *Conference Proceedings of METAL 2015, 24th International Conference on Metallurgy and Materials*, June 3-5, 2015, Hotel Voronez I, Brno, Czech Republic, EU. Ostrava, Tanger, 135-139. ISBN 978-80-87294-62-8.

[4] Snigdha, G., Bharath, B. N., & Viswanathan, N. N. (2019). BOF process dynamics. *Mineral Processing and Extractive Metallurgy, 128*(1-2), 17-33. https://doi.org/10.1080/25726641.2018.1544331

[5] Tripathi, P., Kumar, D. S., Sah, R., & Sekhar, V. R. (2017). An improved lance design for hot metal de-sulphurisation, Ironmaking & Steelmaking, 44(6), 421-429. https://doi.org/10.1080/03019233.2016.1214379

[6] Lakshmi, M. V. & Chandra Mouli, K. V. V. N. R. (2017). Study of Oxygen Blowing Lance and It's Drive Mechanism in LD Converters. *International Journal & Magazine of Engineering, Technology, Management and Research, 4*(4), 53-62. ISSN 2348-4845.

[7] Naito, K., et al. (2013). Behavior of Top-blowing Lance Jets in BOF. *Nippon Steel Technical Report*, 104, 33-41.

[8] Wang, H., Zhu, R., Gu, Y. L., & Wang, C. J. (2014). Behaviours of supersonic oxygen jet injected from four-hole lance during top-blown converter steelmaking process. *Canadian Metallurgical Quarterly, 53*(3), 367- 380. https://doi.org/10.1179/1879139514Y.0000000128

[9] Liu, F., Sun, D., Zhu, R. et al. (2019) Characteristics of Flow Field for Supersonic Oxygen Multijets with Various Laval Nozzle Structures. *Metallurgical and Materials Transactions B, 50*, 2362-2376. https://doi.org/10.1007/s11663-019-01652-7

[10] Cao, L. L., Liu, Q., Wang, Z., & Li, N. (2018). Interaction behaviour between top blown jet and molten steel during BOF steelmaking process. *Ironmaking & Steelmaking, 45*(3), 239-248. https://doi.org/10.1080/03019233.2016.1255373

[11] See http://ispatguru.com/oxygen-blowing-lance-and-its-role-in-basic-oxygen-furnace/

[12] Mikolajek, J. (1985) *Steelworks equipment I.* Textbook VSB Ostrava, 206p.

[13] See http://www.nagpalengineering.com/ products/water-cooled-oxygen-lance-bof-eaf/

[14] Feng, C., Zhu, R., Han, B. et al. (2020). Effect of Nozzle Exit Wear on the Fluid Flow Characteristics of Supersonic Oxygen Lance. *Metallurgical and Materials Transactions B, 51*, 187-199. https://doi.org/10.1007/s11663-019-01722-w

[15] Mikolajek, J., Trávníček R., & Kovář L. (1989). Modeling of the oxygen blowing process in the hearth furnace. *Interim research report VSB Ostrava on economic contract* No. 165/88, 88p.

[16] Kitamura, S., Naito, K. & Okuyama, G. (2019). History and latest trends in converter practice for steelmaking in Japan. *Mineral Processing and Extractive Metallurgy, 128*(1-2), 34-45. https://doi.org/10.1080/25726641.2018.1504661

Authors' contacts:

Ladislav Kovar, doc. Dr. Ing.
VSB-Technical University of Ostrava,
Faculty of Mechanical Engineering,
Department of Production Machines and Design,
17. listopadu 15/2172, 708 33 Ostrava – Poruba, Czech Republic
+420 597 324 585, ladislav.kovar@vsb.cz

Pavel Novak, Ing.
(Corresponding author)
VSB-Technical University of Ostrava,
Faculty of Mechanical Engineering,
Department of Production Machines and Design,
17. listopadu 15/2172, 708 33 Ostrava – Poruba, Czech Republic
pavel.novak.st1@vsb.cz

Tomas Hapla, Ing.
VSB-Technical University of Ostrava,
Faculty of Mechanical Engineering,
Department of Production Machines and Design,
17. listopadu 15/2172, 708 33 Ostrava – Poruba, Czech Republic
tomas.hapla@vsb.cz

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20.....

14pt
14pt

Article Title Only in English (Style: Arial Narrow, Bold, 14pt)

14pt

Ivan Horvat, Thomas Johnson, Marko Marić (Style: Arial Narrow, Normal, 10pt)

14pt

Abstract: Article abstract contains maximum of 150 words and is written in the language of the article. The abstract should reflect the content of the article as precisely as possible. TECHNICAL JOURNAL is a trade journal that publishes scientific and professional papers from the domain(s) of mechanical engineering, electrical engineering, civil engineering, multimedia, logistics, etc., and their boundary areas. This document must be used as the template for writing articles so that all the articles have the same layout. (Style: Arial Narraw, 8pt)

8pt

Keywords: keywords in alphabetical order (5-6 key words). Keywords are generally taken from the article title and/or from the abstract. (Style: Arial Narraw, 8pt)

10pt
10pt

1 ARTICLE DESIGN

(Style: Arial Narrow, Bold, 10pt)

10pt

(Tab 6 mm) The article is written in Latin script and Greek symbols can be used for labelling. The length of the article is limited to eight pages of international paper size of Letter (in accordance with the template with all the tables and figures included). When formatting the text the syllabification option is not to be used.

10pt

1.1 General Guidelines

(Style: Arial Narrow, 10pt, Bold, Align Left)

10pt

The document format is Letter with margins in accordance with the template. A two column layout is used with the column spacing of 10 mm. The running text is written in Times New Roman with single line spacing, font size 10 pt, alignment justified.

Article title must clearly reflect the issues covered by the article (it should not contain more than 15 words).

Body of the text is divided into chapters and the chapters are divided into subchapters, if needed. Chapters are numbered with Arabic numerals (followed by a period). Subchapters, as a part of a chapter, are marked with two Arabic numerals i.e. 1.1, 1.2, 1.3, etc. Subchapters can be divided into even smaller units that are marked with three Arabic numerals i.e. 1.1.1, 1.1.2, etc. Further divisions are not to be made.

Titles of chapters are written in capital letters (uppercase) and are aligned in the centre. The titles of subchapters (and smaller units) are written in small letters (lowercase) and are aligned left. If the text in the title of the subchapter is longer than one line, no hanging indents.

10pt

Typographical symbols (bullets), which are being used for marking an item in a list or for enumeration, are placed at a beginning of a line. There is a spacing of 10pt following the last item:

- Item 1
- Item 2
- Item 3

10pt

The same rule is valid when items are numbered in a list:

1. Item 1
2. Item 2
3. Item 3

10pt

1.2 Formatting of Pictures, Tables and Equations

(Style: Arial Narrow, 10pt, Bold, Align Left)

10pt

Figures (drawings, diagrams, photographs) that are part of the content are embedded into the article and aligned in the centre. In order for the figure to always be in the same position in relation to the text, the following settings should be defined when importing it: text wrapping / in line with text.

Pictures must be formatted for graphic reproduction with minimal resolution of 300 dpi. Pictures downloaded from the internet in ratio 1:1 are not suitable for print reproduction because of unsatisfying quality.

10pt

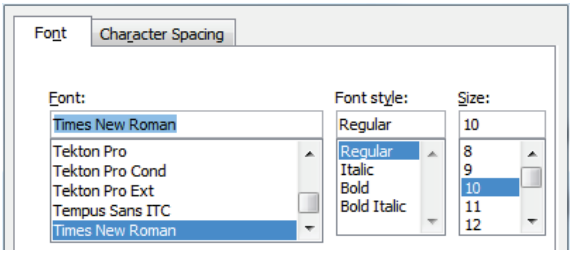


Figure 1 Text under the figure [1]
(Style: Arial Narraw, 8pt, Align Centre)

10pt

The journal is printed in black ink and the figures have to be prepared accordingly so that bright tones are printed in a satisfactory manner and are readable. Figures are to be in colour for the purpose of digital format publishing. Figures in the article are numbered with Arabic numerals (followed by a period).

Text and other data in tables are formatted - Times New Roman, 8pt, Normal, Align Center.

When describing figures and tables, physical units and their factors are written in italics with Latin or Greek letters,

while the measuring values and numbers are written upright.

10pt

Table 1 Table title aligned centre

(Style: Arial Narraw, 8pt, Align Centre)

	1	2	3	4	5	6
ABC	ab	ab	ab	ab	ab	ab
DEF	cd	cd	cd	cd	cd	cd
GHI	ef	ef	ef	ef	ef	ef

10 pt

Equations in the text are numbered with Arabic numerals inside the round brackets on the right side of the text. Inside the text they are referred to with equation number inside the round brackets i.e. ".... from Eq. (5) follows ...." (Create equations with MathType Equation Editor - some examples are given below).

10pt

$$F_{\text{avg}}(t, t_0) = \frac{1}{t}\int_{t_0}^{t_0+t} F\left[q(\tau), p(\tau)\right] \mathrm{d}\tau, \tag{1}$$

$$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cdot\cos\frac{\alpha\text{-}\beta}{2}, \tag{2}$$

$$(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}. \tag{3}$$

10pt

Variables that are used in equations and also in the text or tables of the article are formatted as *italics* in the same font size as the text.

10pt

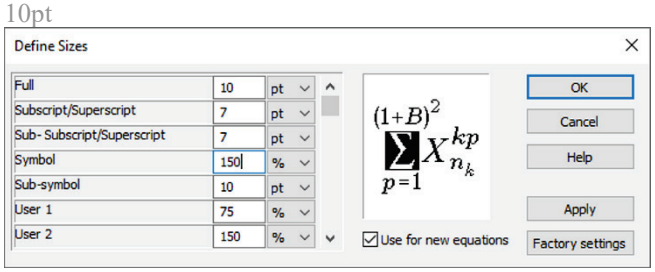


Figure 2 The texts under figures

(Style: Arial Narraw, 8pt, Align Centre)

10pt

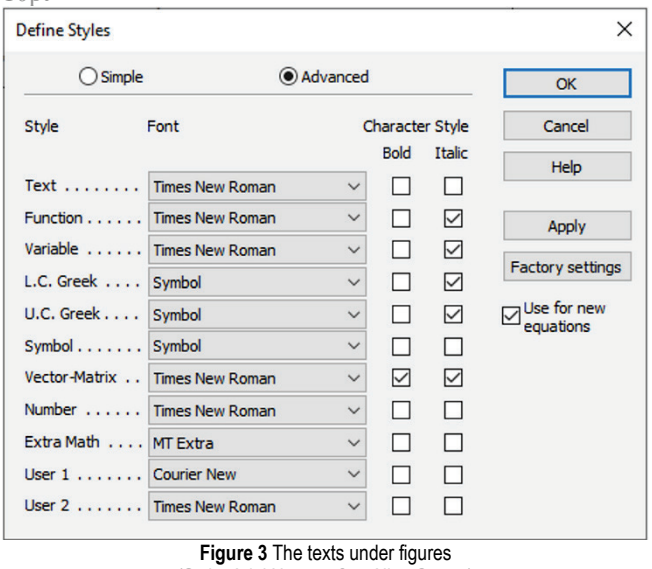


Figure 3 The texts under figures

(Style: Arial Narraw, 8pt, Align Centre)

Figures and tables that are a part of the article have to be mentioned inside the text and thus connected to the content i.e. „ … as shown in Fig. 1…“ or „data from Tab. 1…“ and similar.

10pt

2 PRELIMINARY ANNOTATION

10pt

Article that is offered for publication cannot be published beforehand, be it in the same or similar form, and it cannot be offered at the same time to a different journal. Author or authors are solely responsible for the content of the article and the authenticity of information and statements written in the article.

Articles that are accepted for publishing are classified into four categories: original scientific papers, preliminary communications, subject reviews and professional papers.

Original scientific papers are articles that according to the reviewer and the editorial board contain original theoretical or practical results of research. These articles need to be written in such a way that based on the information given, the experiment can be repeated and the results described can be achieved together with the author's observations, theoretical statements or measurements.

Preliminary communication contains one or more pieces of new scientific information, but without details that allow recollection as in original scientific papers. Preliminary communication can give results of an experimental research, results of a shorter research or research in progress that is deemed useful for publishing.

Subject review contains a complete depiction of conditions and tendencies of a specific domain of theory, technology or application. Articles in this category have an overview character with a critical review and evaluation. Cited literature must be complete enough to allow a good insight and comprehension of the depicted domain.

Professional paper can contain a description of an original solution to a device, assembly or instrument, depiction of important practical solutions, and similar. The article need not be related to the original research, but it should contains a contribution to an application of known scientific results and their adaptation to practical needs, so it presents a contribution to spreading knowledge, etc.

Outside the mentioned categorization, the Editorial board of the journal will publish articles of interesting content in a special column. These articles provide descriptions of practical implementation and solutions from the area of production, experiences from device application, and similar.

10pt

3 WRITING AN ARTICLE

10pt

Article is written in the English language and the terminology and the measurement system should be adjusted to legal regulations, standards (ISO 80 000 series) and the SI international system of units. The article should be written in third person.

Introduction contains the depiction of the problem and an account of important results that come from the articles that are listed in the cited literature.

Main section of the article can be divided into several parts or chapters. Mathematical statements that obstruct the reading of the article should be avoided. Mathematical statements that cannot be avoided can be written as one or more addendums, when needed. It is recommended to use an example when an experiment procedure, the use of the work in a concrete situation or an algorithm of the suggested method must be illustrated. In general, an analysis should be experimentally confirmed.

Conclusion is a part of the article where the results are being given and efficiency of the procedure used is emphasized. Possible procedure and domain constraints where the obtained results can be applied should be emphasized.

10pt

4 RECAPITULATION ANNOTATION

10pt

In order for the articles to be formatted in the same manner as in this template, this document is recommended for use when writing the article. Finished articles written in MS Word for Windows and formatted according to this template must be submitted using our The Paper Submission Tool (PST) (https://tehnickiglasnik.unin.hr/authors.php) or eventually sent to the Editorial board of the Technical Journal to the following e-mail address: tehnickiglasnik@unin.hr

The editorial board reserves the right to minor redaction corrections of the article within the framework of prepress procedures. Articles that in any way do not follow these authors' instructions will be returned to the author by the editorial board. Should any questions arise, the editorial board contacts only the first author and accepts only the reflections given by the first author.

10pt

5 REFERENCES (According to APA)

10pt

The literature is cited in the order it is used in the article. Individual references from the listed literature inside the text are addressed with the corresponding number inside square brackets i.e. "… in [7] is shown …". If the literature references are web links, the hyperlink is to be removed as shown with the reference number 8. Also, the hyperlinks from the e-mail addresses of the authors are to be removed. In the literature list, each unit is marked with a number and listed according to the following examples (omit the subtitles over the references – they are here only to show possible types of references):

9pt

[1] See http://www.bibme.org/citation-guide/apa/

[2] See http://sites.umuc.edu/library/libhow/apa_examples.cfm

[3] (Style: Times New Roman, 9pt, according to APA)

[4] Amidzic, O., Riehle, H. J., & Elbert, T. (2006). Toward a psychophysiology of expertise: Focal magnetic gamma bursts as a signature of memory chunks and the aptitude of chess players. *Journal of Psychophysiology, 20*(4), 253-258. https://doi.org/10.1027/0269-8803.20.4.253

[5] Reitzes, D. C. & Mutran, E. J. (2004). The transition to retirement: Stages and factors that influence retirement adjustment. *International Journal of Aging and Human Development, 59*(1), 63-84. Retrieved from http://www.baywood.com/journals/PreviewJournals.asp?Id=0091-4150

[6] Jans, N. (1993). *The last light breaking: Life among Alaska's Inupiat Eskimos*. Anchorage, AK: Alaska Northwest Books.

[7] Miller, J. & Smith, T. (Eds.). (1996). *Cape Cod stories: Tales from Cape Cod, Nantucket, and Martha's Vineyard*. San Francisco, CA: Chronicle Books.

[8] Chaffe-Stengel, P. & Stengel, D. (2012). *Working with sample data: Exploration and inference*. https://doi.org/10.4128/9781606492147

[9] Freitas, N. (2015, January 6). People around the world are voluntarily submitting to China's Great Firewall. Why? Retrieved from http://www.slate.com/blogs/future_tense/2015/01/06/tencent_s_wechat_worldwide_internet_users_are_voluntarily_submitting_to.html (Style: Times New Roman, 9pt, according to APA)

10pt

10pt

Authors' contacts:

8pt

Full Name, title
Institution, company
Address
Tel./Fax, e-mail

8pt

Full Name, title
Institution, company
Address
Tel./Fax, e-mail

Note: Gray text should be removed in the final version of the article because it is for guidance only.

CALL FOR ABSTRACTS

ICPADM2021

13th International Conference on the Properties and Applications of Dielectric Materials

11–15 JULY 2021 | JOHOR BAHRU, MALAYSIA

Emerging Dielectrics for Energy Sustainability

KEY AREAS

- ❑ Aging of HV insulation
- ❑ Bio-dielectrics
- ❑ Conduction and breakdown in dielectrics
- ❑ Dielectrics phenomena and applications
- ❑ Eco-friendly dielectrics
- ❑ Emerging dielectric materials
- ❑ Gaseous breakdown and discharges
- ❑ HV insulation design via computational analysis
- ❑ HVDC insulation systems
- ❑ Monitoring and diagnostics of electrical insulation
- ❑ Nano-technology and nano-dielectrics
- ❑ New diagnostic applications for dielectrics
- ❑ New functional dielectrics for electrical systems
- ❑ Partial discharges
- ❑ Space charge and its effects
- ❑ Surface and interfacial phenomena

KEY DATES

Receipt of Abstract
SEPTEMBER 15, 2020

Notification of Abstract Acceptance
NOVEMBER 15, 2020

Receipt of Full Paper
JANUARY 15, 2021

Notification of Paper Acceptance
MARCH 15, 2021

For more details, please visit :
https://attend.ieee.org/icpadm-2021/

Hosted by: UTM UNIVERSITI TEKNOLOGI MALAYSIA

Sponsored by: IEEE DEIS

FSB University of Zagreb Faculty of Mechanical Engineering and Naval Architecture

CaPLM Croatian Association for PLM

12th International Scientific Conference

Augmented Reality in

an Atmospere of Covid 19 Crisis

Management of Technology Step to Sustainable Production

MOTSP 2021

September 8 - 10, 2021

Poreč/Parenzo, Istria, Croatia

International Conference *Management of Technology – Step to Sustainable Production* (MOTSP 2021), will take place from 8th-10th September 2021 in Poreč, Istria, Croatia as a joint project organized by the Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb and *Croatian Association for PLM.* The main objective of this Conference is to gather international experts from academic entities, research laboratories and industries related to the field of Management of Technology and Sustainable Production. The Conference will also provide a platform for sharing knowledge, ideas and results between science and industry.

The management of technology, stimulation of innovation and invention, augmented reality, management of COVID 19 and transfer of technology are important challenges of the developed countries and countries in transition.

ORGANIZATION

Faculty of Mechanical Engineering and Naval Architecture
University of Zagreb, Ivana Lučića 5, 10000 Zagreb, Croatia
Croatian Association for PLM, Prilaz V. Brajkovića 12, 10020 Zagreb, Croatia

MAIN TOPICS

Management of Technology
Production, Operation Management
Strategic, Engineering Management
Industrial Engineering
Rapid Prototyping and Manufacturing
Computer Integrated Manufacturing CAD, CAM, CAPP, CAQ
Intelligent Manufacturing
Product Lifecycle Management (PLM)
Total Cost Assessment
Ergonomics
Sustainable Production
Social Responsibility
Total Cost Assessment
Financial Sustainability
Sustainable Transport Management
Transport Technology and Organization
Corona pandemic as a scientific challenge

GENERAL INFORMATION

Deadline for abstract and full paper submission:
March 30th , 2021

Submission of full paper and abstracts (up to 150 words) using abstract submission form at **motsp.eu**
Notification of acceptance of full papers:
April 30th, 2021

Chairman:
Prof. Predrag Ćosić, PhD, FAMENA, Department of Industrial Engineering; pcosic3@gmail.com

Secretary:
Assistant Professor Tihomir Opetuk,
tihomir.opetuk@motsp.eu
Assistant Professor Zdenka Keran,
zdenka.keran@motsp.eu
e-mail: info@motsp.eu
URL: www.motsp.eu

INFORMATION FOR PARTICIPANTS

CONFERENCE VENUE
Valamar Diamant Hotel 4*, Brulo 1, 52440 Porec, Croatia
Url: https://www.valamar.com
Tel: +385 52 408 205

Participants and accompanying persons of **MOTSP 2021** have a special offer from the Valamar Hotel Diamant 4*and Valamar Hotel Residence 3* Poreč. MOTSP2021 sugest to make a **reservation of accommodation ONLY** through motsp21@valamar.com.

CONFERENCE LANGUAGE
English
Presentation: Presentations should take 15 minutes. Authors will have the possibility of using computers for the presentations (Power Point).

REGISTRATION
Registration should be acknowledged at the reception desk on Hotel Valamar Diamant 4*
September 7, 2021 18:00 – 20:00
September 8, 2021 08:00 – 09:00

Conference Fees
Early registration fee until May 15 - **EUR 460**
Late registration fee - **EUR 490**
Deadline for registration fee payment: **May 15, 2021**

All accepted papers would be published in the Technical Journal. Paper publishing would be included by conference fee.

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 15 – BROJ / NUMBER 1

OŽUJAK 2021 / MARCH 2021 – STRANICA / PAGES 1-167

Sveučilište
Sjever

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

ISSN 1846-6168 (PRINT) / ISSN 1848-5588 (ONLINE)

TEHNICKIGLASNIK@UNIN.HR – HTTP://TEHNICKIGLASNIK.UNIN.HR