International Journal of
Computational Learning & Intelligence
Vol. 04 | Issue 03 | 2025

RESEARCH ARTICLE OPEN ACCESS

A Smart Prediction System for Forest Fires Using Cellular Automata and Machine Learning

S Mounika . V Rizwana . S Hussain Basha . S Mohammed Ismail . T Anusha

Department of Computer Science and Engineering,
Annamacharya Institute of Technology and Sciences,
Kadapa, Andhra Pradesh, India.

DOI: **10.5281/zenodo.15210665**
Received: 27 January 2025 / Revised: 21 February 2025 / Accepted: 27 March 2025
©Milestone Research Publications, Part of CLOCKSS archiving

Abstract - Forest fires represent a widespread and devastating natural disaster affecting millions of hectares of forest land each year and posing significant risks to both human lives and property. Timely and accurate predictions of forest fire behavior are crucial for developing effective risk management strategies and improving firefighting responses. This study introduces an innovative Forest Fire Spread Behavior Prediction (FFSBP) model, which integrates two key components: the Forest Fire Spread Process Prediction (FFSPP) model and the Forest Fire Spread Results Prediction (FFSRP) model. The FFSPP model leverages a combination of advanced methodologies to forecast the dynamics of fire spread, while the FFSRP model aims to predict the final outcomes of fire events.Moreover, the FFSRP model demonstrates strong predictive capabilities, particularly for smaller and medium-scale fire scenarios. These results highlight the potential of the FFSBP model as a powerful tool for improving the accuracy of forest fire predictions and supporting more effective risk management and firefighting operations.

Index Terms : Forest fire prediction, forest fire behavior, cellular automata,Wang Zhengfei model, machine learning.

I. INTRODUCTION

The increasing frequency and intensity of global forest fires, fueled by global warming and more extreme weather events, present significant threats to human life, property, and ecosystems. Every year, vast expanses of forested land are destroyed by fire, resulting in severe financial and human losses. Accurate prediction of forest fire behavior is essential for preventing or minimizing the impact of such disasters. Forest fire prediction can be classified into three main areas: fire risk weather prediction, forest fire occurrence prediction, and forest fire behavior prediction. These areas address specific factors: (i) fire risk weather prediction focuses on meteorological elements (ii) forest fire occurrence prediction considers weather, available fuels, and ignition sources and (iii) forest fire behavior prediction accounts for meteorological conditions, fuels, and terrain features to predict how fires evolve.This study specifically focuses on predicting forest fire behavior, providing valuable insights into the overall fire dynamics and outcomes. Unlike traditional approaches, which primarily predict the likelihood of fire occurrence, our work aims to model the overall process and results of fire development.

II. LITERATURE SURVEY

Cellular Automata (CA) models are widely recognized for their ability to represent dynamic systems within discrete, finite states. These models have gained popularity for their simplicity and efficiency in simulating complex systems like forest fire spread. CA allows for the modeling of fire dynamics from a local perspective, making it an effective tool for forest fire prediction. While these models perform well in predicting forest fire behavior, they often consider only a subset of the many variables affecting fire spread. Furthermore, the application of these models can be limited by the availability of relevant data. This study seeks to improve predictive accuracy by combining the Wang Zhengfei model with CA. By selecting easily obtainable factors—such as fuel state, weather conditions, and terrain—this approach aims to generate a more comprehensive prediction of forest fire behavior in practical settings.

Machine learning (ML) has become increasingly valuable in predicting forest fire impact, particularly by identifying complex, non-linear relationships between multiple variables. ML models are particularly useful in forecasting the impact of forest fires, including the extent of the affected areas. However, existing models face challenges, such as difficulty in obtaining certain input data (e.g., combustion parameters) or dealing with an excess or deficiency of data, leading to increased data collection efforts or suboptimal prediction performance.Recognizing the potential of ensemble learning to enhance model performance, this study adopts ensemble methods to improve predictions. By combining multiple ML models, ensemble learning can address complex tasks and data sets. This paper incorporates multicollinearity tests for data selection and grid search methods for model optimization, aiming to improve the accuracy of fire impact predictions.

This study makes several key contributions to addressing these gaps. First, it integrates CA and ML methods to predict forest fire behavior, considering a comprehensive set of factors such as combustibility, weather conditions, and topography. Second, it combines the CA model with the Wang Zhengfei model to improve the accuracy and visual representation of forest fire spread.

III. METHODOLOGY

Wildfires pose a significant environmental challenge, causing extensive damage to ecosystems, wildlife, and human settlements. Predicting fire spread is essential for effective response and mitigation efforts. Various techniques have been explored to model and anticipate fire behavior, aiding firefighters and decision-makers in developing strategies to minimize destruction. The methodology represents a workflow for fire prediction using machine learning and spatial analysis. It starts with a fire dataset, which is split into training and testing data. Feature extraction is performed at fire points and randomly generated points, considering factors such as distance to roads, rivers, lakes, and settlements, as well as land cover, NDVI, DEM (Digital Elevation Model), aspect, slope, runoff, soil moisture, wind speed, and temperature variables. The extracted features are used for model development, where a weighted overlay of raster layers is applied to improve fire prediction. The final step involves implementing precautionary measures, such as establishing more fire stations in high-risk areas.

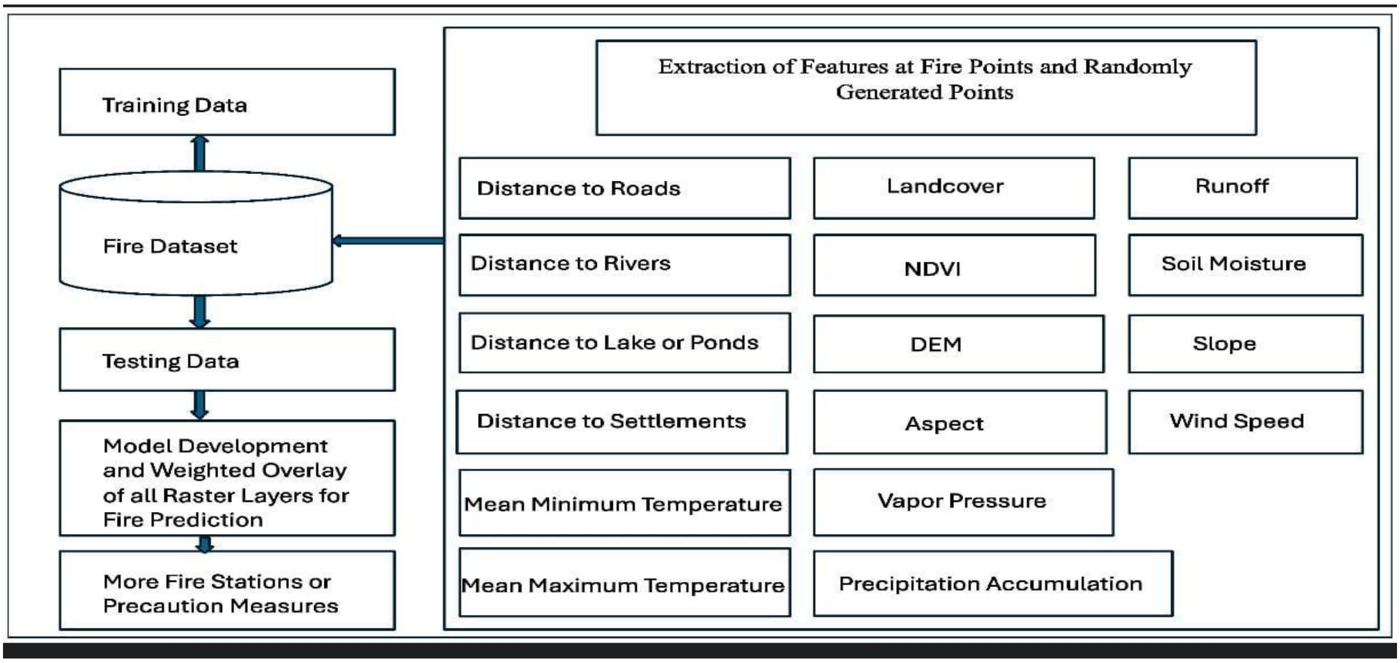

Fig 1: Architecture of Forest Fire Prediction

This study utilizes real-world fire data collected from historical fire data, including key factors such as weather conditions and fire danger indices. The dataset was preprocessed by encoding date information numerically and analyzing key statistical features. To enhance predictive accuracy, two machine learning models—XGBoost (XGB) and Gradient Boosting (GBoost)—were optimized using the GridSearchCV method. Recent advancements in technology have led to the incorporation of data-driven approaches, utilizing satellite imagery, climate data, and machine learning algorithms. These innovations help improve response times and allow for more precise forecasting of fire behavior. Additionally, real-time monitoring systems contribute to better decision-making by providing up-to-date information on fire progression. The significance of fire prediction models lies in their ability to reduce damage and improve safety measures. As research continues, enhancing the accuracy of these models remains a priority, ensuring that emergency response teams are equipped with the necessary tools to mitigate the impact of wildfires effectively.

IV. RESULTS

The performance of forest fire prediction is evaluted using Random Forest, Artificial Neural Network, Support Vector Machine, Gradient Boosting and XGBoost. Among these XGBoost performs the best due to its ability to handle inbalanced datasets effectively.

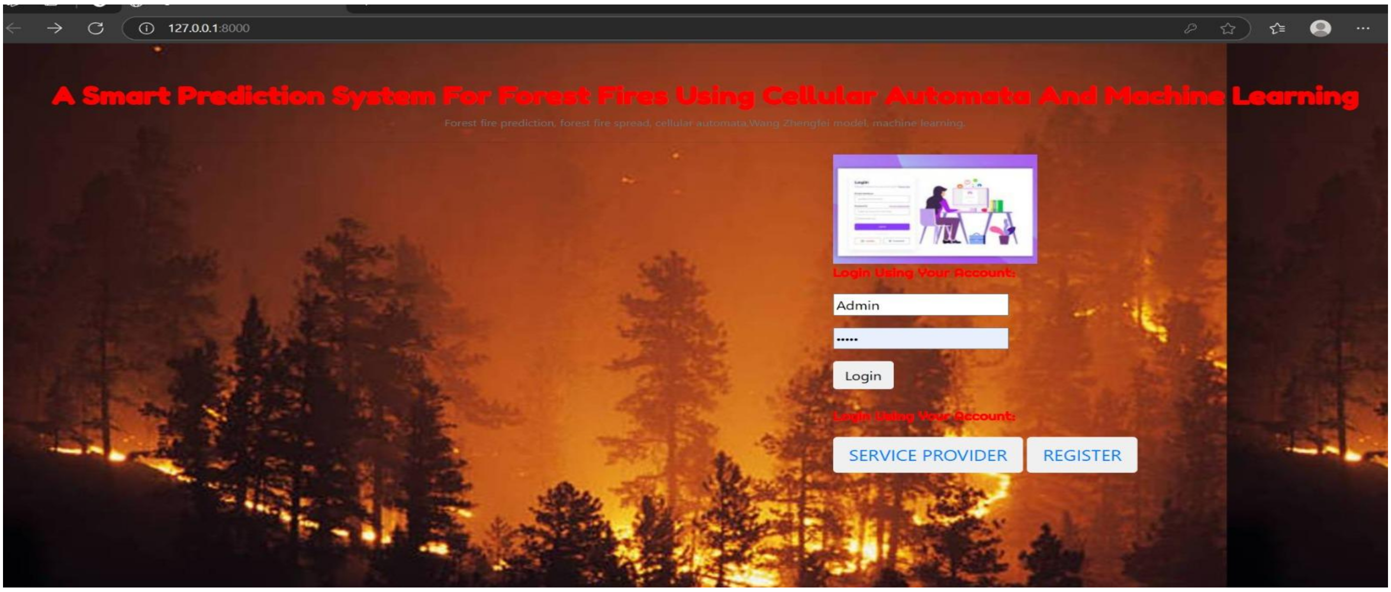

Fig 2: UI of Forest Fire Prediction

LR analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Fig 3: Registration UI

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

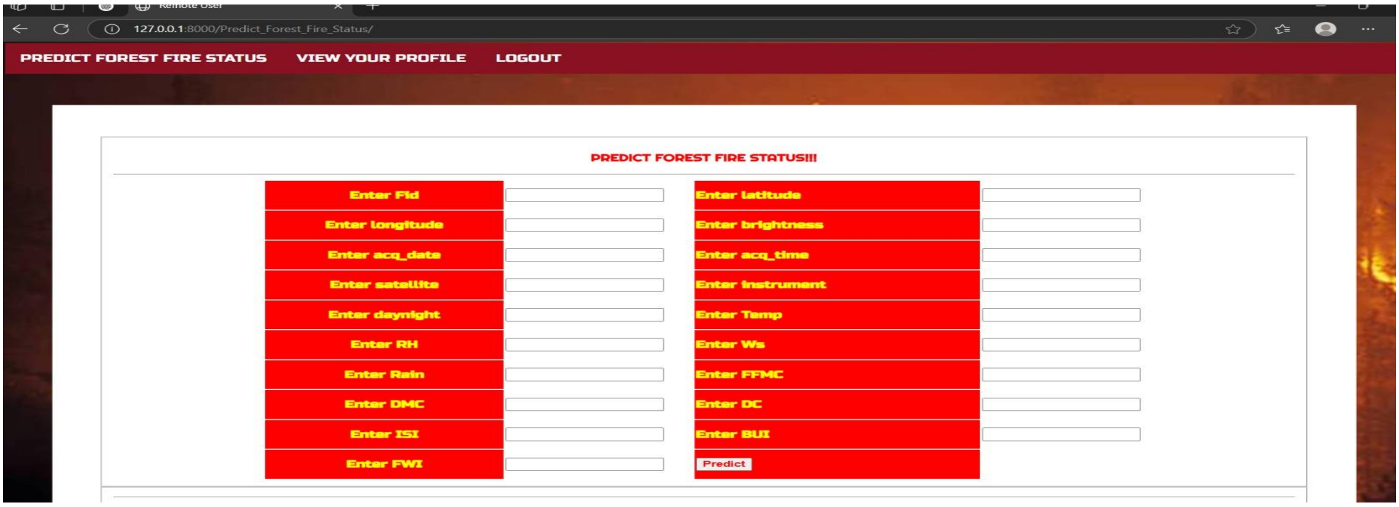

Fig 4: Prediction of Results

SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

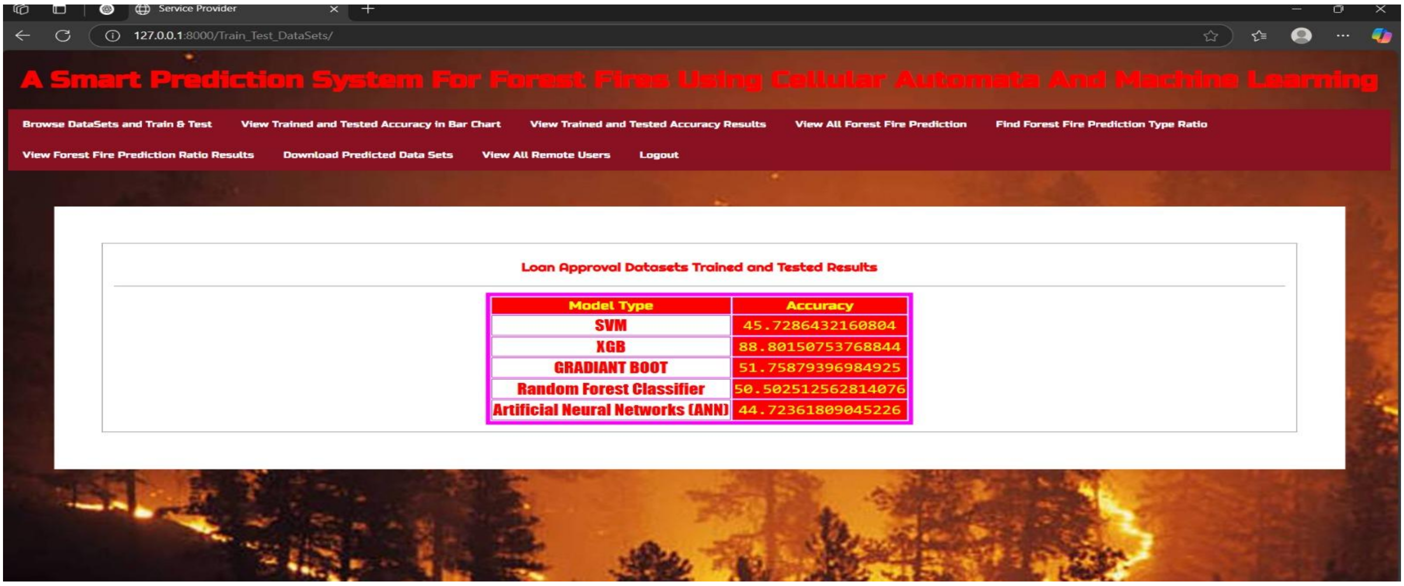

Fig 5: Accuracy based on Algorithms

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

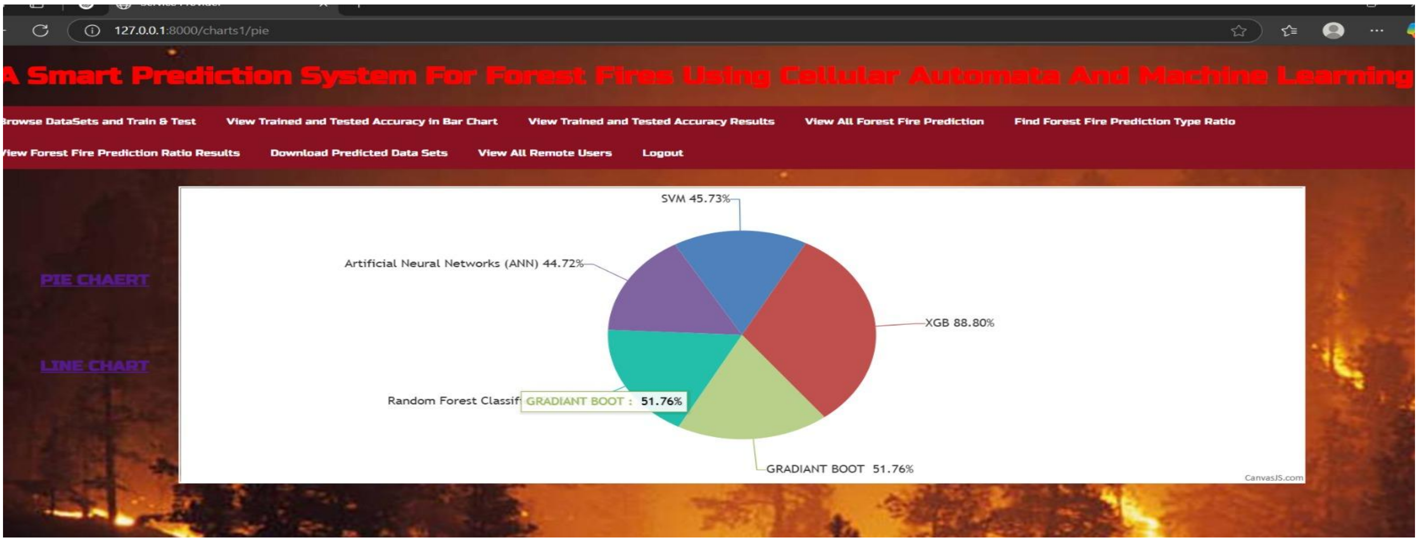

Fig 6: Pie chart Representation accuracy

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm widely used in forest fire prediction due to its efficiency, speed, and accuracy. It is an ensemble learning method that improves predictive performance by combining multiple decision trees.

V. CONCLUSION

It introduces an improved model for predicting forest fire spread following an accidental outbreak. Drawing from relevant literature and real-world conditions, key influencing factors that are readily accessible in practical applications are identified. These factors include fuel characteristics, meteorological conditions, and topographical features. Fuel characteristics refer to vegetation type, meteorological factors encompass temperature, humidity, wind conditions, and the Fire Weather Index (FWI), while topography involves geographical location and elevation changes. A novel approach is developed by integrating machine learning techniques to enhance predictive accuracy. The model is evaluated using historical fire data, including the “3.29 Forest Fire” in China and real fire incidents from Montesinho National Forest Park in Portugal. Comparisons indicate that the proposed model demonstrates greater accuracy than existing simulation tools such as Farsite and Prometheus, particularly for small- and medium-scale fire scenarios. The proposed model effectively captures fire propagation patterns by leveraging advanced machine learning techniques. When applied to historical fire events, it exhibits improved predictive performance compared to widely used models in North America.The integration of multiple machine learning algorithms, including XGB and GBoost through stacking, enhances prediction accuracy.

REFERENCES

1. Deng, X., Zhang, Z., Zhao, F., Zhu, Z., & Wang, Q. (2023). Evaluation of the regional climate model for the forest area of Yunnan in China. *Frontiers in Forests and Global Change, 5*, Article 1073554.
2. Flannigan, M., Stocks, B., Turetsky, M., & Wotton, M. (2009). Impacts of climate change on fire activity and fire management in the circumboreal forest. *Global Change Biology, 15*(3), 549–560.

3. Gao, C., Lin, H., & Hu, H. (2023). Forest-fire-risk prediction based on random forest and backpropagation neural network of Heihe area in Heilongjiang province, China. *Forests, 14*(2), 170.
4. Ntinopoulos, N., Spiliotopoulos, M., Vasiliades, L., & Mylopoulos, N. (2022). Contribution to the study of forest fires in semi-arid regions with the use of Canadian fire weather index application in Greece. *Climate, 10*(10), 143.
5. Pang, Y., Li, Y., Feng, Z., Feng, Z., Zhao, Z., Chen, S., & Zhang, H. (2022). Forest fire occurrence prediction in China based on machine learning methods. *Remote Sensing, 14*(21), 5546.
6. Rubí, J. N. S., de Carvalho, P. H. P., & Gondim, P. R. L. (2023). Application of machine learning models in the behavioral study of forest fires in the Brazilian federal district region. *Engineering Applications of Artificial Intelligence, 118*, Article 105649.
7. Shi, C., & Zhang, F. (2023). A forest fire susceptibility modeling approach based on integration machine learning algorithm. *Forests, 14*(7), 1506.
8. Turetsky, M. R., Kane, E. S., Harden, J. W., Ottmar, R. D., Manies, K. L., Hoy, E., & Kasischke, E. S. (2011). Recent acceleration of biomass burning and carbon losses in Alaskan forests and peatlands. *Nature Geoscience, 4*(1), 27–31.
9. Xie, D. W., & Shi, S. L. (2014). Prediction for burned area of forest fires based on SVM model. *Applied Mechanics and Materials, 513–517*, 4084–4089.
10. Artés, T., Cencerrado, A., Cortés, A., & Margalef, T. (2013). Relieving the effects of uncertainty in forest fire spread prediction by hybrid MPI-OpenMP parallel strategies. *Procedia Computer Science, 18*, 2278–2287.
11. Ahmed, S. T., Kaladevi, A. C., Shankar, A., & Alqahtani, F. (2025). Privacy Enhanced Edge-AI Healthcare Devices Authentication: A Federated Learning Approach. *IEEE Transactions on Consumer Electronics*.
12. Singh, K. D., & Ahmed, S. T. (2020, July). Systematic linear word string recognition and evaluation technique. In *2020 international conference on communication and signal processing (ICCSP)* (pp. 0545-0548). IEEE.
13. Syed Thouheed Ahmed, S., Sandhya, M., & Shankar, S. (2018, August). ICT's role in building and understanding indian telemedicine environment: A study. In *Information and Communication Technology for Competitive Strategies: Proceedings of Third International Conference on ICTCS 2017* (pp. 391-397). Singapore: Springer Singapore.
14. Sreedhar Kumar, S., Ahmed, S. T., & NishaBhai, V. B. (2019). Type of supervised text classification system for unstructured text comments using probability theory technique. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(10).
15. Ahmed, S. T., Basha, S. M., Arumugam, S. R., & Kodabagi, M. M. (2021). *Pattern Recognition: An Introduction*. MileStone Research Publications.