

Early Prediction of Fire Accident in Petroleum Industries by Statistical Machine Learning Algorithm

S R Mugunthan

Associate Professor, Department of Computer Science and Engineering, Sri Indu College of Engineering and Technology, Hyderabad, India
E-mail: srmugunth@gmail.com

Abstract

Due to unpredictability of climatic conditions across the world, early fire forecasting has become more challenging and critical for many oil and gas sectors. It is extremely hard for anyone to predict fires with any degree of certainty, especially in the gas or oil sectors. Until now, the models in use have not been adequate. However, this is critical in order to maintain workers and property safe. As a result, this research work investigates the different approaches available for fire hazard assessment and prediction in order to deal with fire dangers. Also, this research work presents the statistical machine learning methods to detect fire accidents in petroleum industries based on risk index models and risk assessment parameters by performing a statistical process. Moreover, this research work develops a statistical machine learning method to enhance the accuracy in predicting the fire occurrence. Finally, the proposed algorithm is measured by utilizing the performance metrics such as accuracy, proposed risk index, and sensitivity.

Keywords: Machine learning, SVM



1. Introduction

Generally, the non-intrusive and cost-effective methods involve in the identification and management of operational issues that are chosen for the online monitoring of combustion processes. Gravimetric properties, for example, the curvature and luminance of the flame, the temperature distribution, or the frequency of oscillation offer details on combustion efficiency and reliability by making them an excellent candidate for fuel monitoring and control strategy development [1, 2]. Figure 1 shows the sample fire accident occurred at petroleum industry.



Figure 1. Fire Accident at Petroleum Industry

To identify potential associations between image features and operational and performance parameters, these studies have investigated the possible relationships between various swirl numbers and fuel flame emissions as well as operational parameters such as swirl number and many gas emissions in the context of pulverized fuel flames [3-5]. The further research on this subject may help in the introduction of biomass fuels into conventional power plants, especially in

Southern Europe, where it is limited, this research is particularly critical in the EU, where it is almost impossible to get enough energy resources [6, 7]. The CO₂ emissions can be reduced by using co-firing, which is a promising technique to minimize the existing Middle East countries' and Arabia's reliance on non-renewable energy supplies, especially in light of the current political circumstances in such nations [8, 9]. Early prediction can be succeeding through logistic regression approach in statistical SVM. Figure 2 shows the logistic regression approach for SVM [10].

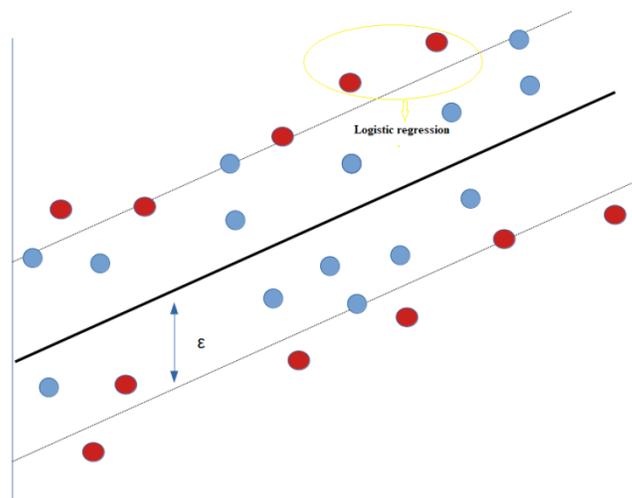


Figure 2. Logistic Regression Approach for SVM

Machine learning is used to build various prediction models for flame detection. It is also used to anticipate accidents, which may benefit the gas and oil sectors by encouraging the development of better designs [11, 12]. When measuring, an influx of information should be constantly monitored to figure out the exact score and provide the anticipated result. When the proof of construct is developed, the organization can perceive the possible result of the solution and pick up the areas where there are certain research gaps [13-15]. In addition, it allows many researchers or shareholders in an international market to track and quantify the potential solutions that help to reduce unnecessary risk. Determine design requirements and project deadlines well

ahead of time. Due to the job's intricacy, developing an algorithmic plan is challenging [16, 17]. To get the best results, a visual interface should be developed to analyze them. To calculate precision and accuracy, the method must be quantitative. The measurement accuracy score emphasizes the outcomes to assist businesses in decision-making and serves as a route to avoid workplace accidents [18].

Motivation of this research

The primary motivation of this paper is to use the past statistics of oil maintenance yard fires and explosion in any country to identify the most dangerous areas, vulnerable facilities, and equipment, as well as the sources of ignition, the types of accidents, and substance of the accidents. Additionally, recommendations are provided to improve security management in oil workshops in the United Arab Emirates, Arabia, and many African countries.

2. Organization of the Research

This research article is organized in several sections as follows; Section 3 discusses about the recent research works in fire detection techniques in oil and gas industries. Section 4 delivers the proposed methodology to predict the fire in early stage at petroleum industries. Section 5 discusses about the various parameters that help to measure the proposed framework. The proposed research work is concluded in section 6.

3. Preliminaries

Mathematical connections between features of the Dutch road infrastructure and road safety were discovered by the duo of Martine Reurings and Theo Janssen in 2006 when they worked on a project called Infrastructure and Road Safety. Using exposure and risk metrics, the features of the Dutch road infrastructure are connected to road safety. RIPCORDER-ISEREST Work



Package 2 is investigating models used for accident prediction and traffic safety impact evaluations [19]. Recently, smys et al have described how data mining might be used to identify the cause of fire accidents that are more likely to have such incidents [20].

Ronza A et al. formulated a technique to represent the frequency of a sequence scenario using the product of the probability of the event's occurrence and the event's frequency. This method is more reliable based on a large variety of recorded fire incidents in the gas and oil industries [21].

Jan K. Wachter et al. developed many practices that were tested in a complex model for measuring and preventing safety incidents, which included using accident rates as an unprejudiced safety metric practically after their theoretical study on relevant research. Based on the data, there appears to be a significant negative relationship between the presence of ten individual safety management practices, as well as the composite of these practices, and accident rates; as well, there appears to be a significant negative relationship between levels of safety-focused worker emotional and cognitive engagement with accident rates [22].

Research Gap

Since the related latest literature on the events was reviewed, it is possible to get an understanding of them. However, when statistically looking at the data, this information will have very less impact on the model's ability to inaccurately forecast many periods. The proposed research study attempts to connect the disparate examining various datasets with the use of a vector-based classification model that has been built by utilizing grid-search principles to provide a reliable early prediction model. The proposed study intends to have a positive impact on workplace safety by finding out whether or not workplace accidents can be predicted early.

4. Methodologies

4.1 Hazard Identification

In this context, problem solutions should provide a course of action which should lead to the accomplishment of goals, and which should serve as the foundation for establishing and executing control measures. Employees and employees who are too near to the sequences may no longer notice and identify the risks, or may view the events as minor since to their knowledge, where no one has been injured [23, 24]. Figure 3 shows the overall block diagram of proposed work.



Figure 3. Overall block diagram of proposed framework

4.1.1 Need of Data Cleansing

Uncovering the source of the event necessitates the usage of main data cleaning, since occurrences with unknown origins need it. An excellent alternative is to use text mining as a problem-solving tool to identify the best strategy for predicting the underlying cause of the occurrence [25].

4.1.2 Cause Factor Identification

It should be necessary to construct a multi-factor classification method to arrive at a measure of the weighting associated with the potential causes of the event, and the same approach may be used to identify the source of the incident [26, 27]. When using the earlier data and the

known variables, the model is built using the past data as a foundation and is then trained to get the future outcome. All of the many causes of an issue will be identified and the relative importance of each will be determined. The learning parameter will be adjusted every time when the event is calculated. Using the learning parameter, the reason for the event will be predicted.

4.2 Risk Index Assessment

To identify each structure and factory's danger of fire, this research work utilizes the benchmark datasets that include both past and present data records for fire occurrence statistics. Many factors are linked to fire incidence in this data collection. Two distinct methods that are used to perform fire risk assessment are discussed in this article. In order to establish fire occurrence prediction, this research work utilizes:

1. Risk index assessment
2. Statistical machine learning with SVM

Next, this research work constructs a risk index that rates the chance of a fire occurring; we name this index "proposed report index" (PRI). In the case of fire risk indexing, forecast accuracy and lift value are used as performance metrics. After carrying out the calculations, a prediction model is used to carry out a fire risk assessment [28, 29].

4.3 Early prediction from past data statistics

Alfred Binet, a pioneer in the study of child intelligence, described statistics as learning from data. Machine learning is used to help computers learn and become more intelligent by absorbing and learning from data. Therefore, statistical machine learning is used for machine learning through statistics. While statistics often rely on the notion of inference, there are three basic techniques: hypothesis testing, estimation, and inference. Normality assumption is also used

when dealing with statistical data. On the other hand, statistical machine learning is inherently involved in improving the performance of current machine learning by utilizing statistics-based inference and normalcy assumptions; it results in statistical machine learning improving the overall performance of machine learning [30]. Regression is one of the most often used forms of statistics. In addition, deep learning is one of the most used machine learning algorithms. We used logistic regression and deep neural networks for statistical machine learning in this study. There are many ways in which statistical machine learning may be carried out. For the most part, these models are used for classification, prediction, and clustering. This paper's goal is to create a prediction model for assessing fire danger. Fire risk forecasting and evaluation using statistical machine learning is performed using these models.

Since the answer variable contains frequency data, we employ count data analysis techniques. When it comes to analysing the fire occurrence data, the Poisson and binomial distributions may be used to both discrete and continuous probability distributions. We suggested creating a prediction model of the likelihood of fire, and so we developed a binomial distribution model.

Step 1:

When Y is expressed as a binomial distribution, it is a binomial random variable with n and p.

$$P(Y = y) = \binom{n}{y} p^y (1 - p)^{n-y}, y = 0, 1 \dots n$$

Where n is the number of Bernoulli iteration number and p is probability of success. The mean and variance of the Y are np and np(1-p) respectively.

Step 2:

Each y has a “0” or “1” are represented as “fire” or “no fire” respectively. Therefore, we can create logistic creation in SVM model based on the variable for the forecast of fire risk. The proposed model can be defined as,

$$\log\left(\frac{p}{1-p}\right) = \alpha_0 + \alpha_1x_1 + \alpha_2x_2 + \dots + \alpha_kx_k + \tau$$

Where τ parameter adjusts constant

4.4 Safety workplace

To establish the safeness score, data from previous safety incidents is used. Once the analysis has done its work, it moves through the data and assigns a category's score as a crucial parameter in the prediction model [15]. Coordinate descent may be used to develop a custom vector-based data mining method for each safety category. The percentage score should range 0 to 100. Scores are based on the data that has been recorded over the last 30 days. Scores are computed according to the criteria and schedules of the organization. The weighted average of all categories is used for estimating the overall score. Figure 4 shows the detailed block diagram of proposed work.

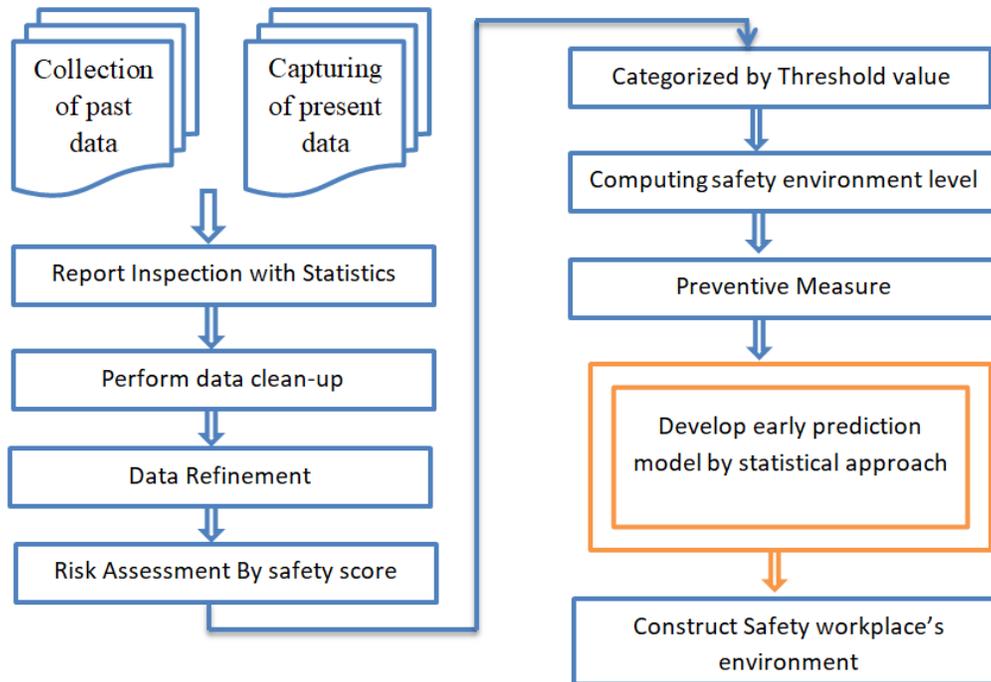


Figure 4. Detailed Block Diagram of Proposed Work

5. Results & Discussion

The proposed framework incorporates PRI in order to accurately predict early fire detection. Logistic regression is used for fire risk forecasting, and since that requires tuning of missing variables, a perfect early-detection method helps in finding the missing fire detection in petroleum industries. As for the new PRI, it is possible to look at various features with variable special values that are utilized for p (probability) values here. We define and classify Safe Score in accordance with the following ranges shown in the table 1:

Table 1. Category based zonal separation

| Category | Range | Safety risk | Zone category |
|----------|-------|-------------|---------------|
|----------|-------|-------------|---------------|

| | | | |
|--------------|----------|------------------------------|-------------|
| Category - 1 | 86-100 | Good Safe | Green Zone |
| Category - 2 | 60-85 | Situation Requires Attention | Orange Zone |
| Category - 3 | Below 59 | Potential Safety Issue | Red Zone |

0 signifies a fire was present in the response variable and 1 means no fire was present in the response variable using probability values with a hypothesis sector. We also tried out by using the logistic regression and SVM to categories the results that were obtained. In order to achieve better classification accuracy, we have divided the dataset into 80% training and 20% testing. This proposed framework has worked well with many hidden nodes and layers, and has thus experienced many iterations and errors. “Relu” activation is found near the neural network's final stages. We've constructed enough forecasting models based on our trained dataset.

The performance of the model has evaluated with various metrics. The result of prediction accuracy of proposed model is high compared to other algorithm. The risk index parameters are used to improve the performance of the statistical model. This PRI has used as good efficient risk assessment tool for the early detection of fire in an industry. The components of the fire indexing and defining a formula for the PRI as,

$$PRI = \frac{H_i \times H_p \times H_{ig}}{B_s \times F_e \times F_a \times F_c \times S_{cs} \times F_{ps}} * \%$$

Some performance metrics are as follows,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FP}$$



Where,

| | | |
|----------|---------------------------------------|---|
| H_i | Intrinsic hazards | Floor, structure fire load and etc |
| H_p | Process hazards | Hazardous material |
| H_{ig} | Ignition hazards | Fire based on gas, electrical facilities |
| B_s | Safety parameters for Building | Fire drill, smoking control and hot work |
| F_e | Fire protection equipment | Fire extinguisher, sprinkler and etc |
| F_a | Fire alarm system | Alarm segment |
| F_c | Fire compartment system | Evacuation system |
| S_{cs} | Control system for Smoke | Fire brigade |
| F_{ps} | Public service for fire | |

Figure 6 shows the overall performance measures. Besides, Our proposed SVM gives highest performance compared to other algorithm.

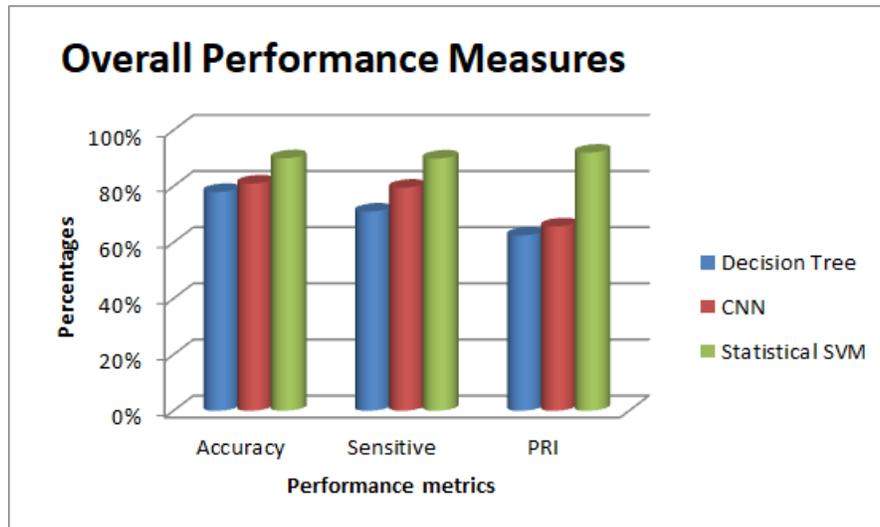


Figure 6. performance Measures of Proposed System

6. Conclusion

Thus, the proposed framework has successfully performed early fire prediction using risk assessment with statistical SVM approach of efficient optimized risk indexing. An early fire risk prediction is bit difficult and focuses very rarely. This article has focused on the recent advances in the important theoretical assessment of fire prediction perspectives such as computation and imposition of fire insurance premiums with regards to real-world issues with fire risk assessment. Traditionally, the process of evaluating the fire danger in buildings and industries rely on the expertise of a group of fire specialists. However, the proposed research has allowed the computation of these hazard ratings to be objective and measurable. The future research study will be dealt with predicting the fire danger in buildings and industries by unmanned vehicle (UAV) in real-time monitoring in a more efficient manner. Additionally, this research work can be applicable and will also be relevant to detect forest fire danger and other fields.

References

- [1] Srinivas, Kethavath, and Mohit Dua. "Fog Computing and Deep CNN Based Efficient Approach to Early Forest Fire Detection with Unmanned Aerial Vehicles." In International Conference on Inventive Computation Technologies, pp. 646-652. Springer, Cham, 2019.
- [2] Bindhu, V., and G. Ranganathan. "Hyperspectral Image Processing in Internet of Things model using Clustering Algorithm." Journal of ISMAC 3, no. 02 (2021): 163-175.
- [3] Yannis, G., Dragomanovits, A., Laiou, A., Richter, T., Ruhl, S., La Torre, F., (2016). Use of Accident Prediction Models in Road Safety Management - An International Inquiry. Transportation Research Procedia, 14, 4257–4266. <https://doi.org/10.1016/j.trpro.2016.05.397>
- [4] Chen, Joy Iong-Zong. "Design of Accurate Classification of COVID-19 Disease in X-Ray Images Using Deep Learning Approach." Journal of ISMAC 3, no. 02 (2021): 132-148.



- [5] Bourhim, El Mostafa, and Abdelghani Cherkaoui. "Exploring the potential of virtual reality in fire training research using A'WOT hybrid method." In *Intelligent Systems, Technologies and Applications*, pp. 157-167. Springer, Singapore, 2020.
- [6] Smys, S., Abul Bashar, and Wang Haoxiang. "Taxonomy Classification and Comparison of Routing Protocol Based on Energy Efficient Rate." *Journal of ISMAC* 3, no. 02 (2021): 96-110.
- [7] Wachter, J. K., &Yorio, P. L. (2014). A system of safety management practices and worker engagement for reducing and preventing accidents: An empirical and theoretical investigation. *Accident Analysis and Prevention*, 68, 117–130. <https://doi.org/10.1016/j.aap.2013.07.029>
- [8] Srishilesh, P. S., Latha Parameswaran, RS Sanjay Tharagesh, Senthil Kumar Thangavel, and P. Sridhar. "Dynamic and Chromatic Analysis for Fire Detection and Alarm Raising Using Real-Time Video Analysis." In *International Conference On Computational Vision and Bio Inspired Computing*, pp. 788-797. Springer, Cham, 2019.
- [9] Verma, S., & Chaudhari, S. (2017). Safety of Workers in Indian Mines: Study, Analysis, and Prediction. *Safety and Health at Work*, 8(3), 267–275. <https://doi.org/10.1016/j.shaw.2017.01.001>
- [10] Pushpa, B., and M. Kamarasan. "Morlet Wavelet Threshold Based Glow Warm Optimized X-Means Clustering for Forest Fire Detection." In *International Conference On Computational Vision and Bio Inspired Computing*, pp. 1094-1105. Springer, Cham, 2019.
- [11] Laska, M., Herle, S., Klamma, R., &Blankenbach, J. (2018). A Scalable Architecture for Real-Time Stream Processing of Spatiotemporal IoT Stream Data—Performance Analysis on the Example of Map Matching. *ISPRS International Journal of Geo-Information*, 7(7), 238. <https://doi.org/10.3390/ijgi7070238>
- [12] Manickam, M. Thanga, M. Yogesh, P. Sridhar, Senthil Kumar Thangavel, and Latha Parameswaran. "Video-Based Fire Detection by Transforming to Optimal Color Space." In



- International Conference On Computational Vision and Bio Inspired Computing, pp. 1256-1264. Springer, Cham, 2019.
- [13] Eitrich, T., & Lang, B. (2006). Efficient optimization of support vector machine learning parameters for unbalanced datasets. *Journal of Computational and Applied Mathematics*, 196(2), 425–436. <https://doi.org/10.1016/j.cam.2005.09.009>
- [14] Singh, A. Robert, Suganya Athisayamani, S. Sankara Narayanan, and S. Dhanasekaran. "Fire Detection by Parallel Classification of Fire and Smoke Using Convolutional Neural Network." In *Computational Vision and Bio-Inspired Computing*, pp. 95-105. Springer, Singapore, 2021.
- [15] Davoudi Kakhki, F., Freeman, S. A., & Mosher, G. A. (2019). Evaluating machine learning performance in predicting injury severity in agribusiness industries. *Safety Science*, 117(July 2018), 257–262. <https://doi.org/10.1016/j.ssci.2019.04.026>
- [16] Sunitha, P. J., and K. R. Joy. "Deep CNN-Based Fire Alert System in Video Surveillance Networks." In *Computational Vision and Bio-Inspired Computing*, pp. 599-615. Springer, Singapore, 2021.
- [17] Ramasubramanian, Sreesruthi, Senthil Arumugam Muthukumaraswamy, and A. Sasikala. "Enhancements on the Efficacy of Firefighting Robots Through Team Allocation and Path-Planning." In *Data Intelligence and Cognitive Informatics*, pp. 139-151. Springer, Singapore, 2021.
- [18] Alkheder, S., Taamneh, M., & Taamneh, S. (2017). Severity Prediction of Traffic Accident Using an Artificial Neural Network. *Journal of Forecasting*, 36(1), 100–108. <https://doi.org/10.1002/for.2425>
- [19] Verslycke, T., Reid, K., Bowers, T., Thakali, S., Lewis, A., Sanders, J., & Tuck, D. (2014). The Chemistry Scoring Index (CSI): A hazard-based scoring and ranking tool for chemicals and products used in the oil and gas industry. *Sustainability (Switzerland)*, 6(7), 3993–4009. <https://doi.org/10.3390/su6073993>



- [20] Smys, S., and Jennifer S. Raj. "Assessment of Fire Risk and Forest Fires in Rural Areas Using Long Range Technology." *Journal of Electronics* 2, no. 01 (2020): 38-48.
- [21] Ronza, A., Félez, S., Darbra, R. M., Carol, S., Vílchez, J. A., & Casal, J. (2003). Predicting the frequency of accidents in port areas by developing event trees from historical analysis. *Journal of Loss Prevention in the Process Industries*, 16(6), 551–560. <https://doi.org/10.1016/j.jlp.2003.08.010>
- [22] Wachter, J. K., & Yorio, P. L. (2014). A system of safety management practices and worker engagement for reducing and preventing accidents: An empirical and theoretical investigation. *Accident Analysis and Prevention*, 68, 117–130. <https://doi.org/10.1016/j.aap.2013.07.029>
- [23] Bourhim, El Mostafa, and Abdelghani Cherkaoui. "Usability Evaluation of Virtual Reality-Based Fire Training Simulator Using a Combined AHP and Fuzzy Comprehensive Evaluation Approach." In *Data Intelligence and Cognitive Informatics*, pp. 923-931. Springer, Singapore, 2021.
- [24] Ghorpade, Pallavi S., and Shilpa K. Rudrawar. "Review on Evacuation Systems for Indoor Fire Situation." In *International Conference on Computer Networks and Inventive Communication Technologies*, pp. 28-37. Springer, Cham, 2019.
- [25] Hamdan, Yasir Babiker. "Construction of Statistical SVM based Recognition Model for Handwritten Character Recognition." *Journal of Information Technology* 3, no. 02 (2021): 92-107.
- [26] Balasubramaniam, Vivekanadam. "Artificial Intelligence Algorithm with SVM Classification using Dermoscopic Images for Melanoma Diagnosis." *Journal of Artificial Intelligence and Capsule Networks* 3, no. 1 (2021): 34-42.
- [27] Sungheetha, Akey, and Rajesh Sharma. "Design an Early Detection and Classification for Diabetic Retinopathy by Deep Feature Extraction based Convolution Neural

- Network." Journal of Trends in Computer Science and Smart technology (TCSST) 3, no. 02 (2021): 81-94.
- [28] Karunakaran, P., and Yasir Babiker Hamdan. "Early Prediction of Autism Spectrum Disorder by Computational Approaches to fMRI Analysis with Early Learning Technique." Journal of Artificial Intelligence 2, no. 04 (2020): 207-216.
- [29] Chen, Joy Iong Zong, and P. Hengjinda. "Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method-A Comparative Study." Journal of Artificial Intelligence 3, no. 01 (2021): 17-33.
- [30] Balasubramaniam, Vivekanadam. "Artificial Intelligence Algorithm with SVM Classification using Dermoscopic Images for Melanoma Diagnosis." Journal of Artificial Intelligence and Capsule Networks 3, no. 1: 34-42.

Author's biography

S R Mugunthan is presently working as an Associate Professor, in the Department of Computer Science and Engineering, in Sri Indu College of Engineering and Technology, Hyderabad, India. His area of research includes ubiquitous networks, Big Data Analytics, Industrial System and Collective Intelligence in all fields of Computing and Communication Technologies.

