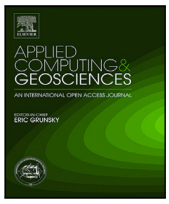
Applied Computing and Geosciences 27 (2025) 100266

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences



Recent advances in explainable Machine Learning models for wildfire prediction

Abira Sengupta *, Brendon J. Woodford

School of Computing, University of Otago, PO Box 56, Dunedin, 9054, New Zealand

ARTICLE INFO

Keywords:
Forest fires
Machine learning
Hyper-parameter optimisation
SHAP values
Explainable artificial intelligence

ABSTRACT

Climate change has caused increasingly frequent occurrences of forest fires around the world. Machine Learning (ML) and Artificial Intelligence models have emerged to predict both the onset of wildfires and evaluate the extent of damage a wildfire would cause. However, understanding what factors lead to generating models that exhibit optimal performance and providing insight into the importance of features on model outcomes is the subject of ongoing research. To help answer these questions, we propose a framework which adopts recent advances in methods for obtaining optimal models along with the application of SHAP (SHapley Additive exPlanations) values to obtain the most important features which affect the performance of wildfire prediction models. We use this framework as a classification task to predict the likelihood of wildfire occurrence based on environmental conditions, using a data set which represents instances of forest fires in Algerian, and as a regression task to predict the burned area once a wildfire has begun, using a data set from Portugal that recorded the area burned after a fire event. Insights provided by this framework allow us to assess the efficacy of specific ML models for wildfire prediction, ultimately making recommendations as to which ML models are more suited towards these challenging tasks.

1. Introduction

Wildfires are causing significant damage to both property and lives. The severity of wildfires is increasing due to numerous factors such as rising temperatures, reduced rainfall, more severe and widespread droughts (Dai, 2013) regional droughts are also correlated with sea surface temperature variations (Shabbar and Skinner, 2004), regional humidity variation (Dessler et al., 2008), a rise in dry vegetation, increasing greenhouse gas levels, and extreme weather conditions. Global warming and climate change issues are major factors in wildfires. Global temperature has increased by ≈2 °C per decade over three decades (Hansen et al., 2010). Recently, summertime drought brought on by prolonged water stress due to global warming has increased forest land dryness and, in turn, the frequency of wildfires (Tyukavina et al., 2022).

In turn, these fires produce gas emissions, which increase carbon dioxide and reduce carbon capture through vegetation loss and ultimately global warming. Besides the natural factors, anthropogenic factors, such as the activities of local people and tourism are also factors of wildfires (Martínez et al., 2009). It is challenging for environmental organisations to predict wildfires due to both natural and man-made wildfire conditioning factors.

The California Air Resources Board estimates that the state's wildfires in 2020 emitted around 112 million metric tonnes of carbon dioxide (California Air Resources Board, 2024). According to the European Union's Copernicus Atmosphere Monitoring Service, wildfires produced 1.76 billion tonnes of carbon worldwide in 2021 (The European Space Agency, 2021). Wildfires have a wide-ranging economic and social impact on local people. Wildfires raise the air's concentration of dangerous gases and particulate matter, which are directly associated with a number of diseases like bronchitis, asthma, and reduced lung volume. People also suffer from the consequences of wildfires, such as mortality, homelessness, displacement, psychological stress, etc.

Therefore, systematic research on prediction approaches for wildfires and fire extents in this area is required. Several studies have been conducted in recent years to detect and predict wildfires. There are three high-tech methodologies that are commonly used to predict wildfire occurrence: physical-based methods, statistical methods, and machine-learning methods.

Physics-based methods use mathematical formulas that rely on surface fuels and heat transport principles to spread fires (Mell et al., 2007). This method makes use of fire behaviour software like FireStation (Lopes et al., 2002), which enables the identification of crucial

* Corresponding author.
E-mail address: sengupta.abira0609@gmail.com (A. Sengupta).

https://doi.org/10.1016/j.acags.2025.100266
Received 31 August 2024; Received in revised form 18 June 2025; Accepted 21 July 2025
Available online 30 July 2025

regions that could experience extremely high fire behaviour. Physics-based models have some limitations: they typically produce results with extremely poor accuracy, exhibit prediction bias in the areas in which they are intended for usage, and are frequently challenging to develop and apply since they necessitate a high number of expert rules. In addition, complicated data formats and high resource requirements frequently make simulating complex environmental factors challenging. In addition to requiring extensive parameters on a variety of inputs and a high degree of software utilisation, forecasts from these systems rely significantly on analysis assumptions, which makes them computationally expensive and difficult to set up.

The second approach is statistical methods, which can also be applied to model large/spatial areas while overcoming the simulation complexity. In addition, these methods can benefit from modern technologies and can be applied at different scales and resolutions (Goodchild, 1991; Finney et al., 2009). However, these methods could be sensitive to the type of analysed data and may require numerical manipulation to satisfy convergence criteria.

The third approach is to use Machine Learning (ML), a field of Artificial Intelligence (AI) that is popular due to its ability to handle complex and high-dimensional data, as well as its scalability and cost-effectiveness. The application, modification, or development of ML methods is made to understand the complex relationship among multiple variables related to wildfires. Wildfires have been assessed using diverse methods, including Support Vector Machine (SVM) (Jaafari and Pourghasemi, 2019), Decision Trees (DT) (Abid and Izeboudjen, 2020), Random Forest (RF) (Gibson et al., 2020), Artificial Neural Network (ANN) (Liang et al., 2019), Logistic regression (LR) (de Bem et al., 2018), and Convolutional Neural Networks (CNN) (Bhatt and Chouhan, 2024). In most of these cases, the proposal of new methodologies involved the empirical comparison of the performance of the models when applied to wildfire prediction. However, less attention has been paid to efficient methods for establishing optimal Hyper-Parameter (HP) values for model generation and assessing the importance of these HPs on model learning.

To this end, there have been recent advances in methods for obtaining optimal HPs for model learning and for assessing the importance of how the choice of values for the HPs influences ML model performance. For example, Watanabe et al. (2023) proposed a new method called Pearson Divergence-ANOVA (PED-ANOVA) as a computationally efficient means of quantifying HP importance. Furthermore, investigating recent methods for providing insight into the importance of features on ML model output, such as Shapley (SHAP) values (Lundberg and Lee, 2017) would enhance the interpretation of these models. Recent work by Kondylatos et al. (2022) has shown the value of SHAP values for the purpose of explaining which variables are important for driving the predictions of wildfire models. Therefore, the main goal of this study is threefold:

(1) We propose a framework that adopts a well-known HP tuning method to obtain the values required for the optimal performance of an ML model. To this end, we employ the Optuna Hyper-Parameter Optimisation (HPO) framework (Akiba et al., 2019), which uses the PED-ANOVA algorithm developed by Watanabe et al. (2023) to help us not only obtain optimal ML models but also provide insight into the contribution of each hyper-parameter in ML model learning for both classification and regression tasks.

(2) The extraction of SHAP values (Lundberg and Lee, 2017) from the trained ML models to determine the importance of the features from wildfire data sets to understand which features have the most impact on ML model output across multiple ML models.

(3) A statistical analysis of the performance of these models to establish which ML models are better suited for different types of wildfire analysis tasks.

We apply our framework for the detection of forest fires, using classification tasks to detect fire occurrences based on environmental conditions and regression tasks to predict the burned area of forest fires once they have occurred. The data set used for the detection of wildfires was the Algerian data set (June 2012 - September 2012) (Abid and Izeboudjen, 2020), while the Portugal data set (January 2000 - December 2003) (Cortez and Morais, 2007) for predicting the burned area due to forest fires.

2. Materials and methods

2.1. Description of the two data sets

2.1.1. Algerian forest fire dataset

For this case study, we use two publicly available data sets. For classification purposes, we used a dataset that combined data from two Algerian regions: Sidi Bel-abbas in the northwestern part of the country and Bejaia in the northeast. Although meteorological observations from 2007 to 2018 were acquired, only those observations for the summer of 2012, specifically from June to September, were included. This was because the likelihood of a fire was highest during this period, and 2012 was the year with the greatest number of reported fire occurrences between 2007 and 2018 (Abid and Izeboudjen, 2020). Therefore, only the data for 2012 were made available by the authors of that study. The resulting data set consisted of 244 instances.

Table 1 details ten of the thirteen features in the Algerian data set. Features that relate to the day, month, and year the data instance was acquired but were omitted. Of note are the features related to the Forest Fire Weather Index (FWI), which is a Canadian system for assessing fire danger (Stocks et al., 1989). It consists of six components: (1) Fine Fuel Moisture Code (FFMC) that refers to the moisture content of surface litter, which affects ignition and fire spread, (2) Duff Moisture Code (DMC) that represents the moisture content of shallow organic layers, influencing fire intensity, (3) Drought Code (DC) represents the moisture content of deep organic layers, impacting fire intensity, (4) The Initial Spread Index (ISI) measures fire velocity and spread, (5) Buildup Index (BUI) indicates available fuel, and (6) FWI is an index calculated from the Relative Humidity (RH), Temperature (Temp), level of Rainfall (Rain), and Wind Speed (WS). Fig. 1 shows how these features relate to each other.

The Canadian system for calculating the components of the FWI (Stocks et al., 1989) is applied as follows (Eqs. (1) to (6)):

$$FFMC_t = FFMC_{t-1} + (0.83 \times P - 0.92) \times e^{-0.0345 \times FFMC_{t-1}} + 0.28 \times (T - H) \times (1 - e^{-0.115 \times FFMC_{t-1}}) \quad (1)$$

where P is the Rainfall in mm, T is the temperature in °C and H is the relative humidity measured as a percentage.

$$DMC_t = DMC_{t-1} + (0.5 \times T) - f(P) \quad (2)$$

where $f(P)$ is a Rainfall function that reduces DMC.

$$DC_t = DC_{t-1} + k(T) - g(P) \quad (3)$$

where $k(T)$ represents a long-term drying function and $g(P)$ represents the effect of Rainfall.

$$ISI = 0.208 \times W_s \times e^{(0.05039 \times FFMC)} \quad (4)$$

where W is the wind speed in km/h.

$$BUI = \begin{cases} 0.8 \times DC \times DMC/(DC + 0.4 \times DMC) & \text{if } DMC > 0 \\ DC & \text{if } DMC = 0 \end{cases} \quad (5)$$

$$FWI = e^{(0.1 \times BUI)} \times ISI \quad (6)$$

The last row describes the output column titled 'Classes', denoting categorical values of '*no-fire*' and '*fire*'. Although the FWI system categorises the FWI values into six fire danger classes ranging from *Nil* to *Extreme*, in Abid and Izeboudjen (2020) they considered the FWI values below < 1.0 as '*no-fire*' and all FWI values ≥ 1.0 as '*fire*'. This scheme resulted in 138 instances of the '*fire*' class and 106 instances of the '*no-fire*' class.

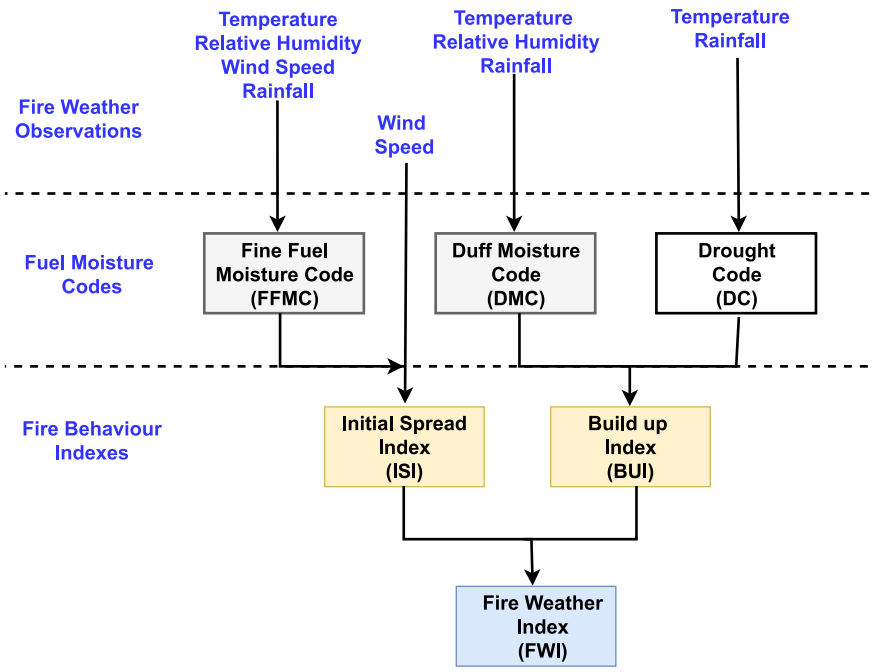


Fig. 1. Structure diagrams for the Fire Weather Index (FWI) System use the Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), and Drought Code (DC) to assess fire ignition, spread, and intensity (Anon, 1993).

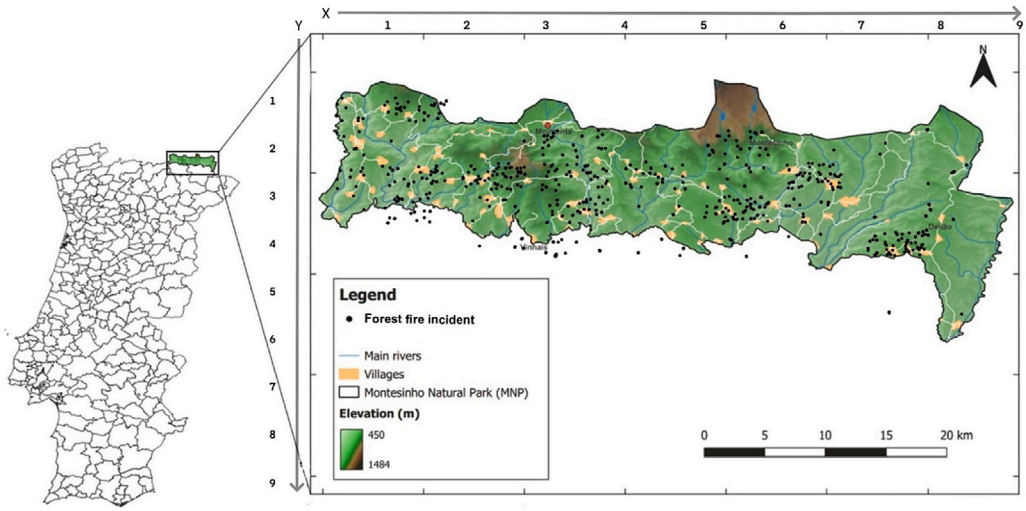


Fig. 2. Montesinho Natural Park, in northeastern Portugal, covers approximately 742 km^2 and is known for its diverse landscapes and high levels of biodiversity. The park's elevations range from 438 metres to 1486 metres at its highest point, Pico Montesinho. Each grid reference represents 5 km^2 of land.
Source: Sourced, reproduced, and adapted from Bhatt and Chouhan (2024, p. 4).

Table 1
Breakdown of features used from the Algerian data set.

Description	Feature	Metric	Range
Fire Weather Index	FWI	FWI system	0–31.1
Fine Fuel Moisture Code	FFMC	FWI system	28.6–92.5
Duff Moisture Code	DMC	FWI system	1.1–65.9
Drought Code	DC	FWI system	6.9–220.4
Initial Spread Index	ISI	FWI system	0–18.5
Buildup Index	BUI	FWI system	1.1–68.0
Temperature at noon	Temp	°C	22–42
Relative Humidity	RH	%	21–90
Wind Speed	WS	km/h	6–29
Rain	Rain	Total mm/m^2	0–16.8
Classes	Class	Categorical	'*no-fire*' or '*fire*'

2.1.2. *Montesinho National Park forest fire dataset*

For predicting the burned area of a forest (Cortez and Morais, 2007), acquired data from the burned area in the Montesinho National Park, from the Trás-os-Montes northeast region of Portugal. Fig. 2 depicts Montesinho National Park. The park has a Mediterranean climate with yearly temperatures between 8 °C and 12 °C, and is home to a wide variety of plants and animals. Areas affected by forest fires are marked with black dots. The data were collected from January 2000 to December 2003 and detailed in Table 2, resulting in 517 data instances. Note that two of the features related to when the data was collected (month and day) are not presented in this table. The Grid Reference X coordinate, or the column index of every 5 km^2 cell from west (1) to east (9), is represented by the *x*-axis labels (1–9). The Grid Reference Y coordinate, or the row index of every 5 km^2 cell from south (1) to north (9), is represented by the *y*-axis labels (1–9).

Table 2
Breakdown of features used from the Montesinho National Park data set.

Description	Feature	Metric	Range
X-axis index	X	Grid reference	X-axis spatial coordinate within the Montesinho park map: 1–9
Y-axis index	Y	Grid reference	Y-axis spatial coordinate within the Montesinho park map: 1–9
Fine Fuel Moisture Code	FFMC	FWI system	18.70–96.20
Duff Moisture Code	DMC	FWI system	1.1–291.3
Drought Code	DC	FWI system	7.9–860.6
Initial Spread Index	ISI	FWI system	0–56.1
Maximum Temperature	Temp	°C	2.2–33.3
Relative Humidity	RH	%	15–100
Wind Speed	Wind	km/h	0.4–9.4
Rain	Rain	Total mm/m^2	0–6.4
Burnt area	Area	Hectares	0–1090.84

2.2. Brief description of each of the ML algorithms

In this section, we present our ML model-based methods for wildfire prediction. ML techniques have made a substantial contribution to the efficient outcomes of classification and prediction systems in recent years. While ML relies solely on inventory data, it is not dependent on expert knowledge. In this study, we adopted six ML algorithms for wildfire prediction: SVM, RF, Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting (LGBM), and Multi-Layer Perceptron (MLP), summarised below:

2.2.1. Support vector machine

Support Vector Machine (SVM), a well-known supervised learning model, and related kernel-based methods have been widely used for both classification and regression tasks. SVM is a classifier that, using N-dimensional data, finds the hyperplane (decision border) in an $N-dimensional$ space that divides the boundary of each class. SVM determines the hyperplane such that the distance between the nearest point of each class and the decision boundary is maximised. The hyperplane is defined as having the form $w^T x + b = 0$, where x is the input vector, b is the bias, and w is the weight vector if the data can be divided by a line. The margin of separation is the distance between the hyperplane and the nearest data point (d), also referred to as a support vector. The goal is to identify the best hyperplane to minimise the margin. If the data are not linearly separable, kernel SVM approaches such as Radial Basis Functions (RBF) utilise a set of transformations to move it to a higher-dimensional space where the hyperplane can be found more easily (Hearst et al., 1998; Tien Bui et al., 2016). In this work, we adopt SVM for both Classification (SVC) and Regression (SVR).

2.2.2. Random forest

Decision Trees (DTs) are another type of universal function approximator that falls within the category of supervised learning techniques (Breiman et al., 2017). DTs can be applied to both classifications where predicted responses are discrete and regression if expected responses are continuous problems.

One of the more popular extensions of the DT is to use bagged DTs with an ensemble model called a Random Forest (RF), which is made up of numerous separately trained DTs (Suthaharan, 2015; Breiman, 2001). In an RF model, each component (DT) chooses which class to classify the input data into; the class with the most votes is the final classification. Regression can also be performed with RFs; the average of the individual-tree outputs gives the final result. The fundamental idea behind the RF technique is that, at each node of each tree, a random subset of features is chosen; bagging, which resamples the initial set of data points, is used to select samples for training each component tree. As the number of trees in a forest increases, the generalisation error converges to a limit. The strength of each individual tree in the forest and the correlation between them determine the generalisation error of a forest of DT classifiers (Suthaharan, 2015; Breiman, 2001).

2.2.3. Adaptive boosting

Among all the theoretically provable boosting techniques, the most successful one in practical applications has been Adaptive Boosting (AdaBoost). It combines the predictions of several weak learners to create a powerful classifier or regressor.

Its success can be attributed to two things: first, it is very simple; second, AdaBoost has a feature called "adaptivity" that other boosting algorithms do not have (Shrestha and Solomatine, 2006; Freund and Schapire, 1997). AdaBoost automatically adapts to the strengths of the weak hypotheses generated by the weak learner. It works by increasing the weight of observations that were previously misclassified. This can, in principle, reduce the classification error, leading to a high level of precision.

2.2.4. Extreme gradient boosting

Extreme Gradient Boosting, or XGBoost, is a scalable and extremely effective gradient-boosting algorithm that is frequently used for ML applications, including regression and classification. XGBoost, as a classifier, is an enhanced implementation of the gradient boosting framework that is optimised for performance and speed (Raghunath et al., 2022). It improves on conventional boosting techniques by maximising model accuracy and computing speed. The XGBoost Classifier (XGBC) constructs an ensemble of decision trees sequentially, with each tree aiming to rectify the mistakes of previous ones. Large-scale classification jobs benefit greatly from their regularisation approaches, which reduce over-fitting. It also has several features, such as support for parallel computation, tree pruning, and handling of missing data. As a regressor, XGBoost is a more advanced version in terms of speed and accuracy. An ensemble of regression trees is constructed by the XGBoost Regressor (XGBR), and each tree is trained to reduce the residual errors of the trees that came before it. It improves generalisationability by introducing regularisation to control model complexity. XGBoost is also very helpful for large-scale regression issues because it has features like missing value handling, tree pruning, and efficient parallel processing. Regression problems in a variety of applications have come to rely on XGBoost due to its adept handling of huge datasets and complex patterns.

2.2.5. Light gradient boosting machine

LightGBM (LGBM) is a gradient learning framework based on decision trees and the concept of boosting; it is a relatively new model. Its primary distinction from the XGBoost model is that it uses a leaf-wise growth approach with depth limitations, speeds up training, and uses histogram-based techniques to minimise memory usage. The histogram algorithm's fundamental concept is to create a histogram with a width of k by discretising continuous floating-point eigenvalues into k bins. The histogram approach does not require additional storing of presorted results, and it can also keep the value after feature discretisation, which is usually sufficient to save with an 8-bit integer, reducing memory consumption to 1/8 of the original. This rough partitioning

does not reduce the accuracy of the model. To ensure high efficiency and avoid over-fitting, LightGBM sets a maximum depth limit at the top of the leaf (Fan et al., 2019).

2.2.6. Multi-layer perceptron

MultiLayer Perceptrons (MLPs) are neural networks that process a single input with multiple independent weights by combining many neurone units in parallel. To accommodate generic functions, additional degrees of freedom can be provided by adding a second layer of hidden neurone units. Simple classification and regression issues can be resolved with MLPs. In a classification work, for example, the output is the expected class for the input data; in a regression work, on the other hand, the output is the regressed value for the input data. MLP can distinguish data that is not linearly separable or separable by a hyperplane. MLP networks are flexible, general-purpose, nonlinear models made up of several units arranged into multiple layers (Khalil Alsmadi et al., 2009).

2.3. Methodology

Our methodology for analysis of the two forest fire data sets is depicted in Fig. 3. Key steps in this methodology are detailed below.

2.3.1. Data pre-processing

Data pre-processing was informed by the type of ML algorithm and the task to which it had been applied. No data normalisation was performed for the XGBC, RF, LGBM, and AdaBoost algorithms, but for the SVM and MLP algorithms the data was standardised to a mean of zero and a standard deviation of one (Hastie et al., 2009) and for the output class of the Algerian data set, the values of '*no-fire*' and '*fire*' were converted to the numerical values of 0 and 1, respectively. In addition, for the Portugal Forest Fire data set, we removed the columns referring to the dates as suggested by Cortez and Morais (2007) to improve model learning and performance and transformed the numerical output values using a logarithmic scale to reduce skewness of the range of values for this column which was also reported by Cortez and Morais (2007). A lead time of one day was used for both classification and prediction due to the limited size of the data sets.

2.3.2. Experimental design

As one objective of this work was to establish the better hyperparameters we adopted the Python Optuna framework (Akiba et al., 2019). We adopted Optuna as it claims to be an agnostic framework that is not tied to any particular ML or deep learning framework. The HPs in question for each model are detailed in Table 3. To obtain a more realistic evaluation of ML model performance given the HPs selected by the Optuna framework, 10-Fold Cross-Validation (CV) was used to assess the test performance of each ML model generated across each trial. The work of Naderpour et al. (2021) and Sarkar et al. (2024) demonstrated the use of 10-Fold CV when evaluating the performance of ML models to predict the risk of wildfire. Stratified 10-Fold CV was adopted for the classification models to reduce class imbalance.

To elaborate, for each type of ML algorithm, we ran it over 80 trials, thus generating 80 10-Fold CV ML models. The mean test Accuracy for classification or the average Root Mean Square Error (RMSE) value for regression obtained from 10-Fold CV was used as the basis for the Tree-structured Parzen Estimator (TPE) HPO method (Bergstra et al., 2013) to adjust the HPs for the next ML model training. Out of these 80 candidate ML models, the best ML model selected was based on the highest average test Accuracy or the lowest average RMSE, depending on if the model was developed for classification or regression, respectively. Performance metrics using 10-Fold CV were obtained from this optimal ML model of Accuracy, Precision, Recall, and F1-Score values (Tharwat, 2021) for classification. RMSE and Mean Absolute Error (MAE) were adopted as performance metrics for prediction. These performance metrics are more appropriate for assessing prediction performance for a wider range of ML models than the R^2 metric, which is only applicable for linear regression, as evidenced by Bhatt and Chouhan (2024) in their study.

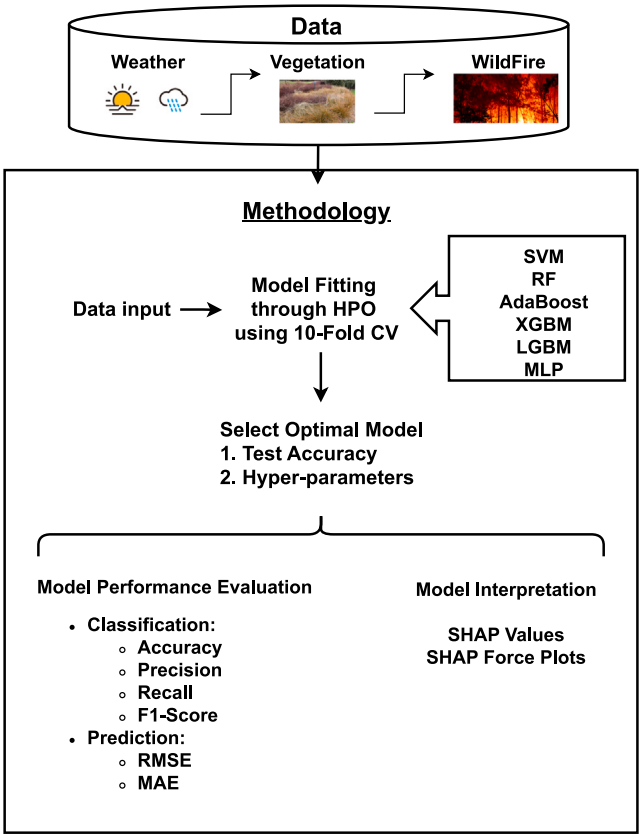


Fig. 3. Methodology used for the analysis of wildfire data. The approach uses a 10-fold cross-validation (CV) technique to fit the model using Hyperparameter Optimisation (HPO). Based on its capacity to balance performance across the dataset, the best model is chosen. The model's accuracy is then tested, and the most important hyperparameters are fine-tuned to improve predictions. In order to understand wildfire behaviour and the factors that contribute to it, the analysis consists of both classification and regression tasks, with evaluation metrics emphasising model performance.

All experiments used the MLP, SVM, and AdaBoost algorithms from the Python `scikit-learn` library (Pedregosa et al., 2011). The Python XGBoost and LightGBM implementations were adopted from Chen and Guestrin (2016) and Ke et al. (2017) respectively and conducted on an Intel i7-13700 PC Desktop System with 32 GB of RAM running Windows 10.

3. Results and discussion

Table 3 shows details of the HPs used for producing the optimal ML models for both classification (left column) and regression (right column).

Table 4 presents the test results of the best-performing ML models for classification using 10-Fold CV on the Algerian data set. Here the performance evaluation metrics of all ML models ranged between 0.96–0.98, with only a small variance in results between the ten folds.

More specifically, SVC was the lowest-performing ML model, whereas MLP had the best performance according to all four performance evaluation metrics. Furthermore, Fig. 4 depicts the contributions of each of the hyper-parameters for learning based on the PED-ANOVA method described in Watanabe et al. (2023). The values beside each bar are the percentage of contribution a specific HP has on the mean test Accuracy. In most cases, for each algorithm, only one or two HPs were the more influential.

Table 3
Hyper-parameter settings for each machine-learning model. The left column shows the optimal hyper-parameters for the classification models (SVC, DT, RF, AdaBoost, XGB, LGBM, MLP), and the right column shows those for the corresponding regression models (SVR, DT, RF, AdaBoost, XGB, LGBM, MLP).

Model	Hyper-parameters for classification	Model	Hyper-parameters for regression
SVC	C = 0.981672, degree = 1, gamma = 'auto', kernel = 'linear'	SVR	C = 0.737379, degree = 16, gamma = 'auto, kernel = 'rbf'
DT	criterion = 'entropy' max_depth = 3 min_samples_leaf = 9 min_samples_split = 13	DT	criterion = 'squared_error' max_depth = 4 min_samples_leaf = 2 min_samples_split = 12
RF	class_weight = 'balanced' max_depth = 4 max_samples = 0.920910 min_samples_leaf = 0.1060876 min_samples_split = 0.223397 n_estimators = 43	RF	criterion = 'Poisson' max_depth = 2 max_samples = 0.651752 min_samples_leaf = 0.189317 min_samples_split = 0.449878 n_estimators = 23
Ada-Boost	learning_rate = 1.562151 n_estimators = 48	Ada-Boost	learning_rate = 0.014532 n_estimators = 110
XGBC	eta = 0.320501 eval_metric = 'auc' gamma = 0.437727 max_depth = 2 max_leaves = 2 min_child_weight = 5 n_estimators = 27	XGBR	eta = 0.027458 eval_metric = 'rmse' gamma = 0.771812 max_depth = 5 max_leaves = 5 min_child_weight = 1 n_estimators = 99
LGBM	class_weight = 'balanced' feature_fraction = 0.489883 learning_rate = 0.614967 max_depth = 5 min_data_in_leaf = 82 n_estimators = 81 num_leaves = 34 objective = 'binary'	LGBM	feature_fraction = 0.458930 learning_rate = 0.082612 max_depth = 1 min_data_in_leaf = 2 n_estimators = 18 num_leaves = 28 objective = 'regression'
MLP	activation = 'tanh' alpha = 0.000544 hidden_layer_sizes = 9 learning_rate = 'invscaling' learning_rate_init = 0.007015 max_iter = 486 momentum = 0.994338 solver = 'lbfgs'	MLP	activation = 'logistic' alpha = 0.000842 hidden_layer_sizes = 14 learning_rate = 'adaptive' learning_rate_init = 0.003694 max_iter = 265 momentum = 0.071646 solver = 'sgd'

Table 4
Classification model test results using 10-Fold CV.

Model	Accuracy	Precision	Recall	F1-Score
SVC	0.9673 ± 0.0503	0.9692 ± 0.0492	0.9673 ± 0.0503	0.9673 ± 0.0505
RF	0.9753 ± 0.0328	0.9770 ± 0.0308	0.9753 ± 0.0328	0.9752 ± 0.0330
AdaBoost	0.9838 ± 0.0368	0.9843 ± 0.0362	0.9838 ± 0.0368	0.9837 ± 0.0370
XGBC	0.9838 ± 0.0368	0.9862 ± 0.0306	0.9838 ± 0.0368	0.9834 ± 0.0378
LGBM	0.9878 ± 0.0258	0.9881 ± 0.0255	0.9878 ± 0.0258	0.9878 ± 0.0258
MLP	**0.9878 ± 0.0186**	**0.9888 ± 0.0171**	**0.9878 ± 0.0186**	**0.9878 ± 0.0186**

To better understand the effect of the outcome of an ML model on a selected data point, we adopt SHAP values (Lundberg and Lee, 2017), which can be visualised as SHAP "force" plots, which show how the SHAP values "forces" either increase or decrease the model outcome. The prediction initially starts from a base value, which is calculated from the average of all probabilities for each sample contained in the dataset. To assess the impact of features on ML model output, we passed the same input data from the Algerian data set, which corresponds to a '*fire*'-instance as shown in Table 5 through each ML model and obtained the resulting SHAP values. Fig. 5 depicts these SHAP "force" plots. The size of the bar shows the magnitude of the effect of the feature. Features and their contributions in blue push the model outcome lower towards the base-value whereas feature values and contributions in pink increase the prediction. The sum of all SHAP values explains why the model's prediction differed from the base-value.

Table 5
Input values used to extract SHAP values from the ML classifiers.

Temp	RH	WS	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Class
32	55	14	0	89.1	25.5	88.5	7.6	29.7	13.9	'*fire*'

It can be seen from Fig. 5 that FFMC and ISI consistently appear as a positive influence in all the "force" plots to different degrees signifying the high importance of these features on the prediction of the onset of forest fire. In Fig. 5(a), (e), and (f) DMC, Rain, RH, WS, BUI, and DC all have a negative impact on the model to produce a ('*fire*') outcome, but to a much lesser degree. These results are consistent with work described in Taylor et al. (2013) where FFMC and ISI were identified as key variables for the development of different logistic regression models for assessing wildfire risk in Canada. Later work

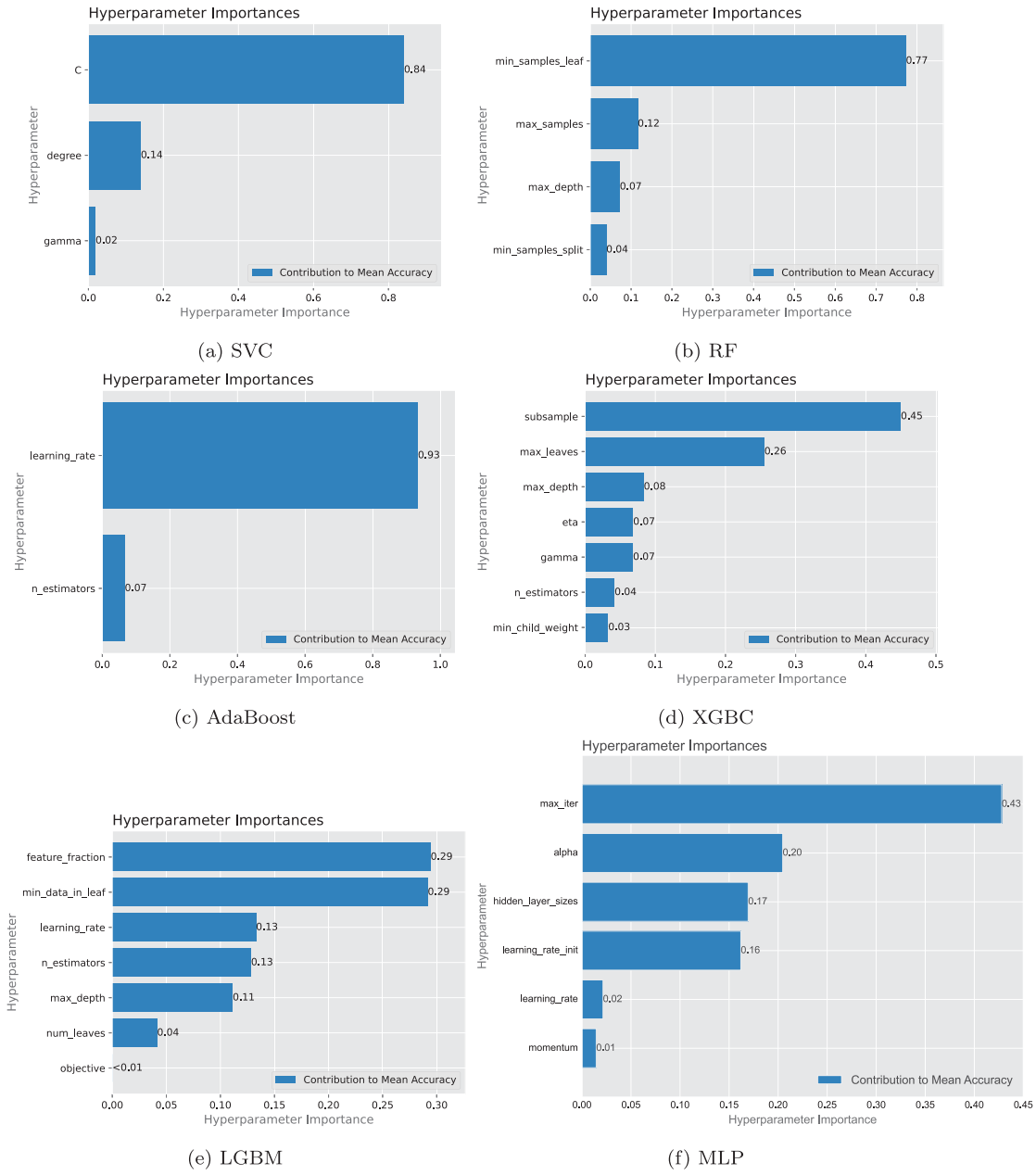


Fig. 4. HP importances for the six classification models. In (a) the most important HP for SVC is C; (b) the most significant HP for RF is `min_samples_leaf`; (c) the key HP for AdaBoost is `learning_rate`; (d) the most influential HP for XGBC is `subsample`; (e) the most critical HP for LGBM is `feature_fraction`; and in (f) the most significant HP for MLP is `max_iter`.

by Davies and Legg (2016) identified that as a result of the statistical analysis of fire incidents in Scotland, both `FFMC` and `ISI` were key factors in modelling the probability of fire occurrence.

Results of models built using the Montesinho data set, as shown in Table 6, this time the better-performing ML model according to the RMSE metric was the MLP, and the SVR model was the worst-performing model. When examining which HPs contribute most to model learning and performance as presented in Fig. 7 there was consistently only one HP for each ML algorithm, which the PED-ANOVA method had identified as being of significant importance to producing low average RMSE values.

Results of the ML models for the regression task are summarised in Table 6. It can be seen that the better-performing ML models for regression were different to those used for classification. In this case, MLP had the lowest RMSE value compared with the rest of the ML models, but SVR had the lowest MAE value.

Table 6
Regression model test results using 10-Fold CV.

Model	RMSE	MAE
SVR	2.2046 ± 0.6196	**1.0691 ± 0.1277**
RF	1.9471 ± 0.4608	1.1492 ± 0.1121
AdaBoost	1.9587 ± 0.4638	1.1550 ± 0.1121
XGBR	2.0523 ± 0.5942	1.1120 ± 0.1136
LGBM	1.9490 ± 0.4598	1.1525 ± 0.1130
MLP	**1.9420 ± 0.4479**	1.1599 ± 0.1112

It would be interesting to see how the regression models perform when estimating FWI values that fall into the numerical class limit of

Fig. 5. SHAP force plots for the same input across six different classifiers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Extreme according to the Canadian FWI System. In the MP data set, 59 observations were classed as *Extreme* i. e. had FWI values of ≥21.0. Fig. 6. As can be seen all ML regressors tend to underestimate the actual FWI values. Moreover, SVR, AdaBoost, and LGBM overestimate some of the FWI values when there is a sharp increase or rapid decrease in FWI value. This result suggests that further work needs to be conducted in modelling trends of very high FWI values without degrading the efficacy of the same models to estimate FWI values representing lower fire danger numerical class limits.

We then extracted SHAP values using the input values as shown in Table 7 below to see the impact of the features on ML model outcome:

Compared with the SHAP "force" plots in Fig. 5, the SHAP "force" plots in Fig. 8 depict more complex interactions by the features to produce an outcome. Most models considered FFMC to have the lowest

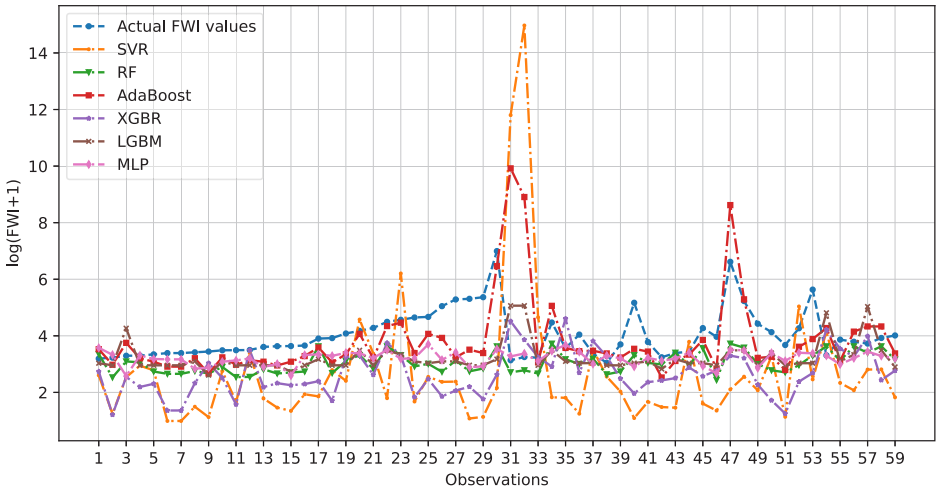


Fig. 6. The results of comparing the actual FWI values plotted in blue and the estimations from the ML models plotted in the remaining colours. log(FWI + 1) of the FWI values was used for the *y*-axis to reduce the range of FWI values presented. For example, a FWI value of 21.0 is plotted as a y-value of 3.09.

Table 7
Input values are used to extract SHAP values from the ML regressors.

X	Y	FFMC	DMC	DC	ISI	Temp	RH	Wind	Rain	Burnt
7.0	4.0	93.5	139.4	594.2	20.3	5.1	96.0	5.8	0.0	15.94

contribution to the prediction of the amount of area burnt by a forest-fire, whereas Wind and Temp, RH, and ISI were the more impactful features regardless if it was a negative effect or a positive effect on the model outcome.

To provide statistical support for these ML classification model results, we performed a Friedman test (Friedman, 1937) at the 95% level of significance to establish whether there was a statistically significant difference between the mean test Accuracy values obtained by the classifiers. The *null* hypothesis was that the performance of the ML was equivalent i. e. had the same average test score rankings. The results of the Friedman test (χ^2 = 8.516, *p*-value = 0.129) indicated that there was no statistically significant difference between classifier performance. Since the *p*-value was greater than 0.05 we could not reject the *null* hypothesis. However, MLP was the better-performing classifier, so it should be considered as a preferred candidate ML model for the purpose of detection the onset of wildfire.

For the ML regression models, a Friedman test at the 95% level (χ^2 = 18.228, *p*-value = 0.002) indicated that there was a statistically significant difference between regressor performance, so this time we could reject the *null* hypothesis. A subsequent Conover post-hoc test with Bonferroni correction (Conover and Iman, 1979) based on the average test RMSE score rankings of the regressors revealed that there was a clear difference between the performance of the SVR and XGBR, as shown in Fig. 9. The ensemble-based models of AdaBoost and LGBM were the highest-ranked regressors, with RF being the next best ranked in performance. MLP was better than both XBGR and SVR. Therefore, our recommendation would be to select AdaBoost or LGBM for this type of regression task, as they both exhibit similar RMSE and MAE values.

Next, an investigation of feature importance on ML model outcome was conducted for both the classification models and regression models with the intention of establishing any similarities or differences between feature ranking by each model. To achieve this we adopted feature importances as measured by passing the same 100 randomly sampled instances from the two data sets into the models and extracting the feature importances based on the extracted SHAP values. The median (Med) feature importance for a specific feature across all ML models was then obtained to provide an overall ranking of feature importance.

The ranking of a feature by each ML model for classification is shown in Table 8 for the Algerian data set. As can be seen, FFMC is the most highly ranked feature across all models, with RH being the lowest ranked feature.

Similarly, the ranking of a feature by each ML model for regression is shown in Table 9 for the Forest Fire data set. This time, seen Temp is the most highly ranked feature across all models, with Rain being the lowest ranked feature.

The results obtained by our framework have similar outcomes to previous work in research that has sought to provide knowledge gained from ML models for the purpose of wildfire prediction.

Recent work by Zaidi (2023), which also built an MLP from the Algerian data set, highlighted the importance of DC, ISI, and RH in the predictions of their model. The best MLP generated from our framework (see Fig. 8(f)) considered only ISI and DC as having the higher feature importances. In terms of the model, ML was built using the Montesinho data set. For example, Bhatt and Chouhan (2024) concluded that Temp, Wind, ISI, DC, RH, FFMC, DMC, and Rain were used to generate an optimal model for prediction. We see this is also the case with our ML regressors, where all features consistently appeared as part of the major contributors (either positive or negative) to the ML model output, as shown in Fig. 8 but in some cases, X was also considered as an important feature.

But the main difference in our study is the use of a comprehensive investigation of which HPs should have more attention paid to them, and show how feature importance varies between ML models depending on the ML algorithm adopted and the type of wildfire analysis task.

4. Threats to validity and limitations

We acknowledge that there are some uncontrolled factors that might have impacted the results reported in our study.

Both data sets contained a limited number of variables, which primarily related to FWI and weather conditions. The inclusion of topographical data would allow for modelling the effect of fire spread across land. Similarly, inputs from satellite imagery data can categorise land type which has an effect on wildfire ignition (Hantson et al., 2016).

The ML models demonstrated high performance for both classification and regression tasks, but only based on specific regions within a country. What impact on model performance when the analysis is scaled to an entire country, or lead times are specified so models forecast FWI up to three days in advance, are questions for future research.

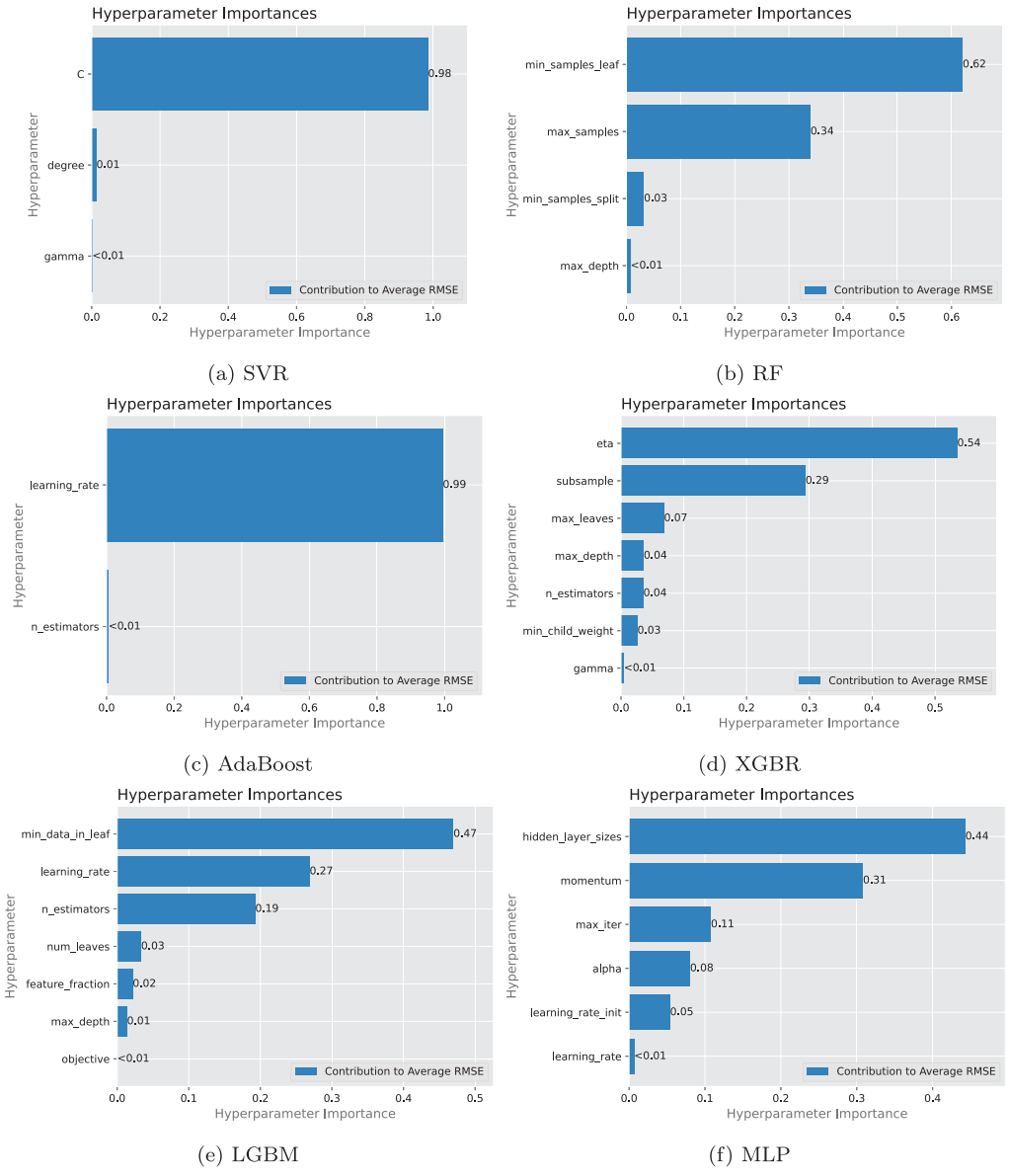


Fig. 7. HP importances for the six regression models. In (a) `C` is the most important HP for SVR; (b) `min_samples_leaf` is the most important HP for RF; (c) The most significant HP for AdaBoost is `learning_rate`; (d) For XGBR, `eta` is the most significant HP; For LGBM, the most significant HP is `min_data_in_leaf`; for MLP, it is `hidden_layer_sizes`, which are shown by (e) and (f), respectively.

Table 8
Median feature importance across all classification models using SHAP values.

Feature \ Model	SVC	RF	Ada Boost	XGBC	LGBM	MLP	Med.
FFMC	1	1	1	1	2	1	1
ISI	2	2	7	2	1	2	2
FWI	3	5	10	3	3	3	3
DMC	4	4	8	6	6	4	5
DC	8	6	6	5	9	5	6
WS	5	10	5	8	5	9	7
Rain	6	7	4	7	10	6	7
BUI	7	3	9	4	7	10	7
Temp	10	8	3	3	8	7	8
RH	9	9	2	10	4	8	9

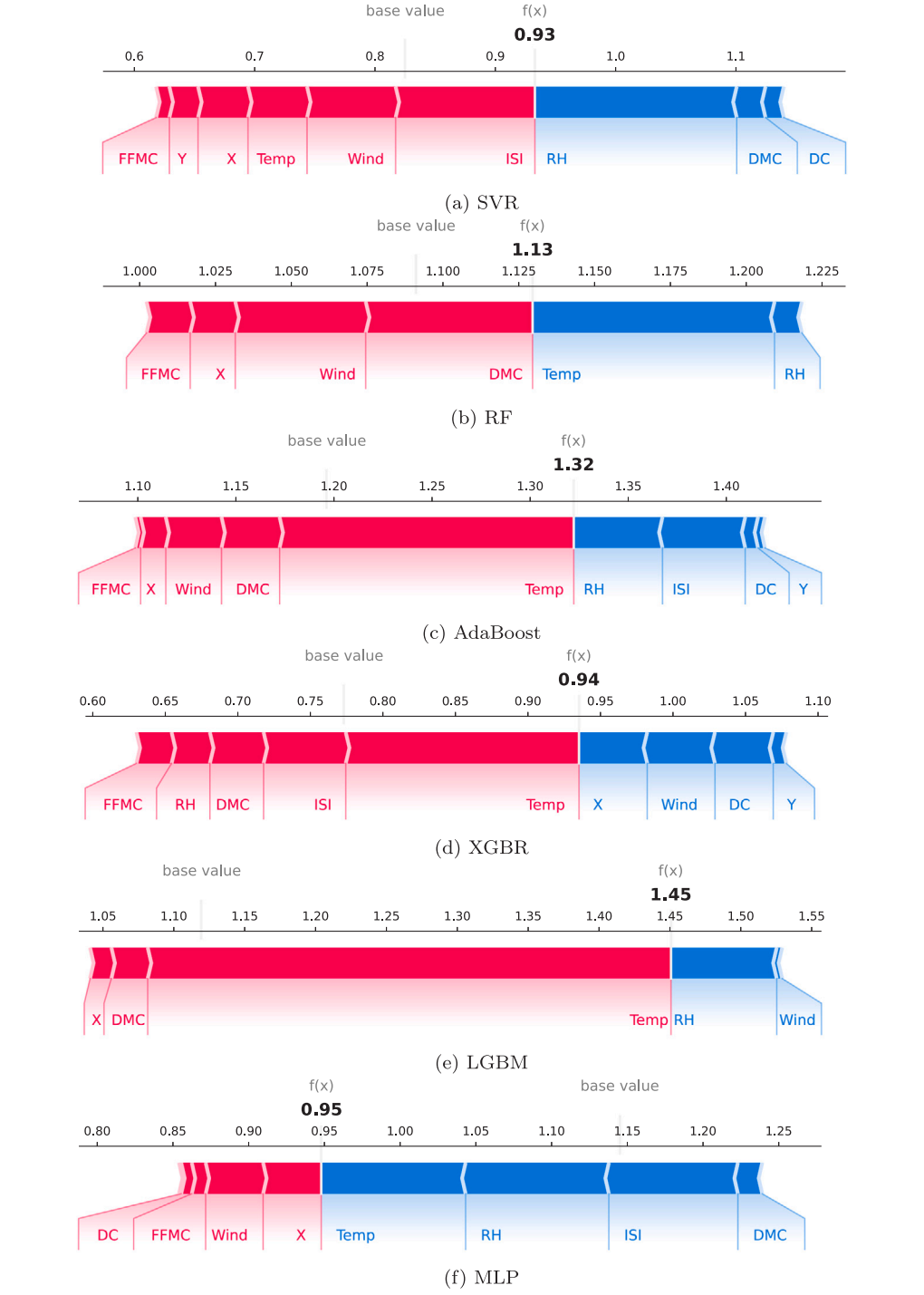


Fig. 8. SHAP force plots for the same input across six different regressors.

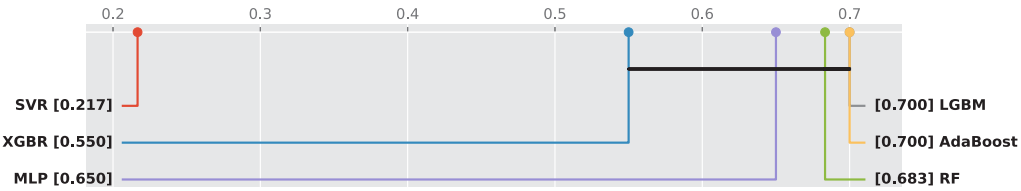


Fig. 9. Critical difference diagram illustrating the comparison of average test RMSE score ranks for the six ML models. The average RMSE ranks are as follows: SVR at 0.217, XGBR at 0.550, MLP at 0.650, RF at 0.683, LGBM and AdaBoost both at 0.700.

Table 9
Median feature importance across all regression models using SHAP values.

Feature \ Model	SVC	RF	Ada Boost	XGBR	LGBM	MLP	Med.
Temp	3	1	1	1	1	2	1
Wind	5	3	5	4	2	3	4
X	2	4	3	8	5	1	4
DMC	6	2	4	3	4	7	4
RH	4	6	7	6	3	4	5
Y	1	7	9	7	6	10	7
DC	7	8	6	7	8	6	7
FFMC	8	5	8	2	7	8	8
ISI	9	9	2	9	9	5	9
Rain	10	10	10	10	10	9	10

Yet, in focusing on region-specific modelling, we demonstrated the need for this type of modelling due to varying climate and vegetation factors. This issue necessitates frequent recalibration of the ML models to address these environmental changes. By doing this, the efficacy and reliability of these ML models as a useful tool for wildfire analysis would be maintained.

Next, accurate climate forecasting is critical for predicting wildfires, but climate change complicates this, limiting the ML models' effectiveness. In the future, it is expected that data produced by more recent climate forecasting models will mitigate this problem.

Furthermore, appropriate resolution of location-specific granular data would benefit model accuracy through the inclusion of more specific location-based data acquired from hand-held sensors or wireless sensor networks. The adoption of higher resolution satellite or drone-based images could also enhance FFMC values, but potentially incur a higher financial cost to obtain.

Another threat was that the ML models constructed could be subject to over-fitting. This might have been the case as the Algerian data set represented data from only one year. Models built from such limited data might be prone to over-fitting. Our methodology accounted for this potential risk by adopting 10-Fold CV when establishing the best HPs for a specific ML model type. The objective of using CV was to reduce the risk of over-fitting and increase the ability of the ML model to generalise to unseen data. Although 10-Fold CV did provide enhanced ML model performance, this strategy did not take into account the temporal nature of the data. More appropriate CV strategies for dealing with spatiotemporal data like those discussed by Oliveira et al. (2021) would address this issue.

Related to this threat is the high performance observed across all classification models, with metrics exceeding 96% suggests that the classification task may be relatively simple given the dataset structure. The high correlation between fire risk and the available weather variables, as measured by FWI, may be one explanation for this. High predictive performance can be achieved by ML models since the dataset mainly consists of meteorological factors that are known to affect the occurrence of wildfires.

Furthermore, the manner in which negative samples i. e., cases in which there was no fire, were gathered may have an impact on the model's performance. Models may find it simpler to inflate accuracy if negative samples differ significantly from positive ones (for example, a clear separation in weather conditions).

Finally, the ML models were built using the currently available Python libraries that implement existing learning algorithms. The reliance on using these Python libraries not only limited the scope of applying other appropriate ML algorithms for the purpose of wildfire analysis but also did not consider non-ML-based approaches. Wildfire prediction using physical models (Sullivan, 2019) or statistical models (Taylor et al., 2013) might yield slightly different results with the same dataset, pointing to a limitation in using ML for wildfire prediction. Despite this, our framework was consistently applied across different geographic locations, supporting model reusability. Future testing of our framework in diverse environmental and geographical contexts can further validate our research approach.

Although our framework performed well on the Montesinho-Portugal (regression) and Algerian (classification) datasets, both are located in semi-arid or Mediterranean climate zones. Boreal North America, the U.S. West Coast, Australia, Southeast Asia, and other places have very different wildfire dynamics, which are influenced by fuel types, ignition sources, and weather patterns. Consequently, our feature-importance rankings and ideal hyper-parameters might not be directly applicable in certain situations.

5. Conclusion

In this work, we have proposed a framework which combines current research on generating optimal ML models and a popular method for explaining ML model predictions for the purpose of both classification and regression tasks for analysing forest fires.

Major findings of our work identified which HPs which the most influential in generating ML models and highlighted that the MLP classifier was better suited to detecting the onset of a fire whereas AdaBoost or LGBM would be more appropriate ML models for predicting damage affected or burned area as a result of wildfire. Furthermore, we demonstrated the use of SHAP values for explaining ML model output and how these SHAP values can be used to rank the importance of features on model outcomes. Similarities were found between our findings and previous research that sought to make ML models more transparent through the adoption of SHAP values for model interpretation.

In terms of future work, we intend to investigate what effect the temporal information, which is present in both the wildfire data sets, has on ML model performance and apply our framework. Furthermore, involving more data sets from other countries to validate our framework is a research direction which we intend to explore in the near future. In relation to this, one of the goals is to augment the existing data sets with topographical and satellite data to provide a richer data set for wildfire risk prediction. This will involve acquiring other publicly available wildfire datasets from other continents (for example, datasets from the Copernicus, Google Earth Engine) and validate across regions to measure ML model performance drop-off in order to develop strategies to mitigate this issue. For example, to improve generalisation across climate regimes, we will also investigate domain-adaptation strategies (such as transfer learning (Farhad Mortezapour et al., 2024)).

Finally, although our present study makes use of ML techniques that work well on these tabular, relatively small wildfire datasets, deep learning architectures can automatically extract richer hierarchical features and often achieve higher predictive accuracy when scaling to much larger or higher-dimensional data sources (such as multi-spectral satellite imagery, continuous sensor streams, or long time-series data) (Badhan et al., 2024). For this reason, subsequent work will involve the application of Convolutional Neural Networks (CNN)s for spatial pattern identification and Recurrent Neural Networks (RNN)s and Long Short-Term Memory (LTSM)/Transformers for

investigating temporal dynamics in subsequent work to see if they provide additional benefits on large wildfire datasets (Farhad Mortezapour et al., 2024). In order to ensure both precision and explainability, we will also look into hybrid techniques that combine the representational strength of deep neural networks with the interpretability of tree-based models.

CRediT authorship contribution statement

Abira Sengupta: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Brendon J. Woodford:** Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Publicly available data.

References

Abid, F., Izeboudjen, N., 2020. Predicting forest fire in Algeria using data mining techniques: Case study of the decision tree algorithm. In: Advanced Intelligent Systems for Sustainable Development (AI2SD'2019) Volume 4-Advanced Intelligent Systems for Applied Computing Sciences. Springer, pp. 363–370. http://dx.doi.org/10.1007/978-3-030-36674-2_37.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '19, Association for Computing Machinery, New York, NY, USA, pp. 2623–2631. http://dx.doi.org/10.1145/3292500.3330701.

Anon, 1993. Fire Weather Index System Tables for New Zealand. Tech. rep., National Rural Fire Authority in association with the New Zealand Forest Research Institute, Wellington, New Zealand, p. 48.

Badhan, M., Shamsaei, K., Ebrahimian, H., Bebis, G., Lareau, N.P., Rowell, E., 2024. Deep learning approach to improve spatial resolution of GOES-17 wildfire boundaries using VIIRS satellite data. Remote. Sens. 16 (4), http://dx.doi.org/10.3390/rs16040715.

Bergstra, J., Yamins, D., Cox, D.D., 2013. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In: Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28. ICML '13, JMLR.org, Atlanta, GA, USA, pp. I–115–I–123. http://dx.doi.org/10.5555/3042817.3042832.

Bhatt, S., Chouhan, U., 2024. An enhanced method for predicting and analysing forest fires using an attention-based CNN model. J. For. Res. 35 (1), 67. http://dx.doi.org/10.1007/s11676-024-01717-7.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J., Olshen, R., Stone, C.J., 2017. Classification and Regression Trees. Routledge, http://dx.doi.org/10.1201/9781315139470.

California Air Resources Board, 2024. Available online: https://ww2.arb.ca.gov/ghg-inventory-data (accessed 29 July 2024).

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16, Association for Computing Machinery, New York, NY, USA, pp. 785—-794. http://dx.doi.org/10.1145/2939672.2939785.

Conover, W.J., Iman, R.L., 1979. Multiple-comparisons procedures. Informal report. http://dx.doi.org/10.2172/6057803.

Cortez, P., Morais, A.d.J.R., 2007. A data mining approach to predict forest fires using meteorological data. URL https://api.semanticscholar.org/CorpusID:36868619.

Dai, A., 2013. Increasing drought under global warming in observations and models. Nat. Clim. Chang. 3 (1), 52–58. http://dx.doi.org/10.1038/nclimate1633.

Davies, G.M., Legg, C.J., 2016. Regional variation in fire weather controls the reported occurrence of Scottish wildfires. Peer J e2649 (4), http://dx.doi.org/10.7717/peerj.2649.

de Bem, P., de Carvalho Júnior, O., Matricardi, E., Guimarães, R., Gomes, R., 2018. Predicting wildfire vulnerability using logistic regression and artificial neural networks: a case study in Brazil. Int. J. Wildland Fire 28 (1), 35–45. http://dx.doi.org/10.1071/WF18018.

Dessler, A., Zhang, Z., Yang, P., 2008. Water-vapor climate feedback inferred from climate fluctuations, 2003–2008. Geophys. Res. Lett. 35 (20), http://dx.doi.org/10.1029/2008GL035333.

Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., Zeng, W., 2019. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. Agricult. Water. Manag. 225, 105758. http://dx.doi.org/10.1016/j.agwat.2019.105758.

Farhad Mortezapour, S., Thinagaran, P., Norwati, M., Raihani, M., 2024. A comprehensive overview and comparative analysis on deep learning models. J. Artif. Intell. 6 (1), 301–360. http://dx.doi.org/10.32604/jai.2024.054314.

Finney, M., Grenfell, I.C., McHugh, C.W., 2009. Modeling containment of large wildfires using generalized linear mixed-model analysis. For. Sci. 55 (3), 249–255. http://dx.doi.org/10.1093/forestscience/55.3.249.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. System Sci. 55 (1), 119–139. http://dx.doi.org/10.1006/jcss.1997.1504.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Amer. Statist. Assoc. 32 (200), 675–701. http://dx.doi.org/10.2307/2279372.

Gibson, R., Danaher, T., Hehir, W., Collins, L., 2020. A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest. Remote Sens. Environ. 240, 111702. http://dx.doi.org/10.1016/j.rse.2020.111702.

Goodchild, M.F., 1991. Geographic information systems. Prog. Hum. Geogr. 15 (2), 194–200. http://dx.doi.org/10.1177/030913259101500 20.

Hansen, J., Ruedy, R., Sato, M., Lo, K., 2010. Global surface temperature change. Rev. Geophys. 48 (4), http://dx.doi.org/10.1029/2010RG000345.

Hantson, S., Arneth, A., Harrison, S.P., Kelley, D.I., Prentice, I.C., Rabin, S.S., Archibald, S., Mouillot, F., Arnold, S.R., Artaxo, P., Bachelet, D., Ciais, P., Forrest, M., Friedlingstein, P., Hickler, T., Kaplan, J.O., Kloster, S., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Melton, J.R., Meyn, A., Sitch, S., Spessa, A., van der Werf, G.R., Voulgarakis, A., Yue, C., 2016. The status and challenge of global fire modelling. Biogeosciences 13 (11), 3359–3375. http://dx.doi.org/10.5194/bg-13-3359-2016.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning Data Mining, Inference, and Prediction. In: Springer Series in Statistics, Springer, http://dx.doi.org/10.1007/978-0-387-84858-7.

Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B., 1998. Support vector machines. IEEE Intell. Syst. Appl. 13 (4), 18–28. http://dx.doi.org/10.1109/5254.708428.

Jaafari, A., Pourghasemi, H.R., 2019. 28 - factors influencing regional-scale wildfire probability in Iran: An application of random forest and support vector machine. In: Pourghasemi, H.R., Gokceoglu, C. (Eds.), Spatial Modeling in GIS and R for Earth and Environmental Sciences. pp. 607–619. http://dx.doi.org/10.1016/B978-0-12-815226-3.00028-4.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30, 3146–3154. http://dx.doi.org/10.5555/3294996.3295074.

Khalil Alsmadi, M., Omar, K.B., Noah, S.A., Almarashdah, I., 2009. Performance comparison of multi-layer perceptron (back propagation, delta rule and perceptron) algorithms in neural networks. In: 2009 IEEE International Advance Computing Conference. IEEE, pp. 296–299. http://dx.doi.org/10.1109/IADCC.2009.4809024.

Kondylatos, S., Prapas, I., Ronco, M., Papoutsis, I., Camps-Valls, G., Piles, M., Fernández-Torres, M.-Á., Carvalhais, N., 2022. Wildfire danger prediction and understanding with deep learning. Geophys. Res. Lett. 49 (17), e2022GL099368. http://dx.doi.org/10.1029/2022GL099368.

Liang, H., Zhang, M., Wang, H., 2019. A neural network model for wildfire scale prediction using meteorological factors. IEEE Access 7, 176746–176755. http://dx.doi.org/10.1109/ACCESS.2019.2957837.

Lopes, A., Cruz, M.G., Viegas, D., 2002. *FireStation* - an integrated software system for the numerical simulation of fire spread on complex topography. Environ. Model. Softw. 17 (3), 269–285. http://dx.doi.org/10.1016/S1364-8152(01)00072-X.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, pp. 4768–4777. http://dx.doi.org/10.5555/3295222.3295230.

Martínez, J., Vega-Garcia, C., Chuvieco, E., 2009. Human-caused wildfire risk rating for prevention planning in Spain. J. Environ. Manag. 90 (2), 1241–1252. http://dx.doi.org/10.1016/j.jenvman.2008.07.005.

Mell, W., Jenkins, M.A., Gould, J., Cheney, P., 2007. A physics based approach to modeling grassland fires. Int. J. Wildland Fire 16 (1), 1–22. http://dx.doi.org/10.1071/WF06002.

Naderpour, M., Rizeei, H.M., Ramezani, F., 2021. Forest fire risk prediction: A spatial deep neural network-based framework. Remote. Sens. 13 (13), http://dx.doi.org/10.3390/rs13132513.

Oliveira, M., Torgo, L., Santos Costa, V., 2021. Evaluation procedures for forecasting with spatiotemporal data. Mathematics 9 (6), http://dx.doi.org/10.3390/math9060691.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830. http://dx.doi.org/10.5555/1953048.2078195.

Raghunath, K.K., Kumar, V.V., Venkatesan, M., Singh, K.K., Mahesh, T., Singh, A., 2022. XGBoost regression classifier (XRC) model for cyber attack detection and classification using inception V4. J. Web Eng. 21 (4), 1295–1322. http://dx.doi.org/10.13052/jwe1540-9589.21413.

Sarkar, M.S., Majhi, B.K., Pathak, B., Biswas, T., Mahapatra, S., Kumar, D., Bhatt, I.D., Kuniyal, J.C., Nautiyal, S., 2024. Ensembling machine learning models to identify forest fire-susceptible zones in Northeast India. Ecol. Inform. 81, 102598. http://dx.doi.org/10.1016/j.ecoinf.2024.102598.

Shabbar, A., Skinner, W., 2004. Summer drought patterns in Canada and the relationship to global sea surface temperatures. J. Clim. 17 (14), 2866–2880. http://dx.doi.org/10.1175/1520-0442(2004)017<2866:SDPICA>2.0.CO;2.

Shrestha, D.L., Solomatine, D.P., 2006. Experiments with AdaBoost.RT, an improved boosting scheme for regression. Neural Comput. 18 (7), 1678–1710.

Stocks, B.J., Lynham, T., Lawson, B., Alexander, M., Wagner, C.V., McAlpine, R., Dube, D., 1989. The Canadian forest fire danger rating system: An overview. For. Chron. 65 (4), 258–265. http://dx.doi.org/10.5558/tfc65450-6.

Sullivan, A.L., 2019. Physical modelling of wildland fires. In: Manzello, S.L. (Ed.), Encyclopedia of Wildfires and Wildland-Urban Interface (WUI) Fires. Springer International Publishing, Cham, pp. 1–8. http://dx.doi.org/10.1007/978-3-319-51727-8_58-1.

Suthaharan, S., 2015. Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning, first ed. In: Integrated Series in Information Systems, no. 36, Springer Publishing Company, Incorporated, http://dx.doi.org/10.5555/2851117.

Taylor, S.W., Woolford, D.G., Dean, C.B., Martell, D.L., 2013. Wildfire prediction to inform fire management: Statistical science challenges. Statist. Sci. 28 (4), 586–615. http://dx.doi.org/10.1214/13-STS451.

Tharwat, A., 2021. Classification assessment methods. Appl. Comput. Inform. 17 (1), 168–192. http://dx.doi.org/10.1016/j.aci.2018.08.003.

The European Space Agency, 2021. New long-term dataset to analyse global fire trends. Available online: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-3/New_long-term_dataset_to_analyse_global_fire_trends (accessed 3rd Aug 2025).

Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13, 361–378. http://dx.doi.org/10.1007/s10346-015-0557-6.

Tyukavina, A., Potapov, P., Hansen, M.C., Pickens, A.H., Stehman, S.V., Turubanova, S., Parker, D., Zalles, V., Lima, A., Kommareddy, I., Song, X.-P., Wang, L., Harris, N., 2022. Global trends of forest loss due to fire from 2001 to 2019. Front. Remote. Sens. 3, http://dx.doi.org/10.3389/frsen.2022.825190.

Watanabe, S., Bansal, A., Hutter, F., 2023. PED-ANOVA: Efficiently quantifying hyperparameter importance in arbitrary subspaces. In: Elkind, E. (Ed.), Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. IJCAI-23, International Joint Conferences on Artificial Intelligence Organization, California, pp. 4389–4396. http://dx.doi.org/10.24963/ijcai.2023/488.

Zaidi, A., 2023. Predicting wildfires in Algerian forests using machine learning models. Heliyon 9 (7), e18064. http://dx.doi.org/10.1016/j.heliyon.2023.e18064.