



Integrating deep learning with geospatial intelligence for real-time forest fire risk assessment

Reem Salman¹, Nizar Hamadeh^{2*}, Ali Karouni², Elias Rachid³

¹ EDST, Lebanese University, Beirut, Lebanon

² Faculty of Technology, Lebanese University, Saïda, Lebanon

³ Ecole Supérieure D'ingénieurs de Beyrouth, Saint-Joseph University, Beirut, Lebanon

* Corresponding author's e-mail: nizar.hamadeh@ul.edu.lb

ABSTRACT

Climate change continues to intensify the frequency of forest fires and other natural disasters around the globe. Intense and recurring extreme weather conditions, particularly in sensitive areas such as the Mediterranean, which is highly susceptible to intense and frequent fires, amplify the frequency and severity of forest fires. This study applies an integrated approach combining machine learning (ML), remote sensing, and geographic information systems (GIS) to predict forest fire risk using the XGBoost algorithm. The model incorporates several factors including topographic features (slope, aspect, and elevation), meteorological features (relative humidity, temperature, dew point, solar radiation, wind speed, precipitation), anthropogenic influences (distance from urban centers, roads, and cultivated land), and the types of fuel available. The final output, the forest fire risk map, divides the region into three risk zones: high, moderate, and low, to enhance community and stakeholder participation and to advance preparedness measures in emergencies. The forecasting model demonstrated remarkable predictive capabilities, achieving an overall accuracy of 0.94 and an AUC of 0.99, underscoring its ability to improve fire management systems, response strategies, and emergency response readiness.

Keywords: machine learning, remote sensing, GIS, XGBoost, forest fire.

INTRODUCTION

Forests are one of the largest terrestrial ecosystems combatting climate change on earth covering 31% of the Planet's land surface (Can et al., 2022). They play an essential role in ensuring the sustainability of human civilization and maintaining biodiversity and ecological balance (Molina et al., 2019). In recent years, climate change, extreme weather, and deforestation events have all contributed to the escalation of this trend. Over 90% of these fires are caused by humans, where the combined effects of anthropogenic factors and climate change have contributed to the continuous reduction in wild areas (Venkatesh et al., 2020). Consequently, several fire management initiatives have been implemented, costing billions of dollars aiming to prevent or mitigate the devastating impacts of forest fire. The Mediterranean Basin (MB)

provides crucial habitat for many endangered animals because of its highly diversified natural vegetation. Forest Fires pose a serious concern in MB countries given their high probability of occurrence and severe impacts. Contingency plans including proactive fire prevention strategies and response measures are often implemented to mitigate the risk of fire occurrence. Forest fires are driven by a combination of several factors like topography, weather, ignitions sources and fuel contents (Jain et al., 2020). Between 2001 and 2023, 1,830 hectares of tree cover were lost in Lebanon as a result of the escalating challenge with the frequently recurring forest fires (Global Forest Watch, 2024). The Mediterranean environment of Lebanon, characterized by hot, dry summers and strong winds, makes it vulnerable for forest fires. Major advancements in technology have enhanced the ability to model and monitor forest fire allowing for more effective

fire management strategies. Developing forest fire risk mapping (FFRM) requires identifying forest fire inventory as the target variable and other explanatory variables. GIS technology has shown its essential role of developing FFRM by effectively processing and overlaying geospatial data (El Hayfani et al., 2020, Teodoro and Duarte, 2013). Furthermore, remote sensing and satellite imagery have emerged as essential resources for data collection and creating FFRM as well (Mohajane et al., 2021). Integration of machine learning (ML) with geographic information systems (GIS), and remote sensing to assess fire risk is a data driven approach that has been adopted for modeling, leading to increase interest in applying ML techniques to forest fire science and management (Ghali et al., 2020). For Lebanon, tentative forest fire risk maps have been created (Faour et al., 2004; Faour et al., 2006; Masri et al., 2003). However, these initiatives haven't yet leveraged the potential of ML techniques to enhance the accuracy and the efficiency of the model through developing a comprehensive FFRM for Lebanon. In recent years, there has been an obvious trend in applying ML algorithms in a wide range of natural hazards assessment fields such as forest fire modeling (Yu et al., 2017, Van Beusekom et al., 2018, Adab et al., 2017), flood (Tehrany et al., 2015b), water quality modelling (Pham et al., 2021), rainfall modeling (Pham et al., 2019, Pour et al., 2020), landslides (Chen et al., 2019, Pham et al., 2018a). Studies have shown that the use of ensemble/hybrid models, combining statistical methods and ML techniques, increases the modeling performance (Truong et al., 2018, Akay et al., 2021). However, existing studies has mostly focused on identifying zones prone to forest fire without using predictive modeling that estimated the risk based on anthropogenic and dynamic environmental factors. Even though ML approaches are widely employed in many studies to classify risk zones, there remains a significant opportunity to enhance Lebanon's forest fire risk assessment by integrating ensemble ML models with GIS spatial analytics and remote sensing. Most studies primarily focus on fire occurrences mapping rather than proactive risk forecasting, which is crucial for effective disaster preparedness and response. This study aims to develop a dynamic and data driven approach for Lebanon enhancing prediction accuracy, and enabling more effective disaster preparedness and response by integrating ensemble ML (XGBoost) with remote sensing and GIS. Also, it leverages ML techniques to break down

the nonlinear relationship between fire events and meteorological parameters. By combining Meteorological, topographic and anthropogenic factors, this study contributes to proactive forest fire management strategies in Lebanon.

MATERIALS AND METHODS

Study area

The study area lies along the eastern Mediterranean coast at the intersection of Europe, Asia, and Africa, with a total surface area of 10,452 km². Its landscape includes elevated mountain chains that rise above 2,500 meters. It is composed of eight governorates (South, North, Nabatieh, Beirut, Bekaa, Akkar, Mount Lebanon and Baalbeck El-Hermel) and subdivided into 26 districts. With a total coastline of 225 kilometers and a maximum depth of 80 kilometers, it is bordered by Palestine to the south by Syria to the east and north, and by the Mediterranean Sea to the west. The Lebanon mountain range has the majority of the country's forests with the most biologically diverse slopes being the wettest (Zohary et al., 1973). According to the 2017 Land use/Land Cover (CNRS, 2017), Akkar, Mount Lebanon, and North governorates have the largest concentration of forests and shrublands 34.4%, 40.6%, and 33.9% respectively. They cover 29.8% of the country's area while

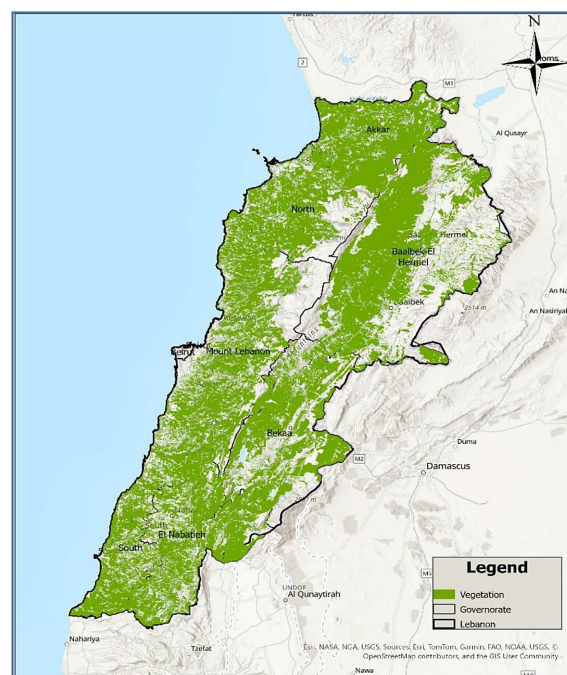


Figure 1. Vegetation map for Lebanon

grasslands account for only 4.1% of the entire area supporting a rich biodiversity (Figure 1). Finally, the governorate of Beirut is entirely covered by Urban area (Georgia et al., 2022). Based on the zohary (2023), Lebanon is situated in the Mediterranean climatic zone (Csa), which is known by hot, dry summers and warm, rainy winters. Also, according to the variation of topography and climate, it is classified into four bioclimatic zones: western mid mountain, coastal, inland plateau, and high mountain. As stated in Lebanon’s 4th national communication to the United Nations Framework Convention on Climate Change (UNFCCC), by 2040, temperature is expected to rise by 1 °C along the coast and 2 °C inland, while rainfall will reduce by 10% to 20% (World Bank, 2023). Subsequently, key sectors like forestry and agriculture will be impacted by these climatic changes, especially in the case of extreme weather events (World Bank, 2023). Lebanon’s socioeconomic issues are intensified by the climate change, further exacerbating the country’s long-term vulnerability.

Methodology

In this research, a supervised ML design model is developed to identify and map areas at high risk of forest fire throughout Lebanon by analyzing a spatial grid (fishnet) that divides the area

into distinct cells. It will perform a classification task by daily assigning the forest fire risk for each cell into one of three classes: Low, Medium or High. The study analyzed 13 key factors contributing to fire risk by integrating ML with GIS technology and Remote sensing through satellite imagery. Different data sources were employed to support the identification and mapping process, providing a comprehensive evaluation of fire risk throughout the country. Data mining is a process that employs a variety of data analytic tools to discover patterns and correlations within data, enabling informed decision making and reliable forecasting. Decision tree was utilized to prioritize the variables affecting forest fire risk, while facing some challenges in data collection. Weather data was collected via online platform, while the rest obtained through field data collection and satellite imagery enriching the study and enhancing its accuracy. A summary of the methodology is presented in Figure 2.

As you can notice in Figure 2, it shows an approach for data processing and analysis, essential for assessing the risk of forest fire. The first step is the collection of native data gathered from different sources as shown in Figure 3. The dataset consists of 2675166 rows with the following attributes (columns), Meteorological factors

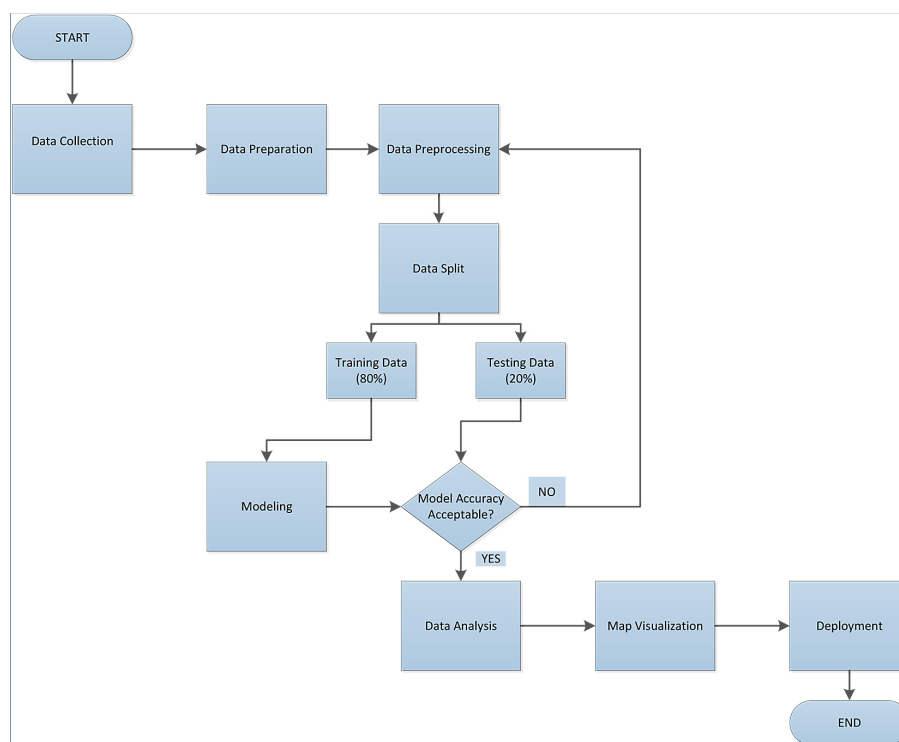


Figure 2. Methodology flow chart

(wind speed, temperature, solar radiation, relative humidity, precipitation, dew point), anthropogenic factors (proximity to urban, proximity to roads, proximity to agriculture), and fuel types. Then, second step it undergoes data preparation, which involves organizing and formatting information into a usable structure and unified data set. This step assures that all datasets are spatially aligned and combined, thereby allowing them to be combined smoothly using GIS. Furthermore, multiple layers are overlaid using spatial integration to provide a cohesive dataset that captures spatial relationships. Following preparation, as a third step the data undergoes data processing, an important step for refining and cleaning the information to ensure accuracy and reliability. This process ensures the enhancement and accuracy, enriching the study and addressing challenges associated with data collection. For example, it involves checking and handling null values using the K-nearest neighbors (KNN) algorithm, addressing the imbalanced dataset, and identifying and removing outliers that could skew the model results. The KNN method imputes null values by using the nearest neighbors in the feature space ensuring the alignment between the imputed values and the underlying patterns of the dataset. After that, the data undergoes a data splitting as a fourth step, which is crucial for evaluating the model's reliability and performance. In this step, the dataset is typically divided into two subsets: one subset comprised 80% of the total data and is used for model training, and the other subset included 20% of the data and is used for data verification and testing. This split ensures that the final model is not over fitted to the training data, allowing it to generalize well to unseen data and enhancing practical applicability. Following data splitting, the fifth step is the data modeling and analysis. It involves performing an overlay analysis with the administrative database, specifically municipalities layer. This step classify each municipality into different risk levels based on the factors influencing fire risk that is low, medium and high risk, which serves as a measure of the overall vulnerability to fire events. Finally, these are used for additional analysis and visualization for the fire risk on a map. The generated map helps in decision making and resource allocation for fire prevention and response. For the final step in the workflow, deployment entails integrating the model and GIS maps into a user-friendly platform (dashboard). It can provide continuous alert

to emergency responders and municipalities. By incorporating GIS integration, it allows spatial visualization, providing operational anticipation tool supporting contingency planning and fire mitigation strategies.

Data extraction

Due to the lack of official fire records and a centralized fire monitoring system in Lebanon, NASA's Fire Information for Resources Management System (FIRMS) was used. This system is used extensively throughout the world as a crucial input for automated fire perimeter detection (Pinto et al., 2021). It offers information on the location of fire events in almost real time, based on observations from NASA's moderate resolution imaging spectroradiometer (MODIS) (NASA FIRMS, n.d.). Despite being the sole source of information, the number of fire events reported is still far less than what has actually happened in Lebanon over the previous 20 years. Consequently, a dashboard that compiles all fire incidents reported by the Lebanese Civil Protection Agency was created as part of a new adaptive strategy. As a result, a new adaptive approach was adopted by developing a dashboard that aggregates all fire events reported by the Lebanese Civil defense on the X platform since April 2024. Sentinel-2 and Landsat data, supported by Google Earth Engine (GEE), were used to map burned regions in Lebanon, combining supervised classification with spectral indices. Fire polygons were initially identified using NASA's and then manually adjusted for accuracy. This procedure ensured the reliable mapping of fire-affected regions. Figure 3 presents the different datasets used in the study including meteorological, topographic and satellite data, highlighting their roles in mapping the factors that are crucial for fire risk analysis. For this study, elevation, aspect, and slope data were obtained from the advanced spaceborne thermal emission and reflection radiometer (ASTER) though NASA's 3D graphic of the earth digital elevation model (DEM) produced by NASA's Terra satellite. It provides high resolution elevation data with a spatial resolution of 30 meters, making it valuable for geospatial analysis. Key topographic factors like slope, elevation and aspect can be obtained by integrating this dataset into GIS software, aiding in the analysis and modeling. The weather bit API was used to get real-time meteorological data in order to fill the

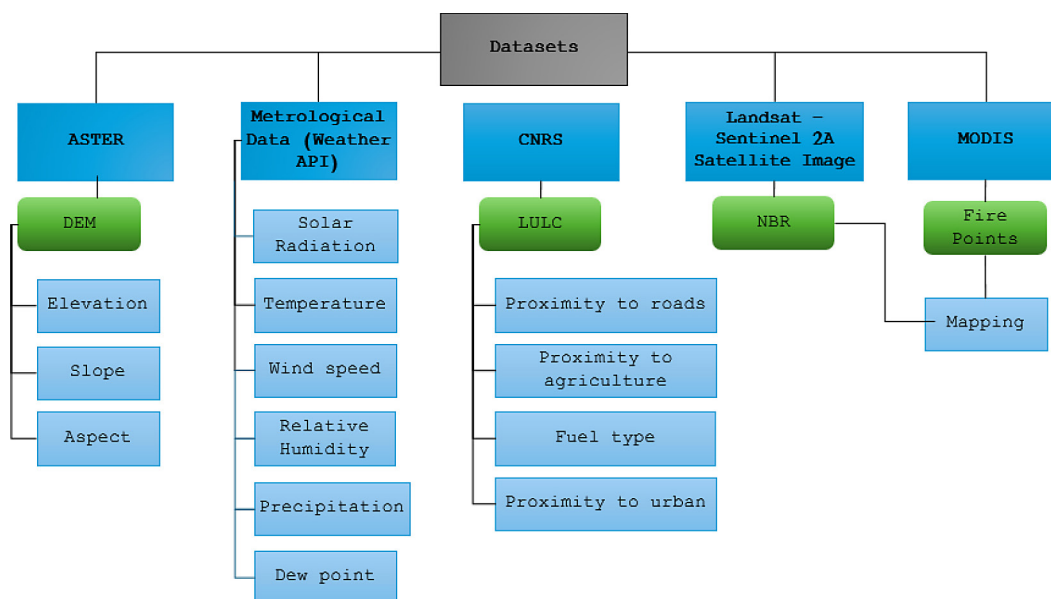


Figure 3. Sources used to extract the attributes needed for FFRM (dataset)

gaps in climatic data from different weather stations around Lebanon. This API pulls data from more than 47,000 operational active weather stations worldwide, along with meso-re-analysis and Doppler radar data. For creating a weather map for Lebanon, a defined cell grid location is established, and a script gets weather data for each grid cell. After being saved, this data is utilized to generate raster layers that depict the different weather parameters. Despite being a paid service, the weather bit API was selected because of its reliability, offering accurate information. As for the influential factors, here is a list for the 13 factors included in the risk assessment:

Topography

Slope (degrees)

Slope speaks to the steepness of the land, usually measured in an angle. It is easier for fire to spread on steeper slopes because the flames are lower and the potential fuel source “uphill” is already being pre-heated. It is estimated that wildfire spread can double for every 10-degree increase in slope. This is because sloped land tends to channel both wind and flames upward, increasing fire activity by pre-heating vegetation (Rothermel, 1983).

Aspect (degrees)

Aspect describes the direction of slope as North (0°), East (90°), South (180°), West (270°), with the degrees marking their respective cardinal

direction (Esri, n.d.). In the northern hemisphere, north-facing slopes get less sun while south-facing slopes receive sunlight and tend to be warmer. This difference in exposure will influence wildfire behavior: south- and west-facing slopes have drier fuel loads because of more sun and high evapotranspiration, particularly during summer months. West-facing slopes warm up in the afternoon and evening, and east-facing slopes warm up early in the day due to cooler nights (Esri, n.d.).

Elevation (meters)

Elevation refers to the height of the land in relation to sea level, this metric greatly influences the risk of wildfires. For instance, Lebanon is made up of a flat coastal zone, hilly region, and a mountainous area which means its elevation has an impact on local temperature, humidity, and type of vegetation. Wildfire activity is typically less in higher elevations due to cooler temperatures and greater humidity which creates less suitable conditions for wildfires to spread (Rothermel, 1983).

Meteorological

Temperature (°C)

As vegetation burns, it increases air temperatures, which enables greater moisture capacity – resulting in greater reduction of relative humidity – making the vegetation drier and more flammable (Countryman, 1974). While in the night period temperatures drops, moisture rise leading

to the thermogenic sinks with respect to fire fuels; grab grasses and leaves which soak the moisture decreasing fire activity (Rothermel, 1983).

Wind speed (m/s)

Wildfire phenomena are highly dependent on winds as they are a source of oxygen which is supplied to the flames and blown towards unburnt areas which dries and pre-heats the vegetation resulting into faster fire spread (Albini, 1976). Winds aid in the broad transportation of embers from the main fire and ignite elsewhere in the spot fires. On the other hand, the winds on the uphill or in the funnel region, the drivers of the canyon, tend to behave like a chimney which further intensifies the speed at which the fire spreads (Sharples et al., 2012).

Relative humidity (%)

Fire ignition and spread potential increases along with increase in temperature and in combination with the decrease in humidity (Viegas, 1998).

Precipitation (mm)

Sprinkled fuels wildfire risk is reduced due to raising its moisture content and therefore, making them less likely to ignite (Bradstock et al., 2009).

Dew point (°C)

The dew point is the temperature at which condensation begins.

Land use and land cover (LULC)

Fuel types

Anthropogenic shifts in land use and land cover change pose perhaps the greatest risk to forest ecosystems and their vulnerability to wildfire (Agarwal, 2002; Kumar et al., 2018). Many studies have emphasized the significance that LULC, land use and land cover, classifications have on wildfire risk assessment, as varied land covers represent different types of fuels which have various characteristics of flammability (Szpakowski and Jensen, 2019).

Proximity to urban areas (m)

Areas near urban centers face a danger of wildfire when compared with distant rural locations. This high risk is primarily the result of human behavior, such as the reckless disposal of incendiary waste, construction activity, and thermal emissions from buildings and infrastructure. The further one moves from urban centers, the less they are influenced by such ignition sources and the lower the chances of fire outbreak (Raval and Motiani, 2022).

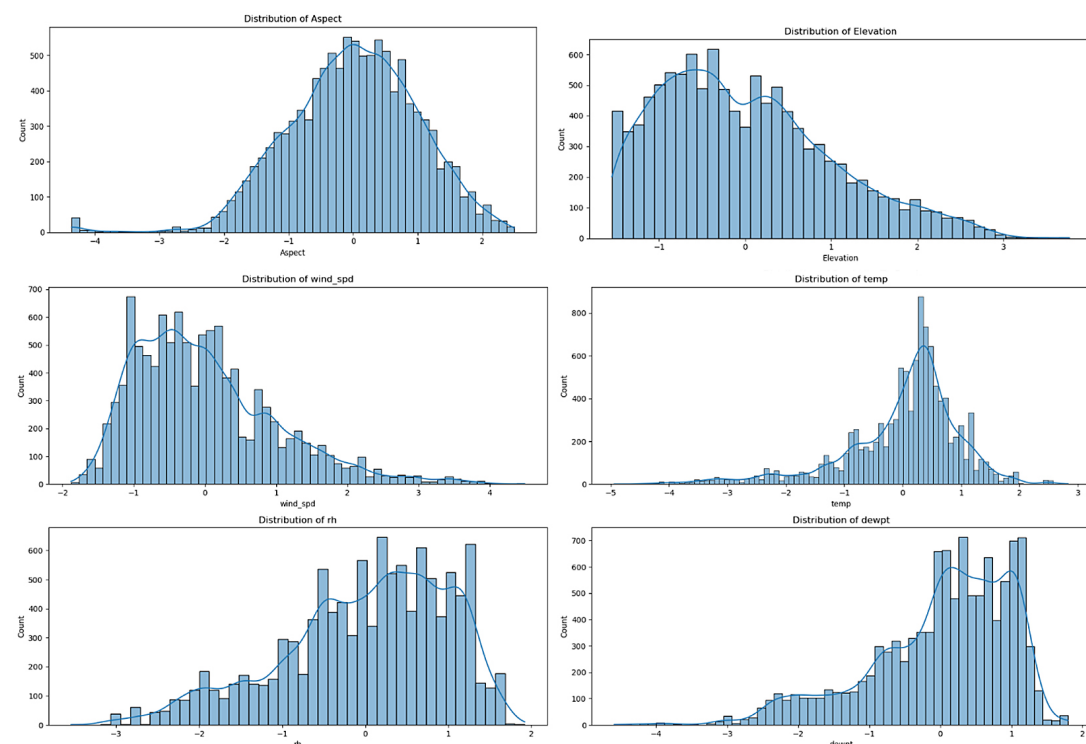


Figure 4. Normalized distribution of some key factors

Proximity to agricultural lands (meters)

Such areas are closer in proximity to agricultural lands and are more susceptible to wildfire than regions that are farther away. This is primarily the case because farming activities frequently involve heated machinery, the deliberate burning of crop residues, and other practices that may accidentally start fires.

Data preparation

A deep data analysis was performed to identify patterns and understand the relationships between different factors affecting fire risk. In order to identify the factors significantly contributing to fire risk, the study tended to explore the correlation between each of the individual attributes and the target variable (fire occurrence). Then outliers were identified and removed to maintain the integrity of the dataset. The dataset was then standardized using feature-engineering techniques, specifically normalization (Figure 4) to ensure a unified measurement scale. Fire occurrence was assigned 0 (no fire), 1 (near to fire) for grid cells adjacent to fire affected cell, representing areas surrounding an active grid cell, 2 (has fire) to ensure consistency. By ensuring that every feature was on a similar scale, this procedure served to reduce bias and enhance the accuracy and performance of the model.

XGboost

In this paper, XGBoost ML method for forecasting forest fire risk was implemented to ensure accuracy and resilience. It is an advanced version of Gradient Tree Boosting, increasing its scalability and efficiency for big datasets. It is widely used for forecasting modeling in fields like climate science, finance and forest fires prediction. The influence of the system has been extensively recognized in a variety of data mining and ML challenges. Consider the challenges hosted by Kaggle, a platform for ML competitions. Among the 29 winning solutions recently published on Kaggle’s blog, 17 utilized XGBoost. Eight of these solutions relied solely on XGBoost, while the majority combined it with other ensemble methods. These results show that our approach may provide state-of-art solution. The main reason for XGBoost’s success is its scalability across various scenarios. The system is capable of handling billions of examples in distributed or

memory-limited settings and operates more than ten times faster on a single machine than other widely used solutions.

Compared to random forest, XGBoost typically delivers superior accuracy because of boosting, whereas it requires more tuning and longer training times. Random Forest, on the other hand, offers greater stability and faster training that is more suitable for baseline models and simpler applications. Which affects our case study, because it entails complex interactions between several environmental factors. When compared to LightGBM, XGBoost is slower but typically more robust and generalized. In summary LightGBM is better for large scale and time sensitive tasks, while XGBoost is the best suited for scenarios where peak predictive performance is crucial and computing factors allow fine-tuning. Which aligns well with our study’s objective of predicting forest fire risk with high accuracy by properly tuning and handling of complex spatial and meteorological data.

XGBoost builds an additive extension of the object function by minimizing a loss function. Given that the only focus of XGBoost is on decision trees as base classifiers, the tree’s complexity is controlled using a variant of the loss function. The objective function in XGBoost includes two parts, loss function (1) and regularization term (2):

$$Lxgb = \sum_{i=1}^N L(y_i, F(x_i)) + \sum_{m=1}^M \Omega(hm) \quad (1)$$

$$\Omega(h) = \gamma T + \frac{1}{2} \lambda ||\omega||^2 \quad (2)$$

where: $L(y_i, F(x_i))$ – loss function that measures the model’s performance that is the difference between the actual value y_i and the predicted value $F(x_i)$; Ω – regularization term which controls the complexity of the model; T – number of leaves of the tree; ω – output scores of the leaves; λ – penalizing parameter, penalizes the squared weight of the leaves (leaves with large weights) to prevent overfitting; γ – regularization parameter that regulates the tree’s complexity.

Finally, the split gain equation is used to determine the best split at each node. The information gain for each split is calculated as:

$$Gain = \frac{1}{2} \left(\frac{(G_L + G_R)^2}{H_L + H_R + \lambda} - \frac{G_L^2}{H_L + \lambda} - \frac{G_R^2}{H_R + \lambda} \right) - \gamma \quad (3)$$

where: G_L, G_R – gradients sum for the right and left child nodes; H_L, H_R – Hessians sum for the right and left child nodes.

XGBoost improves decision tree-based learning through regularization strategies and computational optimizations. It combines a loss function with a pre-pruning technique, where the parameter gamma (γ) regulates tree complexity by figuring out the minimal gain needed for node complexity. Another regularization technique that improves generalization is shrinkage, which reduces the step size. Additionally, it employs randomization techniques such as subsampling of training data at different levels, to reduce overfitting and speed up training. Moreover, it optimizes split finding by employing a compressed column-based structure that pre-sorts data to avoid redundant sorting processes. This enables efficient parallel execution, minimizing computational complexity while maintaining high predictive performance.

The below parameters were tuned for XGBoost in this study:

- the minimum loss reduction: gamma,
- the learning rate: learning_rate,
- the maximum depth of the tree: max_depth,
- the subsampling rate: subsample,
- the fraction of features to be calculated at each split: colsample_bylevel.

The XGBoost model, recognized for its efficiency and speed, was used for managing missing data and capturing complex features relationships. The model can generate predictions without explicit imputation, because it handles missing values by figuring out the optimal split direction when encountering those allowing predictions to continue smoothly. Additionally, its gradient boosting mechanism creates trees sequentially, correcting errors from earlier trees and capturing

complex patterns in the data. The method effectively captures complex relationships while avoiding overfitting by the combination of feature selection, regularization techniques and efficient split finding algorithms.

Data analysis

Shapely additive explanations (SHAP) values utilized to improve the interpretability of the model and get a better understanding of how each factor contributed to the prediction outcomes. SHAP is a visualization tool used to interpret the results of ML models, especially complex ones like XGBoost. SHAP is a visualization tool used to interpret the result of ML models, especially complex ones like XGBoost, by assigning a value for each factor for a particular prediction. This technique aids in visualizing and measuring how various input factors affects the model’s prediction of a certain class. SHAP values breakdown each factor’s contribution to a particular prediction, improving the interpretability of the model’s output.

As shown in Figure 5, according to SHAP summary plot, the most influential factors for the ‘low’ fire risk class (class 0) are temperature, elevation and dew point. While high dew point and temperature values drove predictions away from the low risk category, high elevation values were linked to a higher likelihood of low fire risk.

As shown in Figure 6, according to SHAP summary plot for the ‘moderate’ fire risk class (class 1), elevation, solar radiation and dew point were the most influential factors. Higher solar radiation levels and lower elevation values were

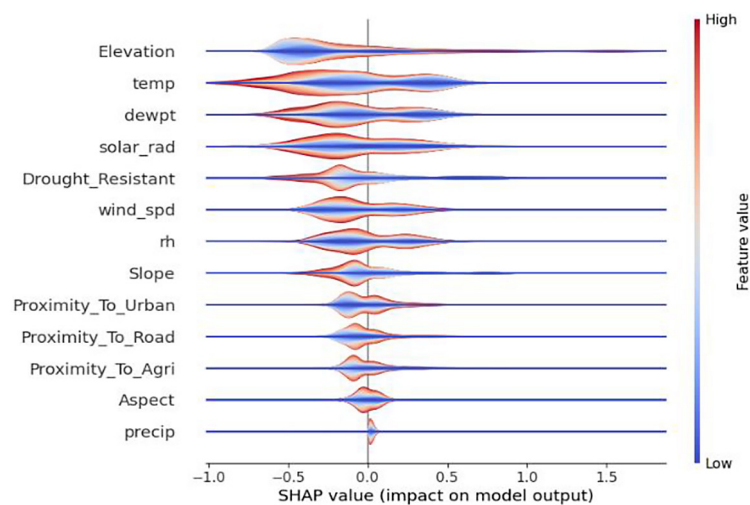


Figure 5. SHAP feature impact for low fire risk (class 0)

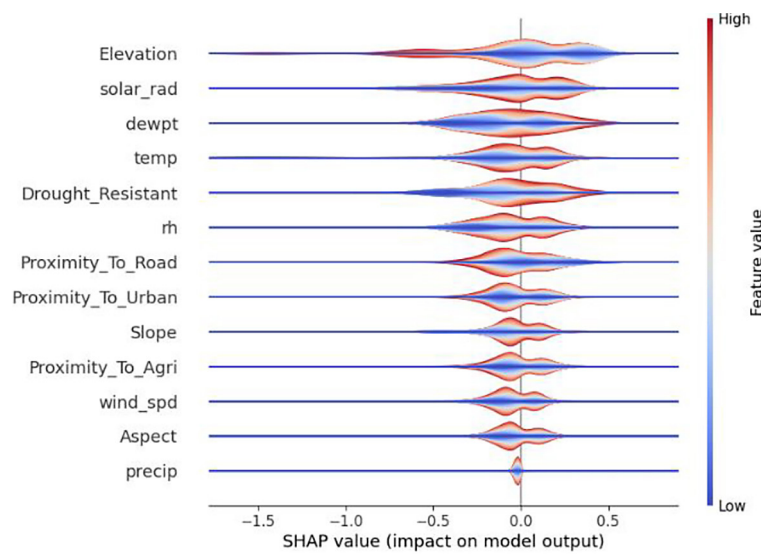


Figure 6. SHAP feature impact for moderate fire risk (class 1)

strongly linked to a higher likelihood of falling into the moderate risk category. In addition, higher dew pint values contributed positively to the forecast of moderate fire risk. These trends highlight the complex interaction between meteorological and topographic factors influence the likelihood of fire risk in the class.

As for Figure 7, the SHAP summary plot for the ‘high’ fire risk class (class 2) showed that elevation, temperature and drought resistance were the most influential factors. While low drought resistance contributed significantly to classify areas as high risk, high temperature and low elevation values were linked to a higher probability of high fire risk. These findings demonstrate that the

model can accurately capture significant spatial and climatic trends and are aligned with the environmental factors typically favorable to severe fire occurrences.

RESULTS AND DISCUSSIONS

Data splitting: train and test sets

Two subsets of the dataset were randomly selected: 20% for testing and 80% for training. This 80/20 split was selected to give the model a significant amount of data for learning while keeping enough data for assessing how well

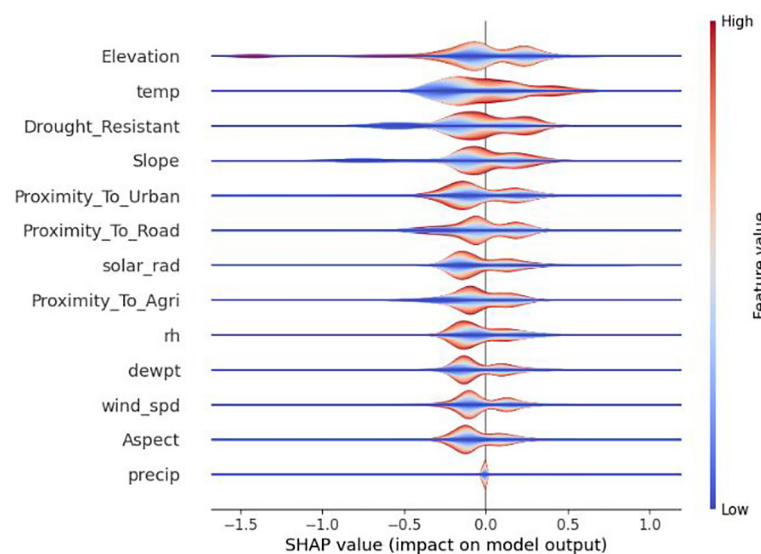


Figure 7. SHAP feature impact for high fire risk (class 2)

it performed on samples that had never been seen before. During the training phase, a k-fold cross-validation strategy was used to further enhance the model’s capacity for generalisation and lower the possibility of overfitting. In particular, the training data was split into five equal parts ($k = 5$). The model was then trained and validated five times, using the remaining four folds as the training set and a different fold as the validation set each time. To produce a more solid and trustworthy estimate of the model’s performance, the outcomes from each iteration were then averaged.

Applying and evaluating XGboost ML model

After installing XGboost library in Python, and after having our data loaded and prepared, XGboost model could be trained. At that point, we could make predictions using the obtained fit model on the test dataset.

For this experiment, the scikit-learn package for the XGBoost was used. The optimum parameters were estimated by applying Optuna for the training set and numerous parameter values was explored. Optuna is an open-source framework for hyperparameter optimization developed to automate the process of determining the best hyperparameters for ML models. It is a helpful approach for complex models like XGBoost, and other deep learning models because it employs an adaptable, effective and scalable method for hyperparameter tuning. The best set of parameters extracted are shown in Table 1.

Table 1. Default values and optimum values for each parameter after optimization

Parameter	Default value	Best values
Gamma	0	1
Learning_rate	0.1	0.0172
Max_depth	3	20
Colsample_by_level	1	0.48
Subsample	1	0.849

To thoroughly evaluate the performance of the model, a set of evaluation metrics were established, including precision, accuracy, recall and the area under receiver operating characteristic curve (ROC-AUC). These comprehensive evaluation measures were essential for capturing the model’s capability to distinguish between places that are prone to fire and those that are not. The evaluation metrics show that the model demonstrates remarkable accuracy and reliability. The confusion metrics shows high precision across the three different classes: 0 (no fire), 1 (near to fire), 2 (has fire) showing strong identification of actual fire risk scenarios. As observed in Table 2, various evaluation measures were derived, allowing for better insights. However, in our case, we are primarily concerned with minimizing false negatives (FN), because missing an actual fire could have serious consequences, potentially leading to uncontrollable fires. In this regard, metrics like ROC and recall are critical for predicting natural hazards. The ROC curve provides a comprehensive analysis of the model’s ability to distinguish between fire and non-fire events, while recall represents the proportion of actual fire occurrences correctly identified by the model. The recall values reported in the classification report for classes 0, 1, and 2 were 0.93, 0.94, and 0.96, respectively, demonstrating the model’s effectiveness in accurately predicting fire events with a mean squared error (MSE) of 0.0875 (Table 2).

Moreover, the ROC curve demonstrates that each class achieved an almost flawless AUC of 0.99, emphasizing the remarkable capability of the model to distinguish between different fire risks without producing false classification (Figure 8).

The integration of advanced technologies is essential for developing FFRM. The resulting map effectively identifies areas with varying levels of fire risk by leveraging GIS for spatial analysis, remote sensing for data collection and the power of ML techniques like XGBoost. The use of these technologies ensures that the map

Table 2. Model’s performance report

Parameter	Precision	Recall	F1-score
Low risk (0)	0.97	0.93	0.95
Moderate risk (1)	0.92	0.94	0.93
High risk (2)	0.94	0.96	0.95
Accuracy	0.94		
MSE	0.0875		

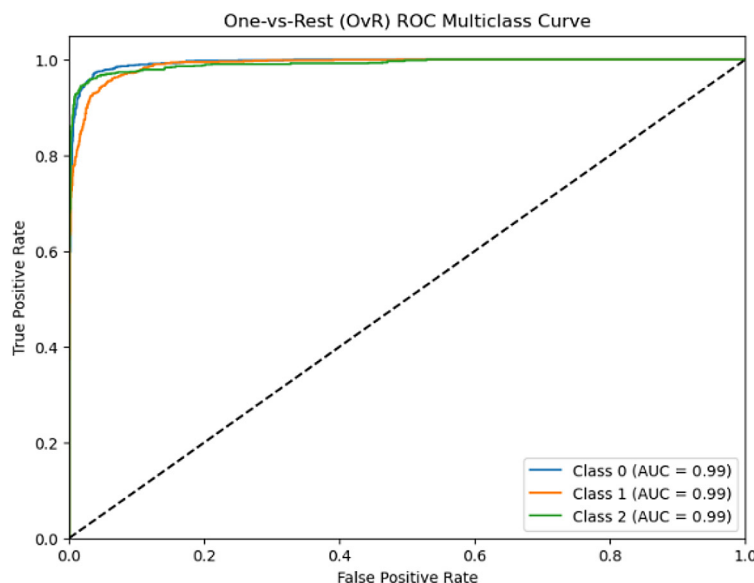


Figure 8. Model testing ROC-AUC

accurately reflects real world conditions by capturing critical details such as vegetation changes (Figure 9). Additionally, this map is crucial tool for strategic planning and offers value beyond merely identifying risk. It can be used by emergency response teams, stakeholders and local authorities to allocate resources efficiently and prioritize fire risk areas for preventive actions. By addressing risks before escalating into active fires, this strategic use of the map greatly enhances the response operations.

Validation of forest fire risk susceptibility map

Model Validation is critical to ensure that its predictions accurately mirror real-world fire incidents. This includes evaluating the model’s outputs by comparing the anticipated fire risk areas with the actual fire occurrences. The effectiveness and accuracy of the model in detecting high risk areas can then be assessed by this comparison. Additionally, it aids in refining the model, ensuring that it offers reliable insights for risk assessment.

In this study, the model was run to predict the fire risk zoning for some days during the summer of 2024, as a spot check, to verify the accuracy of the FFRM. Then the anticipated fire risk zones were validated by the actual fire events by overlaying the events on the risk map (Hagos et al., 2022; Ogato et al., 2020). The percentage of the area for each fire risk zone was then calculated using the spatial analyst tool in GIS as described in Equation 4.

$$\text{Percentage of risk area (\%)} = \left(\frac{\text{estimated area}}{\text{total area}} \right) \times 100 \quad (4)$$

The FFRM generated from this evaluation categorizes the area into three different zones, high, medium and low risk alongside with fire incidents from the specified dates during summer 2024 (Figure 9).

The risk maps in Figure 9 illustrate the spatial distribution of forest fire risk across Lebanon on three distinct dates: June 6, 2024, June 30, 2024, and July 12, 2024. These maps highlight how environmental conditions, seasonal variations, and human activities contribute to dynamic changes in fire risk levels. The validation of the model was done by comparing the predicted against the actual fire occurrence. To assess the accuracy of the model, the spatial distribution of anticipated high-risk areas was analyzed against actual fire incidents.

June 6, 2024 Validation: the map indicates that the majority of Lebanon’s wooded areas were at moderate to low risk of fire. High-risk zones (red) were primarily concentrated in the central and southern Lebanon. Validation using recorded fire occurrences showed a small number of fires, aligning with the model’s prediction.

June 30, 2024 Validation: The fire risk increased significantly, particularly in northern and central Lebanon. Previously medium-risk areas were reclassified as high-risk, indicating intensifying fire hazards. The model’s accuracy was further validated by actual fire incidents, as 18 fires occurred within high-risk zones, compared to 2

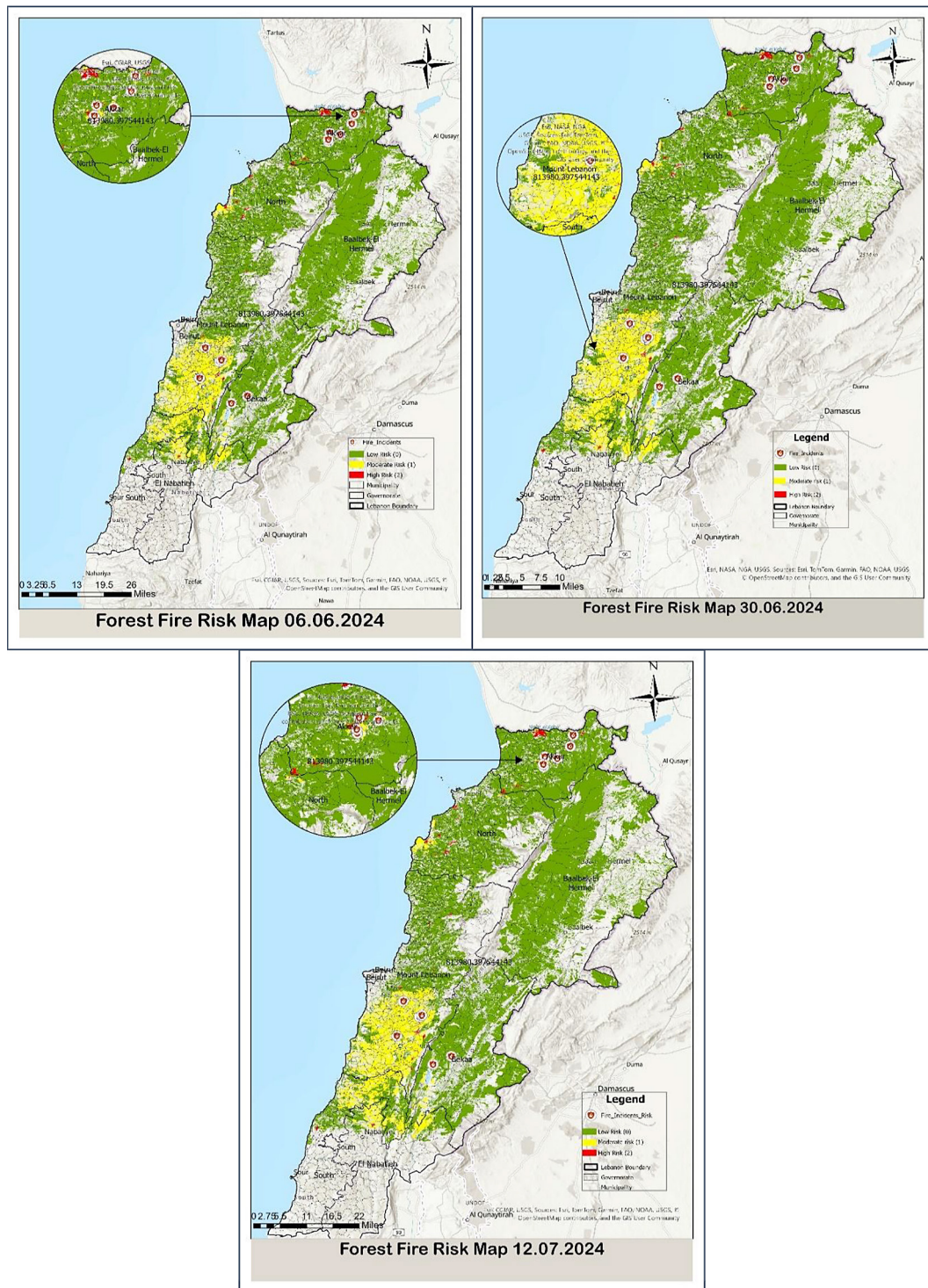


Figure 9. Forest fire risk maps of the specified dates

in medium-risk and 1 in low-risk areas (Table 1). July 12, 2024 Validation: fire risk remained high, with severe conditions especially in central and southern regions where dense vegetation and dry weather conditions persisted. Fire incidents during this period confirmed the intensity of the fire season. Validation showed 16 fires in high-risk zones, 3 in medium-risk, and 2 in low-risk areas (Table 1).

The validity of the model was confirmed by comparing its predictions with real-time fire occurrences, revealing a strong correlation between the results and actual fire events. The analysis of fire incidents further demonstrated the model's reliability. This analysis underscores the importance of continuous monitoring integrating real time data for enhancing fire risk assessment and

Table 3. Distribution of forest fire risk per area in Lebanon for the specified dates

Days	Fire susceptibility	Frequency	Area (Hectares)	Area (%)	Number of fire incidents
Day 1 (06.06.2024)	URBAN	9546	102948.962368	15.92	Null
	Low risk	21679	475299.397048	73.51	1
	Medium risk	5369	64564.01078	9.98	1
	High risk	304	3808.395165	0.59	10
Day 2 (30.06.2024)	URBAN	9489	101,382.62	15.68	Null
	Low risk	18867	343,493.43	53.14	1
	Medium risk	6835	110,754.83	17.13	2
	High risk	1707	90,989.88	14.06	18
Day 3 (12.07.2024)	URBAN	10790	132,007.67	20.42	Null
	Low risk	25172	475,551.35	73.55	2
	Medium risk	157	10,373.04	1.60	3
	High risk	779	28,688.70	4.43	16

emergency response strategies. Public access to this forest fire risk map can play a vital role in community awareness and readiness. By providing an online interactive dashboard, rapid response will be enhanced by taking actionable insights for decision-making taking into consideration the capacities and vulnerabilities particularly within the high-risk areas. This proactive community involvement helps in lowering the susceptibility to forest fires. The following table categorizes the area into three different levels of fire susceptibility (Table 3).

Limitations

The validation period was relatively short, which may introduce seasonal bias. During this period, most of the fire events reported by field missions were too small in size and intensity to be detected by satellite sensors, which may affected the completeness and accuracy of the fire mapping and validation process. Despite the model’s high accuracy, it does not include uncertainty estimation, which will be addressed in future research to increase the reliability of the prediction model.

Future work

Future studies should consider comparing XGBoost with other ML models like SVM, Random Forest, Adaboost and other techniques. This comparison can assist in identifying the best algorithm in terms of speed, accuracy and scalability for forest fire risk forecasting in Lebanon. Exploring these models will improve the knowledge of their strengths and limitations and aid

in developing more optimized and adaptable models. While the validation was conducted on three different dates (June 6, June 30, and July 12, 2024), these were chosen to reflect different phases of fire season in Lebanon. However, this limited timeline may introduce seasonal bias. To further enhance model resilience and better capture inter seasonal variability, future research is recommended to extend the validation period, supported by developing an active, real-time dashboard that shows the forest fire risk map on a daily basis, enabling continuous monitoring and evaluation by linking it to the missions of First Responders. Forecast reliability and interpretability would be enhanced by integrating prediction uncertainty into the XGBoost model. This could be accomplished through some methods like, Quantile regression or applying bootstrapping to train several models on resampled data and examine prediction distributions.

CONCLUSIONS

The model has successfully demonstrated the development and application of a comprehensive FFRM for Lebanon, by the integration of advanced technologies including GIS, remote sensing and XGboost ML model. By analyzing 13 factors affecting forest fire risk across Lebanon: temperature, relative humidity, wind speed, solar radiation, precipitation, dew point, proximity to roads, urban and agriculture, elevation, aspect, slope and fuel type, the model achieved high prediction accuracy of 94%, enabling accurate classification of fire risk areas and provided a

robust tool for mitigation efforts and strategic fire management in Lebanon. To validate the model's performance, fire risk anticipations for three dates were compared with actual fire events. The analysis focused on the alignment between the recorded fire events and the predicted high-risk zones, providing an evaluation for the accuracy of the model. These results validate the model's capability in improving emergency response and resources allocation in addition to its potential to contribute to enhancing forest fire management in Lebanon. Hence, this map might assist disaster risk reduction committees, emergency response teams, and other stakeholders in preventing or mitigating forest fire risk. Based on the above, ensemble ML method is employed to accurately forecast fire risk in Lebanon. It proposes a holistic approach for handling the complexity of forest fire risk in Lebanon. Moreover, the final dashboard is an invaluable asset for enhancing community awareness and proactive engagement for mitigating the risk of fires before they could escalate. Nevertheless, the application of ensemble machine learning models, highlighted the effectiveness of integrating different techniques to enhance reliability and accuracy of the fire risk forecasting. Future work should focus on continually updating real-time data and improving the predictive models by adding more variables and exploring newer ML techniques to contribute in the enhancement of forest fire science.

REFERENCES

- Lai, C., Zeng, S., Guo, W., Liu, X., Li, Y., & Liao, B. (2022). Forest fire prediction with imbalanced data using a deep neural network method. *Forests*, *13*, 1129. <https://doi.org/10.3390/f13071129>
- Molina, J. R., Herrera, M. A., & Silva, F. R. (2019). Wildfire-induced reduction in the carbon storage of Mediterranean ecosystems: An application to brush and forest fires impacts assessment. *Environmental Impact Assessment Review*, *76*, 88–97. <https://doi.org/10.1016/j.eiar.2019.01.004>
- Venkatesh, K., Preethi, K., & Ramesh, H. (2020). Evaluating the effects of forest fire on water balance using fire susceptibility maps. *Ecological Indicators*, *110*, 105856. <https://doi.org/10.1016/j.ecolind.2020.105856>
- Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, *28*(4), 478–505. <https://doi.org/10.1139/er-2020-0019>
- Vilagrosa, A., Bautista, S., & Vallejo, V. R. (2018). Redirecting fire-prone Mediterranean ecosystems toward more resilient and less flammable communities. *Journal of Environmental Management*, *215*, 114–122. <https://doi.org/10.1016/j.jenvman.2018.03.063>
- Global Forest Watch. (2024). Lebanon fires dashboard. Retrieved July 30, 2024, from <https://www.globalforestwatch.org/dashboards/country/LBN/?category=fires>
- Bui, T. D., Bui, Q. T., Nguyen, Q. P., Pradhan, B., Nampak, H., & Trinh, P. T. (2017). A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agricultural and Forest Meteorology*, *233*, 32–44.
- Vecín-Arias, D., Castedo-Dorado, F., Ordóñez, C., & Rodríguez-Pérez, J. R. (2016). Biophysical and lightning characteristics drive lightning-induced fire occurrence in the central plateau of the Iberian Peninsula. *Agricultural and Forest Meteorology*, *225*, 36–47. <https://doi.org/10.1016/j.agrformet.2016.05.003>
- El Hafyani, M., Essahlaoui, A., Van Rompaey, A., Mohajane, M., El Hmaid, A., El Ouali, A., ... & Serrhini, N. E. (2020). Assessing regional scale water balances through remote sensing techniques: A case study of Boufakrane River Watershed, Meknes Region, Morocco. *Water*, *12*(2), 320. <https://doi.org/10.3390/w12020320>
- Teodoro, A. C., & Duarte, L. (2013). Forest fire risk maps: A GIS open source application – A case study in Norwest of Portugal. *International Journal of Geographical Information Science*, *27*, 699–720. <https://doi.org/10.1080/13658816.2012.721554>
- Mohajane, M., Costache, R., Karimi, F., Pham, Q. B., Essahlaoui, A., Nguyen, H., ... & Oudija, F. (2021). Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecological Indicators*, *129*, 107869. <https://doi.org/10.1016/j.ecolind.2021.107869>
- Wang, Z., Wang, B., Liu, M., Tang, Y., Qian, Y., & Liu, J. (2022). Simulation of forest fire spread based on artificial intelligence. *Ecological Indicators*, *139*, 108653. <https://doi.org/10.1016/j.ecolind.2022.108653>
- Ghali, R., Jmal, M., Souidene Mseddi, W., & Attia, R. (2020). Recent advances in fire detection and monitoring systems: A review. In *Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'18)*, 1, 332–340. Springer International Publishing.
- Faour, G. (2004). Forest fire fighting in Lebanon using remote sensing and GIS. *Technical Report*.

15. Faour, G., Kheir, R. B., & Verdeil, É. (2006). Caractérisation sous système d'information géographique des incendies de forêts: L'exemple du Liban. *Forêt Méditerranéenne*, 27(4), 339–352.
16. Masri, T., Khawlie, M., Faour, G., & Awad, M. (2003, June). Mapping forest fire prone areas in Lebanon. In *Proceedings of the EARSeL 23rd Symposium of "Remote Sensing in Transition—4th International Workshop on Remote Sensing and GIS Applications to Forest Fire Management* 109–113.
17. Yu, B., Chen, F., Li, B., Wang, L., & Wu, M. (2017). Fire risk prediction using remote sensed products: A case of Cambodia. *Photogrammetric Engineering & Remote Sensing*, 83(1), 19–25. <https://doi.org/10.14358/PERS.83.1.19>
18. VanBeusekom, A. E., Gould, W. A., Monmany, A. C., Khalyani, A. H., Quiñones, M., Fain, S. J., et al. (2018). Fire weather and likelihood: Characterizing climate space for fire occurrence and extent in Puerto Rico. *Climatic Change*, 146(1–2), 117–131. <https://doi.org/10.1007/s10584-017-2045-6>
19. Adab, H. (2017). Landfire hazard assessment in the Caspian Hyrcanian forest ecoregion with the long-term MODIS active fire data. *Natural Hazards*, 87(3), 1807–1825. <https://doi.org/10.1007/s11069-017-2850-2>
20. Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *CATENA*, 125, 91–101. <https://doi.org/10.1016/j.catena.2014.10.017>
21. Pham, Q. B., Mohammadpour, R., Linh, N. T. T., Mohajane, M., Pourjasem, A., Sammen, S. S., Anh, D. T. (2021). Application of soft computing to predict water quality in wetland. *Environmental Science and Pollution Research*, 28(1), 185–200. <https://doi.org/10.1007/s11356-020-10344-8>
22. Pham, Q. B., Yang, T. C., Kuo, C. M., Tseng, H. W., & Yu, P. S. (2019). Combining random forest and least square support vector regression for improving extreme rainfall downscaling. *Water*, 11(3), 451. <https://doi.org/10.3390/w11030451>
23. Pour, S. H., Wahab, A. K. A., & Shahid, S. (2020). Physical-empirical models for prediction of seasonal rainfall extremes of Peninsular Malaysia. *Atmospheric Research*, 233, 104720. <https://doi.org/10.1016/j.atmosres.2019.104720>
24. Chen, W., Yan, X., Zhao, Z., Hong, H., Haoyuan, B., Bui, D. T., & Pradhan, B. (2019). Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes, and RBF-Network models for the Long County area (China). *Bulletin of Engineering Geology and the Environment*, 78(1), 247–266. <https://doi.org/10.1007/s10064-018-1256-z>
25. Pham, B. T., Prakash, I., & Bui, D. T. (2018). Spatial prediction of landslides using a hybrid machine learning approach based on Random Subspace and Classification and Regression Trees. *Geomorphology*, 303, 256–270. <https://doi.org/10.1016/j.geomorph.2017.12.008>
26. Truong, X., Mitamura, M., Kono, Y., Raghavan, V., Yonezawa, G., Truong, X., Do, T., Tien Bui, D., & Lee, S. (2018). Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree. *Applied Sciences*, 8, 1046. <https://doi.org/10.3390/app8071046>
27. Akay, H. (2021). Flood hazards susceptibility mapping using statistical, fuzzy logic, and MCDM methods. *Soft Computing*, 1–17. <https://doi.org/10.1007/s00542-021-06363-5>
28. Moayedi, H., Mehrabi, M., Bui, D. T., Pradhan, B., & Foong, L. K. (2020). Fuzzy-metaheuristic ensembles for spatial assessment of forest fire susceptibility. *Journal of Environmental Management*, 260, 109867. <https://doi.org/10.1016/j.jenvman.2019.109867>
29. Zohary, M. (1973). *Geobotanical foundations of the Middle East*.
30. CNRS-L. (2019). *Land Cover/Land Use Map of Lebanon* (Scale 1:20,000).
31. NASA FIRMS. (n.d.). FIRMS: Fire Information for Resource Management System. NASA. Retrieved August 26, 2024, from <https://firms.modaps.eosdis.nasa.gov/>
32. NASA. (n.d.). ASTER Global Digital Elevation Model (GDEM). NASA Earthdata. Retrieved from <https://earthdata.nasa.gov/>
33. Esri. (n.d.). How aspect works. Retrieved August 26, 2024, from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-aspect-works.htm>
34. Raval, P., & Motiani, R. (2022, August). Review of fire risk factors for fire risk assessment in urban areas: The case of Ahmedabad, India. In *International Conference on Innovation in Smart and Sustainable Infrastructure* (pp. 607–623). Springer Nature Singapore.
35. Agarwal, C. (2002). A review and assessment of land-use change models: Dynamics of space, time, and human choice.
36. Rothermel, R. C. (1983). *How to predict the spread and intensity of forest and range fires* (Vol. 143). US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment.
37. Kumar, M., Denis, D. M., Singh, S. K., Szabó, S., & Suryavanshi, S. (2018). Landscape metrics for assessment of land cover change and fragmentation of a heterogeneous watershed. *Remote Sensing Applications: Society and Environment*, 10, 224–233.

- <https://doi.org/10.1016/j.rsase.2018.02.007>
38. Vadrevu, K. P., Eaturu, A., & Badarinath, K. (2010). Fire risk evaluation using multicriteria analysis—a case study. *Environmental Monitoring and Assessment*, 166, 223–239. <https://doi.org/10.1007/s10661-009-0990-7>
39. Szpakowski, D. M., & Jensen, J. L. (2019). A review of the applications of remote sensing in fire ecology. *Remote Sensing*, 11(22), 2638. <https://doi.org/10.3390/rs11222638>
40. Majdalani, G., Koutsias, N., Faour, G., Adjizian-Gerard, J., & Mouillot, F. (2022). Fire regime analysis in Lebanon (2001–2020): Combining remote sensing data in a scarcely documented area. *Fire*, 5(5), 141. <https://doi.org/10.3390/fire5050141>
41. Pinto, M. M., Trigo, R. M., Trigo, I. F., & DaCamara, C. C. (2021). A practical method for high-resolution burned area monitoring using Sentinel-2 and VIIRS. *Remote Sensing*, 13(9), 1608. <https://doi.org/10.3390/rs13091608>
42. Hagos, Y. G., Andualem, T. G., Yibeltal, M., & Mengie, M. A. (2022). Flood hazard assessment and mapping using GIS integrated with multi-criteria decision analysis in upper Awash River basin, Ethiopia. *Applied Water Science*, 12(7), 148. <https://doi.org/10.1007/s13201-022-01616-4>
43. Ogato, G. S., Bantider, A., Abebe, K., & Geneletti, D. (2020). Geographic information system (GIS)-based multicriteria analysis of flooding hazard and risk in Ambo Town and its watershed, West Shoa Zone, Oromia Regional State, Ethiopia. *Journal of Hydrology: Regional Studies*, 27, 100659. <https://doi.org/10.1016/j.ejrh.2020.100659>
44. Ibini, F. A. (1976). *Estimating wildfire behavior and effects*. USDA Forest Service, General Technical Report INT-30.
45. Bradstock, R. A., Gill, A. M., & Williams, R. J. (2009). *Flammable Australia: Fire regimes, biodiversity and ecosystems in a changing world*. CSIRO Publishing.
46. Countryman, C. M. (1974). *Can southern California wildland conflagrations be stopped?* USDA Forest Service General Technical Report PSW-7.
47. Müller, M., Schunk, C., & Seidl, R. (2022). Fuel moisture dynamics and wildfire risk under future climate conditions. *Environmental Research Letters*, 17(5), 054012. <https://doi.org/10.1088/1748-9326/ac68b6>
48. Sharples, J. J., McRae, R. H. D., & Wilkes, S. R. (2012). Wind–terrain effects on the propagation of wildfires in rugged terrain: Fire channelling. *International Journal of Wildland Fire*, 21(3), 282–296. <https://doi.org/10.1071/WF10055>
49. Viegas, D. X. (1998). Weather, fuel status and fire occurrence: Predicting large fires. In *Proceedings of III International Conference on Forest Fire Research* 31–48.
50. Westerling, A. L., Hidalgo, H. G., Cayan, D. R., & Swetnam, T. W. (2006). Warming and earlier spring increase western US forest wildfire activity. *Science*, 313(5789), 940–943. <https://doi.org/10.1126/science.1128834>