

ISSN (Print):2321-404X, (Online):2321-4384

# **International Journal of Soft Computing & Artificial Intelligence**

**Volume No. 12**

**Issue No. 1**

**January - April 2024**



**ENRICHED PUBLICATIONS PVT. LTD**

**S-9, IInd FLOOR, MLU POCKET,  
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,  
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,  
PHONE: - + (91)-(11)-47026006**

# **International Journal of Soft Computing & Artificial Intelligence**

## **Aim & Scope**

IJSCAI provides a new forum for dissemination of knowledge on both theoretical and applied research on Artificial Intelligence and Soft Computing with an ultimate aim to bridge the gap between these two non-coherent disciplines of knowledge. The Journal offers survey and review articles from experts in the field, promoting insight and understanding of the state of art, and latest trends in the field. The content includes original research and innovation ideas, applications from all over the world.

All published papers are also available freely with online full-text content. This forum accelerates interaction between the above two bodies of knowledge, and fosters unified development in next generation computational models for machine intelligence. Special Issues devoted to topics of current interest will be published occasionally.

## Editor In Chief

**Dr. Sakkaravarthi Ramanathan**  
mail: editor@iraj.in

## Publication and Distribution In-Charge

**Prof. Vinod Tyagi,**  
National Organizing Secretary, Institute of Research and Journals  
Email-ijaseat@yahoo.in, Mob-+ 91-9040435740

## International Editorial Board

<b>Ching-Hsien (Robert) Hsu,</b> Professor, Department of Computer Science and Information Engineering, Chung Hua University, Taiwan	<b>Dr. P. K. Agarwal,</b> Professor, Department of Civil Engineering, Maulana Azad National Institute of Technology, Bhopal, India
<b>Prof. Anand nayar,</b> Assistant Professor, Department of Computer Applications & IT, KCL Institute of Management and Technology	<b>Dr. R. A. Savanur,</b> Professor, Dept. of Mech. Engineering, BLDEA's CET., BIJAPUR
<b>Prof. Mohammed Farooque Khalil,</b> Sr Lecturer ,Department of Computer Science Jazan University, Kingdom of Saudi Arabia (KSA)	<b>Dr.Yaswanth Kumar Avulapati,</b> Department of Computer Science (M. C. A), S. V. U. College Of Commerce Management and Computer Science, S. V. University, Tirupati, India
<b>Dr. Rawan Abulail,</b> Professor, Dept. of CS & IT, Philadelphia University	<b>Dr. Krishnan,</b> A K.S.R College of Engineering, India
<b>Dr. Duraiswamy,</b> K K.S. Rangaswamy College of Technology, India	<b>Dr. Wathiq Mansoor,</b> American University in Dubai, United Arab Emirates
<b>Dr. Abdel-Hamid,</b> Mourad United Arab Emirates University	<b>Dr. Ivan Bruha,</b> Mc Master University, Canada
<b>Dr. Viliam Makis,</b> University of Toronto, Canada	<b>Dr. Ramesh Agarwal,</b> Washington University in St. Louis, United States
<b>Dr. Peter Milligan,</b> Queen's University Belfast, United Kingdom	<b>Dr. Periasamy,</b> P. S K. S. R College of Engineering, India
<b>Mr. Gnanajeyaraman Rajaram,</b> PVP CETW, India	<b>Dr. Bommanna Kanagaraj,</b> PSNA College of Engineering and Technology, India

**Dr. Yuchou Chang**  
University of Wisconsin, United States

# International Journal of Soft Computing & Artificial Intelligence

(Volume No. 12, Issue No. 1, January - April 2024)

## Contents

Sr. No.	Article / Authors Name	Pg. No.
1	Signature Verification using Convolutional Neural Network - <i>Anagha Bhat, Bharathi Gummanur, Likhitha Priya</i>	1 - 6
2	Study on the Big Data Log Analysis: Goals, Challenges, Issues, and Tools - <i>Mostafa Mohamed Shendi, Hatem Mohamed Elkadi, Mohamed Helmy Khafagy</i>	7 - 20
3	A Real Time Monitoring Dashboard for A Smart and Sustainable Parking Area - <i>Samiha Najah, Siham Al Hinai, Mitha Al Ashkhari, Adhiyaba Al Maqbali</i>	21 - 26
4	Finding an Optimal Free-Obstacle Path for A Mobile Robot using A* Algorithm - <i>Amer S. Alharthi, Saad B. Alotaibi</i>	27 - 34
5	Use of Fuzzy Set Theory in the Medical Diagnosis - <i>Kiran Pal, V. Kumar, H. D. Arora</i>	35 - 43

# Signature Verification using Convolutional Neural Network

<sup>1</sup> Anagha Bhat, <sup>2</sup> Bharathi Gummanur, <sup>3</sup> Likhitha Priya

<sup>1,2,3</sup>Student, Dayananda Sagar College Of Engineering, Shavige Malleshwara Hills, 91st Main Rd, 1st Stage, Kumaraswamy Layout, Bengaluru, Karnataka 560078

E-mail: <sup>1</sup>anaghabhat.97@gmail.com, <sup>2</sup>bharathi.purple@gmail.com, <sup>3</sup>likitha.priya98@gmail.com

## ABSTRACT

*A number of existing authentication systems use biometric information, more specifically, signatures, for authentication. However, authentication based on signatures is not infallible, as it is possible to forge signatures with a convincing degree of similarity. There needs to be a reliable method of detecting fake signatures in order to avoid forgery and fraud. In the proposed method, we use a convolutional neural network as a binary classifier. The datasets used were sourced from SigComp 2011 and Kaggle. The network is made to learn and extract features from the signatures that are pre-labelled either as fake or genuine. This network is then tested on a previously unseen set of signatures. This method gave varying results; the accuracy achieved from the method described varied in the range 15-60%. There could be a significant improvement in the results in case of availability of larger datasets for training.*

**Keywords - Signature Verification, Convolutional Neural Networks**

## I. INTRODUCTION

A plethora of important processes require authentication, from something as trivial as creating an account on social media to something as crucial as bank transactions. Verifying the individual's identity is the core of the process of authentication. This can be done in several ways. These exist at a range of levels, with the complexity of the process depending on the intricacy of the security needed. For example, email verifications for social media accounts, pins for credit/debit card transactions, passwords for user accounts, signatures in legal documents, as well as biometric recognition that involves verifying an individual's irises, gait, keystroke and fingerprints. Out of the above mentioned methods, biometric verification appears to be the most complex and sophisticated method, requiring expensive methods and equipment.

However, some kinds of biometric information, such as fingerprints, retinal scan, iris scan, etc. are important personal data that some users might not wish to disclose. Additionally, many existing systems use signatures for verification, and in many cases, this verification is done manually, by an individual who is designated to identify a forged signature. An example for this is a bank where signatures are verified by specific employees. Unlike biometrics, signatures can be replicated, thereby bypassing security measures taken. This increases the chances of signatures being forged by fraudulent individuals with malicious intentions. Since signature verification is still commonly in use, there is a need for a

system that can detect such forgeries. This process of identification of forgeries can be accomplished by a variety of methods. Forged signatures have a number of „tells“ that help identify them as fake. Common tells are the length of the signature, breaks, blots, shaky lines

[1] and the aspect ratio of the signature. These patterns can be detected using certain machine learning algorithms and using neural networks.

[2] involves signature verification performed on Persian signatures and an approach which utilizes a CNN to achieve the task of feature extraction. The classifier employed here is a multilayer perceptron network. This method is also reliable when it comes to variations in size or position of the signature.

[3] utilizes CNNs for authentication. The approach described uses transfer learning in the process of training, providing promising results with accuracies above 95%. The dataset used was the ICDAR 2011 SigComp dataset.

In [4], the proposed approach involved a Siamese network. The scenario was writer-independent and cross domain datasets were experimented on. Other parameters that were varied and experimented on were scripts from multiple languages as well as handwriting. The problem identified in [5] is writer identification based on handwritten characters in the Chinese language. The approach described utilises a deep CNN in a WI format. A technique known as Drop Stroke is used for the amplification of the data in the dataset. Additionally, a path-signature feature is used with the deep CNN, first described in [6].

In[7], there is a focus on writer-independent (WI) systems. Given that using a writer-dependent (WD) system leads to a significant improvement in performance, the proposed method aimed at determining whether similar results could be achieved using a writer-independent system. This approach utilised a single Support Vector Machine (SVM) as the classifier and used multiple datasets. For each dataset, the approach described was able to surpass the other methods mentioned, including the writer-dependent classification method.

[8] consists of an interesting approach, one that involves both writer-independent as well as writer-dependent scenarios. First, deep convolutional neural networks are made to assimilate features from the signatures. This is done in a writer-independent scenario. Then, the features obtained are utilised on a different group of users in a writer-dependent scenario. Hence, the classification of signatures is WD.

The situation in [9] is the specific case of signatures drawn by hand on a screen. The movement of the device is also taken into account by employing accelerometer sensors. A CNN is employed for the purpose of feature extraction, whereas an autoencoder is utilized for classification. The common „tells“ in a forgery are considered to be the features, which are fed into the classifier.

Considering a slightly different situation, the objective in [10] is different, although the problem statement is similar: the goal is person re-identification from images in which the subject is a human. The approach used involves the combination of two different neural networks, one for identification and the other for verification, by concurrently computing both the identification loss as well as the verification loss. Although the condition is different, utilizing a similar method, i.e. by combining two different models, in signature verification might produce significant results.

This paper is organized as follows: it begins with a detailed description of our approach, the implementation of which consists of two stages: i) preprocessing, and ii) training and running the neural network. This is followed by an analysis of the results, and, lastly, by a conclusion along with a discussion regarding future enhancements.

## II. IMPLEMENTATION AND DESIGN

The approach described utilizes a binary classifier to classify a signature as real or fake, based on the features extracted from the input data. The aim is to train the model on a set of real and forged signatures, hence enabling it to pick up on the above-mentioned „tells“, allowing it to identify such a signature as fake. In addition, these „tells“ intuitively seem to be universal across all languages, and can therefore be used to classify signatures, irrespective of the language.

### Figure 1: Flow diagram

#### Datasets

One of the datasets used was extracted from the ICDAR Signature Dataset which was provided as a part of SigComp 2011, and another was obtained from Kaggle, with the source unknown. The data consists of a series of images in .PNG format, with the genuine signatures and the forged signatures in separate directories. The training data consists of 5,486 images while the test data consists of 1307 images.

#### Preprocessing

The task of preprocessing is extremely essential as it enables the removal of noise as well as the standardization of the images, thus making the process of pattern recognition easier and more efficient. The following are the stages involved in preprocessing:

- **Reformatting file names:** Since the data was obtained from multiple datasets, to ensure uniformity, the file names are altered a common format. Each file, based on whether it is a genuine or a forged signature image, is renamed using a prefix to identify the signature type. „G“ indicates genuine and „F“

indicates forged. This prefix is followed by a numerical sequence that consists of two parts. The first part is the serial number assigned to the individual to whom the signature belongs. The second part acts like a serial number to the image sample for each signature type (genuine or fake) for each individual. This helps to fetch the images in order to label them accordingly as a part of the input to the model.

- **Resizing:** Since it is preferable to pass data of the same shape to the model and since the size of the input layer depends on the same, the images must be reshaped. The maximum value for each dimension (width and height) is found, the image is resized by converting the larger of the dimensions to a size of 200 pixels while keeping the aspect ratio constant. This ensures a maximal limit to the size of the images, and this limit is reinforced in the next stage.
- **Padding:** To ensure that the smaller dimension is also equal to 200 pixels, the image is padded with pixels of value [255, 255, 255]. In other words, this is the same as adding white pixels. This standardizes it to an image of dimensions 200 x 200.
- **Conversion to a binary image:** The image has a multitude of grayscale values. In our approach, we have converted such an image into a binary image, which consists only of pixel values that represent either black or white, eliminating any intermediate grey values. This is done by setting a threshold value such that any pixel values above this threshold are converted to white whereas the values below it are converted to black.

Figure 2 shows the different stages of preprocessing.

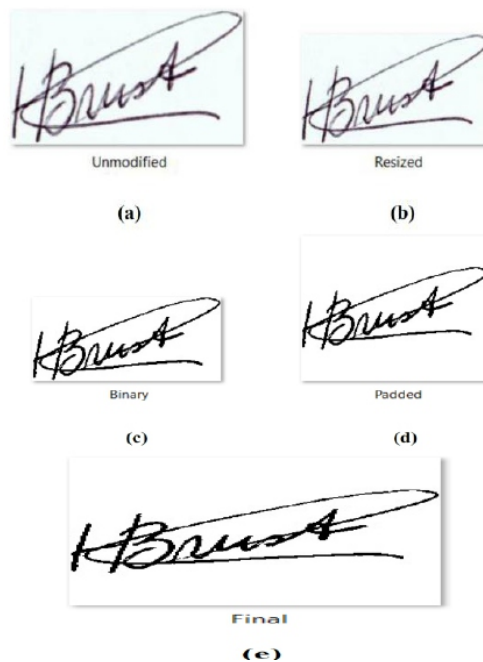


Figure 2 (a)-(e): Stages in preprocessing



### Convolutional Neural Network

This model uses TFLearn which is a deep learning library built on top of tensorflow. Initially, the learning rate is specified. Then, the images are read in grayscale. The training data is appended to a list. Next, a corresponding set of labels indicating whether the signatures are genuine or forged is obtained. The training data is now a list of lists, each list consisting of the image data in the form of a numpy array and the corresponding one hot encoded label([1,0] for genuine and [0,1] for forged). Following this, the structure of the convolutional neural network is designed. The convolutional neural network used consists of 6 layers. The number of features in the input data, i.e. the shape of the input data, is specified in the tf learn. Input data method. Next, tflearn.conv\_2d is used to define the structure of the network, including the number of nodes per layer along with the activation function used in that layer.

The first five layers use the rectifier as the activation function and the last layer uses softmax. The first and the third layers handling linear problems and the second and the fourth layers handling non linear problems. In the first four layers, max pooling is done to reduce the dimensionality of the feature map. The sixth layer uses the Adam optimizer. Next, the model is made to learn by training it on the input data by using the fit() method. The model is then saved. The test data, similar to the training data, consists of a list of lists, but the label isn't in the form of a one hot encoded. It is in the form of anything the user can use to test if the results are accurate. The predict() method is used to observe the model's performance on the test data. The 2D plotting library matplotlib is used to print the images scaled down to manageable size and corresponding labels are printed to demonstrate whether the model predicted the results accurately.

### III. RESULTS

Initially, it was observed that the training accuracy was in the range 40 to 50 percent, as the loss was in the range 12-14. On varying the learning rate, the training accuracy went up to peak at 61.95 percent, as the loss reached a minimum of 9.20, when the learning rate was set to 0.1. Decreasing the learning rate to the range e-2 and lower reduces the training accuracy and increases the loss to the initial values mentioned above. However, increasing the learning rate further led to a decrease in accuracy and an increase in loss, as mentioned above. However, the testing accuracy in these cases varied from 15% to 60%. It was observed that there seemed to be no correlation between accuracy and the value of the learning rate. A record of accuracy versus the number of epochs is shown in Table 1.

Epochs	Accuracy
5	0.35
15	0.55
30	0.40
45	0.50
60	0.55

**Table 1: A measure of accuracy against the number of epochs.**

#### IV. CONCLUSION

It can be seen that a neural network approach aided by image modification or formatting can be used to identify forged and genuine signatures as mentioned above. However, this approach doesn't provide the best results. A slight modification to the above mentioned approach can be used to obtain correct, desired results.

The following approaches could show a significant improvement in the performance of the model:

- Usage of more data to train the model.
- Training the model separately on the real and forged signatures for each individual to enable the model to form patterns for each individual's forged and real signatures.
- Usage of grayscale images as input to the model might enhance results by enabling detection of one of the "tells"- pressure applied by the individual while signing.

#### REFERENCES

- [1] T. Amo. "How to Tell If a Signature Is Forged." Internet: <https://bizfluent.com/how-7599353-tell-signature-forged.html>, Jan. 25, 2019 [May 13, 2019].
- [2] H. Khalajzadeh, M. Mansouri and M. Teshnehlab, "Persian Signature Verification using Convolutional Neural Networks," *International Journal of Engineering Research and Technology*, vol. 1, April- 2012.
- [3] G. Alvarez, B. Sheffer, and M. Bryant, "Offline Signature Verification with Convolutional Neural Networks," *Tech. rep., Stanford University, Stanford* 2016.
- [4] S. Dey, A. Dutta, J. I. Toledo, S. K. Ghosh, J. Lladós, and U. Pal. "SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification," *arXiv:1707.02131*, July 2017.
- [5] W. Yang, L. Jin, M. Liu, "Character-level Chinese Writer Identification using Path Signature Feature Drop Stroke and Deep CNN", *Proc. of the 13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 546-550, 2015.
- [6] K.T. Chen, "Integration of paths--A faithful representation of paths by noncommutative formal power series," *Trans. of the American Mathematical Society*, pp. 395-407, 1958.
- [7] V. L. Souza, A. L. Oliveira, and R. Sabourin, "A writer-independent approach for offline signature verification using deep convolutional neural networks features," In: *7th Brazilian Conference on Intelligent Systems (BRACIS)*. pp. 212–217. IEEE, 2018.
- [8] L. G. Hafemann, R. Sabourin, and L. S. Oliveira, "Writer-independent feature learning for offline signature verification using deep convolutional neural networks," In *IEEE IJCNN (2016)*. IEEE, pp. 2576–2583, 2016.
- [9] S. Nam, H. Park, C. Seo and D. Choi, "Forged Signature Distinction Using Convolutional Neural Network for Feature Extraction." *Appl. Sci.*, 8, 153, 2018.
- [10] Z. Zheng, L. Zheng and Y. Yang, "A Discriminatively Learned CNN Embedding for Person Reidentification", *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, v.14 n.1, pp.1-20, January 2018.

# A Study on the Big Data Log Analysis: Goals, Challenges, Issues, and Tools

<sup>1</sup>Mostafa Mohamed Shendi, <sup>2</sup>Hatem Mohamed Elkadi, <sup>3</sup>Mohamed Helmy Khafagy

Information System Department, Cairo University, Egypt Computer Science Department, Fayoum University, Egypt E-mail: mostafa\_shendi@yahoo.com

## ABSTRACT

*Security, availability, and performance are becoming more frequently requested and sophisticated. Traditional solutions are unable to protect organization's assets or keep their services running smoothly. These solutions need to focus more on customer needs and satisfaction. Organizations need to perform real-time analysis on a huge amount of data from various types to discover anomalous fragments within a reasonable response time. Business can widen the scale of processed data, accelerate the speed of threat detection, keep their services up and running by monitoring the servers' status, predict failure before it happens and increase customer satisfaction by providing efficient service in a timely manner. Processing the huge amount of the system's log files using relational database technology has been facing a bottleneck. To analyze such large data sets we need parallel processing system and reliable data storage mechanism. Big Data is the solution to overcome these issues. The main purpose of this paper is to highlight the characteristics of Big Data and present a review of log file analysis in Big Data environment as a first step towards getting the maximum benefits of big data in logs analytics.*

**Keywords - Big Data, Hadoop, Mapreduce, HDFS, Log Analysis, Log Files.**

## I. INTRODUCTION

Nowadays many businesses are going online. Each field has its own way of bringing their applications and businesses online. We are seeking a mechanism to handle huge volumes of log files generated in different structures and from different devices and applications. Many studies and papers - discussed previously from different perspectives - highlights how big data can be useful in analyzing the logs and user's behaviors on a web page, and detecting anomalous behaviors, errors and exceptions.

### Log files:

Logs are computer-generated files that capture network and server's operation data containing the whole information regarding the user's activity. These activities are written in various log files such as a weblog, firewall log, network log, Router log, etc. these log files have millions of entries. Log analysis needs a long time to be investigated. Analysis of log files has been very important in resolving many issues [1]. Servers, routers and other devices logs may be our best line of defense.

## **Big Data**

Basically, Big Data is defined as data sets that could not be perceived, acquired, managed and processed within a reasonable time by traditional IT and software or hardware tools. Big data exceeds the processing capacity of conventional database systems [2]. The data is huge and massive, it moves at a very high speed, and does not have to fit in the structures of existing database architectures. To gain value from this data, there must be an alternative way to process it. Big data analytics is the key to unlock the insights from all data types as it enables us to analyze all structured, semi-structured and unstructured data together. It is powerful because it enables the organization to combine, integrate and analyze all data at once regardless of their source, type, size or format in order to generate the insights needed to address a wide range of business challenges.

Hadoop-MapReduce framework provides parallel distributed processing and reliable data storage for large volumes of log files. Here Hadoop's characteristic of moving computation to the data rather than moving data to computation helps to improve response time [3].

Results from the reviewed papers show that the majority of researchers applied MapReduce as the main component of Hadoop for analyzing the log files and HDFS as the data storage. Previous researchers have also used some tools and algorithms with the Hadoop Framework for analysis purposes.

The output of this paper will provide a comprehensive review of Big Data usages in analyzing different types of log files with different methods, tools, and techniques.

Hadoop takes log files analysis to the next level by speeding and improving security forensics and providing a low-cost platform to detect compliance violations [4].

## **II. CONSIDERATIONS AND CHALLENGES IN BIG DATA LOGS ANALYSIS**

In the Last few years, big data has been acquired in several domains. A web-based application such as internet text, social computing, recommender systems, online communities, markets prediction, social network analysis, and internet search documents encounter big data frequently. Also, millions of records are generated from the devices over a network which should be reviewed and handled in appropriate response time.

To handle the challenges, we need to manage various computational and information security complexities and methods to analyze Logs. The logs are generated in different formats with different structures and contains useless data that need to be cleaned.

When working with large logs, practitioners often face issues such as scarce storage, incapable analysis tools, inaccurate capture and replay of logs, and inadequate privacy. Researchers have devised some practical solutions, but important challenges remain:

### **Data Distribution**

Performing computation on large volumes of log files has been done earlier but what makes Hadoop different is its simplified programming model and its efficient, automatic distribution of data and portability [5].

### **Isolation of Processes**

Each individual record is processed by a task isolated from other tasks, limiting the communication overhead between the processes by Hadoop. This makes the whole framework more reliable. In MapReduce, Mapper tasks process records in isolation. Individual node failure can be worked around by restarting tasks on another machine.

### **Type of Data**

Log files are rows consisting of semi-structured or unstructured records. Hadoop is compatible with most types of data and more suitable even in case that the log files are mapped to a structured data due to RDBMS data size limitation. It works well for simple text files as well.

### **Fault Tolerance**

Hadoop cluster solved the problem of data loss. Blocks of the input file are replicated by a factor on multiple machines in the Hadoop cluster. Therefore, even if any machine goes down, another machine where the same block is residing will take care of further processing.

### **Data Locality and Network Bandwidth**

Log files are spread across HDFS as blocks, therefore the node which operates on a subset of files will be detected by the locality of the node, and the purpose is reducing the strain on network bandwidth and avoid unnecessary network traffic [6].

## **III. LITERATURE REVIEW TECHNIQUES AND TOOLS**

Numerous efforts have been expended to get the maximum benefits of using big data in the analysis of log files on different structures. Siddharth et Devesh [7] used a weblog analyzer tool for collecting the statistics over web sites, Pig scripts are written on the classified log files to satisfy the certain query. The log file is processed for exceptions and errors.

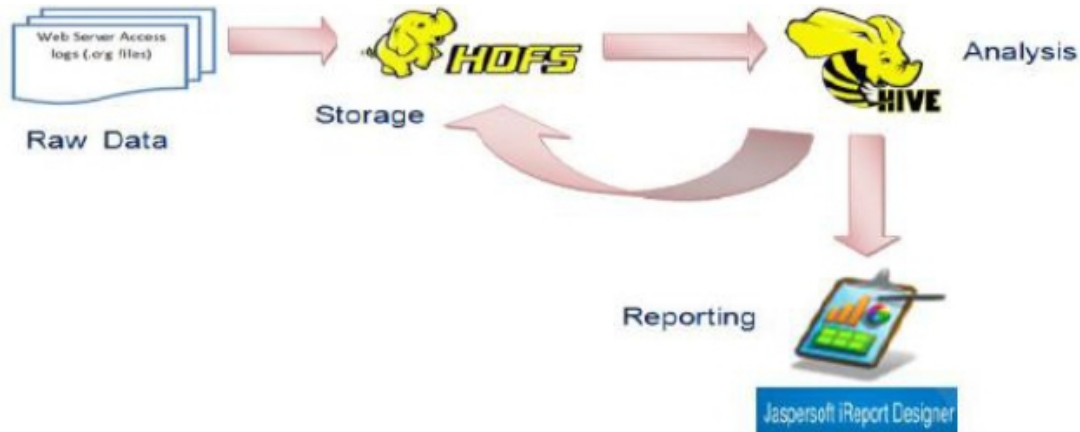
The tool worked on the basic information present in the log file as follows:

IP Address /Username - Time Stamp - Page Visited lastly - Success rate - User agent - URL - Request type

The paper used Map reduce technique to process the log files, Hadoop framework for parallel computation of log files, pig tool for taking a huge data from different sources and placing it into HDFS for further processing. The framework makes use of tableau tool for a pictorial representation of log files accessed by the users.

The limitation of the paper is the need to collect, analyze and gather events occurrences to subscribers such as operating system processes, database rules, and human operators.

Harish et Kavitha [8] applied Hadoop-MapReduce programming model for analyzing web server log files where data is stored on multiple nodes in a cluster so that access time required can be reduced and MapReduce work for large datasets giving efficient results to summarize the results for a web application, perform log analysis that improves the business strategies as well as generates statistical reports. The data cleansing is done using Hive query language. Only the logs with the status code are extracted and stored in Hive table for further analysis.



### Raw log data processing and visualizing

The paper analyzed website logs and obtained the following: General Statistics, Activity statistics, Access Statistics, Visits-per-Country and Errors. Then performed log analysis to improve the business strategies and to generate statistical reports. Which provides graphical reports showing hits for web pages, user's activity, and identifies in which part of the website users are interested, traffic sources, etc. Using these reports the business communities can evaluate parts of the website that needs to be improved, the potential customers and max hit traffic source.

Mayur Mahajan et Omkar Akolkar et. [9] Used web page collection algorithm that uses cluster mining to find a group of connected pages at a web site. The proposed algorithm takes web server access log as an input and maps it into form clustering. The output is obtained by mining the web user's logs. Faster analysis of real-time weblogs generated by various web sites with parallel processing using Apache Hadoop.

Recommendations will help the administrators improve the website in various locations according to various factors. Using the analysis report the developers can fix the bugs in the website. Analyzing the real-time weblog which contains errors and exceptions and the number of times it has occurred.

The paper builds its system architecture which consists of Third Party Website, Web Log Analyzer, and User-interface.

The system analyzes the log files and presents it to the user in a more understandable form such as pie-charts and bar-graphs, and gives recommendations to the user for improvement of the website for popularity and reputation risk matters.

The limitation of this work is that real-time weblog analysis needs to be done. The fluctuations can be shown in reports online as per the changes of the log.

Sayalee Narkhede et Tripti Baraskar [10] the structure used in the work consists of web servers on cloud Framework, Hadoop storage and MapReduce programming model and user interface.

Log files are collected from the web servers. HMR log processor is implemented in three phases: log pre-processing, interacting with HDFS and map-reduce programming model implementation.

HMR log file analysis tool provides graphical reports showing hits for web pages and it's sub-elements, user's activity, and traffic sources. Data get stored in the hierarchy on several nodes in a cluster so that access time required can be reduced which saves much of the processing time. Here Hadoop's characteristic of moving computation to the data rather moving data to computation helps to improve response time. Secondly, MapReduce successfully works for large datasets giving efficient results.

Kyung et Jeong et Hyun [11] suggested the need for new Enterprise Security Management (ESM) which is used for integrated management of network systems such as firewall, IPS, VPN, etc. also they have mentioned that the current ESM has a limitation of blocking external cyber-attacks only due to using the networking attack detection method which monitors the traffic inflows from outside to inside.



The research uses big data to enhance the intelligence of security by analyzing the relationship between security and data events created from network, system, application service of main IT infrastructure. They collect and analyze rapidly a large amount of structured/unstructured data by distribution-based saving/processing technology through security log system using Big data. Fetch real-time analysis performance by distribution-based multi-searching, presenting a flexible dashboard for visible analysis, and applying correlation analysis technique of all elements. The integrated security control will be possible by securing availability and expandability in case of applying security log system using Big data.

Data collection Algorithm is expected to enhance customer satisfaction by restricting the inflow of harmful code developed in-house, and real-time monitoring becomes possible. The limitation of the research is the absence of any experimental setup and not presenting any sample results of the proposed work.

Sandesh et Abhishek et... [12] Discussed the vulnerability assessment and presented some free tools to use for smaller modules, list of some tools which are easy to use, so that anyone with basic technical knowledge can use them, allowing even a small enterprise to utilize them to generate results and take appropriate actions: Vulnerability, Threat, Attack and Exploit described different techniques to collect and analyze logs, which are a very important source of information and could be used for various security purposes such as a NetFlow, IDS, firewall log activity, file access, system errors, or login failures.

The limitation of the work was the fact that the researchers didn't propose a complete integrated module from the tools they have mentioned in order to be used by the SME`s companies.

Sharma et Arun [13] focused on the Web applications which uses a combination of server scripts (ASP, PHP, etc.) and client-side scripts (HTML, JavaScript, etc.) to develop the application. The script on the client-side communicates with the presentation of information, while the server-side script handles all the complex operation such as storing and retrieving the information. The paper added a mining layer of the web which is divided into three categories: web content mining, web structure mining and web usage mining to discover hidden information on the Web by data cleaning, filtration, data cube construction, analytical processing, and data mining.

The basic information found in the log file: User name, Path Traversed, Success rate, User Agent and URL



The limitation of the paper is that there is no integration of the proposed work. Therefore, the paper only added a data mining step for data cleansing and filtration and didn't mention the enhancement percentage through it.

Rotsnarani et Mrutyunjaya [14] used Hadoop framework, data mining techniques to extract useful information and clustering method Centroid-based (K-Means Clustering) to cluster the input logs and then process them in Hadoop framework. The analysis result can be used for detecting security threats in the network and to inform the administrators about Intrusion Detection, SQL-Injection, system error, etc.

Log files can be classified into categories depending on the location of their storage. This classification can provide batch analysis in minimum response time and in-memory computing capacity to process logs in an efficient and stable way.

The paper focused on two types of logs: web server logs and application server logs.

The proposed technique consists of scanning logs in real time as well as in off time which can give the best result of analysis of the logs that can be investigated by the IT admins to take preventive actions for the attacks.

The real-time logs can be stored and processed by using flume. In-memory engine for analytics is a better option for faster processing to detect attacks like SQL injection, Cross-Site Scripting (XSS), URL Injection, etc. The limitation of the paper is focusing on the web servers and applications logs only, it would be better to include network logs such as firewall and switches logs, etc... to the proposed work Souza1 et Girish [15] Error Log Analytics discussed the conversion of data from semi-structured to a uniform structured format, such that Analytics can be performed over it. Business Intelligence (BI) functions such as Predictive Analytics are used to predict and forecast the future status of the application based on the current scenario. Proactive measures can be taken rather than reactive measures to ensure efficient maintainability of the applications and the devices.

Error Log is a file that is created during data processing to hold data known to contain errors and warnings. It is usually printed after completion of processing so that the errors can be rectified .



After ETL operations performed, the error log file is brought into a uniform homogeneous format. To perform analytics on this data, it is loaded into a warehouse. Obtaining the sum of errors and warnings on the severity attribute is achieved using the Grouping operation. Cubing operation generates aggregates for all combinations of values in the selected columns. Historical error log data is maintained in the warehouse, which is used in various Analytics and business intelligence techniques.

Bhawna et Kiran [16] review paper explains how Big Data is analyzed by using the technique of Hadoop and why the Big Data Security Analytics is important to mitigate the security threats.

There should be several opportunities for big data security analytics to enter the enterprise security mainstream because of: continuing problems with detection and response of threats, Moore's law and open source and Tons of activity on the supply side.

Researchers must meet the needs of enterprises for developing a high-quality product by adopting security measures with the help of big data analytics with Hadoop. The paper uses big data analytics for analyzing enterprise data, and builds a framework based on Hadoop for dealing the targeted attacks using Big Data security analytics.

The limitation of the paper was the absence of any implementation for the proposed work, also they lacked focusing on the enterprise data security and developing security alerts which will provide employees with the ability to view the activity.

Jangla et Deepa [17] propose a new model based on big data analysis technology to prevent and detect previously unknown attacks. Explain attack predictions earned from applying big data technologies such as classification, text mining, clustering, and association rules.

Propose a system model that uses big data analysis technology for extracting data from various sources to react to previously unknown attacks.

Big data analysis uses various existing analysis techniques such as machine learning, artificial intelligence, data mining, etc.

The limitation of the paper was that it lacked the software ability to detect advanced threats, reacting accordingly and developing preventive measures for the future. The paper researchers should improve the quality and reliability of the security system.

Work Year	Method	Result	Strengths	Weakness
Siddharth et Saraf. "Analysis of Log Data and Statistics Report Generation Using Hadoop" [2015]	-Hadoop-MapReduce & Pig scripts - Pig scripts are written on the classified log files to satisfy a certain query. the log file is processed for exceptions and errors.	The framework makes use of tableau tool for a pictorial representation of log files accessed by the users.	finding the statics of web sites: IP Address /Uername - Time Stamp - Page Visited lastly - Success rate - User agent - URL - Request type	- Just focus on the website logs. - Need to collect, analyze and signal events occurrences to subscribers such as operating system processes, active database rules as well as human operators.
Harish et Kavitha. " Statistical Analysis of Web Server Logs Using Apache Hive in Hadoop Framework " [2015]	- Hadoop-MapReduce & Hive - data cleansing is done using Hive query language	obtained: General Statistics, Activity statistics, Access Statistics, Visits-per-Country, and Errors	evaluate which parts in the website that needs to be improved, and whom are the potential customers, from which geographical region the website is getting maximum hits, etc., which will help in designing future marketing plans	- focus on application logs only - Doesn't have the capability of distinguishing between bad connections, called intrusions or attacks, and good (normal) ones
Mayur et Omkar. "Real-Time Web Log Analysis and Hadoop for Data Analytics on Large Web Logs" [2016]	-Hadoop-MapReduce & cluster mining -Webpage Collection algorithm that uses cluster mining to find a group of connected pages at a web site.	Provide analysis errors report which help the developer fix the bugs in the website	Help the administrator for the improvements of the website in various locations on various factors	- Real-time weblog analysis needs to be done. The fluctuations can be shown in reports online as per the changes of log

<p>Sayalee et Tripti. "HMR LOG ANALYZER: ANALYZE WEB APPLICATION LOGS OVER HADOOP MapReduce" [2013]</p>	<ul style="list-style-type: none"> <li>-Hadoop-MapReduce &amp;Interface</li> <li>- Builds its structure which Consists of major components: Web servers, Cloud Framework implementing Hadoop storage and MapReduce programming model and user interface.</li> </ul>	<p>provide graphical reports showing hits for web pages, user's activity, in which part of website users are interested, traffic sources, etc.</p>	<p>Business communities can evaluate which parts of the website need to be improved, which are the potential customers, from which geographical region website is getting maximum hits, etc.</p>	<ul style="list-style-type: none"> <li>- Limitation of the work scope, need to build the structure on real-time methods</li> <li>- Security filed should be included to classify the coming requests as normal or deny of service.</li> </ul>
<p>Kyung et Jeong "A Study on the Big Data Log Analysis for Security" [2016]</p>	<ul style="list-style-type: none"> <li>- Hadoop-MapReduce &amp; NOSQL (Hive-Mongo database), Mahout-R language</li> <li>- collecting, saving, processing, and analyzing techniques based on intelligent information analysis platform for system construction of security log analysis using Big data</li> </ul>	<p>SIEM (Security Information &amp; Event management)" intelligent security by integration with Big data analysis technique, such as the new generation security information analysis technique</p>	<p>Supporting fast searching and showing visualized method to analyze the logs. enhance customer service satisfaction by inflow harmful code in-house, and real-time monitoring becomes doable</p>	<ul style="list-style-type: none"> <li>- Limitation of selecting a part of the security field among diverse big data methodology.</li> <li>- The extra study of analysis techniques of big data analysis area can now be applied to diverse fields.</li> </ul>
<p>Sandesh et Abhishek "Survey on Log Analysis and Management" [2015]</p>	<ul style="list-style-type: none"> <li>- Hadoop-MapReduce &amp; Splunk – Hive - Chukwa</li> <li>- discusses the vulnerability assessment utilizing some free and easy to use tools for smaller modules.</li> </ul>	<p>described different techniques to collect and analyze logs, which are a very important source of information and could be used for various security purposes such as NetFlow, IDS detection, firewall log activity, file access, system error, or login failure</p>	<p>Discusses the vulnerability assessment utilizing some free and simple tools for smaller modules so that even a small enterprise can utilize them to generate results and take appropriate actions.</p>	<p>They haven't implemented their solution but discussed the tools only, there is no future architecture or design</p>
<p>Sharma et Jhapate "Analytics of Web Log File through Map Reduce and Hadoop" [2016]</p>	<ul style="list-style-type: none"> <li>- Hadoop-MapReduce&amp; Data Mining, Cloud</li> <li>-Use Web content mining, Web structure mining, web usage mining and, log analysis. analyzed data generation from different sites in massive amount is progressive and distributed over the cloud</li> </ul>	<p>Review the basics of log analysis as big data in the web environment. collection of web access information for web pages</p>	<p>elaborates the concept of Hadoop a Big data tool which is made for Handling big data analytics and meets with the expectation of ever growing demands of the data.</p>	<ul style="list-style-type: none"> <li>- There is no integration on this work has been known</li> </ul>
<p>Girija et Sanket "Big Data Analysis of Fetched Logs using Hadoop Framework" [2017]</p>	<ul style="list-style-type: none"> <li>-Hadoop-MapReduce&amp; Data mining (Centroid-based clustering (K-Means Clustering), Density-based clustering)</li> <li>-Hadoop framework to process huge amount of data extract useful information data mining technique is used. Various types of clustering like Connectivity-based clustering.</li> </ul>	<p>scanning logs in real time as well as in off time which can give the best result of analysis of logs that can be investigated by the IT admins to take preventive actions for the attacks</p>	<p>The collective result of the real-time and manually performed analysis will be then displayed on the web page for effective and easy interaction</p>	<ul style="list-style-type: none"> <li>- There is no implemented framework, only scanning of some related work</li> </ul>



<p>Souza et Girish "Error Log Analytics using Big Data and MapReduce" [2015]</p>	<ul style="list-style-type: none"> <li>- Hadoop-MapReduce&amp; Pig, Flume, Hive, Oozie, R, Business Intelligence (BI), Data Mining, Machine Learning, Predictive Analytics Business Intelligence (BI) Predictive Analytics is used to predict and forecast the future status of the application based on the current scenario</li> </ul>	<p>Categories of severity are Error, Warning, And Info. -Abbreviations and Acronyms -Data Preprocessing Cleansing and Integration Pattern -Discovery and Analysis-Predictive Analytics - Proactive measures to ensure efficient maintainability of the applications and the device</p>	<p>predict the future severity value Whenever message IDs are the values displayed in the left branch of the tree, then there is a high chance of the system going to an Error</p>	<ul style="list-style-type: none"> <li>- Doesn't perform direct analysis on the targeted logs</li> <li>- Create Error Log file during data processing on logs</li> <li>- No integration between targeted logs, each log is a separate processed</li> </ul>
<p>Bhawna et Kiran "Big Data Analytics with Hadoop to analyze Targeted Attacks on Enterprise Data" [2015]</p>	<p>Hadoop-MapReduce &amp; HBase, FIG, HIVE, SQOOP and ZOOKEEPER. Review paper discuss how Big Data is analyzed by using the technique of Hadoop</p>	<p>design to meet the production needs of enterprises for developing a high-quality product by applying security measures with the help of Big Data Analytics with Hadoop</p>	<p>Plan to take up data collection, pretreatment, integration, map reduce and prediction using machine learning techniques.</p>	<ul style="list-style-type: none"> <li>- they have a challenge to implement and call for strong support in terms of security policy formulation and mechanisms</li> </ul>
<p>Jangla et Deepa "Development of an intrusion detection system based on Big data for detecting unknown attacks" [2016]</p>	<ul style="list-style-type: none"> <li>- Hadoop&amp; combiner, Partitioned HBase, FIG, HIVE, SQOOP</li> <li>- Propose a new model based on big data analysis technology to prevent and detect previously unknown attacks</li> </ul>	<p>explain attack predictions earned from applying big data technologies such as classification, text mining, clustering, and association rules</p>	<p>system model that uses big data analysis technology for extracting data from various sources to react to previously unknown attacks</p>	<ul style="list-style-type: none"> <li>- Need to improve the ability of software to find advanced threats, react accordingly and develop preventive measures for the future.</li> </ul>

**Table-1 summarized the related research efforts of log analysis using Hadoop Framework. The proposed frameworks involving various types of log analysis proved that Hadoop can handle a variety of data. Mainly, the research work applied MapReduce as the main component of Hadoop for analyzing the log files and HDFS as the data storage. To fulfill the analysis purposes, the researchers have also used other tools and algorithms together with the Hadoop Framework. The results of these researches show that implementing Hadoop framework enables us to successfully minimize the analysis process response time as well as analyzing larger datasets.**

#### IV. DISCUSSION (GOALS OF RESEARCH ON BIG DATA LOGS ANALYTICS)

The analysis of log files in real time and off time plays an important role in any IT infrastructure. One of the benefits is to make them robust enough to withstand against the hackers' attacks, it collects and analyzes logs, could also be used for various security purposes such as NetFlow, IDS, firewall log activity, file access, system error, or login failure and detecting security threats in network and alerting the administrators with the intrusion detection, SQL-Injection, system error, etc. One additional benefit is to enhance the intelligence of security by analyzing the relationship between security and data events created from network, system, application service of main IT infrastructure.

The analysis reports help the developers fix the bugs in the website and handle exceptions and errors and capture the number of times they have occurred. Furthermore, it gives recommendations which will help the administrators to improve the environment in various locations on various factors.

The information about site visitors and their user activities, accessed files, paths in the sites, user navigation between pages, various browser statistics, and operating system statistics will help in studying user behaviors when using highly interactive systems and improve the business strategies as well as to generate statistical reports.

The visualization tool provides graphical reports showing hits for web pages, user's activity, traffic sources, in which part of the website users are interested, the potential customers and from which geographical region website is getting maximum hits, etc., which will help in designing future marketing plans. The previous table shows that most of the related researches uses Hadoop and MapReduce techniques. They focus mainly on logs analysis and extract useful information.

The output of the review is in the form of the graphical reports and charts to allow the administrator to adjust their websites.

**Research techniques have strength in:**

- Fetching the statistics of web sites
- Help administrators for the improvement of websites and handling the exceptions
- Provide graphical reports showing hits for web pages, user's activity
- But still lack important features such as:
- Real-time analytics can provide a better result and corrective actions.
- Scanning different logs at the same time which is very important, this type of integration may lead to more evidence.
- Analyzing log network traffic to identify anomalies and suspicious activities, and to correlate multiple sources of information into a coherent view
- Prediction model should be built, not just classify the events.

We recommend the research to target deficiencies such as handling the large volumes of data and different types of logs generated frequently, selecting a part of security field and provide the corrective actions immediately by using Hadoop, real-time analytics techniques and build a predictive model.

## V. CONCLUSION

This paper presents a comparative study about the log data analytics from the traditional data analysis to the recent big data analysis. The main objective was to show how to get the maximum benefits of big data in logs analytics with different methods, tools, and techniques. It was found that there is a limitation of selecting a part of the security field among diverse big data methodology, there is no integration between different security devices and a shortage in alerting methods. Future work is to design a framework that includes ingesting and Processing Net Flow Data, alerting, flagging, transforms, filtering, Real-Time Processing - aggregations, windowed computations, and machine learning to be presented in different ways not only for detection of the failure and anomalous attack on time, but also to predict it. We plan to implement such a predictive model which is capable of distinguishing between bad connections (called intrusions or attacks) and good (normal) ones, apply the module on different devices logs and integrate it to achieve maximum security, availability, performance and customer satisfaction possible.

## REFERENCE

- [1] Latib, M. A., Ismail, S. A., Sarkan, H. M., & Mohd Yusoff, R. C. (2015). *Analyzing log in big data environment: A review*. *ARNP Journal of Engineering and Applied Sciences*, 10(23), 17777–17784.
- [2] Grable, J., & Lyons, A. (2018). *An Introduction to Big Data*. *Journal of Financial Service Professionals*, 72 No. 5(September), 17–20. <https://doi.org/10.1016/B978-0-12-801967-2.00001-X>
- [3] S. Harish and G. Kavitha, "Statistical Analysis of Web Server Logs Using Apache Hive in Hadoop Framework," pp. 4510–4515, 2015.
- [4] K. Jeon, S. Park, S. Chun, and J. Kim, "A Study on the Big Data Log Analysis for Security," vol. 10, no. 1, pp. 13–20, 2016
- [5] Miranskyy, A., Hamou-Lhadj, A., Cialini, E., & Larsson, A. (2016). *Operational-log analysis for big data systems: Challenges and solutions*. *IEEE Software*, 33(2), 52–59. <https://doi.org/10.1109/MS.2016.33>
- [6] Narkhede, S., & Baraskar, T. (2013). *HMR Log Analyzer: Analyze Web Application Logs Over Hadoop MapReduce*. *International Journal of UbiComp*, 4(3), 41–51. <https://doi.org/10.5121/iju.2013.4304>
- [7] S. Adhikari, D. Saraf, M. Revanwar, and N. Ankam, "Analysis of Log Data and Statistics Report Generation Using Hadoop," pp. 4054–4058, 2014.
- [8] S. Harish and G. Kavitha, "Statistical Analysis of Web Server Logs Using Apache Hive in Hadoop Framework," pp. 4510–4515, 2015.
- [9] I. Technology and R. M. D. Sinhgad, "Real-Time Web Log Analysis and Hadoop for Data Analytics on Large Web Logs," pp. 1827–1832, 2016.
- [10] S. Narkhede and T. Baraskar, "HMR LOG ANALYZER: ANALYZE WEB," vol. 4, no. 3, pp. 41–51, 2013.
- [11] K. Jeon, S. Park, S. Chun, and J. Kim, "A Study on the Big Data Log Analysis for Security," vol. 10, no. 1, pp. 13–20, 2016.
- [12] S. Thosar, A. Mane, S. Raykar, P. Khude, and S. Guru, "Survey on Log Analysis and Management," vol. 3, no. 6, pp. 9–14, 2015.
- [13] C. Sharma and A. Jhapate, "A Survey: Analytics of Web Log File through Map Reduce and Hadoop," vol. 2, no. 5, pp. 133–138, 2016.
- [14] P. G. Chiddarwar, S. Chhajed, S. Deshmukh, P. Dongre, and B. Nile, "Big Data Analysis of Fetched Logs using Hadoop Framework," vol. 6, no. 3, pp. 706–707, 2017.
- [15] L. T. Souza and U. R. Girish, "Error Log Analytics using Big Data and MapReduce," vol. 6, no. 3, pp. 2364–2367, 2015.
- [16] B. Gupta, "Big Data Analytics with Hadoop to analyze Targeted Attacks on Enterprise Data," vol. 5, no. 3, pp. 3867–3870, 2014.
- [17] G. K. Jangla and D. A. Amne, "Development of an Intrusion Detection System based on Big Data for Detecting Unknown Attacks," vol. 4, no. 12, pp. 229–232, 2015.





# A Real Time Monitoring Dashboard for A Smart and Sustainable Parking Area

<sup>1</sup> Samiha Najah, <sup>2</sup> Siham Al Hinai, <sup>3</sup> Mitha Al Ashkhari, <sup>4</sup> Adhiyaba Al Maqbali

<sup>1,2,3,4</sup> Computing Department, Middle East College, Oman

E-mail: <sup>1</sup>samiha@mec.edu.om, <sup>2</sup>salhinai@mec.edu.om, <sup>3</sup>15F14238@mec.edu.om,  
<sup>4</sup>415F14217@mec.edu.om

## ABSTRACT

*The country of Oman is according a high importance to the environment protection. In the last decades, the urbanization rate has seen a great growth which increased the mobility of citizens. The number of cars has increased with a high rate as a result. The pollutant elements emitted by the cars can threaten the air quality in the country especially in the city of Muscat. The project aims to develop a smart solution that supports the realization of a sustainable parking area. The solution can be implemented for any parking and can have a positive environmental impact in the country of Oman. By monitoring the environmental factors, the awareness about the pollution issues can be increased among car owners and appropriate decisions can be taken by senior management. The monitoring tool is a user friendly web application. Different types of sensors are collecting the environmental measurement data and sending it to the cloud. The result will be provided through a clear dashboard. The system will also provide access to the historical data collected with some analysis which can help to understand the progression of the air quality and the trend for the next years. The system being analyzed will have a positive impact for the country of Oman by supporting the Environment researches being conducted and providing them accurate data on the air quality in different cities.*

**Keywords - Smart Environment, Air Pollution, Parking, Iot, Dashboard, Cloud Computing, ISO14001, Sensors.**

## I. INTRODUCTION

Air pollution has been the object of many researches aiming to highlight the most important factors that are threatening the environment safety. This issue is now having more and more importance due to the expansion of the urbanization and industrialization [2]. During the past 4 decades, Oman has knew a very fast urbanization rate growth which reached in 2009, 84% [2]. Dispersed settlement in Muscat and the transition from rural to urban for some other areas are not the only examples of the great urbanization development in Oman. The population also has increased. According to [3], the city of Muscat has around 800,000 citizens in 2018 followed by Seeb with approximatively 240,000 citizens and Salalah with approximatively 165,000 citizens.

The business functions has been distributed between many cities. Add to this fact, the increasing number of citizens contributed to the high mobility that the country is facing the last few years. The citizens are relying on the car as a primary transportation mode if not the exclusive one.

According to [9], the total number of vehicles that registered during the past five years is more than 635,000 as shown by the statistics prepared by Oman News Agency (ONA). Most of these vehicles are private cars (74.1%).

Air pollution can be known as the release of toxic substances to the atmosphere. [11]. Many elements can be considered as pollutantssuch as :

Volatile organic compounds (VOCs), carbon monoxide (CO), ozone (O3), nitrogen oxides (NOx), and Sulphur dioxide (SO2),...Oman is one of the countries that give the environment conservation big importance. Keeping the air quality clean and protect it from various sources of pollution represents a major global concern. The United Nations (UN) Environment Program has considered Oman to be one of the best countries in environmental care based on some insured records.[3] This project is willing to support the Environment protection efforts in the country of Oman by increasing the awareness about the air pollution degree. It will also help the senior management for any company to take decisions regarding in order to protect its surrounding. In this project proposal, we are going to describe a smart solution that aims to display in real time the monitoring data for the most important pollutant elements in a parking using a user friendly dashboard. In a parking, the air pollution is very high especially during pick hours. The system will be installed in the parking area of one organization in order to draw the attention of the senior management and the citizens to importance of the issue.

## II. SYSTEM GOALS

The proposed project will have important positive impact for the country of Oman.

The user friendly dashboard showing the environmental pollutant element details will help the senior management in one organization to control its surrounding quickly. More details about the collected data from the saved history along with analysis can also help them to understand the trend of the issue and how serious the issue will be in the future. The senior management can then take their decisions easily, especially when deciding to expand, regarding planting trees or any other solution that can limit the pollution effect.

The system will also have a positive impact on increasing the citizen's awareness about the environment status. The dashboard will be designed to be used for any user. It will show summarized, clear and understandable information about the air pollutants without need for being a chemist expert. The car owner can then give more importance to this subject by checking whether his car is emitting any kind of pollutants since this will affect his health.

The proposed work will also support the sustainability research in smart cities. It can help the country to have an inventory of environment conditions in different cities in Oman after coordinating with different organizations having installed the system. Many other researches have already been conducted and continuing in the country in order to preserve the good air conditions. This tool can contribute to the success of these researches providing them with accurate data about the current pollution status in different places. The historical information saved can also give an important idea about the trend of the issue.

### **Study of ISO standard regarding environment monitoring systems**

The proposed system can be classified as part of an environment monitoring system (EMS). In order to design such system, it is important to know the developed standards for the environment monitoring systems in general. The proposed system will support the EMS by providing smart features implemented using IoT blocks.

ISO has defined the ISO 14000 family of standards concerning the environmental management. ISO 14001 defines the criteria for an EMS. It illustrates the framework that can be followed by any organization independently of its size and type. [15]

The clause 9 of the standard is "Performance evaluation". It includes three parts: "Monitoring, measurement, analysis and evaluation", "Internal audit" and "Management review".

Monitoring, measurement, analysis and evaluation of the environmental performance must first determines the factors that need to be monitored and the indicators that will be evaluated. It must also indicate when an analysis of the measurements will be done. The organization must have adequate monitoring equipment for its studied environmental factors. The relevant environmental performance information must be communicated internally and externally. The documentation of the monitored data and its analysis is necessary in this sub-clause. [6]

The results of the environmental performance analysis should help the senior management and the decision makers in the organization to take appropriate actions. [6] The proposed system is designed in order to support an EMS to monitor, measure and analyze the environment quality of the organization's surroundings as specified by ISO140001 standard and more precisely the first part of clause 9.

## Design of the proposed system

The system will be developed based on Internet of Things (IoT). It is an advanced technology that is having a considerable success world widely in the last years. [12]

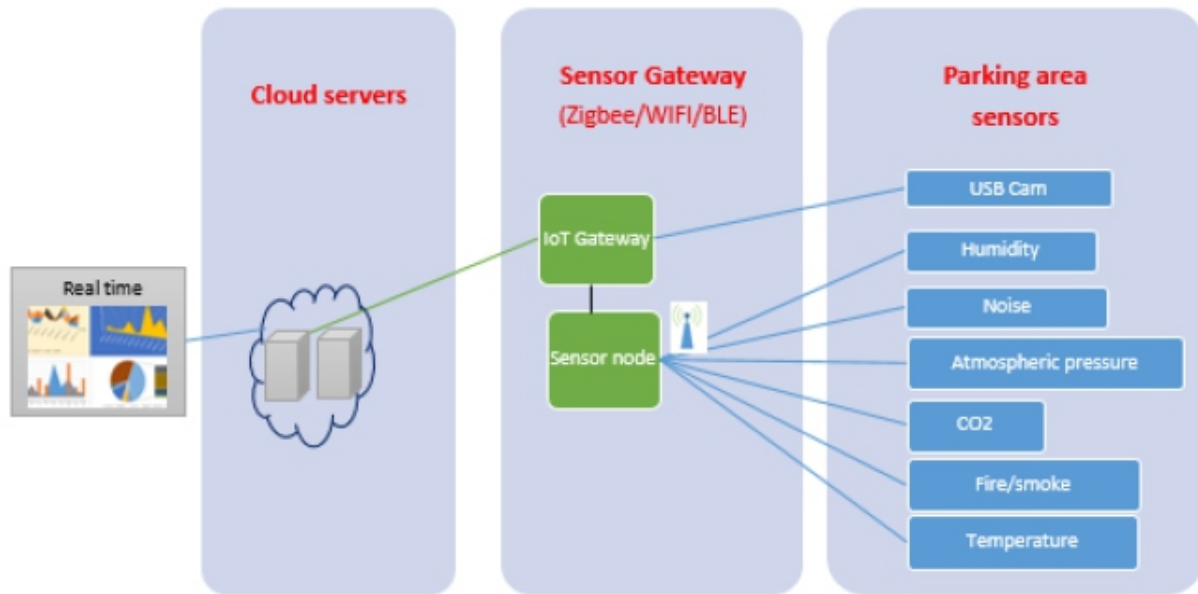
According to ITU-T Y.2060 recommendation (06/2012), IoT is considered to be as a global infrastructure for information community enabling for multi services by the interconnection of virtual and physical things based on evolving communication and information technologies.[7]

The proposed system is formed by four main components:

- The environmental sensors and IoT cameras: The IoT device consists of actuators and sensors, Raspberry Pi and open hardware Arduino.
- Temperature sensor: measures the heat amount by detecting any change and sends the digital information to another device for processing.
- Gas sensors are deployed to monitor the air quality by detecting the changing amount of various types of gases. [13]
- Humidity sensor: it measures the amount of Humidity in the air. When any change in the humidity amount is detected, a digital information will be sent to the processing device. [13]
- USB Cameras can capture the image and streams it in real time to be processed by a computer. The USB technology is used in the data transferring. In our example, it will be connected to raspberry pi before being sent on the network (Sensor gateway 1). [1]

The Sensor Node and the sensor Gateway : The IoT gateway transmits the sensing data to the cloud via the internet. The sensor node is the microcontroller that will aggregate the data collected by the different sensors. Arduino mega 2560 can be used for the sensor node. Raspberry pi will be used for the sensor gateway to forward the gathered data from the sensor node to the network and also to connect to the USB cam.

- The cloud server: The data captured by the sensor is processed in the cloud , resulting the IoT device actuators to actuate through the IoT gateway.
- The web application: the real time dashboard. It is a web application that will get the real time information from the cloud. The data will be stored in a database for later analytics.



**Fig.1. Architecture of the Real Time dashboard based on IoT technology.**

### III. CONCLUSION

The environment monitoring especially in the parking area of any organization is very important to preserve the health and wellness of its employees and customers. ISO has defined standards regarding the implementation of an EMS. The current paper studies the aims and design of an environment monitoring tool which will support the EMS implementation according to ISO 14001 standard (clause 9). The project is based on IoT technology. It will provide smart features and help the decision makers in any organization to take the appropriate actions to save the environment. It will also help the employees and customers of the organization to have valuable information regarding the environment in which they are living.

### REFERENCE

- [1] Abduelhadi1, A., 2017. *Smart Motion Detection*. *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)*, 12(3), pp. 53-58.
- [2] Anon., 2011. *Urbanisation*. [Online] Available at: <http://urbanoman.org/urbanization.htm>
- [3] CURTIS, 2017. *Environmental Law*. [Online] Available at: <https://omanlawblog.curtis.com/2009/03/united-nations-un-environment-programme.html>
- [4] Curtis, M.-P. C. & M. L., 2009. *Oman Law Blog*. [Online] Available at: <https://omanlawblog.curtis.com/2009/03/united-nations-un-environment-programme.html> [Accessed 15 November 2018].
- [5] Dr. V. Bhuvaneswari, D. R. P., 2014. *The Internet of Things (IoT) Applications and Communication Enabling Technology Standards: An Overview*. India, s.n.
- [6] ISO, 2015. *ISO 14001:2015 Clause 9 Performance evaluation*. [Online] Available at: <http://isoconsultantpune.com/iso-140012015-clause-9-performance-evaluation/> [Accessed 20 11 2018].
- [7] ITU-T, 2015. <https://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx>. [Online] Available at: <https://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx> [Accessed 25 11 2018].
- [8] Mahmoud Elkhodr, S. S. a. H. C., 2016. *EMERGING WIRELESS TECHNOLOGIES IN THE*. *International Journal of Wireless & Mobile Networks*, 8(5), pp. 67-82.
- [9] ONA, 2016. *article/79270*. [Online] Available at: <https://timesofoman.com/article/79270>

- [10] Praveen Vijaia, B. S. P., 2016. *Design of IoT Systems and Analytics in the context of Smart City*. Bhubaneswar, Odisha, India, ELSEVIER.
- [11] Roser, H. R. a. M., 2017. *air-pollution*. [Online] Available at: <https://ourworldindata.org/air-pollution>
- [12] Rouse, M., 2016. *internet of things (IoT)*. [Online] Available at: <https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT> [Accessed 25 11 2018].
- [13] Sharma, R., 2018. *Top 15 Sensor Types Being Used in IoT*. [Online] Available at: <https://www.finoit.com/blog/top-15-sensor-types-used-iot/> [Accessed 25 11 2018].
- [14] Vimal Jerald A., A. R. S. D. P. B. T., 2015. *Internet of Things (IoT) based Smart Environment*. *International Journal of Computer Applications*, 128(8), p. 0975–8887.
- [15] Wikipedia, 2018. *ISO 14000*. [Online] Available at: [https://en.wikipedia.org/wiki/ISO\\_14000#ISO\\_14001\\_standard](https://en.wikipedia.org/wiki/ISO_14000#ISO_14001_standard) [Accessed 20 11 2018].

# Finding an Optimal Free-Obstacle Path for A Mobile Robot using A\* Algorithm

<sup>1</sup> Amer S. Alharthi, <sup>2</sup> Saad B. Alotaibi

King Abdulaziz for Science and Technology (KACST), National Center for Robotics and Automation System E-mail: 1aharhi@kacst.edu.sa

## ABSTRACT

*The path planning is a fundamental problem in robotics, it consists to find an optimal free-obstacle path for a robot from a starting position to a given destination (a goal). There are many approaches to solve this problem, one of the most common and widely known is the A Star search. In this paper we will implement the A\* searching technique to solve the path planning problem. We assume the environment of the robot is represented by a grid map in which the initial and goal states (current location and the destination) are clearly localized on the map. Some cells of the map contains obstacles the robot cannot traverse them to reach the goal. The robot can move in different directions (horizontally or vertically) from a cell to another neighbor free cell. In the following sections we will describe how to implement A\* algorithm on static environment with randomly obstacles.*

**Keywords - Path Planning, A\* Algorithm, Free-Obstacle**

## I. INTRODUCTION

Path planning is an important and essential process in the field of robotics. It is the planning for sequence of steps that needs to be followed by the moving robot to reach its final destination. A moving robot can reach its final destination in many different paths. The planning must be carried out to select the path that is shortest, efficient and optimal. Various different algorithms are available to perform the path planning process. Some of the popular algorithms are Breath-first, Dijkstra, A\*, D\*, ARA\*, AD\* and many others. All these algorithms works on graph mapping approach and finds the shortest route by selecting the nearest neighbor through a series of iteration. The iteration methods are different for different cases and are suitable for a particular situation. Each of the algorithms has its own merits and demerits. Some algorithm works well for static obstacles whereas others are good for dynamic obstacles or both. Some algorithm follows global path planning whereas the others follow local path planning. The global path planning requires the entire information of the planning plane and performs the planning with respect to the information. The local path planning does not have any information of planning plane and the path planning is carried out online while the robot is moving.

The A\* algorithm also known as the static algorithm is perfectly suitable for the robot movement in plane having static obstacles. It always aims for the shortest path and takes the weightage of all the nearby cells before following the path[1]. The idea of the paper is to use the A\* algorithm to find the optimal path between the current and goal state for a mobile platform. In addition, it should find it by decreasing the



number of steps between current and goal states. Path planning problem has many methods to find the solution from 200 years, but those methods are not strong enough also, they are not easy to adapt with any dynamic change in the problem formulation as most of them based on the concept of space configuration. The biggest problem of the methods are the need for extra resources (size, time and memory) to get an optimal solution. A\* has been recognized as one of the most strong searching algorithm for complex and misbehaved optimization problems and can work with static and dynamic environment, furthermore it inheres parallel search techniques. the problem will be formulated as a robot moving in environment and want to find a path along with static obstacles from a starting point to ending point using optimization criteria which is Shortest Path, moreover it allows four-neighbor movements with low complexities [1]. In the following section, we will discuss the commonly path planning algorithms.

## II. RELATED WORK

Based on the path planning algorithms that used in real environment (Static or Dynamic), we discuss some of methods in path planning of mobile robots. The path planning is the set of principles as followed by mobile robots to ensure collision free and shortest path movement for both dynamic as well as the static obstacles. Many different algorithms are available to ensure the path planning. One of the proposed method, the authors have taken the commonly available honeybee movement as an analogy and proposed a path planning algorithm. The proposed algorithm (Artificial Bee Colony algorithm) is experimented for both static as well as dynamic obstacles under certain conditions. The authors proposed solution in two cases namely, path planning in static and dynamic environments. The algorithm is proposed for a 2D environment with grids. Each of the cell joining four grid points is called node. In the first case (static environment), A 2D environment with multiple nodes are considered and a population size is initialized with equal number of employed and onlooker bee. The number of paths (Pt) is assigned as half of population size. At start of the algorithm, Pt is determined randomly. An objective function is determined for each Pt. A fitness function is estimated for each Pt by taking the inverse of the objective function. The fitness value is criteria for the obstacle free shortest patch. At every instant 8-neighbour rule is used check the next possible obstacle free paths and objective function is again determined. This process is followed until the required destination is met with shortest time frame. In second case (dynamic environment), the number of paths as determined for the first case are arranged in decreasing order. The sorting is carried out depending on the fitness value. The best path is determined by considering the dynamic obstacles. If the fitness value is increased beyond a specific value then that particular path is discarded and the remaining ones are considered[1].



Other researchers discuss three of the popularly available graph search algorithms for path planning and presents the results. The algorithms are namely Breadth-first, Dijkstra and A\*. Experiments are carried out to determine the shortest path and the situation of multiple goals for a single starting point. Mobile robots generally follow grid based approach to detect the shortest and economic route. The experiments are carried out by assuming certain set of situations. The three graph search algorithms have their own merits and demerits and must be employed depending upon the situations and requirements[2].

The Breadth-first algorithm works on branching method. It starts with the starting cell of the map matrix and locates the nearby cell that is obstacle free and close to the goal. The identification number of the next nearby cell is updated in the OPEN LIST until the goal is reached. The OPEN LIST provides the optimum path upon reaching the goal cell. The total cost is later estimated by multiplying the OPEN LIST path and the cost of the corresponding cells[2].

The Dijkstra algorithm is an upgraded version of the Breadth-first algorithm. It locates the nearby obstacle free cell based on the cell cost. It chooses the least cost nearby cell for movements and updates the OPEN LIST accordingly[2].

A\* algorithm is the most efficient and requires a bit more computation for locating the optimum path. For each of the movement it selects the next new neighbor by considering all the previous located nearby neighbors. At any point it considers the just previous cell for selecting the optimum low cost path[2].

### III. EXPERIMENTS AND RESULTS

In our work we assume the environment of the robot is represented by a grid map in which the initial and goal states are clearly localized on the map. Some the cells in a map contains obstacles and the robot cannot traverse them to reach the goal. The following figure illustrates the environment (AreaMap) of the robot.

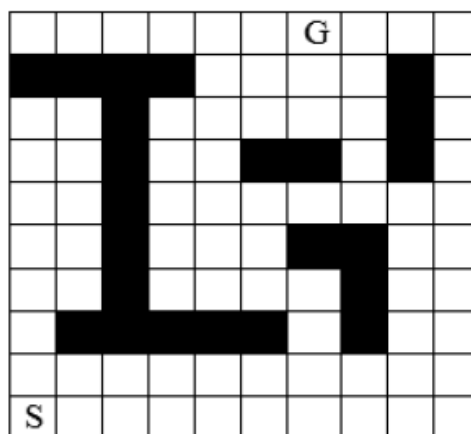


Figure 1

The robot can move in different directions (horizontally or vertically) from a cell to another neighbor free cell. In the following sections we will describe the requirements of A\* algorithm that we will implement on our environment.

### 1- Area Map

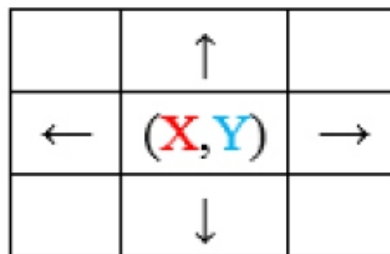
The whole map will be represented by Map width, Map height, All map cells (nodes) Start location in map (X, Y) and Goal location in map (X, Y). To start using the map, it must be created first and be will prepared as the following steps:

1. Fill the array representing the map with nodes and put the X and Y for each node.
2. Register each node neighbors (other nodes) from (East, West, North and South).
3. Set start location and remove any old start location and mark the start location as start.
4. Set goal location and remove any old goal location and mark the goal location as goal.

To get distance between any two nodes as it will be required through the way to the goal, and if the two nodes on the same column (X) or the same row (Y), then the distance will be 1.

### 2- Node

Each node which is a cell in the map has a location determined by column (X) and row (Y) and also has neighbors (other nodes) from East, West, North and South as shown below:



Each node will have its related info and also some information about its neighbors:

1. Its location on X axis (row).
2. Its location on Y axis (column).
3. Is this node the start node or not.
4. Is this node the goal node or not to stop search.
5. Is this node an obstacle (block) node
6. Is this node visited before or not to not visit it again (Boolean visited).
7. Its distance from start node.
8. Its distance based on heuristic from goal.
9. The previous step (node) before visiting this node.
10. The node on the west side of this node (west).

11. The node on the east side of this node (east).
12. The node on the north side of this node (north).
13. The node on the south side of this node (south).
14. List of all neighbors from all sides.

Almost of functions in Node class is a setters and getters functions. The different setters and getters from normal is the ones for west, east, north and south because each setter insert into the neighbors list. Node class implement comparable as it will be so often to compare two nodes with each other, its equal when Xs of nodes equals and Ys equals to each too or not, and also implement equals function to have compareTo function from comparable. To compare must be check if the total distance from current node to goal node less than or greater than the same distance for other node and will return (1, -1 or 0)

1 -> current node distance is greater than the other node.

-1 -> current node distance is less than the other node.

0 -> current node distance equal to the other node.

### 3- Path

Path is simply the steps to go from start node to goal node in terms of Nodes list (ArrayList<Node> steps). And the most important functions are the ones add to the list and check if a node already in the list or not

- getLength – return the length of path which is the list size.
- getStep – get the node object in a specific index.
- getX – get the X axis value for a node in a specific index.
- getY – get the Y axis value for a node in a specific index.
- appendStep – add step to the end of list.
- prependStep – add step to the start of list and will be used with the previous information from Node.

### 4- Heuristic

It is the way of trying to find a solution for any problem by making some compromising (speed, accuracy, optimality or completeness) to find approximate solution for problems don't have exact solution. In other words, heuristic is the way to trying to guess the solution in the way we think after getting a problem or puzzle. Heuristic trying to find the closest node to the each time by applying the following equation:

$$\text{Distance to goal} = (dx)^2 + (dy)^2$$

Where d is delta (dx = goalX – currentX) and same for dy

And after getting the distance for each node visited during process, it will be add to a list to pick the next step from, given that Node implements comparable based on the value of heuristic then the list will be sorted after adding new node to find the closest node on the top of list and finally pick the top of the list to work with. To implement A\* algorithm we need to encapsulates all the previous requirements in one unit and will be used directly from the Test class file and also has the full business of algorithm and how it works. A star It is the class that encapsulates all the previous classes in one unit and will be used directly from the Test class file and also has the full business of algorithm and how it works. AStar has inner class called SortedList based on Array List to add the following functions which will be useful in Astar:

- getFirst -> get the first entry in the list.
- Add -> add to list and sort the list after addition.

AStart has all the required data to be ready for work based on the previous classes as following:

- AreaMap object -> the map will work on.
- ClosestHeuristic object -> will be used to guess the next step.
- closedList -> list of nodes visited to not visit it again.
- SortedList object -> for open nodes (ready to be visited) sorted using heuristic.
- Path object -> for shortest path formation during the process.

AStar has three main functions to make the algorithms

- reconstructPath -> to form the path after reaching the goal using the previous value in each node, and will go from goal previous node to its previous and so on till reaching the start node which has no previous and during each step will add to the path.

Goal -> previous -> previous -> ..... -> start

- printPath -> will go through the whole map (nested loop) to print it with its obstacles, start, goal and path drawing as following:

- if the visited is obstacle, then print „X“
- else if it is the start node, then print „S“
- else if it is the goal node, then print „G“
- else if it is in the detected shortest path, then print „,“,
- else print empty space „,“,
- calcShortestPath -> important function which start with the start node and get its neighbor to check which way will take to the goal based on heuristic and doing the same with the selected neighbor till reaching the goal. The chart below identifies the steps to move from node to node and construct the path and finally print the path.

We have developed A\* algorithm in Java and implemented it on grid map using NetBeans IDE 8.1. Testing of the algorithm was carried out for three simple situations. The results of the experiment show an excellent performance for getting a shortest path. The following figures (cf. figure 3-5 ) show the results in three simple situations.

#### **IV. CONCLUSION**

In this paper, we discussed the path planning of mobile robots, proposes a method and experiments. We done the proposed method in three different randomly grid map. The initial and goal states are clearly localized on the grid map. Some the cells in a grid map contains obstacles and the robot cannot traverse them to reach the goal. In the end, we have an excellent performance for getting a shortest path using A\* algorithm.

#### **ACKNOWLEDGMENT**

This research was supported by King Abdulaziz for Science and Technology (KACST), National Center for Robotics and Automation System.

#### **REFERENCE**

- [1] I. Al-Taharwa, A. Sheta, and M. Al-Weshah, "A mobile robot path planning using genetic algorithm in static environment," *Journal of Computer Science*, vol. 4, no. 4, pp. 341–344, 2008.
- [2] Goel, Priyanka; Singh, Devendra, "Efficient ABC Algorithm for Dynamic Path Planning", *International Journal of Computer Applications*, February 2014, pp. 15.
- [3] T. Y KSEL, A. SEZGIN, "An Implementation of Path Planning Algorithms for Mobile Robots on a Grid Based Map", *Proc of the 4th International Conference on Electrical and Electronics Engineering (ELECO 2005)*, Bursa, Turkey, 2005, pp. 382-387.



# Use of Fuzzy Set Theory in the Medical Diagnosis

<sup>1</sup> Kiran Pal, <sup>2</sup> V. Kumar, <sup>3</sup> H. D. Arora

<sup>1</sup>Research Scholar, Department of Mathematics, Amity University, Noida, India Assistant Professor, Department of Mathematics, DITE, Delhi, India

<sup>2</sup>Faculty of Engineering and Technology Manav Rachna International University, Faridabad

<sup>3</sup>Professor, Department of Mathematics, Amity University, Noida, India

E-mail: kiranpaldite@gmail.com

## ABSTRACT

*World over the problem of scarcity of qualified doctors is alarming. Uncertain nature of calamities and seasonal make it further difficult to manage this. In this work an attempt is made to highlight the use of fuzzy set theory in the field of medical diagnosis. Fuzzy set theory allows to incorporate the opinion of a number of doctors in proper diagnosis of critical diseases. It is also helpful in providing proper diagnosis at an early stage of diseases in remote areas with use of some trained technical medical experts. Therefore, it is proven as an effective tool for the medical diagnosis.*

**Keywords - Medical Diagnosis, Fuzzy Set Theory, Fuzzy Number, MADM, VBD**

## I. INTRODUCTION

Population of globe is growing exponentially and due to which many deadly diseases are spreading because of inadequate food supply and health related facilities. Further, day by day, patients are increasing, however, doctors' numbers are less to diagnose and treat them. Simultaneously, computers evolution and various other innovative technologies faced the problem of diseases diagnosis and thus, increases the doctor efficiency. Computational techniques for Medical diagnosis is an effective tool for doctors to assess the suitable treatment to patients. Among the other available techniques and generalizations, generalized fuzzy set theory is assumed as a useful tool to find out possible treatment for disease.

### Vector-Borne diseases (VBDs)

VBDs are transmitted by mosquitoes, black flies, ticks, snails, and lice, which are having a severe threat to the human being particularly in countryside locations results in affecting the development and health of vulnerable population. The difficulty in accurate and timely diagnosis of VBDs can delay the relative treatment procedure especially in remote regions where doctors as well as hospitals are in scarce number.

## II. FUZZY SET CONCEPT

Lotfi A. Zadeh (1965) introduced fuzzy sets and Zadeh proposed a paradigm shift that firstly accept in Far East and it is adopted now around the globe because of its successful application. Fuzzy sets are used in fuzzy logic and it is extension of classical set theory in which elements membership has been assessed in binary terms with respect to set according to crisp condition that means an element either belongs or does not belong to set. However, fuzzy set theory allows the gradual evaluation of elements membership with the aid of value in real unit interval  $[0,1]$ . Fuzzy sets are an extension of classical set theory since, for a certain universe, a membership function may act as an indicator function, mapping all elements to either 1 or 0, as in the classical notion. Now, fuzzy set can allow its members to have dissimilar grades of membership in the interval  $[0,1]$ . A fuzzy set on a classical set  $X$  is defined as follows:

$$\tilde{A} = \{(x, \mu_A(x)) \mid x \in X\} \quad (1)$$

The membership function  $\mu_A(x)$  accounts the grade of membership of elements  $x$  to fundamental set  $X$ . An element mapping to the value 0 implies that member is not included in given set. Value 1 indicates fully included member. Fuzzy members can be characterized strictly between 0 and 1

Membership function terminology

**Universe of Discourse:** it is defined as the range of all possible values for an input to fuzzy system 'F'.

**Support:** it is defined as the crisp set of all points in the universe of discourse  $U$ , for instance membership function of  $F$  is non-zero.

$$\text{Supp}A = \{x \mid \mu_A(x) > 0, \forall x \in X\} \quad (2)$$

**Core:** it is defined as the crisp set of all points in the universe of discourse  $U$ , for instance membership function of  $F$  is 1.

$$\text{core}A = \{x \mid \mu_A(x) = 1, \forall x \in X\} \quad (3)$$

**Boundaries:** it is defined as the crisp set of all points in the universe of discourse  $U$ , for instance membership function of  $F$  lies between 0 and 1.

$$\text{Boundaries}A = \{x \mid 0 < \mu_A(x) < 1, \forall x \in X\} \quad (4)$$

**Crossover point:** it is the element in  $U$  at which its membership function is 0.5.

**Height:** it is a biggest value of membership functions of fuzzy set.



Normalized fuzzy set: the fuzzy set of Cardinality of the set:  $\text{Height}(A) = 1$

$$\begin{aligned}
 & X : \text{finite} \\
 & |A| = \sum_{x \in X} \mu_A(x) = \sum_{x \in \text{Supp}(A)} \mu_A(x) \quad (5)
 \end{aligned}$$

Relative cardinality:

$$\|A\| = \frac{|A|}{|X|} \quad (6)$$

Convex fuzzy set:  $X \in \mathbb{R} \quad \forall \lambda \in [0, 1]$   
 Fuzzy set A is Convex, if for

$$\mu_A(\lambda x_1 + (1 - \lambda)x_2) \geq \min(\mu_A(x_1), \mu_A(x_2)) \quad (7)$$

### III. INTUITIONISTIC FUZZY SET (IFS) THEORY

The concept of IFS has been introduced by Atanassov as a fuzzy sets generalization that is capable of information capturing that further contains some degree of hesitation.

Over a finite non empty fixed set  $X$ , Atanassov's IFS is a set, i.e.  $A = \{ \langle x, \mu_{\tilde{A}}(x), \gamma_{\tilde{A}}(x) \rangle \mid x \in X \}$  that allocates to every element  $x \in X$  to the set  $\tilde{A}$  i.e. subset of  $X$  with the membership degree  $\mu_{\tilde{A}}(x): X \rightarrow [0,1]$  and non-membership degree  $\gamma_{\tilde{A}}(x): X \rightarrow [0,1]$  fulfilling  $0 \leq \mu_{\tilde{A}}(x) + \gamma_{\tilde{A}}(x) \leq 1$ , for all  $x \in X$ .

For every IFS in  $X$ , a hesitation margin  $\pi_A(x)$ , which is IF index of element  $x$  in IFS  $\tilde{A}$  expressed by  $\pi_A(x) = 1 - \mu_{\tilde{A}}(x) - \gamma_{\tilde{A}}(x)$  denotes a non-determinacy measure. Here  $F(X)$  is a set of all IFSs on  $X$ .

### IV. FUZZY NUMBERS

Wang (2008) introduced the concept of intuitionistic trapezoidal fuzzy numbers (ITFNs) and it may be noted that ITFNs as compared to trapezoidal fuzzy numbers gives more flexible and plentiful information  $\mu(x)+0.5$ .

In order to define a fuzzy interval in the real number  $\mathbb{R}$ , Fuzzy number can be expressed as fuzzy set. Since this interval has ambiguous boundary, interval is also recognized as fuzzy set. A fuzzy interval can be denoted by two end and a peak point  $a_1, a_2$ , and  $a_3$  as  $[a_1, a_2, a_3]$  (Fig 1). The fuzzy number is used for  $\alpha$ -cut operation. Let for fuzzy number  $A$ .  $\alpha$ -cut interval can be expressed as  $A_\alpha$  as discussed below:

$$A_\alpha = [a_1^{(\alpha)}, a_3^{(\alpha)}] \quad (8)$$

Fuzzy number can be normalized and convex. Therefore, normalization condition implies that value of maximum membership is 1.

$$\exists x \in \mathfrak{R}, \quad \mu_A(x) = 1$$

According to convex condition: line by  $\alpha$  cut is continuous and its interval should satisfy the following points:

$$A_\alpha = [a_1^{(\alpha)}, a_3^{(\alpha)}]$$

$$(\alpha' < \alpha) \Rightarrow (a_1^{(\alpha')} \leq a_1^{(\alpha)}, a_3^{(\alpha')} \geq a_3^{(\alpha)}) \quad (10)$$

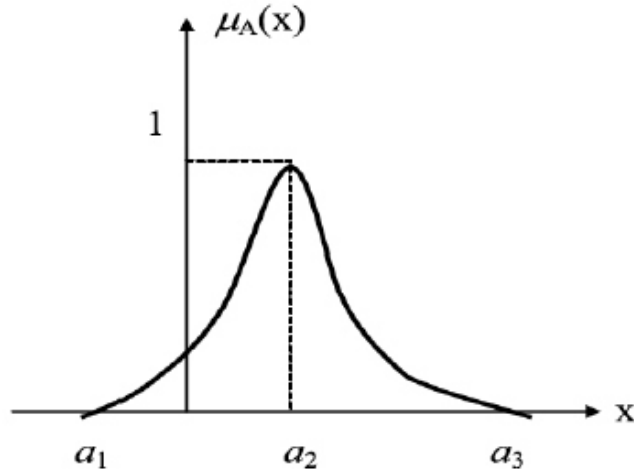


Figure1. Fuzzy Number  $A = [a_1, a_2, a_3]$

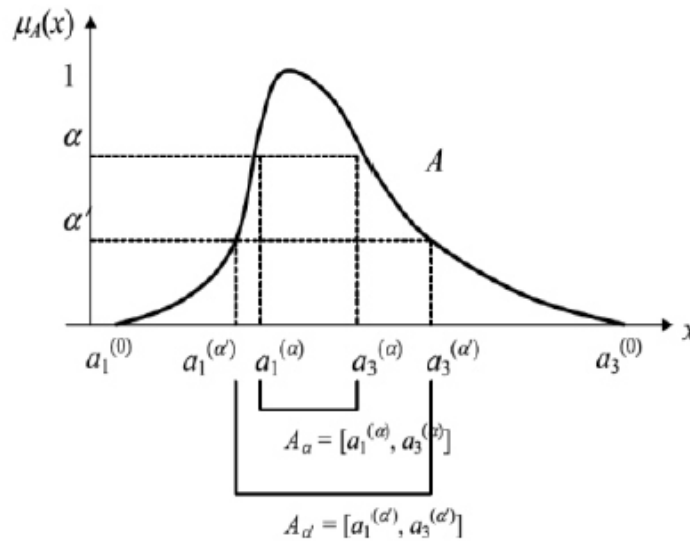


Figure2  $\alpha$ -cut of fuzzy number  $(\alpha' < \alpha) \Rightarrow A_\alpha \subset A_{\alpha'}$

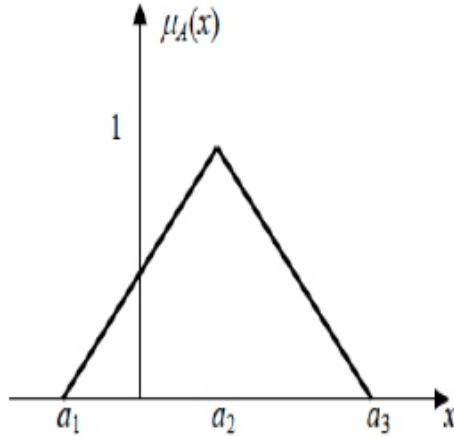
### V. DEFINITION OF TRIANGULAR FUZZY NUMBER

Triangular fuzzy number (TFN) is most popular one among the numerous shapes of fuzzy number. TFN is a fuzzy number represented with three points as follows:

$$A = (a_1, a_2, a_3) \quad (11)$$

This representation is interpreted as membership functions

$$\mu_{(A)}(x) = \begin{cases} 0, & x < a_1 \\ \frac{x - a_1}{a_2 - a_1}, & a_1 \leq x \leq a_2 \\ \frac{a_3 - x}{a_3 - a_2}, & a_2 \leq x \leq a_3 \\ 0, & x > a_3 \end{cases} \quad (12)$$



The  $\alpha$ -cut operation can be obtained by crisp interval  $A_\alpha$  which is attained as follows  $\forall \alpha \in [0, 1]$ . From

$$\frac{a_1^{(\alpha)} - a_1}{a_2 - a_1} = \alpha, \quad \frac{a_3 - a_3^{(\alpha)}}{a_3 - a_2} = \alpha \quad (13)$$

we get

$$a_1^{(\alpha)} = (a_2 - a_1)\alpha + a_1 \quad (14)$$

$$a_3^{(\alpha)} = -(a_3 - a_2)\alpha + a_3 \quad (15)$$

## VI. TRAPEZOIDAL FUZZY NUMBER (TRFN)

Another shape of fuzzy number is TRFN that is originated from the fact that membership degree should be maximum i.e.  $\alpha = 1$ .

TRFN 'A' can be defined as:

$$A = (a_1, a_2, a_3, a_4) \quad (16)$$

The membership function can be expressed as follows (Fig. ).

$$\frac{a_1^{(\alpha)} - a_1}{a_2 - a_1} = \alpha, \quad \frac{a_3 - a_3^{(\alpha)}}{a_3 - a_2} = \alpha \quad (17)$$

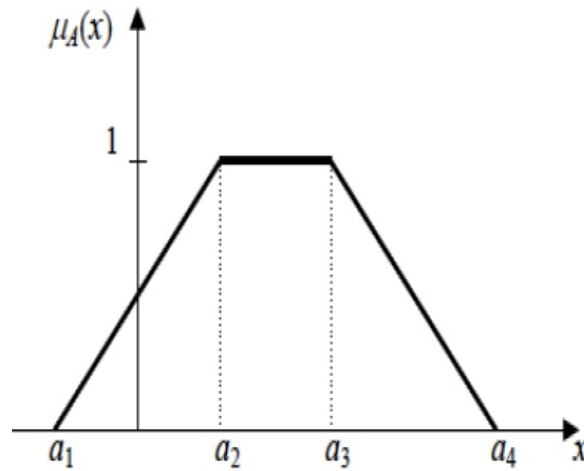


Figure.3 TRFN 'A' = (a<sub>1</sub>, a<sub>2</sub>, a<sub>3</sub>, a<sub>4</sub>)

The  $\alpha$ -cut interval for this shape is shown as:  $\forall \alpha \in [0, 1]$

$$A_\alpha = [(a_2 - a_1)\alpha + a_1, -(a_4 - a_3)\alpha + a_4] \quad (18)$$

When  $a_2 = a_3$ , the TRFN coincides with TFN one.

## VII. MULTIPLE CRITERION DECISION MAKING (MADM)

In the presence of multiple conflicting criteria, MCDM refers to making decisions. Depending on type either selection or design problem, it can be broadly classified into multiple attribute decision making (MADM) and multiple objective decision making (MODM). An MADM method defines that how attribute information could be processed so as to reach at a choice. An MADM require comparisons on the basis of inter-and intra-attribute, and includesuitable explicit trade-offs.

## VIII. LITERATURE ON MADM APPLICATIONS

Atanassov (1986) introduced the notion of an IFS to deal with imperfect and imprecise information. Wang(2008,2008a) explained that ITrFNs are the extension of ITFNs. Wang (2009) developed ITrFWA operator for the propose of better decision making. Wang (2009a) explained the expected values of ITrFN and proposed programming method of MADM. Wu (2013) explains ITFWG operator and its hybrid version. Wan (2010) discussed the expected score of ITrFNs from geometric point of view and proposed ITrFHA operators. Yager (2004-09)modelled the concept of prioritized operators for streamline decision making. Wei (2011) elaborated the generalized concept of prioritized aggregation operators as explained by Yager(2004-09) and also developed some hesitant fuzzy prioritized aggregation operators. Yu (2013) developed some prioritized IF aggregation operators. But these operators lack over certain parameters, they cannot be used when the input is in form of ITrFNs, having difficulty in dealing with problems ofMADM. To overrule this drawback, Zhang (2014) proposed some

ITrF prioritized operators. These operators are not only compatible with the input arguments of ITrFNs but also considering prioritization among the input arguments. Li(2011) and Lin (2007) explains various MADM models and methods using IFSs. Guorong(2011) explains the limitations for MADM flavored with ITrFN. Shu (2006) introduced intuitionistic TFN and their operations. Wan (2010) defined the expected score of ITrFNs across the geometric point and present operators of trapezoidal IF ordered weighted averaging operator and its hybridization.

Zhang [2014] discussed an algorithmic approach of ITFPWA operators and algorithm applicability for the selection of suitable treatment for lung cancer over certain attributes. These attributes are given in the form of ITrFNs. Once the disease has been diagnosed, it is difficult for a doctor to choose the appropriate treatment from the available set of treatments. The present work helps the doctors to select the appropriate treatment for the disease as above mentioned and rank it as per the requirement. For this purpose, we are taking hypothetical case study to explain the algorithm in research work.

Vectors has an ability to produce serious diseases in human or animal populations because these are organisms that can spread pathogens and parasites. With respect to south-east Asia region, common vector-borne diseases VBDs types are chikungunya, dengue, and malaria etc. Chikungunya resembles to dengue along with symptoms are severe joint pain also known as arthritis accompanied by high fever, rash, joint swelling, headache, muscle pain, nausea, and fatigue etc. establishes itself 9-14 days after the mosquito bite and its symptoms are high fever, headache, nausea, vomiting, and muscles pain. Symptoms of dengue: high fever, pain behind the eyes, head ache, body aches and joint pain, and skin rash. Also, symptoms of Malaria are high fever, headache, nausea, vomiting, and muscles pain. To control these VBDs, WHO recommends some instructions such as different countries of the world should be provided moral and technical support as well as new tools and innovative diagnosis should be developed to further fortify this effort. In the world, one sixth of illness and disability endured is because of VBDs, with more than half of population right now evaluated to be in danger of these infections [WHO (2004) and (2009)].

MCDM methodologies have been created and useful to engineering, management, economics, and so forth. Hwang and Yoon (1981) first developed the traditional MCDM approach like TOPSIS (technique for order performance by similarity to ideal solution). The essential idea of TOPSIS is that most favoured alternative ought not just have the only shortest distance from the positive ideal solution, but also have the farthest distance from the negative ideal solution [Hwang and Yoon (1981); Zeleny (1982)]. Also, this method has benefits like simple, rationally comprehensible concept, computational efficiency is good, capability of measuring the relative performance for every alternative [Yeh (2002)]. Then Zadeh

(1965) introduced the fuzzy sets theory in year 1965, which can be only assigns a single membership value to each element amongst zero and one. Balioti et al. (2018) deployed MCDM technique along with TOPSIS method regarding the selection of spillway for a dam in Greece.

In this research, we have deployed TOPSIS method with entropy weight for medical field so as to obtain accurate diagnosis during initial stages that can further help to patient in management of medical expenditures and time savings because of collective opinion of three experts by TOPSIS model can provide quickest and precise diagnosis, which makes it different from other past research work.

In healthcare management, decisions are taken on the basis of single physician's experience and judgment In previous eras, few procedures have been created so as to partially deal with such issues. In any case, the majority of these methodologies center just around a single dimension, though in all actuality, medical experts have requirement of decision supporting strategies to settle on decisions by considering multiple criteria at the same time. To successfully meet clinician data requirements to accomplish the normal patient rational patient care management, it is important to move the decision supporting tools that recently centered around single criteria to the ones which are transparent and systematic methodologies by considering all significant criteria simultaneously. Hence, use of a multi-criteria decision supporting strategy is usually preferred in priority setting, technique ordered preference by similarity to the ideal solution (TOPSIS) [Lai et al. (1994)], višekriterijumsko kompromisno rangiranje (VIKOR) [Opricovic (1998)] etc., and these can be utilizing in managing decision issues under the nearness of various decisions criteria.

MCDM strategies are intended to help individuals settle on better decisions when they manage complex choices containing numerous dimensions. Previously mentioned medical decision making has these abilities, which make MCDM technique an ideal medical decision support system to empower patients and doctors to do tasks essential to execute proficiently in clinical settings. Various application papers that are concentrating on the MCDMs as decision supporting techniques in medicinal services settings have been studied.

## REFERENCE

- [1] Deng H., Yeh C. H., Willis R. J. *Inter-company comparison using modified TOPSIS with objective weights. Computers & Operations Research* 2000 (27) 963–973.
- [2] Jiang W., Wei B., Liu X., Li X., Zheng H. *Intuitionistic Fuzzy Power Aggregation Operator Based on Entropy and Its Application in Decision Making. International Journal of Intelligent Systems* 2018 (33) 49–67.
- [3] Li J., Deng G., Li H., Zeng W. *The relationship between similarity measure and entropy of intuitionistic fuzzy sets. Inform Sci.* 2012 (188) 314–321.
- [4] Tzeng G. H., Lin C. W., Opricovic S., *Multi-criteria analysis of alternative-fuel buses for public transportation. Energy Policy* 2005 (33) 1373–1383



- [5] Wei C., Wang P., Zhang Y. Entropy similarity measure of interval-valued intuitionistic fuzzy sets and their applications. *Inform Sci.* 2011 (181) 4273–4286.
- [6] Yu P. L. A class of solutions for group decision problems. *Management Science* 1973 (19) 936–946
- [7] Zeleny M., *Multiple criteria decision making*. McGraw Hill, New York, 1982.
- [8] Lai Y. J., Liu T. Y., Hwang C. L. TOPSIS for MODM. *Eur. J. Oper. Res* 1994 (76) 486–500.
- [9] Lin L., Yuan H., Xia Z. Q. Multi -criteria fuzzy decision making methods based on intuitionistic fuzzy sets. *Journal of Computer and System Sciences* 2007 (73) 84-88.
- [10] Wan S. P., Dong J. Y. Method of trapezoidal intuitionistic fuzzy number for multi-attribute group decision. *Control and Decision*. 2010 (25) 773-776.
- [11] Wang J. Q. Overview on fuzzy multi-criteria decision-making approach. *Control and Decision* 2008 (23) 601-606.
- [12] Wang J. Q., Zhang Z. H. Programming method of multi-criteria decision-making based on intuitionistic fuzzy number with incomplete certain information. *Control and Decision* 2008a (23) 1145-1148.
- [13] Wang J. Q., Zhang Z. H. Multi-criteria decision-making method with incomplete certain information based on intuitionistic fuzzy number. *Control and Decision* 2009 (24) 226-230.
- [14] Wang J. Q., Zhang Z. Aggregation operators on intuitionistic trapezoidal fuzzy number and its application to multi-criteria decision making problems. *Journal of Systems Engineering and Electronics* 2009a (20) 321-326.
- [15] Wei G. W. Hesitant fuzzy prioritized operators and their application to multiple attribute decision making. *Knowledge-Based Systems* 2012 (31) 176-182.
- [16] Yager R. R. Modeling prioritized multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics, Part B. Cybernetics* 2004 (34) 2396-2404.
- [17] Yager R. R. Prioritized aggregation operators. *International Journal of Approximate Reasoning* 2008 (48) 63-274.
- [18] Yager R. R. Prioritized OWA aggregation. *Fuzzy Optimization and Decision Making* 2009 (8) 245-262.
- [19] Yu X., Xu Z. Prioritized intuitionistic fuzzy aggregation operators. *Information Fusion* 2013 (14) 108-116
- [20] Zhang Z. Intuitionistic Trapezoidal Fuzzy Prioritized Operators and their application to Multiple Attribute Group Decision Making. *British Journal of Mathematics & Computer Science* 2014 1951-1998.
- [21] Wan S. P., Dong J. Y. Method of trapezoidal intuitionistic fuzzy number for multiattribute group decision. *Control and Decision* 2010 (25) 773-776.
- [22] Zhang X., Jin F., Liu P. D. A grey relational projection method for multi-attribute decision making based on intuitionistic trapezoidal fuzzy number. *Applied Mathematical Modelling* 2013 (37) 3467-3477.
- [23] The world health report 2004 – changing history. Geneva: World Health Organization; 2004. Retrieved from <<https://www.who.int/whr/2004/en/>> accessed on 18/6/2019.
- [24] A global brief on vector-borne diseases, World Health Organization. Retrieved from [25] <[https://apps.who.int/iris/bitstream/handle/10665/111008/WHO\\_DCO\\_WHD\\_2014.1\\_eng.pdf;jsessionid=4CF4DCE325DB2FBAC03B0AE0BCFAE09?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/111008/WHO_DCO_WHD_2014.1_eng.pdf;jsessionid=4CF4DCE325DB2FBAC03B0AE0BCFAE09?sequence=1)> accessed on 18/6/2019.
- [26] Hwang C. L., Yoon K. *Multiple attribute decision making – methods and applications: a state-of-the-art survey*. Springer Verlag. New York, 1981.
- [27] Zeleny M. *Multiple criteria decision making*, McGraw-Hill, New York, 1982.
- [28] Yeh C. H. A problem-based selection of multi-attribute decision -making methods, *International Transactions in Operational Research* 2002 (9) 169-181.
- [29] Zadeh L. A. Fuzzy sets. *Information and Control* 1965 (8) 338- 356.
- [30] Balioti V., Tzimopoulos C., Evangelides C. Multi-criteria decision making using TOPSIS method under fuzzy environment. application in spillway selection. *Proceedings* 2018 (2)1-8.



# Instructions for Authors

## Essentials for Publishing in this Journal

- 1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.
- 2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.
- 3 All our articles are refereed through a double-blind process.
- 4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

## Submission Process

All articles for this journal must be submitted using our online submissions system. <http://enrichedpub.com/> . Please use the Submit Your Article link in the Author Service area.

---

## Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd.** The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

### Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

### Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

### Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

### Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

### Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

#### Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).
2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);
3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);
4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

### **Professional articles:**

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

### **Language**

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

### **Abstract and Summary**

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

### **Keywords**

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

### **Acknowledgements**

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

### **Tables and Illustrations**

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

### **Citation in the Text**

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

### **Footnotes**

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.

The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

### **Address of the Editorial Office:**

**Enriched Publications Pvt. Ltd.**  
S-9, IInd FLOOR, MLU POCKET,  
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,  
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,  
PHONE: - + (91)-(11)-45525005