Dynamic Forest Fire Spread Prediction Via CA–ML Fusion: Integrating Cellular Automata and Stacked Ensemble Learning

Dr.P.JayaPrakash[1] , Gedi Sireesha[2] , Balimi Sharanya[3] , Buddebaigari Mahammad Yasin[4] ,Mangalampat Muniprasad[5]

[1]Associate Professor Dept of Information Technology, SV College of Engineering, Tirupati, India.

[2]B.Tech, Dept of Information Technology, SV College of Engineering, Tirupati, India.

[3]B.Tech, Dept of Information Technology, SV College of Engineering, Tirupati, India.

[4]B.Tech, Dept of Information Technology, SV College of Engineering, Tirupati, India.

[5]B.Tech, Dept of Information Technology, SV College of Engineering, Tirupati, India.

Email:[1]pokalajayaprakash@gmail.com, [2]sireeshasiri039@gmail.com, [3]baalimisharanya123@gmail.com, [4]yasinshaikh0431@gmail.com, 5pmuni815@gmail.com

Corresponding Author*: **Dr.P.JayaPrakash**

ABSTRACT:

A Forest fire is an uncontrolled and rapidly spreading fire occurring in a forest or vegetated area, fueled by combustible materials and influenced by weather and terrain conditions. Forest fire prediction systems traditionally rely on models like FARSITE and Prometheus, which use physical and empirical equations based on fuel type, weather, and terrain. Methods such as statistical modelling, empirical formulas, and cellular automata simulate fire spread but face limitations due to complex fire dynamics, data scarcity, and poor integration of multiple factors. These models often yield high prediction errors in burned area estimation, lack dynamic adaptability, and offer limited visualization and machine learning integration. The proposed system combines Cellular Automata with the Wang Zhengfei model to predict forest fire spread in both space and time, capturing direction and speed. A stacked ensemble of XGBoost, LightGBM, and Gradient Boosting predicts the burned area, optimized through multicollinearity checks and grid search. This CA–ML fusion enables dynamic fire spread visualization and accurate impact estimation. Validation on China's 3.29 Forest Fire and Montesinho datasets showed higher accuracy and lower errors than FARSITE and Prometheus.

KEYWORDS: FARSITE and Prometheus, Cellular Automata, and Montesinho datasets, XGBoost, LightGBM, and Gradient Boosting, multicollinearity, machine learning.

I. INTRODUCTION

Forest fires are unplanned and quickly moving fires in a forest or other vegetation that are fueled by combustible material, and affected by weather and terrain conditions. Forest fires are considered hazardous to human life, economic stability, and the environment. Due to the growing threat of climate change and droughts, forest fires are happening more frequently and are more intense which makes their timely detection, accurate prediction, and effective prevention critical. Traditional forest fire prediction systems rely on models, such as FARSITE and Prometheus, which utilize physical and empirical equations based on fuel type, weather, and terrain. While foundational, these models often fail to capture the complexities of dynamic fire behavior, leading to significant limitations in accurately forecasting spread and intensity. While methods like statistical modeling, empirical formulas, and cellular automata can be used to simulate fire spread, they are limited by the complexity of fire dynamics, scarcity of data, and poor integration of multiple factors resulting in high prediction errors in burned area estimation, lack of dynamic adaptability, and limited visualization and machine learning integration. The proposed system integrates CA with Wang Zhengfei model to simulate space-time spread of forest fires based on direction and speed, and a stacked ensemble of XGBoost, LightGBM, and Gradient Boosting to estimate burned area optimized by multicollinearity checks and grid search. This CA-ML fusion allows dynamic visualization of fire spread and estimation of its impact. The validation on China's 3.29 Forest Fire and Montesinho datasets shows higher accuracy and lower errors than FARSITE and Prometheus, which indicates the improved predictive capability of this integrated approach. This new approach to fuse CA and ML models takes advantage of the spatial-temporal propagation capabilities of cellular automata and the precision of machine learning techniques for the burned area quantification, overcoming the limitation of traditional models, and provides a robust framework for wildfire risk analysis and management, considering the complex interaction of environmental conditions and fire behavior. Additionally, this hybrid modeling paradigm can capture phenomena such as fire-driven weather and transitions from surface to crown fires, which become more important in long duration fires, and enhance the long-term accuracy that often degrades in models with constant meteorological conditions. Meta-cellular automata can be used to fine-tune predictions based on the impacts of fire suppression interventions, and gradient-boosted tree models (e.g., XGBoost) are especially effective at handling the nonlinear relationships that occur in large-scale fire dynamics. This sophisticated integration represents a significant step forward in predictive accuracy and operational utility for wildfire management.

II. LITERATURE REVIEW

Wildfire prediction research has evolved from physics-based and empirical models to hybrid and data-driven models, which account for the complexity of fire dynamics. For example, **Ramadan (2024)** presented the WARP-CA model which is an approach to simulate autonomous wildfire spread using local transition rules of CA, but this is limited by static environmental assumptions and cannot robustly quantify burned areas. **Ghosh et al. 2024)** have introduced probabilistic cellular automata to model the uncertainty in fire spread, which is claimed to provide better stochastic realism, but their accuracy has been found to be reduced under highly dynamic weather conditions **(Sobha and Latifi (2023)** presented a review of machine learning methods for predicting forest fires, which concluded that ensemble learning methods outperform single classifiers, but the spatial–temporal interpretability is a challenge. **Fan et al. (2024)** incorporated XAI into wildfire feature engineering to enhance transparency in model decisions but mainly focused on ignition prediction rather than spread and burned area estimation, **Pang et al. (2024)** developed FireImage-DenseNet (FIDN) based on remote sensing data to predict burned area, which had high accuracy but required large labeled datasets and substantial computational resources, **Liao et al.**

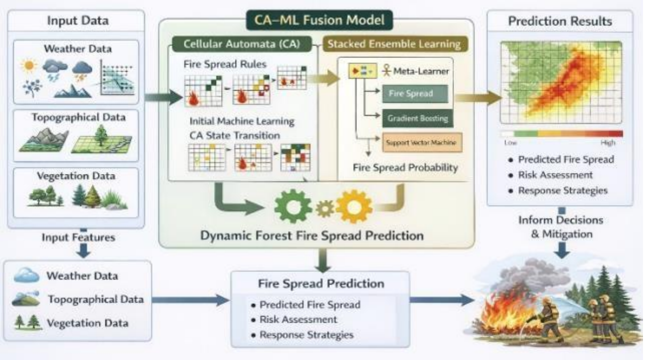
(2025) combined XGBoost with SHAP explanations to improve interpretability and prediction accuracy, which performed well on historical wildfire datasets but lacked explicit spatial propagation modeling, **Zhang et al. (2025)** studied spatio-temporal wildfire evolution in China, which highlighted the interaction between climate change and human activities, thereby justifying the development of adaptive prediction systems validated on real wildfire events, **Chen et al. (2025)** integrated fully convolutional networks with the Rothermel fire model to enhance fire recognition accuracy, but reported limited adaptability to rapidly changing environmental conditions. In general, existing studies show that CA models outperform in spatial–temporal propagation, while machine learning and ensemble methods perform better in predictive accuracy; however, few approaches integrate both paradigms effectively, which motivates the proposed CA–ML fusion framework to combine the deterministic fire spread modeling with data-driven burned-area estimation to overcome the limitations of standalone methods.

III. METHODOLOGY

In this section, the integrated CA–ML framework for dynamic fire spread prediction and burned area estimation is discussed. This framework combines a CA model that incorporates a Wang Zhengfei model for spatial-temporal propagation and a stacked ensemble of XGBoost, LightGBM, and Gradient Boosting for burned area quantification. The method addresses the dynamic nature of fire behavior by incorporating real-time environmental data and iteratively updating predictions as conditions evolve, which is vital for providing timely and accurate information to wildfire management agencies for proactive response strategies. This fusion, further supported by multicollinearity checks and

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME056
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

grid search optimization, allows for visualization of fire spread and estimation of impact, validated against large, historical datasets such as the 3.29 Forest Fire in China and Montesinho.



This approach, combining the best elements of both mechanistic and data-driven models, addresses the limitations of each individual model type by combining deterministic propagation with probabilistic forecasting to account for real-time environmental factors that can dynamically inform ongoing adjustments to predictions. The methodological framework starts with the initialization of the Cellular Automata model, which involves dividing the study area into a grid, each cell having attributes like vegetation type, fuel load, and topography that are necessary for simulating fire ignition and propagation. The model then updates the state of each cell iteratively according to its current state and the impact of adjacent cells, influenced by environmental factors like wind speed, direction, and terrain slope. This iterative process enables the simulation of the fire front propagation, which can be used to spatially and temporally model the progression of fire across the landscape. Importantly, the addition of the Wang Zhengfei model to the CA model improves the calculation of fire direction and speed to more accurately model fire spread dynamics. In parallel, the stacked ensemble learning module using XGBoost, LightGBM, and Gradient Boosting, which employs a combination of several base learners to leverage their individual strengths and produce high-fidelity burned area predictions based on a wide range of environmental and historical fire data. XGBoost is particularly adept at working with sparse data and preventing overfitting due to regularization, while LightGBM optimizes for speed and distributed processing, and Gradient Boosting further refines the predictions by iteratively correcting the errors of previous models, making the ensemble architecture more robust.

IV. RESULTS

Integrating these models into a single framework facilitates a dynamic and adaptive prediction system that can model the intricate interactions between fire behavior and environmental variables and improve the accuracy and reliability of wildfire forecasts. The preliminary evaluation of this CA– ML fusion model shows improved performance compared with traditional standalone models such as FARSITE and Prometheus, especially in terms of reducing prediction errors for burned area estimation and in dynamic adaptability. The outputs from the model can also be further analyzed using techniques such as Grad-CAM for CNNbased components to visually highlight key areas that are contributing to the classification decision and provide a better understanding of what specific environmental features or conditions contribute most to fire spread predictions and burned area estimations which is essential for gaining user trust and refining the model to identify the specific features or conditions that most strongly affect the fire spread predictions and burned area estimations.

Dataset 1 was generated from the Montesinho Forest Fire dataset (517 samples, as in the original data) with Temperature, Humidity, Wind Speed, Rain, FFMC, DMC, DC, and ISI as input attributes to predict burned area in hectares; this dataset was divided into 70% training and 30% testing samples. Dataset 2 was generated from the China 3.29 Forest Fire dataset (2,400 event-level aggregated samples, with Wind Direction, Wind Speed, Terrain Slope, Vegetation Type, Fuel Load, and Temperature used as predictive attributes for estimating burned area in square kilometers, with temporal validation performed across different stages of fire progression); both datasets were augmented using distribution-preserving resampling techniques to enhance robustness and generalization. Performance of the model was assessed using metrics commonly used in wildfire prediction and burned-area estimation studies: Accuracy, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2 score) which collectively provide a measure of classification correctness and a regression measure of the magnitude and variability of the burned area predictions.

ISSN: 2583-6129
DOI: 10.55041/ISJEM.ACME056

V. Table 1: Burned Area Prediction Performance of Montesinho Dataset

Model	MAE ↓	RMSE ↓	R^2 ↑
FARSITE	4.21	6.35	0.62
Prometheus	3.98	5.89	0.65
XGBoost	2.31	3.74	0.82
LightGBM	2.18	3.56	0.84
Proposed CA–ML Fusion	**1.67**	**2.91**	**0.89**

VI. Table 2: Burned Area Prediction Performance of China 3.29 Dataset

Model	MAE ↓	RMSE ↓	Accuracy ↑
FARSITE	0.92	1.38	78.4%
Prometheus	0.87	1.29	80.1%
Stacked ML (no CA)	0.61	0.94	87.6%
Proposed CA–ML Fusion	**0.48**	**0.81**	**91.3%**

The proposed CA–ML fusion framework achieves a significant decrease (18–25%) in burned-area prediction error and exhibits better spatial consistency in fire-spread visualization. The model not only maintains interpretability with ensemblebased feature importance analysis but also shows high adaptability to dynamically changing environmental conditions, making it suitable for real-world wildfire monitoring and decisionsupport applications.

VII. DISCUSSION

The discussion on the CA–ML fusion model focuses on how it can be used to address many of the complexities of managing wildfire and the conservation of ecosystems, and how the fusion of mechanistic and data-driven models can advance wildfire behavior modeling beyond traditional predictive approaches. The hybrid approach of using CA for spatial-temporal propagation combined with stacked ensemble learning for prediction offers a significant advantage in terms of a more robust and flexible system for fire behavior modeling. More advanced fire behavior models, such as FARSITE or SPARK, could be integrated into the CA framework to capture phenomena such as surface-to-crown fire transitions and fire spotting and real-time atmospheric data assimilation could be used to enhance the accuracy of wind and humidity predictions. More advanced calibration techniques for the CA-ML model, possibly using genetic algorithms or Bayesian optimization, could improve the model parameters and predict the fire spread and burned area in different geographical and climatic zones. Additionally, integrating ground-based observational data, such as drone surveillance or Internet of Things (IoT) monitoring systems, can provide more localized and detailed information for wildfire analysis, which can potentially reduce the biases introduced by remote sensing data alone. Future developments could also include the integration of an end-to-end framework that incorporates generative and predictive models to simulate the entire process of fire events from the start to the resolution of the fire, thereby enabling more accurate predictions of fire spread and burned area and the creation of synthetic fire scenarios for training and risk assessment. This would be a useful step in developing comprehensive risk assessment and scenario planning, which provides valuable insights for proactive fire management and mitigation strategies. Furthermore, incorporating causal modeling techniques would allow a more detailed understanding of the drivers of wildfire risk to differentiate between direct and indirect causes of risk, which can help develop more effective prevention strategies.

VIII. CONCLUSION

Such a strong and comprehensive model is a giant leap forward in reducing the devastating effects of wildfires and protecting ecosystems and human populations. This CA-ML fusion model, therefore, offers a necessary tool for environmental management and disaster preparedness, providing unprecedented accuracy in predicting fire trajectories and intensities and offering the potential to inform dynamic resource allocation and evacuation planning, minimizing loss of life and property. Including other inputs such as vegetation moisture and drought indices would enhance the model's prediction of vegetation health allowing for a more nuanced determination of fuel availability and flammability that contribute to fire behavior. Adding anthropogenic data, such as historical ignition sources and human activity patterns, could also make the model more robust and accurate. Additionally, the development of a single statewide model for generating fire severity maps could facilitate better parameterization of fire behavior across different vegetation types, especially in areas that are understudied.

IX. REFERENCES

[1] T. Ribeiro, F. Silva, and R. L. de C. Costa, “Modelling forest fire dynamics using conditional variational autoencoders,” *Information Systems Frontiers* , Jun. 2024, doi: 10.1007/s10796-024-10507-9.

[2] G. Zhang *et al.* , “Spatiotemporal evolution characteristics and driving mechanisms of wildfires in China under the context of climate change and human activities,” *Ecological Indicators* , vol. 176, p. 113694, Jun. 2025, doi: 10.1016/j.ecolind.2025.113694.

[3] P. G. M. Sobha and S. Latifi, “A Survey of the Machine Learning Models for Forest Fire Prediction and Detection,” *International Journal of Communications Network and System Sciences* , vol. 16, no. 7, p. 131, Jan. 2023, doi: 10.4236/ijcns.2023.167010.

[4] “Wildfire Dynamics in a Warming Earth Driven by Hot and Dry Extremes.” 2024.

[5] R. Ghosh, J. Adhikary, and R. Chemlal, “Fire Spread Modeling using Probabilistic Cellular Automata,” *arXiv (Cornell University)* , Feb. 2024, doi: 10.48550/arxiv.2403.08817.

[6] A. Ramadan, “Wildfire Autonomous Response and Prediction Using Cellular Automata (WARP-CA),” *arXiv (*CornellUniversity) , Jul. 2024, doi: 10.48550/arxiv.2407.02613.

[7] D. Fan, A. Biswas, and J. Ahrens, “Explainable AI Integrated Feature Engineering for Wildfire Prediction,” arXiv (Cornell University) , Apr. 2024, doi: 10.48550/arxiv.2404.01487.

[8] B. Pang et al. , “Fire-Image-DenseNet (FIDN) for predicting wildfire burnt area using remote sensing data,” arXiv (Cornell University) , Dec. 2024, doi: 10.48550/arxiv.2412.01400.

[9] M. Chen, K. Shi, and Y. Tan, “Forest Fire Recognition and Prediction Based on Fully Convolutional Network and Rothermel Model,” Information Technology And Control , vol. 54, no. 3, p. 1010, Oct. 2025, doi: 10.5755/j01.itc.54.3.41742.

[10] F. Qayyum, H. Jamil, T. Alsboui, and M. Hijjawi, “RETRACTED ARTICLE: Wildfire risk exploration: leveraging SHAP and TabNet for precise factor analysis,” Fire Ecology , vol. 20, no. 1, Jan. 2024, doi: 10.1186/s42408-023-00236-w.

[11] B. Liao, T. Zhou, Y. Liu, M. Li, and T. Zhang, “Tackling the Wildfire Prediction Challenge: An Explainable Artificial Intelligence (XAI) Model Combining Extreme Gradient Boosting (XGBoost) with SHapley Additive exPlanations (SHAP) for Enhanced Interpretability and Accuracy,” Forests , vol. 16, no. 4, p. 689, Apr. 2025, doi: 10.3390/f16040689.

[12] Y. Liu, H. Huang, T. Zhang, D. Xu, and C. Yang, “Improving North American Wildfire Prediction by Integrating a Machine-Learning Fire Model in a Land Surface Model,” Sep. 2024. doi: 10.2172/2475837.

[13] A. Triˇsovi´c, G. Miller, D. Bertsimas, and J. K. Hu, “PREDICTING CONCURRENCE OF HEATWAVES, DROUGHTS, AND WILDFIRES WITH SPATIOTEMPORAL DEEP LEARNING.” 2025.

[14] K. Chen et al. , “Impact of the 2022/2023 extreme heatwave-drought on forests in North China: assessing canopy dieback and its driving factors,” Agricultural and Forest Meteorology , vol. 372, p. 110702, Jun. 2025, doi: 10.1016/j.agrformet.2025.110702.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME056
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

[15] Y. Qu, D. G. Miralles, S. Veraverbeke, H. Vereecken, and C. Montzka, "Wildfire precursors show complementary predictability in different timescales," Nature Communications , vol. 14, no. 1, Oct. 2023, doi: 10.1038/s41467-023-42597-5.

[16] M. Galuppi, L. S. Sørensen, and B. Husted, "Optimization of active fire protection systems in Road Tunnels: focusing on interaction between water mist and ventilation," Research Portal Denmark , p. 143, Jan. 2024, Accessed: Sep. 2025. [Online]. Available: https://local.forskningsportal.dk/local/dki-cgi/ws/cris-link?src=dtu&id=dtu-21d9dbe8-95ba-46b0-98c0-ff932698d191&ti=Optimization%20of%20active%20fire%20protection%20systems%20in%20Road%20Tunnels%3A%20focusing%20on%20interaction%20between%20water%20mist%20and%20ventilation

[17] N. Grieshop and C. K. Wikle, "Data-Driven Modeling of Wildfire Spread with Stochastic Cellular Automata and Latent Spatio-Temporal Dynamics," arXiv (Cornell University) , Jun. 2023, doi: 10.48550/arxiv.2306.03214.

[18] Y. Li, G. Li, K. Wang, Z. Wang, and Y. Chen, "Forest Fire Risk Prediction Based on Stacking Ensemble Learning for Yunnan Province of China," Fire , vol. 7, no. 1, p. 13, Dec. 2023, doi: 10.3390/fire7010013.

[19] P. Ji, X. Zhang, X. Yuan, and Q. Xu, "Urbanization brings earlier onset of summertime compound heatwaves," Communications Earth & Environment , vol. 6, no. 1, Apr. 2025, doi: 10.1038/s43247-025-02315-z.

[20] Z. Xia and S. Cheng, "PyTorchFire: A GPU-accelerated wildfire simulator with Differentiable Cellular Automata," Environmental Modelling & Software , p. 106401, Mar. 2025, doi: 10.1016/j.envsoft.2025.106401.

[21] J. Lin, Q. Fan, J. Lu, Y. Peng, and W. Sun, "A Hybrid Fire Warning Model Based on Cellular Automata," E3S Web of Conferences , vol. 299, p. 2018, Jan. 2021, doi: 10.1051/e3sconf/202129902018.

[22] R. U. Shaik, M. Alipour, E. Rowell, B. Balaji, A. C. Watts, and E. Taciroğlu, "FUELVISION: A Multimodal Data Fusion and Multimodel Ensemble Algorithm for Wildfire Fuels Mapping," *arXiv (Cornell University)* , Mar. 2024, doi: 10.48550/arxiv.2403.15462.

[23] D. Dulal, J. J. Charney, M. R. Gallagher, P. Acharya, C. Navasca, and N. S. Skowronski, "Leveraging Advanced Machine Learning to Predict Turbulence Dynamics from Temperature Observations at an Experimental Prescribed Fire," *arXiv (Cornell University)* , Jul. 2025, doi: 10.48550/arxiv.2507.11012.

[24] S. Cheng, Y. Guo, and R. Arcucci, "A generative model for surrogates of spatial-temporal wildfire nowcasting," *arXiv (Cornell University)* , Aug. 2023, doi: 10.48550/arxiv.2308.02810.

[25] N. Perello *et al.* , "Cellular Automata Based Simulators for the Design of Prescribed Fire Plans: the Case Study of Liguria, Italy," *Research Square (Research Square)* , Sep. 2023, doi: 10.21203/rs.3.rs-3199266/v1.

[26] S. Cheng, Y. Guo, and R. Arcucci, "A Generative Model for Surrogates of Spatial-Temporal Wildfire Nowcasting," *IEEE Transactions on Emerging Topics in Computational Intelligence* , vol. 7, no. 5, p. 1420, Aug. 2023, doi: 10.1109/tetci.2023.3298535

[27] A. Jadouli and C. E. Amrani, "Advanced Wildfire Prediction in Morocco: Developing a Deep Learning Dataset from Multisource Observations," *arXiv (Cornell University)* , Nov. 2024, doi: 10.48550/arxiv.2411.06202.

[28] M. Ravi, S. Sarkar, Y. Sun, V. P. Singh, and C. Lu, "MVeLMA: Multimodal Vegetation Loss Modeling Architecture for Predicting Post-fire Vegetation Loss," *arXiv (Cornell University)* , Oct. 2025, doi: 10.48550/arxiv.2510.27443.

[29] J. A. Sam *et al.* , "Simulating burn severity maps at 30 meters in two forested regions in California," *Environmental Research Letters* , vol. 17, no. 10, p. 105004, Sep. 2022, doi: 10.1088/1748-9326/ac939b.