Fire Type Classification in the USA Using Supervised Machine Learning Techniques

Ranyah Taha[*,1] (iD), **Fuad Musleh**[2] (iD)**, Abdel Rahman Musleh**[3] (iD)

[1] Computer Science Dept., Al-Iman School, Bahrain

[2] Civil engineering Department, College of Engineering, University of Bahrain, Sakhir, 1054, Bahrain

[3] Electrical and Electronics Engineering Department, College of Engineering, University of Bahrain, Sakhir, 1054, Bahrain

*Corresponding author: Ranyah Taha, raniacs2014@gmail.com

ABSTRACT: Wildfires are a growing global concern, causing widespread environmental, economic, and health impacts. In the USA, fire incidents have become more frequent and intense due to factors such as climate change, prolonged droughts, and human activities. Machine learning plays a vital role in predicting and classifying fires by analyzing vast satellite and environmental datasets with high speed and accuracy. These models support early warning systems and informed decision-making, ultimately helping to reduce damage and improve emergency response strategies. This study evaluates the effectiveness of supervised machine learning algorithms—including Decision Tree (DT), Random Forest (RF), Support Vector Classifier (SVC), K-Nearest Neighbors (KNN), Logistic Regression (LR), and Gradient Boosting Classifier (GBC)—in classifying different fire types. The DT emerges as the top-performing model, achieving the highest results across all evaluation metrics, including 96.69% accuracy, precision, recall, and F1 score. RF follows closely with similarly strong performance, making it a highly reliable alternative. GBC ranks third, showing balanced and consistent results above 92% in all metrics. In contrast, SVC and LR perform less effectively, particularly in precision and F1 score, indicating that they are not ideal choices for fire type classification in this study. The novelty of this study lies in its application of a comparative ML framework to classify fire types using real satellite-based observations specific to the USA. region. By integrating and evaluating multiple ML models on this large-scale, real-world dataset, the study provides valuable insights into model suitability for fire classification tasks and offers practical guidance for deploying predictive tools in environmental monitoring and disaster management systems.

KEYWORDS: Artificial Intelligence, Data Analysis, Fire type Classification, Machine Learning, USA, NASA, Civil Engineering.

1. Introduction

Fires represent a major environmental disaster due to their rapid spread, the complexity of containment efforts, and the extensive damage they inflict on ecosystems, infrastructure, and human health. In the USA, fire incidents—particularly wildfires—have become increasingly frequent and intense, driven by factors such as climate variability, land use changes, and human activity. The severe consequences of these events have underscored the importance of fire detection, classification, and management, making fire monitoring a vital component of forestry, environmental protection, and emergency response strategies [1].

Several critical factors contribute to the occurrence and spread of fires across the United States. Climatic variables—including high temperatures, strong wind speeds, low relative humidity, limited rainfall, and lightning probability—create conditions that significantly increase the risk of fire ignition and propagation. In addition to environmental influences, human-related factors such as population density, land development, and increased recreational or industrial activity in forested and rural regions further elevate fire risk. The combination of these natural and anthropogenic elements makes fire prediction and classification an increasingly urgent priority for disaster management and environmental protection [1].

Artificial Intelligence (AI) plays a transformative role in modern wildfire detection and classification systems, significantly enhancing the ability to anticipate, monitor,

and manage fire events. AI technologies contribute to various aspects of wildfire preparedness and response, including fuel assessment, fire behavior prediction, real-time detection, impact estimation, and strategic fire management. Leveraging tools such as satellite imagery, historical weather data, and computational models, AI enables the automated analysis of complex environmental patterns [2].

In particular, Machine Learning (ML)—a subset of AI—is increasingly utilized for the early prediction and accurate classification of fires by identifying patterns in large-scale datasets. These intelligent systems support timely decision-making and resource allocation, making AI a critical component in reducing wildfire-related risks and improving emergency response strategies [2].

This study utilizes a dataset comprising fire incident records detected throughout the United States in 2021. The data were collected by the VIIRS sensor aboard the SNPP satellite and sourced from the NASA Open Data Portal. The research follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework to ensure a structured approach to data analysis and model development. Since each machine learning method has its own advantages and limitations, a comparative evaluation is necessary to determine the most effective model for classifying fire types. Therefore, this work focuses on assessing the performance of six supervised learning algorithms—DT, RF, SVC, KNN, LR, and GBC— in predicting fire categories. The paper is organized into several sections: a literature review, methodology, data description and preprocessing, model implementation, results, discussion, conclusion, and future recommendations.

2. Literature Review

Several ML algorithms have been instrumental in advancing forest fire forecasting. This section reviews various studies that have applied these methods, as outlined below recent research has extensively explored various ML and AI techniques for forest fire prediction and management.

In [3], the authors addressed critical challenges in forest fire prediction by proposing a robust ML framework specifically designed to handle severely imbalanced datasets, a frequent issue in wildfire modeling. The study utilized Copernicus reanalysis data from 2000 to 2018, incorporating 27 features including temperature, soil moisture, wind speed, and vegetation indices to model fire susceptibility in Canada's boreal forests. To manage the 158:1 non-fire-to-fire ratio, the authors employed a hybrid sampling strategy combining NearMiss3 for undersampling and SMOTE-ENN for oversampling with noise reduction. Among the models tested—RF, XGB, LGBM, and CatBoost—XGB combined

with NearMiss3 at a 0.09 sampling ratio achieved optimal performance, with 98.08% accuracy, 86.06% sensitivity, and 93.03% specificity. Moreover, the study emphasized the balance between computational efficiency— demonstrated by LGBM's histogram-based learning— and model interpretability, using feature importance to highlight soil moisture as a dominant factor in fire prediction.

Similarly, the authors in [4] conducted a detailed evaluation of ML models using meteorological data from Algeria, integrating a temporal-stage approach and correlation-based feature selection (CFS) to enhance predictive accuracy. The study divided the dataset into six-time intervals and focused on weather indicators such as temperature, humidity, and FWI components. Important predictors including FFMC, DMC, and FWI were identified through CFS, significantly improving model accuracy. Among the tested models—DT, RF, SVC, LR, KNN, and GNB—DT and RF both achieved perfect accuracy (100%) during the peak fire season (June–July), outperforming SVC, LR, and KNN, each of which recorded 98%. The authors also observed that variables like wind speed contributed minimally, reinforcing the need for region-specific features in fire prediction. Although GBC was not part of the study, the findings strongly support the use of ensemble and tree-based methods for regionally adapted fire forecasting, particularly within U.S. contexts.

In another effort to improve prediction through model integration, the authors in [5] employed an ensemble-based soft voting strategy combining DT, KNN, and LR to map wildfire susceptibility in Iran's Alborz Mountains. Using MODIS thermal anomaly data and a GPS-corrected fire inventory, the study incorporated 17 variables across anthropogenic, vegetation, topographic, climatic, and hydrological domains. The ensemble model achieved an average AUC of 88%, peaking at 93% in one-fold during 10-fold cross-validation, surpassing the performance of each individual base classifier. The generated susceptibility map classified the landscape into five risk zones, revealing that 21% of the area was at high or very high risk—correlating well with historical fire records. The study underscored the benefits of ensemble learning for improving accuracy and robustness, and suggested that integrating more advanced models like RF or GBC into such frameworks could further improve adaptability across diverse USA terrains.

Expanding the geographical scope, the authors in [6] conducted a large-scale comparative study involving more than 1.04 million fire events from the USA (1992– 2015) and 517 cases from Portugal (2000–2003). The dataset featured a wide range of spatial, temporal, and environmental variables. A variety of models—DL, DT, SGD, ExGBT, and LR—were evaluated for wildfire size

classification, with results showing accuracy ranging from 80% to 82%. DT and ExGBT outperformed others, while GA was employed to derive symbolic representations of wildfire behavior, producing correlation coefficients above 0.80. To enhance balance and interpretability, SMOTE was used to address class imbalance, and SHAP values revealed temperature and weather indices as critical predictive factors. The study demonstrates the value of combining performance-focused models with interpretable AI techniques, especially when handling large, complex wildfire datasets like those found in the U.S.

On a global scale, in [7], the authors used high-resolution (0.25°) global data from 2015 to evaluate wildfire susceptibility based on meteorological variables, fire weather indices, and anthropogenic influences. Models assessed included RF, XGB, and MLP, benchmarked against traditional LR and linear regression. The XGB model yielded the highest performance with an AUC of 97% for wildfire occurrence and a MAE of 3.13 km² for burned area prediction. SMOTE and class-weighted loss functions were used to mitigate data imbalance, while SHAP analysis identified key variables such as historical fire activity, relative humidity, and precipitation.

Although the study aimed for global applicability, regional analysis showed that ML models performed better in Africa and Asia, while in North America, traditional fire indices remained relevant. These findings reinforce the effectiveness of ensemble and deep learning models like XGB and MLP, particularly in high-dimensional, data-rich environments such as the U.S.

In the context of localized prediction, in [8], the authors applied several ML models to Greece's Attica basin, using a custom dataset with 12 meteorological features including temperature, humidity, wind, and rainfall. The study explored binary classification (fire/no fire), multiclass classification (fire severity), and regression (burned area prediction). Among the tested models—RF, XGB, KNN, NN, SVM, LR, and DT—RF performed best for binary classification with 70% accuracy using all features, XGB was most effective with a reduced four-feature set (67.4% accuracy), and KNN achieved the highest R² score of 70% for regression. Validation against the Montesinho dataset supported the generalizability of the approach, suggesting its adaptability to fire-prone regions in the USA.

Similarly, the authors in [9] proposed an ML-driven prediction framework utilizing meteorological variables and FWI data from Portugal's Montesinho Park. The study tested RF, SVM, GBC, LR, and K-means, using stepwise regression and backward elimination for feature selection. Temperature and humidity were identified as the most influential features. SVM and RF performed best

in estimating burned areas. While regression performance was modest (R² = 14%), clustering via K-means (optimized with the elbow method) allowed for localized fire risk assessment. The authors emphasized the value of incorporating spatial and climatic diversity into prediction models—especially relevant to U.S. regions like California and the Pacific Northwest—and suggested further improvements including vegetation types, forest density, and ignition source modeling.

Building on the comparison of classifiers, in [10], the authors evaluated the performance of RF, SVM, DT, and NB and identified RF as the most accurate model for wildfire forecasting. Their findings highlight RF's reliability in supporting early warning and fire response efforts. Similarly, in [11], the authors affirmed RF as the top-performing algorithm among the same set, emphasizing its critical role in risk reduction strategies.

The reviewed literature reflects the increasing reliance on advanced ML techniques for wildfire prediction and classification, particularly ensemble and tree-based models such as RF, XGB, LGBM, CatBoost, DT, GBC, and AdaBoost. These models consistently outperform traditional approaches like LR and linear regression, especially when combined with strategies such as SMOTE, correlation-based and stepwise feature selection, and SHAP for model interpretability. Other algorithms including SVM, KNN, GNB, SGD, MLP, NN, and GA have also demonstrated strong performance in specific tasks, such as burned area regression and symbolic modeling. Unsupervised methods like K-means have been effectively used for spatial clustering and localized risk assessment. The studies emphasize the importance of regional and temporal adaptation, the integration of spatial and environmental data, and handling class imbalance. Although challenges remain in accurately modeling fire extent, ensemble and hybrid methods show strong potential. Overall, the literature confirms the adaptability and scalability of a wide array of ML models for wildfire forecasting across the diverse climatic zones of the U.S.

3. Research Methodology and approach

3.1. Background of the Research Study

This research was conducted using the Google Collab platform as the primary workspace, with Scikit-learn serving as the main Python library for implementing machine learning models. A total of six algorithms—DT, RF, SVC, KNN, LR, and GBC—were employed to explore and analyze the dataset. The study adopted the CRISP-DM methodology, a widely accepted framework for machine learning projects. This methodology comprises six essential phases: identifying the project goals (business understanding), examining the dataset (data understanding), preparing the data for analysis (data

preparation), building and optimizing models (modeling), evaluating the performance of those models (evaluation), and making the model ready for real-world use (deployment) [2]. Utilizing this structured approach ensured clarity and efficiency throughout the process, contributing to the reliable and accurate results illustrated in Figure 1.
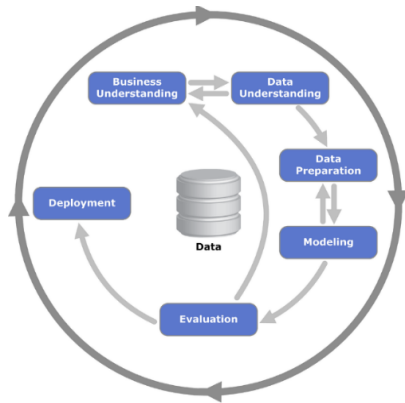

Figure 1: Phases of the CRISP-DM Methodology.

3.2. Dataset Description

The data set used in this study contains records of fire incidents detected across the USA during the year 2021. These observations were captured by the VIIRS sensor on board the Suomi National Polar-orbiting Partnership (SNPP) satellite and obtained through the NASA Open Data Portal [12]. This open-access platform provides researchers with dependable, high-resolution datasets crucial for advancing studies in renewable energy and enhancing grid management strategies. It delivers comprehensive information on solar radiation, meteorological variables, and atmospheric conditions, which are instrumental in building precise energy forecasting models and tackling the unpredictability inherent in renewable energy systems. Furthermore, the platform supports sophisticated simulations and machine learning applications, contributing to more accurate predictive analytics and improved grid efficiency. Its commitment to open data access fosters cross-disciplinary research and innovation, establishing it as a vital resource for environmental and energy research communities [12].

The dataset includes 661,058 records, comprising 360,993 nighttime and 300,065 daytime entries. It features eight input variables and one categorical target variable, which classifies fire events into four categories: Type 0 (presumed vegetation fires), Type 1 (active volcanic activity), Type 2 (fires from stationary land-based sources), and Type 3 (offshore fire detections over water bodies).

This classification framework underscores the dataset's emphasis on distinguishing between different fire origins and behaviors [12]. A summary of the dataset's attributes is provided in Table 1.

Table 1: Dataset Description

Attribute	Definition	Datatypes
Bright_ti4	Measures the brightness temperature in Band 4 of the thermal infrared spectrum (TIR).	Float64
Bright_ti5	Measures the brightness temperature in Band 5 of the TIR.	Float64
Scan	Measures the satellite's scanning ability, including angle, direction, and spatial coverage.	Float64
Track	Describes the satellite's orbital path, alongside its current location and trajectory.	Float64
FRP	Fire radiative power (MW).	Float64
Latitude	Fire pixel latitude(degree).	Float64
Longitude	Fire pixel longitude (degree).	Float64
Day-night	Uses the solar zenith angle (SZA) to determine whether conditions are day or night.	Object
Type	Type attributed to thermal anomaly.	Object

3.3. Dataset Preparation

Following the data exploration phase, the preparation of the dataset is initiated. This stage involves multiple preprocessing steps, including managing missing values, removing duplicate entries, applying normalization techniques, selecting relevant features, encoding categorical variables, and dividing the data into training and testing sets. These steps are essential to ensure the dataset is clean, structured, and ready for effective modeling and further analytical procedures.

3.3.1. Missing Data

To verify the integrity of the dataset, two standard functions were employed: isnull().sum() and duplicated().sum() [13]. The isnull(). sum() function is used to detect and count any missing values across the dataset columns, while duplicated().sum() identifies repeated rows that could compromise data quality. The execution of these checks revealed that the dataset contained neither missing values nor duplicate entries. This confirmation of data completeness and consistency contributes to improved data quality, which is critical for building accurate and reliable machine learning models.

3.3.2. Balancing the Dataset

The distribution of fire types in the dataset reveals a significant imbalance, with Type 0 (presumed vegetation fires) dominating at 86.88% of the total records. In contrast, the other categories are considerably less represented, especially Type 1 (active volcano), which constitutes only 0.10%. To address this disparity and enhance the performance of machine learning models across all classes, the dataset was balanced using the Synthetic Minority Over-sampling Technique (SMOTE) technique prior to training. SMOTE is a popular technique used in imbalanced classification problems to

help balance the dataset by generating synthetic data points for the minority class [14].

3.3.3. Encoding Categorical Data

The dataset underwent label encoding to transform categorical variables into numeric format, an essential preprocessing step since most machine learning algorithms require numerical input [15].

In this study, fire incidents were categorized according to their type: Type 0 representing presumed vegetation fires, Type 1 indicating volcanic activity, Type 2 referring to stationary land-based fires, and Type 3 covering offshore fire detections over water. This conversion was vital to ensure the data was compatible with the classification models, thereby improving the effectiveness and accuracy of the training process.

3.3.4. Splitting Data

Initially, the dataset was split into two parts: 80% for training and 20% for testing. This division allows the model to learn from the majority of the data while reserving a portion for evaluating its performance on unseen examples.

3.3.5. Data Normalization

The numerical features bright_ti4, bright_ti5, scan, track, and frp were normalized to bring their values within a consistent range, such as 0 to 1 or -1 to 1 [16]. This scaling process ensures that each feature contributes equally during model training, preventing any one variable from disproportionately influencing the learning process and supporting more balanced, unbiased model performance.

3.4. Modelling

Six machine learning algorithms—DT, RF, SVC, KNN, LR, and GBC—were implemented to classify the fire types.

Decision Tree (DT) is a non-parametric learning method that uses a tree-like structure to make decisions based on feature thresholds. It recursively splits the dataset into subsets based on the most significant feature at each node, making it interpretable and efficient for handling both categorical and numerical data. However, it is prone to overfitting, particularly on noisy datasets [15].

Random Forest (RF) is an ensemble learning technique that builds multiple decision trees during training and merges their outputs for improved accuracy and robustness. By averaging the results (in classification, via majority voting), RF reduces overfitting and variance compared to individual trees, offering better generalization on unseen data [15].

K-Nearest Neighbors (KNN) is a simple, instance-based learning algorithm that classifies data points based on the majority label among their k-nearest neighbors in the feature space. Though computationally intensive during prediction, KNN is intuitive and works well with non-linear data distributions when appropriate distance metrics and normalization are applied [17].

Logistic Regression (LR) is a statistical model that uses the logistic function to model the probability of a binary or multiclass outcome. Despite its simplicity, LR is a strong baseline model due to its efficiency, interpretability, and solid performance in linearly separable problems [18].

Gradient Boosting Classifier (GBC) is a powerful ensemble method that builds models sequentially, where each new model attempts to correct the errors made by the previous ones. It combines weak learners (typically shallow trees) using gradient descent optimization to minimize the loss function, achieving high predictive accuracy at the cost of increased training time [16].

Support Vector Classifier (SVC) is based on the principles of Support Vector Machines (SVM). It attempts to find the optimal hyperplane that best separates the data into distinct classes by maximizing the margin between support vectors. SVC is especially effective in high-dimensional spaces and is robust to overfitting when the kernel and regularization parameters are properly selected [19] .

3.5. Performance Evaluation

The effectiveness of the supervised machine learning models is evaluated using key performance metrics, including accuracy, recall, F-measure and precision, which collectively provide insight into their classification performance.

3.5.1. Accuracy

It represents the proportion of correctly predicted instances out of the total number of predictions made. It reflects the overall effectiveness of a model in classifying both positive and negative cases correctly shown in equation (1) [15].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

3.5.2. F-measure

It offers a balanced assessment by combining both metrics into a single value, especially useful when the data is imbalanced or when equal consideration of false positives and false negatives is needed shown in equation (2) [15].

$$F-measure = 2 \times \frac{precision \times recall}{precision + recall} \quad (2)$$

3.5.3. Precision

It measures the ratio of correctly predicted positive instances to the total predicted positives. It indicates how many of the instances labeled as positive by the model are actually relevant, helping to evaluate the model's reliability in making positive predictions shown in equation (3) [17].

$$Precision = \frac{TP}{TP + FP} \qquad (3)$$

3.5.4. Recall

It refers to the proportion of actual positive cases that are correctly identified by the model. It is particularly important in situations where missing positive cases is costly or undesirable shown in equation (4) [18].

$$Recall = \frac{TP}{TP + FN} \qquad (4)$$

4. Results

In terms of accuracy, DT attains the top performance with 96.69%, closely followed by RF at 96.37%, both demonstrating strong capabilities in correctly identifying fire types. GBC also delivers notable accuracy at 93.16%, with KNN achieving 91.27%. On the other hand, SVC and LR register comparatively lower accuracy rates of 88.35% and 87.58%, respectively, suggesting relatively less effective classification results, as illustrated in Table 2 and Figure 2.

Looking at precision, DT again leads with 96.70%, indicating a high level of accuracy in its positive predictions and a minimal rate of false positives. RF follows closely with a precision of 96.31%, while GBC achieves 92.76%, both reflecting reliable classification outputs. KNN also shows solid results with 90.57%, whereas SVC and LR lag behind at 83.61% and 83.65%, respectively, highlighting a greater occurrence of incorrect positive classifications.

Regarding recall, which assesses the ability to correctly identify actual fire instances, DT maintains its lead at 96.69%, with RF slightly behind at 96.37%. GBC continues to perform well with 93.16%, while KNN records 91.27%. In contrast, SVC and LR exhibit lower recall rates of 88.35% and 87.58%, indicating a higher chance of failing to detect true fire occurrences.

When considering the F1 score, which harmonizes precision and recall into a single performance metric, DT secures the highest value at 96.67%, confirming its balanced and robust classification ability. RF follows with an F1 score of 96.19%, and GBC reaches 92.67%. KNN also maintains dependable performance with 90.79%. Meanwhile, SVC and LR yield lower F1 scores of 85.50% and 84.71%, respectively, indicating limitations in managing the trade-off between precision and recall.

Table 2: Performance Comparison between models.

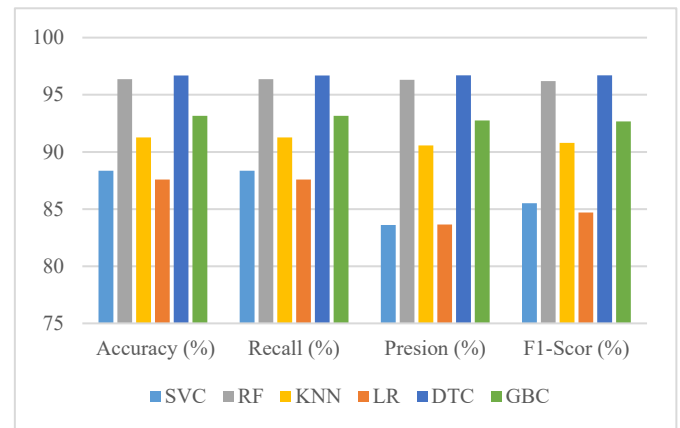| Model	Accuracy (%)	Recall (%)	Presion (%)	F1-Scor (%)
SVC	88.35	88.35	83.61	85.50
RF	96.37	96.37	96.31	96.19
KNN	91.27	91.27	90.57	90.79
LR	87.58	87.58	83.65	84.71
DTC	96.69	96.69	96.70	96.70
GBC	93.16	93.16	92.76	92.67

Figure 2: Performance Plot of Proposed Models

5. Discussion

The findings of the current study, which evaluates six supervised ML models—DT, RF, GBC, KNN, SVC, and LR—for fire type classification, align well with trends observed in the reviewed literature while also offering noteworthy advancements in model performance and application specificity.

In this study, DT achieved the highest accuracy (96.69%), precision (96.70%), recall (96.69%), and F1 score (96.67%), outperforming other models. These results are consistent with the findings of Khosravi et al., who reported perfect classification accuracy for DT and RF during peak fire seasons in Algeria, confirming the effectiveness of tree-based models in wildfire classification tasks. Similarly, RF performed robustly across all metrics in the current study—attaining 96.37% accuracy and 96.19% F1 score—which echoes its dominant position in several previous studies, including those by Tavakoli, Barzani et al., and Al-Bashiti & Naser, where RF either matched or exceeded other ensemble models in terms of predictive accuracy and interpretability.

GBC also demonstrated strong performance in this work, with consistent results across accuracy (93.16%), precision (92.76%), recall (93.16%), and F1 score (92.67%). While GBC was not explicitly evaluated in some past works such as those by Khosravi et al., its potential was highlighted in Chaubey et al. and Alkhatib et al., who supported the integration of ensemble models to improve classification reliability—particularly when using complex and high-dimensional environmental data.

KNN, although not an ensemble method, delivered solid results (accuracy: 91.27%, F1 score: 90.79%), which aligns with Stafylas Demetrios' regression-based analysis, where KNN showed competitive performance in predicting burned area. However, KNN remains sensitive to feature scaling and may not capture complex decision boundaries as effectively as tree-based models, which is reflected in its slightly lower scores compared to DT, RF, and GBC.In contrast, SVC and LR showed the weakest performance across all metrics. SVC recorded 88.35% accuracy and 85.50% F1 score, while LR followed closely behind with 87.58% accuracy and 84.71% F1 score. These outcomes are consistent with earlier studies, such as those by Al-Bashiti and Naser, where LR underperformed relative to ensemble and tree-based models, and by Shmuel and Heifetz, who showed that while traditional models like LR offer baseline predictability, they fall short in handling the nonlinear and complex nature of wildfire dynamics.

Another important point of comparison is how well the current study addresses model balance. Unlike some previous works that focused on peak fire seasons or lacked formal imbalance-handling strategies, this study ensured an equal class distribution prior to training, which likely contributed to the high and consistent scores for DT, RF, and GBC across all evaluation metrics. This balanced approach strengthens the reliability and generalizability of the findings, especially for real-world applications in USA fire forecasting, where underrepresented classes often challenge prediction accuracy.

Furthermore, this study's comparative framework adds value by using a unified dataset and standardized preprocessing, enabling a fair and direct performance comparison. While prior literature often evaluated models on region-specific or task-specific datasets (e.g., ignition, size, burned area), this study provides a focused comparison on fire type classification, offering insights particularly useful for U.S.-based fire management systems aiming for categorical fire event identification.

6. Conclusion and Future Directions

This study assessed the effectiveness of six supervised machine learning algorithms—DT, RF, GBC, KNN, SVC, and LR—in classifying fire types in the United States using satellite-derived data. Among the evaluated models, DT consistently achieved the best results, recording the highest scores in accuracy (96.69%), precision (96.70%), recall (96.69%), and F1 score (96.67%). RF closely followed, while GBC also demonstrated strong and balanced performance across all metrics. In contrast, SVC and LR exhibited comparatively lower predictive capabilities, highlighting their limitations in capturing the complex, nonlinear patterns characteristic of fire behavior.

These findings align with previous research, where tree-based and ensemble models—particularly DT, RF, and XGB—have repeatedly proven effective in wildfire prediction. Their success can be attributed to several key strengths. First, these models are well-suited to capturing nonlinear interactions among environmental variables such as temperature, humidity, wind, and vegetation, which are critical in fire dynamics. Second, they effectively manage heterogeneous and high-dimensional datasets, including those combining meteorological indices, satellite imagery, and geospatial information. Third, they demonstrate robustness to noise, missing values, and outliers, enabling more reliable predictions in real-world conditions.

Moreover, ensemble methods such as RF and XGB offer enhanced generalization through the aggregation of multiple decision paths, thereby reducing the risk of overfitting. These models also support model interpretability through feature importance rankings and SHAP analysis, providing valuable insights into the most influential factors driving fire classifications—an essential feature for transparent and accountable decision-making in wildfire management systems.

By applying a balanced dataset and a standardized evaluation framework, this study provides a robust comparison of model performance, contributing novel insights to the evolving field of ML-driven wildfire forecasting. The findings reaffirm that tree-based and ensemble algorithms are not only highly accurate but also scalable, flexible, and interpretable, making them particularly well-suited for operational deployment in real-world fire risk management applications—especially across the diverse climatic and ecological regions of the USA.

Looking forward, future research should explore the integration of real-time meteorological feeds, higher-resolution spatial data, and advanced ensemble strategies such as model stacking and hybrid architectures. Additionally, incorporating deep learning techniques and spatiotemporal modeling could further enhance predictive precision, enabling more dynamic and proactive wildfire forecasting systems capable of addressing both localized threats and broader regional patterns.

References

[1] A. Modaresi Rad et al., "Human and infrastructure exposure to large wildfires in the United States," Nature Sustainability, vol. 6, no. 11, pp. 1343-1351, 2023, doi:10.1038/s41893-023-01163-z.

[2] S. P. H. Boroujeni et al., "A comprehensive survey of research towards AI-enabled unmanned aerial systems in pre-, active-, and post-wildfire management," Information Fusion, p. 102369, 2024, doi:org/10.1016/j.inffus.2024.102369.

[3] F. Tavakoli, "Dataset Creation and Imbalance Mitigation in Big Data: Enhancing Machine Learning Models for Forest Fire Prediction," University of Waterloo, 2023, http://hdl.handle.net/10012/20046.

[4] H. Khosravi, M. R. Shafie, A. S. Raihan, M. A. B. Syed, and I. Ahmed, "Optimizing Forest Fire Prediction: A Comparative Analysis of machine learning models through feature selection and time-stage evaluation," Preprints. org, 2023, doi: 10.20944/preprints202312.0577.v1

[5] A. Rezaei Barzani, P. Pahlavani, and O. Ghorbanzadeh, "Ensembling of decision trees, KNN, and logistic regression with soft-voting method for wildfire susceptibility mapping," ISPRS Annals of the Photogrammetry, Remote Sensing Spatial Information Sciences, vol. 10, pp. 647-652, 2023, doi:10.5194/isprs-annals-X-4-W1-2022-647-2023, 2023.

[6] M. K. Al-Bashiti and M. Naser, "Machine learning for wildfire classification: Exploring blackbox, eXplainable, symbolic, and SMOTE methods," Natural Hazards Research, vol. 2, no. 3, pp. 154-165, 2022, doi:10.1016/j.nhres.2022.08.001.

[7] A. Shmuel and E. Heifetz, "Global wildfire susceptibility mapping based on machine learning models," Forests, vol. 13, no. 7, p. 1050, 2022, doi:10.3390/f13071050.

[8] D. Stafylas, "Wildfire prediction using machine learning," M.S. thesis, University of West Attica, 2022.

[9] T. Preeti, S. Kanakaraddi, A. Beelagi, S. Malagi, and A. Sudi, "Forest fire prediction using machine learning techniques," in 2021 International Conference on Intelligent Technologies (CONIT), 2021, pp. 1-6: IEEE, DOI:10.1109/CONIT51480.2021.9498448.

[10] R. Alkhatib, W. Sahwan, A. Alkhatieb, and B. Schütt, "A brief review of machine learning algorithms in forest fires science," Applied Sciences, vol. 13, no. 14, p. 8275, 2023, doi:10.3390/app13148275.

[11] F. N. Ismail, B. J. Woodford, S. A. Licorish, and A. D. Miller, "An assessment of existing wildfire danger indices in comparison to one-class machine learning models," Natural Hazards, pp. 1-32, 2024, doi:10.1007/s11069-024-06738-3.

[12] " NASA Open Data Portal https://data.nasa.gov/browse " 2021.

[13] S. Alshakrani, R. Taha, and N. Hewahi, "Chronic kidney disease classification using machine learning classifiers," in 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2021, pp. 516-519: IEEE, doi: 10.1109/3ICT53449.2021.9581345.

[14] G. A. Pradipta, R. Wardoyo, A. Musdholifah, I. N. H. Sanjaya, and M. Ismail, "SMOTE for handling imbalanced data problem: A review," in 2021 sixth international conference on informatics and computing (ICIC), 2021, pp. 1-8: IEEE, doi: 10.1109/ICIC54025.2021.9632912.

[15] F. A. Musleh and R. G. Taha, "Forecasting of forest fires using machine learning techniques: a comparative study," in 6th Smart Cities Symposium (SCS 2022), 2022, vol. 2022, pp. 337-342: IET, doi: 10.1049/icp.2023.0571.

[16] F. Ahmad Musleh, "A Comprehensive Comparative Study of Machine Learning Algorithms for Water Potability Classification," International Journal of Computing Digital Systems, vol. 15, no. 1, pp. 1189-1200, 2024, doi: 10.1049/icp.2023.0571.

[17] R. Taha, S. Alshakrani, and N. Hewahi, "Exploring Machine Learning Classifiers for Medical Datasets," in 2021 International Conference on Data Analytics for Business and Industry (ICDABI), 2021, pp. 255-259: IEEE, doi: 10.1109/ICDABI53623.2021.9655862.

[18] A. M. Zeki, R. Taha, and S. Alshakrani, "Developing a predictive model for diabetes using data mining techniques," in 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2021, pp. 24-28: IEEE, doi: 10.1109/3ICT53449.2021.9582114.

[19] F. A. Musleh, "A comparative study to forecast the total nitrogen effluent concentration in a wastewater treatment plant using machine learning techniques," International Journal of Computing Digital Systems, vol. 14, no. 1, pp. 10447-10456, 2023.

Mrs. Ranyah Taha completed her MSc in Big Data Science and Analytics in 2022 through a joint program between Liverpool John Moores University and the University of Bahrain. She earned her BSc in Computer Science from the University of Bahrain in 2018. Her research focuses on leveraging Data Science and Analytics, particularly Machine Learning and Deep Learning, to build advanced models and extract valuable insights from complex datasets. She has contributed to many research papers and was awarded the NASA International Space Apps Challenge – Space Apps Bahrain 2023 Local Impact Award.

Dr. Fuad Musleh has received his Ph.D. and M.Sc. degrees from the University of Alabama in Huntsville, and his B.Sc. from Jordan University of Science and Technology in Jordan. He is currently serving as an Assistant Professor at the University of Bahrain. His research interests include environmental and water resource engineering, particularly in vegetation–flow interaction and environmental conservation. He also focuses on data analysis applications in environmental systems and sustainability science.

Mr. Abdel Rahman Musleh is currently a senior undergraduate student pursuing a B.Sc. in Electrical Engineering at the University of Bahrain. His academic focus lies in developing cyber-physical systems that integrate artificial intelligence (AI) and machine learning (ML) to enhance system efficiency, reliability, and sustainability.His interests include the application of ML in renewable energy systems and smart grid infrastructure, with a growing involvement in research related to intelligent automation and real-time simulations.