



Article

Data-Driven Approaches for Wildfire Mapping and Prediction Assessment Using a Convolutional Neural Network (CNN)

Rida Kanwal ¹, Warda Rafaqat ¹, Mansoor Iqbal ² and Song Weiguo ^{1,*}

¹ State Key Laboratory of Fire Science, University of Science and Technology of China (USTC), Hefei 230027, China; ridakanwal@mail.ustc.edu.cn (R.K.); warda@mail.ustc.edu.cn (W.R.)

² Department of Electronic Engineering & Information Science, University of Science and Technology of China (USTC), Hefei 230027, China; man2017@mail.ustc.edu.cn

* Correspondence: wgsong@ustc.edu.cn

Abstract: As wildfires become increasingly perilous amidst Pakistan's expanding population and evolving environmental conditions, their global significance necessitates urgent attention and concerted efforts toward proactive measures and international cooperation. This research strives to comprehensively enhance wildfire prediction and management by implementing various measures to contribute to proactive mitigation in Pakistan. Additionally, the objective of this research was to acquire an extensive understanding of the factors that influence fire patterns in the country. For this purpose, we looked at the spatiotemporal patterns and causes of wildfires between 2000 and 2023 using descriptive analysis. The data analysis included a discussion on density-based clustering as well as the distribution of the data across four seasons over a period of six years. Factors that could indicate the probability of a fire occurrence such as weather conditions, terrain characteristics, and fuel availability encompass details about the soil, economy, and vegetation. We used a convolutional neural network (CNN) to extract features, and different machine learning (ML) techniques were implemented to obtain the best model for wildfire prediction. The majority of fires in the past six years have primarily occurred during the winter months in coastal locations. The occurrence of fires was accurately predicted by ML models such as random forest (RF), which outperformed competing models. Meanwhile, a CNN with 1D and 2D was used for more improvement in prediction by ML models. The accuracy increased from an 86.48 to 91.34 accuracy score by just using a CNN 1D. For more feature extraction, a CNN 2D was used on the same dataset, which led to state-of-the-art prediction results. A 96.91 accuracy score was achieved by further tuning the RF model on the total data. Data division by spatial and temporal changes was also used for the better prediction of fire, which can further be helpful for understanding the different prospects of wildfire. This research aims to advance wildfire prediction methodologies by leveraging ML techniques to explore the benefits and limitations of capturing complex patterns and relationships in large datasets. Policymakers, environmentalists, and scholars studying climate change can benefit greatly from the study's analytical approach, which may assist Pakistan in better managing and reducing wildfires.

Keywords: machine learning; wildfire assessment; CNN; random forest; fire occurrence



Citation: Kanwal, R.; Rafaqat, W.; Iqbal, M.; Weiguo, S. Data-Driven Approaches for Wildfire Mapping and Prediction Assessment Using a Convolutional Neural Network (CNN). *Remote Sens.* **2023**, *15*, 5099. <https://doi.org/10.3390/rs15215099>

Academic Editors: Sawaid Abbas, Janet E. Nichol, Faisal M. Qamer and Jianchu Xu

Received: 6 September 2023

Revised: 20 October 2023

Accepted: 21 October 2023

Published: 25 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The global risk of wildfire disasters has surged due to the effects of climate change. These wildfires pose a significant threat to wildland ecosystems, affecting their formation and stability on a global scale [1,2]. While fires have historically played a role in renewing and reshaping landscapes, the intensification of wildfires has led to devastating consequences for people, the environment, and economies [3]. Moreover, the impacts of wildfires extend beyond plants and animals, affecting the air, soil, and water quality [4,5]. The economic losses and threats to human lives and health are substantial [6–9].

The escalation in the frequency and severity of wildfires has been particularly noticeable across several Asian countries [10–12]. Research strongly suggests that with the

ongoing impact of climate change, wildfires are projected to become more frequent and intense worldwide [13–17]. This phenomenon is also evident in Pakistan, where forest cover is notably low, accounting for just approximately 4.5 percent of the land. Consequently, Pakistan grapples with significant forest degradation issues [16].

During the latter half of the 20th century, there was a noticeable increase in fire occurrences in Pakistan and other nations, surpassing historical records. This period also witnessed a rise in large-scale wildfires and an expansion of their affected areas [2].

Understanding the factors that influence wildfire behavior such as ignition ease, rate of spread, management challenges, and overall impact is often referred to as “wildfire hazard” assessment [17]. Several components play a role in the occurrence and progression of wildfires including favorable weather conditions, the presence of combustible materials, the continuity of these materials, and potential ignition sources [18]. These factors can be categorized into topography, weather, and fuel-related factors. Each factor’s spatial and temporal variability contributes differently to the overall fire risk.

Among these factors, climatic conditions have gained substantial attention in the context of wildfire changes [19,20]. The influence of topographic elements such as slope, aspect ratio, and elevation on the burnt area and ignition density has also been extensively studied [21]. Notably, fuel—comprising various plant species—plays a critical role in fire initiation and propagation [22,23]. While natural causes are important, human activities significantly contribute to fire incidents, impacting both ignition likelihood and subsequent fire behavior [24–26].

Human factors including population dynamics, socioeconomic elements, land use changes, and activities such as agriculture have a substantial impact on fuel availability and thus fire variation [27–29]. However, these human-related variables exhibit complex temporal and spatial dynamics.

Considering the availability of thermal sensors and RADAR systems, this research utilized space-borne remote sensing to derive relevant predictor variables (e.g., temperature, precipitation, population) from open-source satellite imagery, addressing the challenges associated with modeling numerous environmental and socioeconomic factors. Satellite remote sensing has emerged as a crucial tool for monitoring ecosystems and identifying potential hazards such as wildfires.

Recent advancements in ML algorithms offer significant potential for various data science applications [30–33]. These algorithms require high-quality training datasets to excel. To address this issue, remote sensing data are enhanced using a CNN. Previously, researchers utilized machine learning techniques, along with remotely sensed fire data and GIS, to develop a wildfire susceptibility map for the Adana and Mersin provinces in Turkey, aiming to identify and predict areas at high risk of wildfires by analyzing multiple factors [34]. Another researcher explored the use of remote sensing data and machine learning algorithms to predict and map wildfire occurrences over multiple time periods, providing valuable insights into the temporal patterns of wildfires through the generation of grid maps [35]. One further study introduced a machine-learning approach to identify and assess dense-fire events and their atmospheric emissions on the Indochina peninsula, utilizing data from 2010 to 2020 and providing insights into the impact of these fires on the atmosphere [36].

This study delved into ML techniques such as decision trees (DTs), RFs, K-nearest neighbors (KNNs), and support vector machines (SVMs) to map wildfires, with the goal to identify the most effective approach for model training and validation. Furthermore, to augment the features used in modeling, CNNs were employed—both 1D and 2D—to demonstrate their capabilities in wildfire assessment [37–42].

Compared to physics-based models, ML models train on satellite data and climate features from monitoring stations, covering different span periods. By incorporating historical data, ML models can learn from past patterns and trends, enabling them to make predictions based on historical patterns of wildfire occurrence and behavior. Physics-based models may require detailed knowledge of physical processes, which may not

be readily available or practical for all regions. ML models offer improved prediction power by extracting patterns and relationships from complex datasets, resulting in more accurate predictions of wildfire occurrence. Being relatively novel, ML models provide fresh approaches to wildfire prediction, potentially uncovering new insights. Additionally, ML models can handle imbalanced datasets, which is crucial for obtaining reliable predictions in wildfire analysis [43]. The previous study emphasized the computational efficiency of ML models for fire forecasting, offering significant advantages over physics-based simulations. The integration of reduced-order modeling and ML ensures efficiency while maintaining accurate predictions. Data assimilation and error covariance tuning improve forecasting accuracy by incorporating real-time data. ML models accelerate fire forecasting, enabling near real-time predictions for effective decision-making and emergency response [44]. The application of existing models in other regions can be impracticable or very difficult due to distinct geographic characteristics and data availability. ML models, trained using a dataset specific to the Brazilian Federal District and enriched with various features, offer adaptability and outperform physics-based models, achieving an accuracy of 91% in predicting wildfire impacts, making them valuable for fire management agencies [45].

Despite the existing research, there remains a gap in thoroughly assessing fire susceptibility using remote sensing data. Previous research has highlighted the application of artificial neural networks and mapping techniques for fire risk assessment and vegetation analysis [46–48]. This study addresses this gap by focusing on a specific region with limited datasets. The research aims to build improved prediction models that can provide tailored recommendations for fire management and prevention, considering geographic, temporal, and seasonal aspects.

This research aimed to comprehensively assess wildfire prediction and management by evaluating model durability and reliability over an extended period, exploring the impact of seasonality on accuracy, capturing patterns linking precipitation and wildfire events, analyzing spatial distribution, developing targeted strategies, assessing advanced machine learning models, integrating real-time data, constructing a comprehensive management framework, implementing preemptive measures, and improving forecasting accuracy through data refinement. These objectives contribute to proactive wildfire mitigation in Pakistan. This study seeks to holistically address the wildfire conundrum in Pakistan by shedding light on its multifaceted dimensions. The overarching objective is to delve into the temporal and spatial facets of wildfire occurrences, aiming to advance our comprehension of their patterns and behaviors. At the heart of this pursuit is the integration of cutting-edge machine learning techniques including a CNN 2D and RF to bolster the accuracy of wildfire prediction models. The driving force behind this research is the aspiration to not only contribute to the academic discourse, but also to offer practical insights that have tangible implications for policymaking, disaster management frameworks, and sustainable community development in the face of escalating wildfire risks.

The primary objectives of this study encompass four key aspects. First, the study aims to undertake comprehensive data collection including historical fire data, and subsequently preprocesses the data by rectifying errors and standardizing formats. This process will involve the acquisition of pertinent details such as location, date, and pertinent factors such as temperature. Second, the study intends to perform model training and evaluation. This involves the identification of influential features, the division of data for training and testing purposes, and the training of a model—potentially employing techniques such as RF. Model performance evaluation will rely on metrics such as accuracy and recall. Third, the study seeks to acquire and refine data from the FIRMS dataset, which is utilized for fire detection. This refined dataset will be subjected to spatial analysis, allowing for a comparison with historical fire data and an assessment of alignment. Finally, the study will focus on identifying causal factors through an analysis of feature importance within the model. Insights garnered from this analysis will be leveraged to propose effective strategies aimed at fire management, prevention, and early detection. Collaboration with experts in the field will further enhance the applicability and efficacy of these recommendations.

The structure of the study comprises various sections: Section 2 gives an overview of the research area and past fire incidents; Section 3 details the techniques used and presents results; Section 4 covers data preparation, a CNN design, accuracy assessment, and key findings; Section 5 discusses the statistics and findings. Finally, Section 6 concludes the study with a brief summary.

2. Study Area

The research region is Pakistan, and the study period was from November 2000 to December 2020. Pakistan is situated in the Arabian Sea in South Asia. A total of 875,175 km² of surface area in Pakistan is made up of 25,683 km² of wildland, and 33.3% is agriculture. There are 292,882 km² of grassland and shrub land overall. The percentage of wilderness in Pakistan is 36.4%. There were 175,000 fires reported in Pakistan, of which 22,311 were wildfires. According to land cover categorization, Figure 1 represents the wilderness area of Pakistan, which is categorized as wildness. The area of interest is shown on the map, along with historical fire occurrences that occurred between 2001 and 2020. One of the most vulnerable regions of Pakistan is its eastern region. The least rainy and most prone to desertification weather, dry weather, is featured. The maximum elevation reaches 28,251 m, whereas the minimum altitude is at sea level, which is 0 m [49]. The northern region of the country is characterized by mountainous terrain, while the southern areas gradually transition into flatter landscapes. Notably, the northeast region exhibits the highest population density in the nation. The population is very evenly dispersed over the nation, with population clusters also being visible close to the largest cities. Southwest Pakistan is the least inhabited part of the country [50].

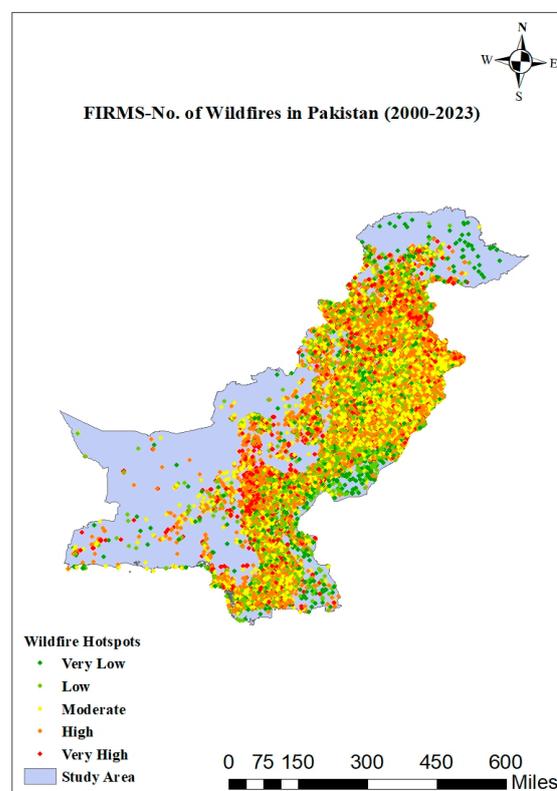


Figure 1. Map depicting the wildfire study area in Pakistan from November 2000 to October 2023.

Pakistan experiences four distinct seasons throughout the year. From December to February, the country experiences a cool and dry winter. The months of March to May mark the arrival of a hot and dry spring season. Then, from June to August, Pakistan experiences a wet monsoon season characterized by rainfall. Finally, the monsoon season gradually

subsides from September to November, known as the waning monsoon season. Previous wildfire studies were not undertaken on a national scale due to the absence of a continuous dataset that spanned the entire country. Due to the excellent remotely sensed datasets, the best model for fire prediction may be studied using a publicly available worldwide dataset collected from remotely sensed data.

After defining the essential terms, we conducted an exploratory investigation into Pakistani wildfires. Our research approach involved using diverse remotely sensed datasets that are publicly accessible. Furthermore, we integrated datasets from other industries that also relied on satellite images as essential features (dependent variables) and influential factors (independent variables). The primary objective of this study was to gain a deeper understanding of Pakistani wildfires. By analyzing information from previous fire incidents, along with utilizing remote sensing data and incorporating datasets from various sources, we aimed to identify underlying patterns and factors contributing to wildfire occurrences in the region. This comprehensive methodology allowed us to explore potential relationships among different variables and contribute to the development of effective strategies for wildfire management and prevention.

To determine the total number of fires, data from the Fire Information for Resource Management System (FIRMS) were utilized. FIRMS provides data that are almost real-time and available approximately three hours after satellite observation. These data are transferred via FIRMS to the Terra and Aqua Earth Observing System (EOS) as part of NASA's Land, Atmosphere, Near Real-Time Capability (LANCE) for EOS. The data originate from the Visible Infrared Imaging Radiometer Suite (VIIRS) and the Moderate Resolution Imaging Spectroradiometer (MODIS). The wildfires currently occurring are highlighted in Figure 1, which displays pixels covering a square kilometer of the landscape. Within each 1 sq. km pixel, the publicly accessible dataset includes records for temperature, fire confidence, and fire type. From November 2000 to October 2023, users can obtain FIRMS data with a one-day temporal resolution and a one-kilometer square geographic precision.

In this study, we applied filtering criteria to remotely sensed fire data. Specifically, we categorized the fire data based on two criteria: confidence level and fire type. For the confidence level, we set a threshold of 60 percent confidence. This means that only fire observations with a confidence level of 60 percent or higher were chosen for further analysis. This filtering step helped ensure that we focused on more reliable fire detections and reduced the inclusion of potential false positives. Regarding fire type, we utilized the MODIS landcover vegetation data to filter the fire observations. We selected only those fire pixels that corresponded to vegetation-covered areas, as indicated by the land cover data. This helped us narrow down the analysis to fires occurring within vegetated regions and excluded non-vegetated areas or other land cover types. By applying these filtering criteria, we aimed to improve the quality and accuracy of the remotely sensed fire data used in our study.

Figure 2 illustrates the overall count of pixels indicating active wildfires in Pakistan during specific seasons from 2000 to 2023. In terms of fire activity, the data suggest that the year 2020 did not stand out significantly compared to the previous two decades. The most severe years in terms of fire activity were 2008, 2016, and 2017. Notably, the number of active fires in 2016 surpassed the combined count of 2015 and 2019. To identify irregularities across seasons, it is necessary to have a comprehensive understanding of fire activity throughout the year. The number of active fires in 2008 was significantly higher than that in earlier years. Satellite-based fire data, however, revealed that until 2023, every year, in December and February of each year for the previous 24 years, MODIS recorded 10,357 active wildfire indications throughout Pakistan.

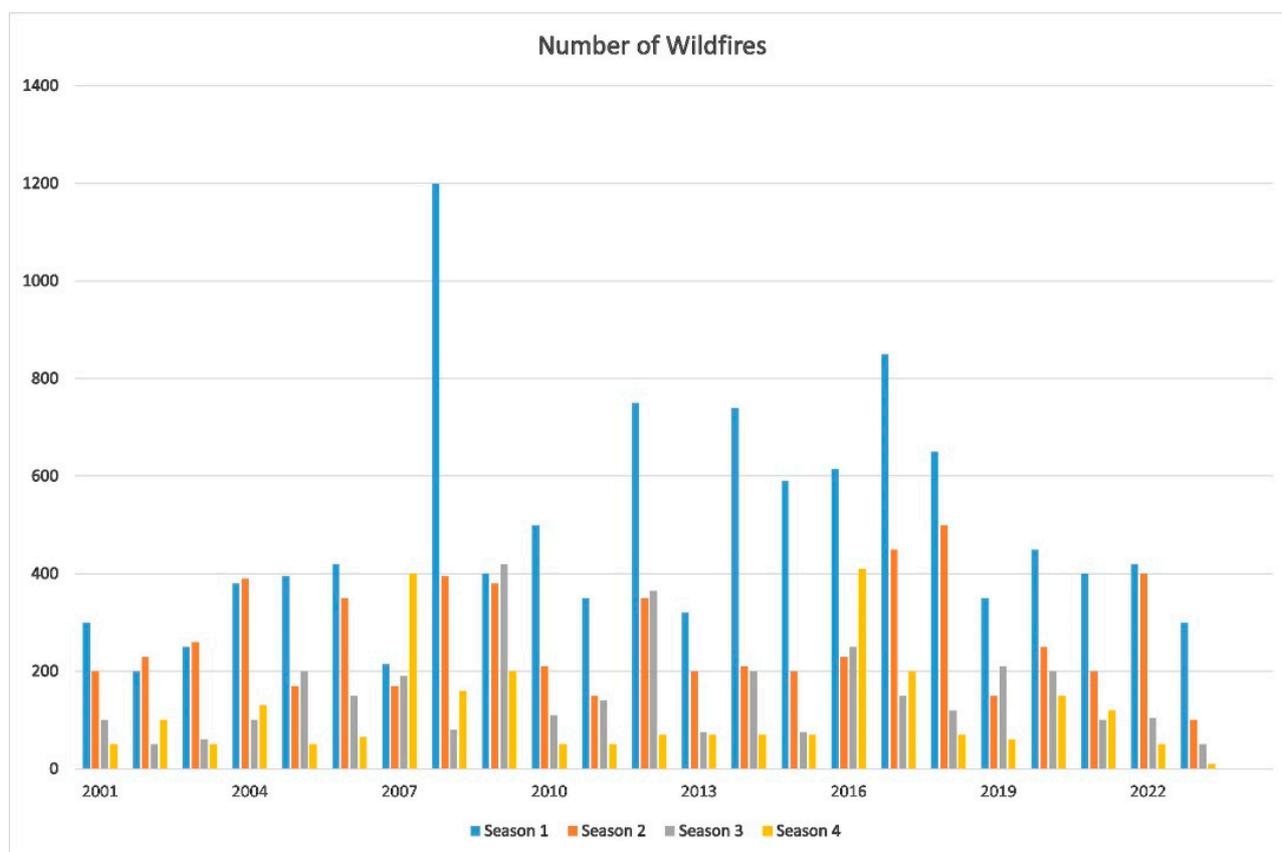


Figure 2. The distribution of the MODIS annual wildfire season is as follows: Seasons 1 and 2 cover December to February and March to May, respectively. Season 3 spans from June to August, and Season 4 covers the period from September to November.

Figure 2 presents the seasonal distribution of active fire sites between January 2001 and October 2023. It can be observed that the majority of fires tended to occur during the winter season, spanning from December to February. The seasons are categorized as follows: Seasons 1 and 2 encompass December to February and March to May, respectively; Season 3 encompasses June to August; Season 4 encompasses September to November.

A diverse range of studies utilizing machine learning techniques in the context of wildfire management and prediction [51] focused on the comprehensive application of deep learning methodologies using satellite remote sensing data for detecting, mapping, and predicting wildland fires. The authors in [52] explored the creation of wildfire susceptibility maps through the integration of interferometric synthetic aperture radar (InSAR) coherence, deep learning, and metaheuristic optimization techniques. The study by [53] contributed to the field by presenting a novel approach to forest fire prediction that leveraged long- and short-term time-series networks [54] and concentrated on active fire mapping in the Brazilian Pantanal region by employing deep learning and China-Brazil Earth Resources Satellite-4A (CBERS-4A) satellite imagery. These studies collectively showcase the increasing significance of machine learning approaches in advancing our understanding of wildfire behavior, enhancing prediction accuracy, and enabling effective fire management strategies using remote sensing data.

3. Methodology

In this section, we discuss how we went about achieving our main goals. The first stage was to find the optimum fire susceptibility model, and the second was to use a CNN to improve the characteristics of the features. The structure was divided into three sections: (a) data preparation and preprocessing; (b) a CNN; (c) model assessment (Figure 3). On a

dataset with varied perspectives, the entire procedure was run four times. For the complete dataset, the data were split by year (six-year gap), the data were sorted regionally based on density, and the data were separated by season. To construct the training dataset, the three main driving factors of topography, weather, and fuel were considered. This information is gathered from various sources such as soil, socioeconomics, vegetation, and past fire incidents. The historical fire occurrences were divided into separate subgroups for training and testing purposes. Machine learning models were then trained using the training dataset, and the performance of the model was evaluated using the test dataset. The most effective ML model was employed to predict wildfire susceptibility.

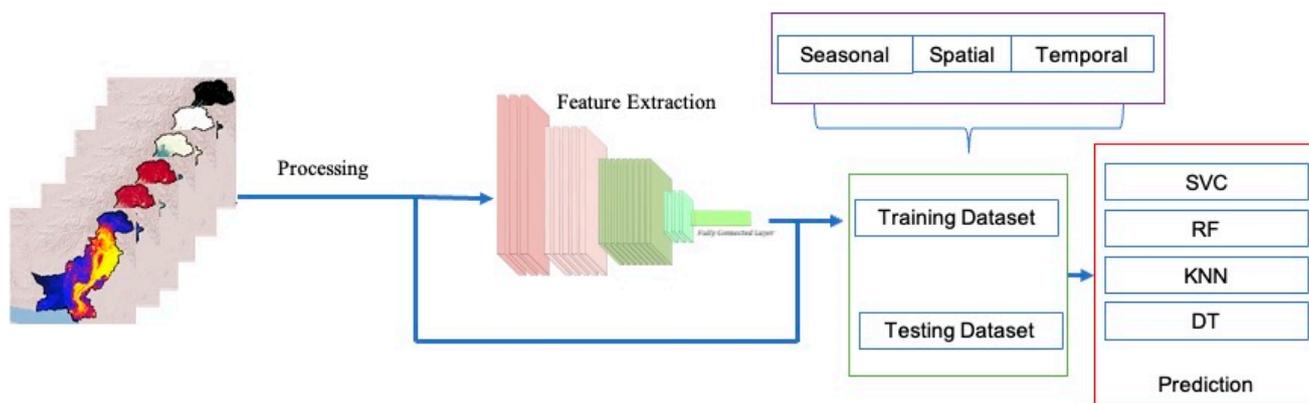


Figure 3. Workflow diagram for wildfire mapping and prediction using the CNN and ML models.

3.1. Feature Extraction Layer

The feature extraction layer in the workflow diagram for wildfire mapping and prediction patterns was extracted from the input data. In the context of this particular workflow, the feature extraction layer was responsible for identifying and extracting meaningful features from the wildfire data such as satellite imagery or other relevant sources. This layer played a crucial role in capturing the distinctive characteristics and patterns associated with wildfire events. Typically, techniques such as CNN or other machine learning algorithms are employed in this layer to automatically learn and extract these features from the input data. The extracted features are then used as inputs to subsequent layers or models for the further analysis, prediction, or mapping of wildfires. The goal of the feature extraction layer is to transform the raw input data into a more compact and representative feature representation that can be effectively utilized by the subsequent stages of the workflow such as classification or prediction models. By properly extracting and selecting relevant features, the feature extraction layer helped to improve the accuracy and effectiveness of the overall wildfire mapping and prediction process.

3.2. Data Extraction and Preprocessing

The passage describes the training dataset used for a specific analysis. The training dataset consists of two types of variables: independent variables (also known as predictors) and dependent variables (also known as response variables). Independent variables were the factors that were used to predict or explain the outcome, while dependent variables were the outcomes or responses that we were interested in understanding or predicting. The entire dataset was quite large and contained a substantial amount of data. It spans a 24-year period and covers information at the country level. Handling such a vast dataset manually would be impractical and time-consuming, so the creation of training datasets was automated. Automation allowed for a more efficient and streamlined process of preparing the necessary data to train the models. By providing the chosen models with more training data samples, their performance improved. This means that having a larger and more diverse set of data to train the models resulted in more accurate and robust predictions or classifications. The models were able to learn from a more extensive range

of examples, leading to better generalization and higher accuracy in their predictions or classifications when applied to new, unseen data.

Fire and no-fire detail: The dependent variable was a binary variable that contrasted with areas where fire incidence occurred and those where it did not. Hence, mapping fire susceptibility can be viewed as a binary classification task in machine learning, with two distinct classes: fire and no-fire. This research served as the foundation for the development of an automated method for identifying where fires and no-fire incidents occurred.

To perform additional research on the driving mechanisms, no-fire instances must also be gathered. Pakistan's wildland land cover collection created a historical database of no-fire occurrences. Random locations in a wilderness area were selected and separated by a distance of one kilometer. To be classified as a no-fire occurrence, sites had to fall outside the firing zones within the 1 km buffer zone. From November 2000 to October 2023, six points were assigned to each day as no-fire points, corresponding to the same time period as the fire data. Additionally, latitude and longitude values were assigned to each month, resulting in a total of 184 points. Both fire and no-fire data had the same spatial resolution. Pakistan's dataset consisted of 22,311 fire points and an equal number of no-fire points.

Contributing factors: Choosing the independent variables, also referred to as predictors or conditioning factors, is a crucial step in predictive modeling. Based on field observations from several research projects as well as open-source and international satellite data, 34 driving elements were selected for this investigation [55–57]. While there are no strict guidelines regarding the inclusion of variables in a model, geomorphological, climatic, and human-related factors are commonly considered conditional components in many studies [58].

Wildfires are mostly caused by three factors: topography, weather, and fuel. The two primary applied wildfire conditioning variables related to fuel conditions are vegetation type and socioeconomic issues. A summary of each dataset used in this study can be found in Table 1. Information from a publicly available database of remotely sensed data, some of which had records with different classifications, was used to build the independent variables. Numerous socioeconomic traits were constructed using remote sensing data. Table 1 includes links to the data sources and data catalogs in the Source of Data column. The imported data in this study were in the form of raster data.

The weather is one of the primary factors generating wildfires, according to numerous studies. The fire weather index (FWI) is a dataset provided by NASA that encompasses various weather information such as temperature, humidity, and transpiration [59]. It is employed widely outside of Canada [60–63]. Instead of merely using a few parameters for data models, in our study, we utilized the Terra Climate dataset [64] and the Global Land Data Assimilation System (GLDAS) [65], which offer a higher spatial resolution, as indicated in Table 1. We incorporated several climatic variables, both derived and primary, into our analysis. It is worth noting that Pakistan has not been extensively explored in terms of the FWI, and the global FWI only encompasses a limited number of factors. Given their superior spatial resolution, we opted for the GLDAS and Terra Climate datasets among the suggested datasets for FWI analysis.

To generate the training dataset, we established fire and no-fire training points along with corresponding conditional factors [66–75]. Subsequently, 34 independent variables were collected and resampled to match the spatial resolution. Information extraction techniques were employed to incorporate predictor values into the dataset. Training samples were then created for each latitude and longitude, resulting in a comprehensive dataset containing fire and non-fire sites, with associated driving variable data stored for each geographic location.

Table 1. Various datasets represent different variables, all obtained from reliable sources. These datasets have a spatial resolution of 1 km grid cells and cover the time period from November 2020 to October 2023.

S. No.	Parameter [Source]	Unit	Spatial Resolution	S. No.	Parameter [Source]	Unit	Spatial Resolution
1	Elevation [66]	Meters	90 m	18	Average surface skin temperature [68]	k	28 km × 28 km
2	Slope [66]			19	Soil moisture [68]	kg/m ²	
3	Aspect [66]	Degree		20	Actual evapotranspiration [67]	mm	
4	Hill shadow [66]			21	Water deficit [67]	mm	
5	Population [76]	person/km (grid cell)	1 km	22	Downward surface shortwave radiation [67]	W/m ²	4.5 km
6	Human modification [68]	Km ²		23	Precipitation accumulation [67]	mm	
7	Travel speed [69]	minutes/meter		24	Minimum temperature [60]	°C	
8	Travel speed walk [69]			25	Maximum temperature [60]		
9	Settlement [73]	classes	100 m	26	Vapor pressure [60]	kPa	
10	Urban cover [67]	%	12 m	27	Tree cover [73]	%	100 m
11	Precipitation [71]	kg/m ² /s		28	NDVI [74]	nm	
12	Transpiration [71]	W/m ²		29	FPAR [75]		500 m
13	Wind speed [71]	m/s		30	LAI [75]	m ²	
14	Soil temperature [71]	k	28 km × 28 km	31	Land cover [73]	classes	100 m
15	Humidity [71]	kg/kg		32	Soil bulk density [70]	kg/m ³	
16	Heat flux [71]	W/m ²		33	Soil taxonomy [72]	classes	250 m
17	Albedo [71]	%		34	Soil texture [71]		

3.3. Convolutional Neural Network

Following the creation of the training set, the classification accuracy was evaluated. The supervised classifications frequently make use of the input training samples. It is essential to carry out multicollinearity in the creation of the training set, and the classification accuracy was evaluated, so it is essential to carry out a multicollinearity analysis before beginning fire modeling. In Figure 4 a multicollinearity analysis was conducted to check for any connected components. The link between the predicted wildland fire factors was then investigated using a multicollinearity analysis. The choice of features is extremely important in the prediction of fire susceptibility.

Multicollinearity analysis was conducted to assess the relationship between the climate, environment, socioeconomic, topography, and vegetation factors. The analysis aimed to determine whether there were high correlations or dependencies among these factors, which could impact the reliability and interpretability of the model. The results of the multicollinearity analysis provide insights into the extent to which these factors are correlated, helping to identify potential redundancies or overlapping information.

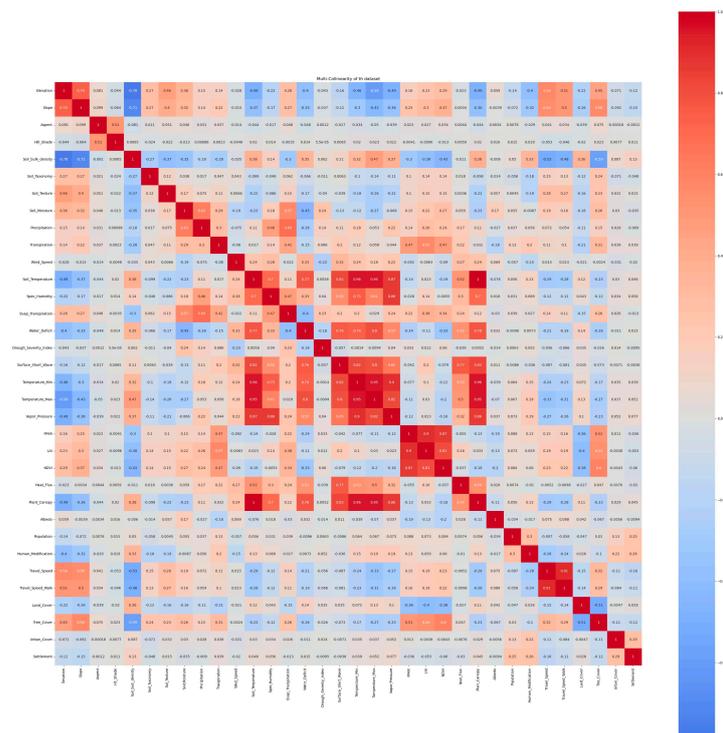


Figure 4. A thorough multicollinearity analysis examines the relationships between the climate, environment, socioeconomic, topography, and vegetation factors. This assessment aims to uncover high correlations that may affect the model reliability and interpretability.

Prior to implementing machine learning techniques, a comprehensive analysis of multicollinearity was conducted on the variables. The results indicated a strong correlation exceeding 0.8 among certain variables. Variables with correlation coefficients surpassing 0.8 or 0.9 were considered for removal. Consequently, these variables were excluded from the machine learning process based on the criterion of high correlation coefficients. This approach was adopted due to the understanding that strong correlations between variables imply the capture of similar information, thereby complicating the ability to discern their individual effects on the dependent variable. By prioritizing variables with lower correlations, the aim was to enhance the reliability and applicability of the models.

Feature Selection: Multicollinearity helped identify redundant features. In the multicollinearity analysis, we found that some of the variables were highly correlated, and they likely captured similar information. In such cases, in our feature importance analysis, we selected only one of these variables as it can be sufficient, reducing the dimensionality of the dataset and improving model performance.

Feature importance analysis: By assessing the importance of features, we gain insights into which variables have the most influence on the model's predictions. This analysis helps prioritize and focus on the most influential factors when interpreting and making decisions based on the model's outputs. In Figure 5 the results of the feature importance analysis provide a ranking or score for each feature, indicating their relative importance in predicting the target variable. This information can guide further analysis, model refinement, or decision-making processes.

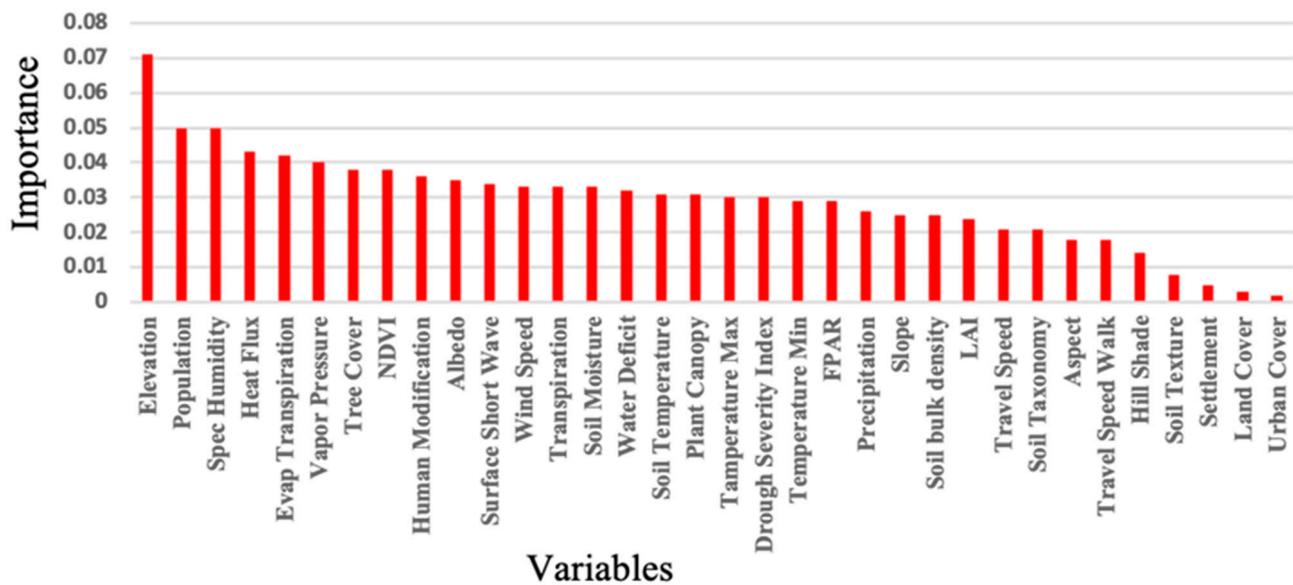


Figure 5. Feature importance analysis. Understanding the impact: unveiling the importance of features in our model aids in prioritizing influential variables for interpretation and decision-making. Each feature is assigned scores or rankings, providing guidance for further analysis and model refinement.

The model utilizes a CNN with a depth of 20 layers. The architecture of the CNN, consisting of 20 layers, is depicted in Figure 6. CNN has emerged as a fundamental concept in the field of deep learning. Its ability to leverage vast amounts of data in the era of big datasets is different from traditional methods, yielding promising results. Consequently, numerous applications have been developed. Furthermore, a CNN can be applied to process both 1D data and 2D images.

$$out(N, C_{out}) = bias(C_{out}) + \sum_{k=0}^{C_{in}-1} weight(C_{out}, k) * input(N, k) \quad (1)$$

In the 1D case, the output value of the layer with input size (N, C_{in}, L_{in}) and output (N, C_{out}, L_{out}) can be precisely described, where N is the batch size, C denotes the number of channels, and L is the length of the sequence (34 in our case). In the 2D case, the output value of the layer with input size $(N, C_{in}, H_{in}, W_{in})$ and output $(N, C_{out}, H_{out}, W_{out})$ can be precisely described, where N is the batch size, C denotes the number of channels, H is the height of the input planes in pixels, and W is the width in pixels.

The maps of wildland fire susceptibility were created using various classification methods including DT, SVM, RF, and KNN. Among these methods, RF employs an ensemble learning technique by combining the results of multiple classification trees. It achieves this by utilizing bootstrapping techniques and selecting subsets of observations to construct random binary trees. This approach involves taking random samples from the original training dataset. Using an RF offers several advantages over a single-participant method. It yields more accurate results due to the diversification and randomness introduced by the subset selection process, leading to enhanced generalization and robustness in the model.

The SVM is a popular and extensively employed classification method known for its non-parametric kernel-based approach. They are the most efficient at handling difficult, robust linear and non-linear classification and regression tasks.

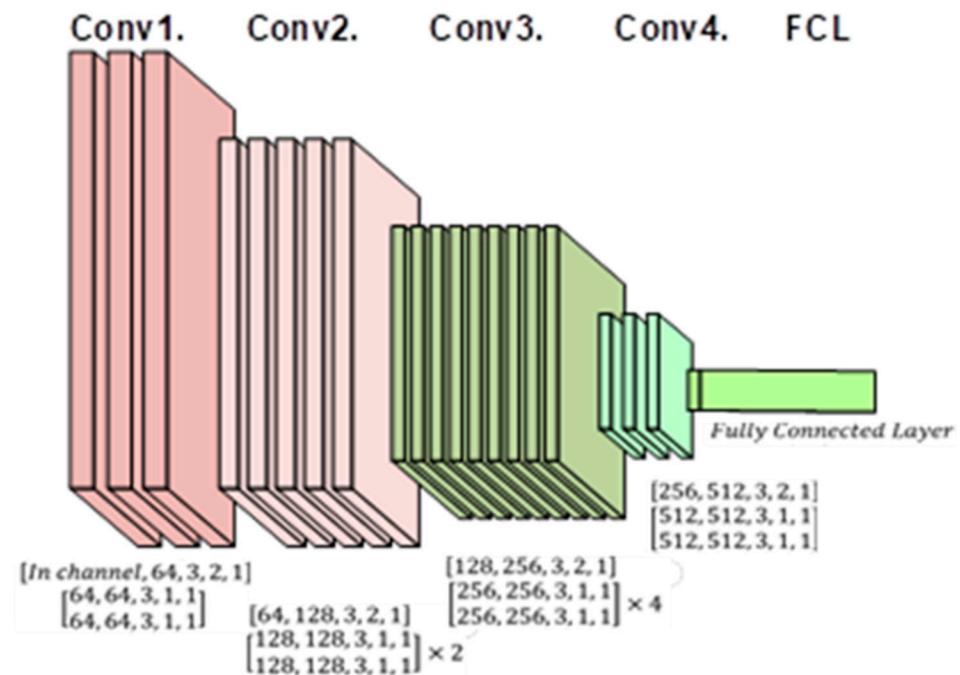


Figure 6. A 20-layer CNN architecture was used for this study with a fully connected layer of size 512 at the end, which will be fed to classifiers.

The KNN approach is a non-parametric model that is commonly used for classification and regression tasks. It adopts a lazy learning strategy, which means that it does not make any assumptions about the data. This makes the KNN model particularly valuable for predicting air pollution when there is a lack of general predictors. The KNN technique begins by randomly selecting a certain number of class centers (k) and then classifies the training data based on their proximity to these centers. The class centers are then iteratively adjusted to the middle of the training data, and the data are reclassified accordingly.

DTs are considered universal function approximators and belong to the category of supervised learning algorithms. However, achieving such universality in its basic form can be challenging. DTs are versatile and can be applied to both classification and regression problems. A DT is composed of a set of if-then-else rules organized as branches connected by decision nodes that ultimately lead to leaf nodes. The decision nodes represent the points at which the tree splits into different branches, indicating the specific decisions made by the algorithm. On the other hand, the leaf nodes represent the output of the model.

3.4. Validation and Evaluation Matrix

Once ML models have been trained to predict the origin of a fire, it becomes essential to evaluate their performance. To accomplish this, an accuracy evaluation was carried out. To validate the model, the sample dataset containing fire and non-fire zones was divided into training and testing datasets. The points were allocated in an 80:20 ratio, with 80% assigned for training and the remaining 20% for testing purposes.

Accuracy assessment: The performance of the ML model was evaluated using a widely recognized accuracy evaluation technique. This involved assessing characteristics such as overall accuracy, area under the curve score (AUC score), and precision score.

To conduct the accuracy evaluation, independent testing datasets were created by dividing the sample dataset at an 80:20 ratio, with 80% allocated for training and 20% reserved for testing.

The accuracy of a binary classification test is determined by its ability to correctly determine the presence or absence of a condition. It is measured as the percentage of

accurate predictions, which includes both true positives (TP) and true negatives (TN), out of all the instances considered.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

The F1 score combines precision and recall, focusing on the model's performance in binary classification tasks. It is particularly useful when there is class imbalance or varying impacts of false positives (FP) and false negatives (FN).

The F1 score combines two important metrics: precision and recall. Precision measures the accuracy of the model's positive predictions, while recall quantifies the model's ability to correctly identify positive instances. While TN are important for evaluating a classification model's overall performance, they do not directly factor into the F1 score calculation, which specifically focuses on the positive class.

Precision is a crucial metric in binary classification, emphasizing the quality of positive predictions made by the model. It is calculated by dividing the number of TN predictions by the sum of TP and FP predictions. Precision ensures that when the model claims something is positive, it is highly likely to be correct, minimizing FP . High precision indicates reliable and accurate positive predictions. Precision is often used alongside recall, which measures the model's ability to capture all positive instances. Precision quantifies the trustworthiness of the model's positive predictions. In summary, precision helps us understand the proportion of the true positive predictions among all instances predicted as positive by the model. It is an important metric, especially when FP has significant consequences.

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

4. Results

In this section, the conclusions of the study are presented based on the methodology utilized. The first part reports the results of the fire occurrence sites obtained from the FIRMS dataset. The second half of this section reviews the findings of various machine learning methods. It is worth noting that the majority of the predictive variables used in this study were sourced from open-source platforms.

This study employed supervised ML methods (RF, DT, SVM, and KNN) with an 80–20 split for training and testing, respectively. The accuracy summary in Table 2 and Figure 7 shows that RF had the highest accuracy (96.56%), DT had the lowest (85.12%), and SVM and KNN both had 87.2%. Notably, all methods showed higher accuracy in predicting the absence of fire than in predicting its occurrence.

The evaluation script using RF algorithms is commonly used to determine the optimal number of maximum trees and max-depth for the RF model. These statistics directly influence the accuracy of the model, making this step crucial. Additionally, the script can provide insights into the number of leaf nodes needed for classifying two classes. Hyperparameter tuning involves selecting the best combination of hyperparameters to maximize model performance. Finding the perfect set of hyperparameters is essential for achieving optimal model performance. Manual tuning of hyperparameters involves manually setting and testing various combinations. However, this method can be time-consuming and impractical when there are numerous hyperparameters to test.

For a CNN 1D and a CNN 2D, the output of the completely connected layer is 512, which is fed directly into the RF. In the case of a CNN, the RF accuracy also significantly increased when the variable rose to 512. When the variable increased from 34 to 512, the RF accuracy rose from 86.48 to 91.34. A total of 512 variables were sent to the RF while using the picture data for a CNN 2D to maintain quick processing and achieve a 96.56 accuracy.

Table 2. An overview of the ML algorithm results for accuracy assessment including the RF, SVC, DT, and KNN models compared with the Conv1D, Conv2D, and no Conv approaches.

Classifier	Feature	Accuracy Score	AUC Score	Precision Score	F1-Score
RF	No-Conv	86.48	86.55	85.11	86.83
	Conv1D	91.34	90.90	90.04	91.71
	Conv2D	96.56	95.90	95.24	96.95
SVC	No-Conv	77.88	77.22	74.95	78.19
	Conv1D	82.19	82.58	80.02	82.52
	Conv2D	87.14	86.99	84.19	87.49
DT	No-Conv	76.41	76.40	70.91	76.72
	Conv1D	81.76	81.36	77.53	82.09
	Conv2D	85.13	82.37	80.03	85.47
KNN	No-Conv	77.18	77.10	71.99	77.49
	Conv1D	83.24	83.19	80.07	83.57
	Conv2D	87.19	88.24	82.27	87.54

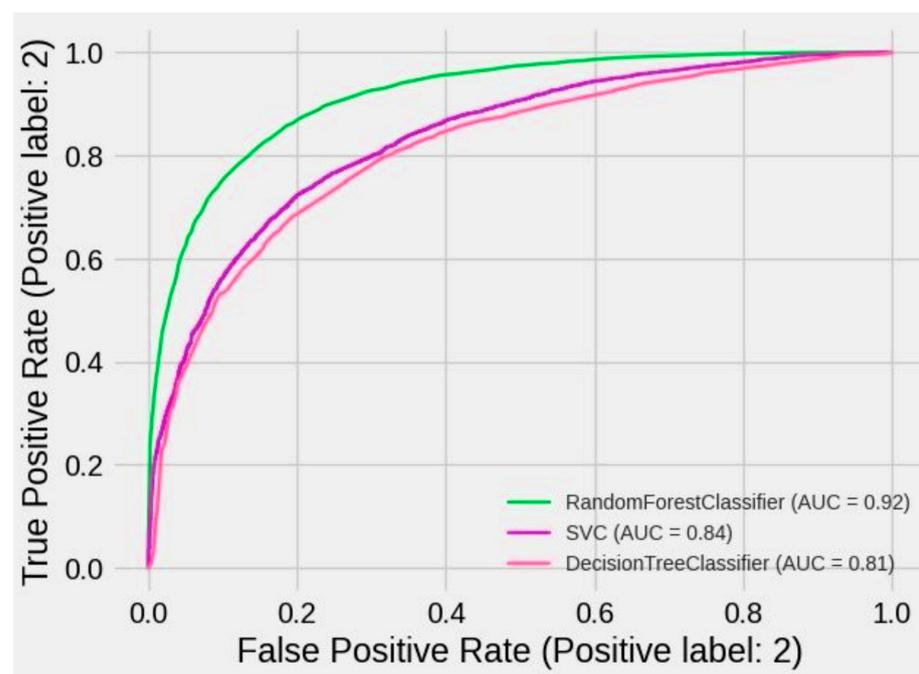


Figure 7. The ROC curves of RF, SVC, and DT.

The RF model's accuracy improved as the number of leaf nodes increased, reaching its peak at 110 leaf nodes and a maximum depth of 50. Beyond the 50 max-depth, the accuracy remained relatively stable. Based on these findings, it can be concluded that 110 leaf nodes were the optimal number of trees for the RF model in this experiment. The RF model with 110 trees and a maximum depth of 50 achieved an impressive accuracy of 96.91 percent, as demonstrated in Table 3.

Table 3. Hyperparameter (leaf node 110 and max-depth 50) tuning of the RF model and scores on the testing dataset.

Classifier	Feature	Accuracy Score	AUC Score	Precision Score	F1-Score
RF	No-Conv	86.73	86.53	85.29	87.13
	Conv1D	91.56	91.04	90.35	91.92
	Conv2D	96.91	95.94	95.54	97.29

It is essential to assess the performance of the RF model compared to other models under different temporal, geographical, and seasonal conditions to ensure its reliability. Extensive research has shown that the RF model consistently produces diverse results based on the specific location. Additionally, it is crucial to consider the temporal validity of the models. To enhance the model's applicability in real-time scenarios, the entire dataset was categorized into different subsets and subcategories. This approach aims to improve the model's adaptability in dynamic environments.

Temporal reliability of the model: One common issue in data-driven problems is the selection of the optimal time period for model training. The stochasticity or heterogeneity of the data presents a challenge, as using long-term data in model construction may result in a smoothing effect on behavior tendencies when applied to current data. Conversely, relying on short-term data could introduce higher levels of uncertainty. It is important to evaluate the durability of the model over an extended period to assess its reliability.

In Table 4 the dataset was divided into four groups with a six-year interval between them. The time periods from 2000 to 2005, 2006 to 2011, 2012 to 2017, and 2018 to 2023 were designated Y1, Y2, Y3, and Y4, respectively. It is important to note that there were fewer samples available in Y1 due to the data accessibility limitations for the year 2000. Across all time periods, the CNN 2D feature extraction technique consistently outperformed other feature techniques in terms of accuracy. Specifically, the highest accuracy score of 96.21 was achieved in Y1 using a CNN 2D feature extraction, while the lowest accuracy score of 86.01 was observed in Y1 with regular tabular data.

Table 4. The temporal reliabilities of the RF models from 2000 to 2005, 2006 to 2011, 2012 to 2017, and 2018 to 2023 were designated Y1, Y2, Y3, and Y4, respectively.

Division	Feature	Accuracy Score	AUC Score	Precision Score	F1-Score
Y1	No-Conv	86.01	83.47	83.87	86.35
	Conv1D	90.87	88.33	88.73	91.23
	Conv2D	96.21	93.67	94.07	96.59
Y2	No-Conv	83.52	83.35	83.20	83.85
	Conv1D	88.38	88.03	88.06	88.73
	Conv2D	93.72	93.55	93.40	94.09
Y3	No-Conv	84.23	84.10	85.74	84.56
	Conv1D	89.09	89.77	90.60	89.44
	Conv2D	94.43	95.02	95.94	94.80
Y4	No-Conv	83.67	83.59	82.44	84.00
	Conv1D	88.53	88.45	87.30	88.88
	Conv2D	93.87	93.79	92.64	94.24

Seasonal reliability of the model: Pakistan experiences four distinct seasons: S1 (mild, dry winter) from December to February, S2 (hot, dry spring), S3 (summer rainy season or southwest monsoon period), and S4 (receding monsoon period) from September to November. To analyze the data, it was divided into these seasonal categories (Table 5).

Notably, the CNN 2D approach consistently outperformed other feature techniques in each seasonal distribution. S3 with a CNN 2D feature extraction had the highest accuracy score of 98.71, whereas S3 with conventional tabular data had the highest accuracy score of 90.51.

Table 5. Seasonal reliability of the RF model.

Division	Feature	Accuracy Score	AUC Score	Precision Score	F1-Score
S1	No-Conv	82.75	80.63	82.86	83.16
	Conv1D	87.61	85.49	87.72	88.04
	Conv2D	92.95	90.83	93.06	93.41
S2	No-Conv	83.89	83.20	83.53	84.30
	Conv1D	88.75	88.06	88.39	89.19
	Conv2D	94.09	93.40	93.73	94.56
S3	No-Conv	90.51	87.43	89.28	90.96
	Conv1D	95.37	92.02	94.14	95.84
	Conv2D	98.71	97.63	99.48	99.20
S4	No-Conv	87.30	80.12	86.71	87.73
	Conv1D	92.16	84.98	91.57	92.62
	Conv2D	97.50	90.32	96.91	97.98

Recognizing the relationship between wildfires and precipitation during the summer rainy season is of utmost importance as it highlights the considerable influence that precipitation can exert on wildfire behavior and frequency. Rainfall plays a vital role in increasing the fuel moisture content, making it less susceptible to ignition and reducing the potential for fire spread. Moreover, higher humidity levels associated with rainy seasons can also influence the flammability of vegetation, further contributing to wildfire prevention. In our study, we found that the use of machine learning models including No-Conv, Conv1D, and Conv2D significantly contributed to the prediction of wildfires during the summer rainy season in Pakistan. Through extensive data collection on factors such as weather patterns, historical wildfire incidents, and fuel moisture content, we were able to train and evaluate our model accurately. Our findings support the notion that there is a strong relationship between wildfires and precipitation. Specifically, during the summer rainy season in Pakistan, increased rainfall leads to a higher moisture content in vegetation and fuels. This increased moisture makes them less susceptible to ignition and reduces the potential for fire occurrence and spread. Based on our analysis, we observed a significant reduction in wildfires during the summer rainy season, which was reflected in the high accuracy score of our prediction model. This aligns with our understanding of the connection between wildfires and precipitation, as rainfall plays a vital role in mitigating the risk of fire incidents.

Spatial assessment of the model: Although the data-driven approach in this study focused solely on a national perspective, the significance of distinctive attributes across various clusters within Pakistan is vividly illustrated in Figure 8. The data's relative characteristics separated in a pronounced manner, highlighting spatial variations. To delve into the impact of these variations on fire occurrences, the study opted for a selection of four distinct density-based clusters. For this investigation, the K-means clustering technique was employed, recognized for its simplicity and effectiveness. The primary objective of K-means clustering is to minimize the collective distance between each entity and its respective centroid, disregarding any predefined grouping. This process aims to identify the best arrangement of n entities into k clusters, assigning each object to the cluster with the nearest average, resulting in a partitioned data space known as Voronoi cells. Specifically, this study employed K-means spatial location clustering to segment the

dataset into four clusters based on density. These clusters are denoted as C1, C2, C3, and C4, with corresponding colors of brown and pink for C1 and C2 and purple and green for C3 and C4, respectively.

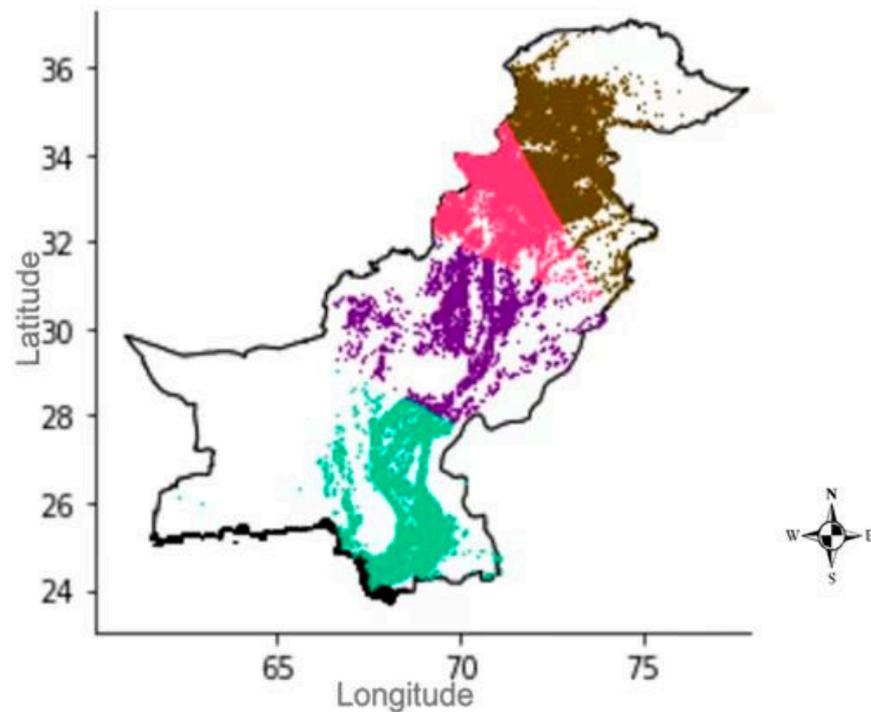


Figure 8. Cluster distribution of data from the north to the south of Pakistan (study region). C1 is represented by the color brown, C2 by pink, C3 by purple, and C4 by green.

The implementation of K-means clustering allowed for the categorization of spatially related data points into meaningful groupings, each representing a unique cluster. The selection of four clusters was based on the characteristics and distribution of the data. The distinctive attributes and patterns within each cluster provide insights into the spatial variations in fire occurrences. This approach enables a deeper exploration of the spatial dynamics of fires across different regions of Pakistan. By partitioning the data into these clusters, it becomes possible to identify areas with similar fire occurrence trends and understand the factors contributing to these patterns.

The resulting clusters, designated C1, C2, C3, and C4, offer a more refined understanding of the spatial distribution of fire incidents within Pakistan. The colors assigned to each cluster—brown and pink for C1 and C2 and purple and green for C3 and C4, respectively—further aid in visually distinguishing the distinct groupings. Through this K-means spatial clustering approach, the study gains valuable insights into how fire occurrences vary across the landscape. This information can be particularly useful for policymakers, land managers, and other stakeholders involved in fire management and prevention strategies. By considering the spatial patterns revealed by the clusters, more targeted and effective approaches to managing fire risks can be developed.

It was observed that in all cluster distributions, a CNN 2D outperformed other feature techniques. In Table 6 the highest accuracy score of 97.66 was observed in C4 with a CNN 2D feature extraction, and an 87.45 accuracy score was observed in C4 with normal tabular data.

Table 6. Spatial relatability of the RF model.

Division	Feature	Accuracy Score	AUC Score	Precision Score	F1-Score
C1	No-Conv	86.10	83.61	83.87	86.61
	Conv1D	90.96	88.47	88.73	91.50
	Conv2D	96.30	93.81	94.07	96.87
C2	No-Conv	82.71	81.75	82.85	83.20
	Conv1D	87.57	86.61	87.71	88.09
	Conv2D	92.91	91.95	93.05	93.46
C3	No-Conv	86.71	85.13	87.06	87.23
	Conv1D	91.57	89.99	91.92	92.11
	Conv2D	96.91	95.33	97.26	97.49
C4	No-Conv	87.46	87.09	82.90	87.98
	Conv1D	92.32	91.95	87.76	92.87
	Conv2D	97.66	97.29	93.10	98.24

To retain all of the available data, the researchers utilized one-month average values of daily datasets, considering that some figures were missing for specific dates. Consequently, they chose a minimum timeframe of one month for all driving-related components in their analysis.

The second objective of the study was to evaluate and compare various machine learning models to identify the most effective model for predicting fire occurrences. The test dataset was used to assess the performance of four different ML techniques, with the RF model demonstrating the most favorable outcomes. The RF model is characterized by multiple independently trained decision trees, and its classification decision is based on selecting the class with the highest number of votes from the individual trees. This ensemble approach contributes to the RF model's high precision. The researchers fine-tuned the RF model by adjusting parameters such as the maximum depth and the number of trees, which further improved its accuracy. The optimal configuration that yielded the best results consisted of 110 trees and a maximum depth of 50. Additionally, the study analyzed the importance of each variable to understand its contribution to the classifier's accuracy. Eliminating any of these variables led to a decrease in the classifier's accuracy rate.

Furthermore, researchers have explored the integration of a CNN model with RF to achieve even better prediction accuracy than other state-of-the-art models [77]. In the case of a CNN 1D and a CNN 2D, the output of the completely connected layer was set to 512, which was directly fed into the RF. By increasing this variable from 34 to 512, the RF accuracy improved significantly, rising from 86.48% to 91.34%. This combination of a CNN and RF led to an overall accuracy of 96.56%, making it a successful approach for wildfire prediction.

Comparing the performance of the CNN alone with the ensemble approach, we observed a notable increase in precision and F1-score by incorporating traditional classifiers. This suggests that the ensemble approach better captures the intricate patterns in the wildfire data, resulting in more accurate predictions. In conclusion, while using a CNN directly for classification is a robust approach, the combination of a CNN with traditional classifiers offers a performance boost in terms of the key evaluation metrics. The ensemble approach's improved precision, F1-score, and accuracy highlight the benefits of synergizing deep learning and classical machine learning techniques in wildfire prediction.

The study proposes several succinct recommendations: first, the formulation of customized fire management plans for different sub-areas based on density clusters, utilizing appropriate datasets for training; second, the adoption of a seasonal approach to wildfire mitigation, incorporating insights into fire behavior across seasons for effective strategies; third, considering population and critical facility relocation from fire-prone regions to

mitigate financial and public health risks; fourth, the enhancement of existing wildfire control decision support systems in Pakistan through the integration of the study's data, model insights, and platform development; finally, the utilization of freely available global databases to assess catastrophic risks and augment wildfire management approaches in emerging nations, fostering cross-border knowledge exchange for improved prevention and control efforts.

5. Discussion

The primary aim of this study was to improve the process of extracting data related to fire incidents. The aim of this study was to improve the process of extracting data related to fire incidents and no-fire conditions from historical inventories. These data were intended to be used in machine learning models for predicting wildfires and analyzing the factors that contribute to their occurrence. Here, we referenced previous novel studies that utilized both remotely sensed fire data and tree-based machine learning algorithms [78,79]. Our research goes beyond previous studies by incorporating a CNN, which is an advanced deep learning technique, to map and predict wildfires. Previous research often focused on identifying fire incidence sites using a 1 km-resolution dataset from FIRMS, but it neglected the inclusion of no-fire points. Additionally, previous research conducted on the wildland–urban interface (WUI) has revealed that wildfires have been observed to extend beyond the 1 km buffer zone surrounding the initial ignition points. This occurrence is attributed to the dispersion of firebrands over longer distances, leading to the ignition of spot fires. In our study, the classification of no-fire occurrence zones was based on a systematic approach. We defined a no-fire occurrence zone as areas that fell outside the firing zones within a 1 km buffer zone. The purpose of this classification was to differentiate between areas with documented fire incidents and areas that did not experience fires during the specified time period. The selection of the 1 km buffer zone as the basis for classifying no-fire occurrences was determined by several factors. First, it allowed us to establish a clear spatial boundary for analysis and comparison with fire occurrence zones. By defining a buffer zone around the fire zones, we can establish a contextually relevant reference area for identifying no-fire occurrences. This buffer zone allows us to capture the immediate vicinity surrounding the fire zones, enabling a focused analysis of the factors influencing fire occurrence and spread in nearby areas. The buffer zone helps in differentiating between areas with documented fire incidents and areas that remained unaffected by fires. It provides a clear spatial boundary within which we can classify and study the no-fire occurrence zones. By excluding areas within the 1 km buffer zone from the fire occurrence zones, we identified and analyzed the areas that fell outside this buffer zone as no-fire occurrence zones. Moreover, certain areas within the WUI exhibited an extended buffer zone of up to 5 km, which significantly increases the risk of wildfires encroaching upon residential properties [80,81]. While we acknowledge that wildfires can spread beyond a 1 km buffer zone, it is important to note that our study focused on analyzing the immediate impact and characteristics of fire incidents. The purpose was to identify factors contributing to fire susceptibility within a relatively close proximity to the fire zones. By examining the areas within this buffer zone, we aimed to gain insights into the factors influencing fire occurrence and spread in the immediate vicinity of fires.

Our study introduced advanced techniques such as the utilization of a CNN, which has demonstrated significant potential in image analysis and classification tasks. By applying this advanced technique to wildfire mapping and prediction assessment, we aimed to improve the accuracy and reliability of the results. Some of the new studies published recently in this field have aimed to contribute to the field of wildfire mapping and prediction assessment by exploring different approaches, metrics, and indices to improve our understanding and assessment of fire severity. These addressed specific regions and ecosystems and provide valuable insights into the application of remote sensing techniques and indices for assessing fire impacts [82–84]. Our research adopted a comprehensive approach by considering various contributing factors to wildfire occurrence and severity. We recognize

that wildfires are complex events influenced by multiple factors such as weather conditions, vegetation type, terrain, and human activities. By taking into account these different factors, our study provides a more holistic understanding of the dynamics and patterns of wildfires. Moreover, our findings uncover novel relationships, patterns, or correlations that were not previously explored in the literature. The utilization of a CNN allows us to analyze and interpret large amounts of data, potentially revealing hidden insights and connections between contributing factors and wildfire occurrences. These novel findings contribute to the advancement of knowledge in wildfire mapping and prediction assessment, improving our ability to predict, prevent, and manage wildfires.

We build upon the literature by introducing advanced techniques, adopting a comprehensive approach, and potentially uncovering novel findings. Through our research, we aim to contribute to the advancement of wildfire mapping and prediction assessment, ultimately enhancing our understanding and ability to mitigate the devastating impacts of wildfires.

6. Conclusions with Constraints and Limitations

In the realm of data-driven approaches, the selection of an optimal time period for model training is a common challenge, influenced by data stochasticity and heterogeneity. Striking a balance between long-term data utilization, which may smooth behavior trends, and short-term data reliance, which introduces uncertainty, is pivotal. Evaluating a model's durability over an extended period is crucial to assessing its reliability. This study addressed this concern by dividing the dataset into four distinct time intervals, Y1 to Y4, each spanning six years. Notably, the CNN 2D feature extraction consistently showcased superior performance in accuracy across all time periods, with the highest accuracy score of 96.21 achieved in Y1.

Moving to seasonal applicability, Pakistan experiences four distinct seasons, and this study delved into the impact of seasonality on model performance. The CNN 2D approach consistently outperformed other feature techniques in each seasonal category. The highest accuracy score of 98.73 was achieved during S3, the summer rainy season, utilizing a CNN 2D feature extraction. These findings underscore the adaptability of a CNN 2D in capturing season-specific fire occurrence patterns, an essential aspect for accurate prediction.

Our study found that machine learning models (No-Conv, Conv1D, Conv2D) improved wildfire prediction in Pakistan's summer rainy season by considering weather patterns, historical incidents, and fuel moisture content, revealing a strong relationship between wildfires and precipitation. The prediction model achieved a high accuracy score of 98.71, confirming the effectiveness of considering wildfire-precipitation dynamics in minimizing fire risk and spread.

Moreover, the study investigated the spatial distribution of fire incidents within Pakistan, revealing significant attributes in various clusters through K-means clustering. This approach offered valuable insights into the diverse landscape of fire occurrences. The spatial clusters, labeled C1, C2, C3, and C4, enhanced our understanding of the fire dynamics in different regions, and the consistently superior performance of a CNN 2D was evident across all clusters. By utilizing these spatial patterns, policymakers and stakeholders can formulate targeted strategies for effective fire management and prevention.

Addressing the increasing wildfire risk in Pakistan, which is exacerbated by urban expansion encroaching upon wildland interfaces, demands immediate attention from policymakers. Mitigation strategies should encompass urban planning, zoning regulations, and community awareness initiatives to safeguard human settlements and infrastructure. This study highlights the potential of advanced machine learning models such as a CNN 2D and RF in predicting wildfires, emphasizing the role of technology in assessing fire risk. The integration of real-time remote sensing and weather monitoring data could further enhance prediction accuracy, empowering proactive wildfire management strategies.

In conclusion, this study's evaluation of historical fire events in Pakistan provides a solid foundation for constructing a comprehensive wildfire management framework.

By harnessing insights from data-driven analysis and advanced modeling techniques, policymakers can implement preemptive measures to protect communities and natural resources while mitigating the impacts of wildfires. Collaborative efforts among scientific communities, government bodies, and local stakeholders are pivotal in addressing the escalating wildfire threat in Pakistan. While data limitations have been acknowledged, refining the variable values, employing finer-resolution datasets, and aligning satellite data with real fire occurrences are crucial steps to enhance wildfire forecasting accuracy and management understanding. The alignment of remotely sensed data with forest department-maintained records promises further improvements in prediction reliability and overall management strategies.

Author Contributions: R.K.: Conceptualization, methodology, data curation and writing-original draft preparation. W.R.: Visualization, investigation, software and validation. M.I.: Writing-reviewing and editing. S.W.: Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the National Key R&D Program of China (2021YFC3000300).

Data Availability Statement: All the data that support this study are open access and can be accessed using websites or data repositories. The sources of the datasets are described accordingly.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Verde, J.C.; Zêzere, J.L. Assessment and validation of wildfire susceptibility and hazard in Portugal. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 485–497. [[CrossRef](#)]
2. FerreirA-leiTe, F.; Lourenço, L.; Bento-Gonçalves, A. Large Forest fires in mainland Portugal, brief characterization. *Méditerranée. Rev. Géographique Pays Méditerranéens/J. Mediterr. Geogr.* **2013**, *121*, 53–65. [[CrossRef](#)]
3. Tedim, F.; Remelgado, R.; Borges, C.; Carvalho, S.; Martins, J. Exploring the occurrence of mega-fires in Portugal. *For. Ecol. Manag.* **2013**, *294*, 86–96. [[CrossRef](#)]
4. Brown, J.K.; Smith, J.K. *Wildland Fire in Ecosystems: Effects of Fire on Flora*; Gen. Tech. Rep. RMRS-GTR-42-vol. 2. Ogden, UT; U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: Fort Collins, CO, USA, 2000; Volume 42, 257p.
5. Neary, D.G.; Ryan, K.C.; DeBano, L.F. *Wildland Fire in Ecosystems: Effects of Fire on Soils and Water*; Gen. Tech. Rep. RMRS-GTR-42-vol. 4. Ogden, UT; US Department of Agriculture, Forest Service, Rocky Mountain Research Station: Fort Collins, CO, USA, 2005; Volume 42, 250p.
6. Xu, R.; Yu, P.; Abramson, M.J.; Johnston, F.H.; Samet, J.M.; Bell, M.L.; Haines, A.; Ebi, K.L.; Li, S.; Guo, Y. Wildfires, global climate change, and human health. *N. Engl. J. Med.* **2020**, *383*, 2173–2181. [[CrossRef](#)] [[PubMed](#)]
7. Sandberg, D.V. *Wildland Fire in Ecosystems: Effects of Fire on Air*; US Department of Agriculture, Forest Service, Rocky Mountain Research Station: Fort Collins, CO, USA, 2003.
8. Johnston, L.M.; Wang, X.; Erni, S.; Taylor, S.W.; McFayden, C.B.; Oliver, J.A.; Stockdale, C.; Christianson, A.; Boulanger, Y.; Gauthier, S.; et al. Wildland fire risk research in Canada. *Environ. Rev.* **2020**, *28*, 164–186. [[CrossRef](#)]
9. Martell, D.L. *Forest Fire Management, in Handbook of Operations Research in Natural Resources*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 489–509.
10. Shvidenko, A.Z.; Schepaschenko, D.G. Climate change and wildfires in Russia. *Contemp. Probl. Ecol.* **2013**, *6*, 683–692. [[CrossRef](#)]
11. Vadrevu, K.P.; Lasko, K.; Giglio, L.; Schroeder, W.; Biswas, S.; Justice, C. Trends in vegetation fires in south and southeast Asian countries. *Sci. Rep.* **2019**, *9*, 7422. [[CrossRef](#)] [[PubMed](#)]
12. Reddy, C.S.; Bird, N.G.; Sreelakshmi, S.; Manikandan, T.M.; Asra, M.; Krishna, P.H.; Jha, C.S.; Rao, P.V.N.; Diwakar, P.G. Identification and characterization of spatio-temporal hotspots of forest fires in South Asia. *Environ. Monit. Assess.* **2019**, *191*, 791. [[CrossRef](#)]
13. Hantson, S.; Pueyo, S.; Chuvieco, E. Global fire size distribution is driven by human impact and climate. *Glob. Ecol. Biogeogr.* **2015**, *24*, 77–86. [[CrossRef](#)]
14. Jolly, W.M.; Cochrane, M.A.; Freeborn, P.H.; Holden, Z.A.; Brown, T.J.; Williamson, G.J.; Bowman, D.M. Climate-induced variations in global wildfire danger from 1979 to 2013. *Nat. Commun.* **2015**, *6*, 7537. [[CrossRef](#)]
15. Barbero, R.; Abatzoglou, J.T.; Pimont, F.; Ruffault, J.; Curt, T. Attributing increases in fire weather to anthropogenic climate change over France. *Front. Earth Sci.* **2020**, *8*, 104. [[CrossRef](#)]
16. Oliveira, S.L.; Pereira, J.M.; Carreiras, J.M. Fire frequency analysis in Portugal (1975–2005), using Landsat-based burnt area maps. *Int. J. Wildland Fire* **2011**, *21*, 48–60. [[CrossRef](#)]
17. Stacey, R.; Gibson, S.; Hedley, P. *European Glossary for Wildfires and Forest Fires*; European Union-INTERREG IVC: Cham, Switzerland, 2012.

18. Álvarez-Díaz, M.; González-Gómez, M.; Otero-Giraldez, M.S. Detecting the socioeconomic driving forces of the fire catastrophe in NW Spain. *Eur. J. For. Res.* **2015**, *134*, 1087–1094. [[CrossRef](#)]
19. Flannigan, M.D.; Wotton, B.M. *Climate, Weather, and Area Burned, in Forest Fires*; Elsevier: Amsterdam, The Netherlands, 2001; pp. 351–373.
20. Tymstra, C.; Jain, P.; Flannigan, M.D. Characterisation of initial fire weather conditions for large spring wildfires in Alberta, Canada. *Int. J. Wildland Fire* **2021**, *30*, 823–835. [[CrossRef](#)]
21. Nunes, A.; Lourenço, L.; Meira, A.C. Exploring spatial patterns and drivers of forest fires in Portugal (1980–2014). *Sci. Total Environ.* **2016**, *573*, 1190–1202. [[CrossRef](#)] [[PubMed](#)]
22. Cao, X.; Cui, X.; Yue, M.; Chen, J.; Tanikawa, H.; Ye, Y. Evaluation of wildfire propagation susceptibility in grasslands using burned areas and multivariate logistic regression. *Int. J. Remote Sens.* **2013**, *34*, 6679–6700. [[CrossRef](#)]
23. Holsinger, L.; Parks, S.A.; Miller, C. Weather, fuels, and topography impede wildland fire spread in western US landscapes. *For. Ecol. Manag.* **2016**, *380*, 59–69. [[CrossRef](#)]
24. Calviño-Cancela, M.; Chas-Amil, M.L.; García-Martínez, E.D.; Touza, J. Interacting effects of topography, vegetation, human activities and wildland-urban interfaces on wildfire ignition risk. *For. Ecol. Manag.* **2017**, *397*, 10–17. [[CrossRef](#)]
25. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Aryal, J. Forest fire susceptibility and risk mapping using social/infrastructural vulnerability and environmental variables. *Fire* **2019**, *2*, 50. [[CrossRef](#)]
26. Gigović, L.; Pourghasemi, H.R.; Drobnjak, S.; Bai, S. Testing a new ensemble model based on SVM and random forest in forest fire susceptibility assessment and its mapping in Serbia's Tara National Park. *Forests* **2019**, *10*, 408.
27. Tehrany, M.S.; Jones, S.; Shabani, F.; Martínez-Álvarez, F.; Tien Bui, D. A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. *Theor. Appl. Climatol.* **2019**, *137*, 637–653. [[CrossRef](#)]
28. Keeley, J.E.; Syphard, A.D. Historical patterns of wildfire ignition sources in California ecosystems. *Int. J. Wildland Fire* **2018**, *27*, 781–799. [[CrossRef](#)]
29. Rodrigues, M.; Jiménez, A.; de la Riva, J. Analysis of recent spatial-temporal evolution of human driving factors of wildfires in Spain. *Nat. Hazards* **2016**, *84*, 2049–2070. [[CrossRef](#)]
30. Choubin, B.; Abdolshahnejad, M.; Moradi, E.; Querol, X.; Mosavi, A.; Shamshirband, S.; Ghamisi, P. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Sci. Total Environ.* **2020**, *701*, 134474. [[CrossRef](#)] [[PubMed](#)]
31. Ghorbanzadeh, O.; Meena, S.R.; Abadi HS, S.; Piralilou, S.T.; Zhiyong, L.; Blaschke, T. Landslide Mapping Using Two Main Deep-Learning Convolution Neural Network Streams Combined by the Dempster-Shafer Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 452–463. [[CrossRef](#)]
32. Gholamnia, K.; Gudiyangada Nachappa, T.; Ghorbanzadeh, O.; Blaschke, T. Comparisons of diverse machine learning approaches for wildfire susceptibility mapping. *Symmetry* **2020**, *12*, 604. [[CrossRef](#)]
33. Iqbal, M.; Sameem MS, I.; Naqvi, N.; Kanwal, S.; Ye, Z. A deep learning approach for face recognition based on angularly discriminative features. *Pattern Recognit. Lett.* **2019**, *128*, 414–419. [[CrossRef](#)]
34. Iban, M.C.; Sekertekin, A. Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces, Turkey. *Ecol. Inform.* **2022**, *69*, 101647. [[CrossRef](#)]
35. Yoon, H.J.; Voulgaris, P. Multi-time predictions of wildfire grid map using remote sensing local data. In Proceedings of the 2022 IEEE International Conference on Knowledge Graph (ICKG), Orlando, FL, USA, 30 November–1 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 365–372.
36. Zhong, Y.; Ning, P.; Yan, S.; Zhang, C.; Xing, J.; Shi, J.; Hao, J. A machine-learning approach for identifying dense-fires and assessing atmospheric emissions on the Indochina Peninsula, 2010–2020. *Atmos. Res.* **2022**, *278*, 106325. [[CrossRef](#)]
37. Gould, J.S.; McCaw, W.L.; Cheney, N.P.; Ellis, P.F.; Knight, I.K.; Sullivan, A.L. *Project Vesta: Fire in Dry Eucalypt Forest: Fuel Structure, Fuel Dynamics and Fire Behaviour*; Csiro Publishing: Clayton, Australia, 2008.
38. Phelps, N.; Woolford, D.G. Comparing calibrated statistical and machine learning methods for wildland fire occurrence prediction: A case study of human-caused fires in Lac La Biche, Alberta, Canada. *Int. J. Wildland Fire* **2021**, *30*, 850–870. [[CrossRef](#)]
39. Cruz, M.G.; McCaw, W.L.; Anderson, W.R.; Gould, J.S. Fire behaviour modelling in semi-arid mallee-heath shrublands of southern Australia. *Environ. Model. Softw.* **2013**, *40*, 21–34. [[CrossRef](#)]
40. Nadeem, K.; Taylor, S.W.; Woolford, D.G.; Dean, C.B. Mesoscale spatiotemporal predictive models of daily human-and lightning-caused wildland fire occurrence in British Columbia. *Int. J. Wildland Fire* **2019**, *29*, 11–27. [[CrossRef](#)]
41. Woolford, D.G.; Martell, D.L.; McFayden, C.B.; Evens, J.; Stacey, A.; Wotton, B.M.; Boychuk, D. The development and implementation of a human-caused wildland fire occurrence prediction system for the province of Ontario, Canada. *Can. J. For. Res.* **2021**, *51*, 303–325. [[CrossRef](#)]
42. Rafaqat, W.; Iqbal, M.; Kanwal, R.; Song, W. Study of Driving Factors Using Machine Learning to Determine the Effect of Topography, Climate, and Fuel on Wildfire in Pakistan. *Remote Sens.* **2022**, *14*, 1918. [[CrossRef](#)]
43. Pérez-Porras, F.J.; Triviño-Tarradas, P.; Cima-Rodríguez, C.; Meroño-de-Larriva, J.E.; García-Ferrer, A.; Mesas-Carrascosa, F.J. Machine learning methods and synthetic data generation to predict large wildfires. *Sensors* **2021**, *21*, 3694. [[CrossRef](#)] [[PubMed](#)]
44. Cheng, S.; Prentice, I.C.; Huang, Y.; Jin, Y.; Guo, Y.K.; Arcucci, R. Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting. *J. Comput. Phys.* **2022**, *464*, 111302. [[CrossRef](#)]

45. Rubí, J.N.S.; de Carvalho, P.H.P.; Gondim, P.R.L. Application of machine learning models in the behavioral study of forest fires in the Brazilian Federal District region. *Eng. Appl. Artif. Intell.* **2023**, *118*, 105649. [[CrossRef](#)]
46. Mahamed, M.; Wittenberg, L.; Kutiel, H.; Brook, A. Fire Risk Assessment on Wildland–Urban Interface and Adjoined Urban Areas: Estimation Vegetation Ignitability by Artificial Neural Network. *Fire* **2022**, *5*, 184. [[CrossRef](#)]
47. Mahamed, M.; Wittenberg, L.; Kutiel, H.; Brook, A. A novel urban vegetation mapping approach for fire risk assessment: A Mediterranean case study. *Urban Ecosyst.* **2023**, *26*, 1263–1274. [[CrossRef](#)]
48. Polinova, M.; Wittenberg, L.; Kutiel, H.; Brook, A. Reconstructing pre-fire vegetation condition in the wildland urban interface (WUI) using artificial neural network. *J. Environ. Manag.* **2019**, *238*, 224–234. [[CrossRef](#)]
49. Kattel, D.B.; Yao, T.; Ullah, K.; Rana, A.S. Seasonal near-surface air temperature dependence on elevation and geographical coordinates for Pakistan. *Theor. Appl. Climatol.* **2019**, *138*, 1591–1613. [[CrossRef](#)]
50. Begum, B.A.; Biswas, S.K.; Pandit, G.G.; Saradhi, I.V.; Waheed, S.; Siddique, N.; Seneviratne, M.S.; Cohen, D.D.; Markwitz, A.; Hopke, P.K. Long-range transport of soil dust and smoke pollution in the South Asian region. *Atmos. Pollut. Res.* **2011**, *2*, 151–157. [[CrossRef](#)]
51. Ghali, R.; Akhloufi, M.A. Deep Learning Approaches for Wildland Fires Using Satellite Remote Sensing Data: Detection, Mapping, and Prediction. *Fire* **2023**, *6*, 192. [[CrossRef](#)]
52. Nur, A.S.; Kim, Y.J.; Lee, C.-W. Creation of Wildfire Susceptibility Maps in Plumas National Forest Using InSAR Coherence, Deep Learning, and Metaheuristic Optimization Approaches. *Remote Sens.* **2022**, *14*, 4416. [[CrossRef](#)]
53. Lin, X.; Li, Z.; Chen, W.; Sun, X.; Gao, D. Forest Fire Prediction Based on Long- and Short-Term Time-Series Network. *Forests* **2023**, *14*, 778. [[CrossRef](#)]
54. Higa, L.; Marcato Junior, J.; Rodrigues, T.; Zamboni, P.; Silva, R.; Almeida, L.; Liesenberg, V.; Roque, F.; Libonati, R.; Gonçalves, W.N.; et al. Active Fire Mapping on Brazilian Pantanal Based on Deep Learning and CBERS 04A Imagery. *Remote Sens.* **2022**, *14*, 688. [[CrossRef](#)]
55. Meira Castro, A.C.; Nunes, A.; Sousa, A.; Lourenço, L. Mapping the causes of forest fires in portugal by clustering analysis. *Geosciences* **2020**, *10*, 53. [[CrossRef](#)]
56. Nolan, R.H.; Boer, M.M.; Collins, L.; Resco de Dios, V.; Clarke, H.; Jenkins, M.; Kenny, B.; Bradstock, R.A. Causes and consequences of eastern Australia's 2019–20 season of mega-fires. *Glob. Chang. Biol.* **2020**, *26*, 1039–1041. [[CrossRef](#)]
57. Edwards, R.B.; Naylor, R.L.; Higgins, M.M.; Falcon, W.P. Causes of Indonesia's forest fires. *World Dev.* **2020**, *127*, 104717. [[CrossRef](#)]
58. Tien Bui, D.; Tuan, T.A.; Hoang, N.D.; Thanh, N.Q.; Nguyen, D.B.; Van Liem, N.; Pradhan, B. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides* **2017**, *14*, 447–458. [[CrossRef](#)]
59. Lawson, B.D.; Armitage, O. *Weather guide for the Canadian forest fire danger rating system*; Natural Resources Canada: Edmonton, AB, Canada, 2008; Available online: <https://scf.mnrcan.gc.ca/pubwarehouse/pdfs/29152.pdf> (accessed on 4 September 2023).
60. Horel, J.D.; Ziel, R.; Galli, C.; Pechmann, J.; Dong, X. An evaluation of fire danger and behaviour indices in the Great Lakes Region calculated from station and gridded weather information. *Int. J. Wildland Fire* **2014**, *23*, 202–214. [[CrossRef](#)]
61. de Jong, M.C.; Wooster, M.J.; Kitchen, K.; Manley, C.; Gazzard, R.; McCall, F.F. Calibration and evaluation of the Canadian Forest Fire Weather Index (FWI) System for improved wildland fire danger rating in the United Kingdom. *Nat. Hazards Earth Syst. Sci.* **2016**, *16*, 1217–1237. [[CrossRef](#)]
62. Romero, R.; Mestre, A.; Botey, R. A New Calibration for Fire Weather Index in Spain (AEMET). In *Advances in Forest Fire Research*; University of Coimbra: Coimbra, Portugal, 2014; p. 1044. Available online: <http://hdl.handle.net/10316.2/34013> (accessed on 21 November 2014).
63. Tian, X.R.; Zhao, F.J.; Shu, L.F.; Wang, M.Y. Changes in forest fire danger for south-western China in the 21st century. *Int. J. Wildland Fire* **2014**, *23*, 185–195. [[CrossRef](#)]
64. Abatzoglou, J.T.; Dobrowski, S.Z.; Parks, S.A.; Hegewisch, K.C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* **2018**, *5*, 170191. [[CrossRef](#)] [[PubMed](#)]
65. Rodell, M.; Houser, P.R.; Jambor, U.E.A.; Gottschalck, J.; Mitchell, K.; Meng, C.J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; et al. The global land data assimilation system. *Bull. Am. Meteorol. Soc.* **2004**, *85*, 381–394. [[CrossRef](#)]
66. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*. [[CrossRef](#)]
67. Pesaresi, M.; Huadong, G.; Blaes, X.; Ehrlich, D.; Ferri, S.; Gueguen, L.; Halkia, M.; Kauffmann, M.; Kemper, T.; Lu, L.; et al. A global human settlement layer from optical HR/VHR RS data: Concept and first results. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2102–2131. [[CrossRef](#)]
68. Kennedy, C.M.; Oakleaf, J.R.; Theobald, D.M.; Baruch-Mordo, S.; Kiesecker, J. Managing the middle: A shift in conservation priorities based on the global human modification gradient. *Glob. Chang. Biol.* **2019**, *25*, 811–826. [[CrossRef](#)]
69. Weiss, D.J.; Nelson, A.; Vargas-Ruiz, C.A.; Gligorić, K.; Bavadekar, S.; Gabilovich, E.; Bertozzi-Villa, A.; Rozier, J.; Gibson, H.S.; Shekel, T.; et al. Global maps of travel time to healthcare facilities. *Nat. Med.* **2020**, *26*, 1835–1838. [[CrossRef](#)]
70. Hengl, T. *Soil Bulk Density (Fine Earth) 10 × kg/m-Cubic at 6 Standard Depths (0, 10, 30, 60, 100 and 200 cm) at 250 m Resolution*; Zenodo, Ed.; CERN: Geneva, Switzerland, 2018.

71. Hengl, T. *Soil Texture Classes (USDA System) for 6 Soil Depths (0, 10, 30, 60, 100 and 200 cm) at 250 m*; Zenodo, Ed.; CERN: Geneva, Switzerland, 2018.
72. Hengl, T.; Nauman, T. *Predicted USDA Soil Great Groups at 250 m (Probabilities)*; Zenodo, Ed.; CERN: Geneva, Switzerland, 2018.
73. Buchhorn, M.; Lesiv, M.; Tsendbazar, N.E.; Herold, M.; Bertels, L.; Smets, B. Copernicus global land cover layers—Collection 2. *Remote Sens.* **2020**, *12*, 1044. [[CrossRef](#)]
74. Didan, K. *MOD13A1 MODIS/Terra Vegetation Indices 16-Day L3 Global 500 m SIN Grid V006*; DAAC, Ed.; NASA: Washington, DC, USA, 2015.
75. Myneni, R.; Knyazikhin, Y.; Park, T. *MCD15A3H MODIS/Terra + Aqua Leaf Area Index/FPAR 4-Day L4 Global 500 m SIN Grid V006*; DAAC, Ed.; NASA: Washington, DC, USA, 2015.
76. Center for International Earth Science Information Network-CIESIN-Columbia University. *Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11*; NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, NY, USA, 2018.
77. Zhang, G.; Wang, M.; Liu, K. Deep neural networks for global wildfire susceptibility modelling. *Ecol. Indic.* **2021**, *127*, 107735. [[CrossRef](#)]
78. Mohajane, M.; Costache, R.; Karimi, F.; Pham, Q.B.; Essahlaoui, A.; Nguyen, H.; Laneve, G.; Oudija, F. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecol. Indic.* **2021**, *129*, 107869. [[CrossRef](#)]
79. Florath, J.; Keller, S. Supervised machine learning approaches on multispectral remote sensing data for a combined detection of fire and burned area. *Remote Sens.* **2022**, *14*, 657. [[CrossRef](#)]
80. Kumar, M.; Li, S.; Nguyen, P.; Banerjee, T. Examining the existing definitions of wildland-urban interface for California. *Ecosphere* **2022**, *13*, e4306. [[CrossRef](#)]
81. Radeloff, V.C.; Hammer, R.B.; Stewart, S.I.; Fried, J.S.; Holcomb, S.S.; McKeefry, J.F. The wildland-urban interface in the United States. *Ecol. Appl.* **2005**, *15*, 799–805. [[CrossRef](#)]
82. Holsinger, L.M.; Parks, S.A.; Saperstein, L.B.; Loehman, R.A.; Whitman, E.; Barnes, J.; Parisien, M.A. Improved fire severity mapping in the North American boreal forest using a hybrid composite method. *Remote Sens. Ecol. Conserv.* **2022**, *8*, 222–235. [[CrossRef](#)]
83. Brook, A.; Hamzi, S.; Roberts, D.; Ichoku, C.; Shtober-Zisu, N.; Wittenberg, L. Total Carbon Content Assessed by UAS Near-Infrared Imagery as a New Fire Severity Metric. *Remote Sens.* **2022**, *14*, 3632. [[CrossRef](#)]
84. Giddey, B.L.; Baard, J.A.; Kraaij, T. Verification of the differenced Normalised Burn Ratio (dNBR) as an index of fire severity in Afrotropicate Forest. *S. Afr. J. Bot.* **2022**, *146*, 348–353. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.