Role of Machine Learning Algorithms in Forest Fire Management: A Literature Review

Muhammad Arif1, Khoud K Alghamdi1, Salma A Sahel1, Samar O Alosaimi1, Mashael E Alsahafi, Maram A Alharthi1 and Maryam Arif2

1Department of Computer Science, College of Computer and Information Systems, Umm Alqura University, Saudi Arabia
2Freelance Artificial Intelligence Researcher, Pakistan

Abstract

Forest fire disasters are recently getting lots of attention due to climate change globally. Globally, climate changes are rapidly changing the fire patterns on Earth. Effective fire management requires accurate information about the fire occurrence, its spread, and impact on the environment. Prediction of fire activities in the forest guides the authorities to make optimal, efficient, and sound decisions in fire management. This paper aims to summarize recent trends in the forest fire events prediction, detection, spread rate, and mapping of the burned areas. Furthermore, fire emissions in terms of smoke also put the Earth’s public health and ecological system at greater risk. Hence, future policymaking can be more accurate in saving billions of dollars, improving the healthy environment and ecological cycle for the inhabitants of this Earth. This paper provides a comprehensive review of the usage of different machine learning algorithms in forest fire or wildfire management. Furthermore, we have identified some potential areas where new technologies and data can help better fire management decision making.

Keywords

Machine Learning, Wildfire, Forest Fires, Fire Susceptibility, Forecasting

Introduction

Chris Maser says in his book [1] that “What we are doing to the forests of the world is but a mirror reflection of what we are doing to ourselves and to one another”.

Forest, a precious resource provided by nature, is the house for numerous organisms; we tend to take forests for granted, underestimating how irreplaceable they are for humankind. Our endurance is linked with Forest. Hydrologic cycle, soil conversation, prevention of climate change, and preservation of biodiversity are a few examples of many environmental benefits that are attained through Forests [2]. Regardless of all the benefits, we gain we still are permitting them to vanish. Deforestation is caused due to anthropogenic activities or natural disturbances and affects nature and climate [3].

Man is clearing away the forest area for agriculture, industrial development, or mineral/Oil extraction, and in certain regions, degradation of the forest is practiced to develop modern cities [4]. Not all deforestation is intentional, and forest fires are one of the most reoccurring natural disasters worldwide. Each year, fires burn millions of hectares of forest. Fires are a part of nature and have enormous side effects on economics, ecological and social culture globally and regionally. Fires destroy biodiversity and endanger the ecosystem killing the large mass of inhabitants. Moreover, forest fires lead to the loss of many timbers growing for hundreds of years, which leads to significant material losses to the countries that contain it due to the importance of wood in making many raw materials. Valuable Tree species decline after a fire, creating an enormous down imprint on the nation’s economy [5,6]. Many scientists have studied the causes of these forest fires to reduce their economic and environmental impacts. Furthermore, they are trying to work out how to predict future forest fires to minimize the damages that are caused by them. Scientists and researchers have been in a debate on various causes of forest or wildfire fires. Predicting the source of wildfires could lead to many considerable benefits for hu-
man health, human life, the economy, and the environment. If we can predict the occurrence of fires before their occurrence will help emergency teams intervene and prevent them from happening or reduce its bad effects. Also, predicting the forest fire may help to protect the life of humans and other inhabitants, protecting human health from the damage caused by smoke from forest fires and reducing accidents in vehicles and aircraft moving nearby due to smoke generated by huge fires [7]. Table 1 gives a list of some of the deadliest forest fires in history [8].

Recently Australia faced disastrous bushfires causing severe damage. Over three months, more than seventeen million hectares of forest were burnt, two dozen lives were lost, more than two thousand homes were destroyed, and around a billion animals were killed. In addition to the suffering, loss of life, and property damage, the whole nation experienced massive disruption to infrastructure and the economy, and unprecedented air pollution [9]. A list of annual wildfires from the National Interagency Coordination Center, USA can also be found at [10]. Information about large wildfires in Europe can be found in [11]. Problems related to the worldwide fires are discussed in [12]. From the above-mentioned few facts and figures, we can comprehend the deadliest impacts of Forest fires. The scientific community is investing its efforts to assemble a system to prevent forest fires, and the primary objective is to predict and explain the cause of forest fires. Numerous researchers had contributed to this field. Several computational tools or simulators have been developed to aid in predicting the Fire propagation at the right time and preventing it. In case, Forest Fire erupts, then how to discover the reason, the rate and the direction of fire spread and the flame intensity, etc. The early prediction will reduce the death toll and help Fire Fighters to control the spread of fire. Achieving correct prediction is vital, and for that, we have to analyze as much information as possible [13]. The mean annual maximum temperature is increasing at the rate of 0.184 centigrade per decade from 1979 to 2013 across the vegetated land surface [14]. Hence, the length of fire seasons is increasing globally and getting severe in fire intensity in different fire-prone areas of Earth [15,16] resulting in an increasing number of adult trees mortality.

Scope of ML Algorithms in Forest Fire Management Frame-Works

There are four important areas of research related to forest or wildfire as follows,

- Prediction of Fire Occurrence (time and location)
- Detection of already started fire event
- Prediction of Spread of Wildfire (burned area in future)
- Detection of Burned Area due to fire

The accurate prediction of the occurrence of forest fires is a challenging task [17]. Prediction is not only knowing the exact time of the start of a forest fire but also the location where this fire occurs. Hence it is a combination of temporal and spatial prediction problems, making it more complex and challenging for the researchers. Machine learning algorithms are applied to forecast events in many real-life applications which are highly stochastic in nature, naming a few like stock market crisis forecasting [18-22] stock market investment strategies [23-25], weather extreme events forecasting [26-28], predictive outage estimation [29-31], forecasting health-related issues [32-34].

Machine learning algorithms are also used extensively in the image processing/computer vision based applications [35-37], thermal infrared image based applications [38-40], satellite imagery [41-43], sensor signal processing [44-46], big data processing [47-49] etc. There are other numerous application areas of machine learning algorithms including energy sector [50,51] and activity monitoring using wearable sensors [52,53]. Details of different machine learning algorithms and how they are applied to different types of problems can be found some good textbooks like [54-56]. Forest fire prediction and detection require spatial placement of sensors. Hence optimization of placement locations of the sensor nodes, whether on ground or airborne, is also very important [44,57,58]. Overall framework optimization is also required to reduce the operational and deployment costs without compromising the quality of operations. In the forest fire-related literature, researchers have access to lots of data coming from different resources. A set of features is selected according to the relative importance of the features considering their statistical significance. Machine learning algorithms can be helpful in a better selection of important features [59-61]. Selecting only relevant and essential features may provide optimal performance with less computational and operational costs. Machine Learning algorithms can further be applied to decision making and decision support systems for forest fire pre-

<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
<th>Size</th>
<th>Area</th>
<th>Casualties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miramichi Fire</td>
<td>1825</td>
<td>3 million acres</td>
<td>New Brunswick, Canada</td>
<td>160 to 300</td>
</tr>
<tr>
<td>Peshtigo Fire</td>
<td>1871</td>
<td>1,200,000 acres</td>
<td>Wisconsin, United States</td>
<td>1,200 to 2,500</td>
</tr>
<tr>
<td>Black Friday Bushfires</td>
<td>1939</td>
<td>5 million acres</td>
<td>Victoria, Australia</td>
<td>killed 71</td>
</tr>
<tr>
<td>Greek Forest Fires</td>
<td>2007</td>
<td>670,000 acres</td>
<td>Greece</td>
<td>84</td>
</tr>
<tr>
<td>Black Dragon Fire</td>
<td>1987</td>
<td>18 million acres</td>
<td>China</td>
<td>191</td>
</tr>
<tr>
<td>Cloquet Fire</td>
<td>1918</td>
<td>250,000 acres</td>
<td>Minnesota, United States</td>
<td>453</td>
</tr>
<tr>
<td>Great Hinckley Fire</td>
<td>1894</td>
<td>350,000 acres</td>
<td>Minnesota, United States</td>
<td>418+</td>
</tr>
<tr>
<td>Indonesian forest fires</td>
<td>1997</td>
<td>8 million hecta acres</td>
<td>Indonesia</td>
<td>240</td>
</tr>
</tbody>
</table>
Forests recorded the fire data for ten years from 2008 to 2018 in the region. Features such as temperature, relative humidity, month, distance from leveling, wind speed, distance from agricultural land, amount of burned area, tree types, and distance from the road are recorded. A Bayesian network model is trained on the data to predict possible causes of the forest fire. The Bayesian network produced a high AUC of 0.9, 0.89, 0.89, 0.82 for hunting, picnic, stubble burning, and shepherd fire, respectively.

Simulation of forest fires to study fire behavior is very useful in predicting the rate of fire spread, fire size, and the number of trees lost in the fire. BehavePlus V6 is currently used by the US and Canada to study fire behavior [69]. It is one of the most widely used forest fire modeling systems that predict wildfire behavior and prescribe fire planning. Other similar systems also exist [70,71]. BehavePlus provides much useful information, including the rate of fire spread and spotting distance, fire effect, and fire environment, and includes over 40 fire models. Rapp modeling system (http://redapp.org) is developed in Canada with the support of the Canadian Interagency Forest Fire Centre (CIFFC). Fire is a fire behavior prediction simulator that calculates fire effect on stand characteristics [72]. Stacy A Drury [73] compared predicted fire behaviors from four models (BehavePlus, RedAPP, CanFIRE, and Crown Fire Irrigation and Spread System) with the observed fire behavior on the Alaskan black spruce forest. He studied the rate of fire movement, predicted flame length, energy, and ecological impact, and concluded that Canadian models, including RedAPP and CanFIRE provide more accurate predictions than BehavePlus. Similarly, many tools exist for fire, and smoke models, which are crucial in decision making and planning to tackle the forest or a wild-land fire [74-77]. A lot of research has been done in the last two decades, especially after the increase of computation power during the last few years [77-80]. Satellite active fire data such as Visible Infrared Imager Radiometer Suite (VIIRS) [81], moderate-resolution imaging spectro-radiometer (MODIS) [82] etc., can be

**Figure 1:** Application of ML algorithms to forest fire management.
used to validate the simulation models [83]. Performance of the forest fire prediction models depends on measuring important input parameters such as topology, meteorological conditions, vegetation, fire front situation, etc. Evolutionary algorithms are used to reduce the uncertainty in the forest fire propagation models [13]. Evolutionary statistical systems can reduce the uncertainty in the input parameters of the forest fire prediction models to improve their accuracy [13,84]. We will present a review of how different machine learning algorithms are used to predict the occurrence and spread, decision-making, and planning of the forest or wild-land fires.

**Satellite-based monitoring of forest fire**

Satellite imagery-based fire detection and finding out the fire affected region (hot spots) have their own benefits as a satellite can monitor large areas and detect fires. Sifakis, et al. [85] used the Spinning Enhanced Visible and Infrared Imager (SEVIRI) data collected from the geostationary Meteosat Second Generation (MSG) satellite for fire detection and tracking in Greece during 2007 when there was disastrous wildfire occurred. EUMETSAT’s Active Fire Monitoring Image processing algorithm was used to quickly locate the hot spots with 85% accuracy and helped the authorities correctly calculate the fire extent and progress at the country level. An agglomerative hierarchical clustering algorithm is used on the satellite images to detect the hotspots and predict the fire spread direction [86]. The Moderate Resolution Imaging Spectroradiometer (MODIS) sensors on satellites of NASA collect information in terms of 36 spectral bands from the whole surface of the Earth. Based on a certain threshold on the temperature of a pixel, hotspots are identified [87]. Chiaraviglio, et al. [88] proposed an alpha algorithm to estimate the fire perimeter in situations when cloud or smoke blocks the fire front position. The algorithm implemented in the European Forest Fire Information System (EFFIS) to estimate perimeters of fire from MODIS hotspots on fire data occurred in 2014 in Sweden. Unmanned aerial imagery is useful in natural resources management [89]. Lots of remote sensing data through aircraft and satellite imagery is collected during the fire and after the completion of a fire event. Classification of the post-wildfire burned area to calculate the damage through satellite imagery can be used in future decision making and repair of losses. Monitoring Trends in Burn Severity (MTBS) is a US project to map fire severity [90]. Mallinis and Koutsias [91] evaluated different classification methods to map the burned area after the forest fire. They have used three study sites, including the fire at Parnitha National Park in 2007, fire at Kassandra Peninsula in 2006, and fire at Alexandroupoli in 2009, all in Greece. Aerial imagery is acquired through satellite. Burned area mapping was classified, and the performance of ten classifiers is tested on these study sites. Neural networks (NN), SVM, Logistic regression, and maximum likelihood classifiers mapped the burned area in Parnitha park with 96% accuracy. In Kassandra fire, NN, SVM, and CART performed better with an accuracy between 93% and 94%. Whereas in Alexandroupoli fire, NN and Object-based image analysis were the best with 95% accuracy. Mohler and Goodin [92] devised suitable mapping methods by comparing seven combinations of spectral bands and indices from the MODIS sensor using pixel-based and object-based classifications of burned area. They have found that at least 250 m is the spatial resolution required to map the burned area accurately using red and NIR bands of the MODIS sensor. Meng, et al. [93] proposed a method to produce high-resolution fire severity maps by using VHR satellite imagery data (acquired by WV-2 spaceborne platform) of the wildfire at Long Island, New York, USA. Multiple spectral indices were calculated. The burned mask was generated, and burned pixels were classified into three severity levels (Low, Medium, and High). RF classifier produced good classification results, and the burn severity map showed an overall accuracy of 84% in the sub-crown scale and 82% on the crown scale. A good survey about the developments in the research of mapping post-fire burned areas accurately from satellite observations [94,95]. Pinto, et al. [96] explored the use of deep learning approach on multi-spectral images obtained by Visible Infrared Imaging Radiometer Suite (VIIRS) 750 m bands from satellite imagery. Daily images between 2012 and 2018 of Red, NIR, and MIR reflectance of five regions in California, Portugal, Brazil, Mozambique, and Australia are recorded and resampled to a regular grid using bilinear interpolation. Different datasets collected from different satellites are used for validation. The input data consists of 64 images of size 128 × 128 and 4 data channels (Red, NIR, MIR, and Active Fires) and U-Net architecture with several modifications are used, which contains thirteen layers between input and output layers (called BA-Net by authors). Temporal validation of BA-Net achieved the lowest overall Mean Absolute Error (MAE) for all regions as 1.13 days compared to MCD64A1C6 and FireCC151 algorithms. In terms of spatial validation, the BA-Net Dice score was 0.678 was found to be near to the MCD64A1C6 (Dice score 0.687) and better than FireCC151 (Dice score 0.656).

Sousa, et al. [97] highlighted some of the challenges in identifying the wildfire through image surveillance systems due to outdoor settings having varying lighting conditions, reflection from different surfaces, and smoke may create fog conditions. Authors have augmented different publically available datasets including Portuguese Fire Fighters Portal Database (PFPDB), Corsican fire database and other datasets. Transfer learning is used on pre-trained ImageNet deep learning architecture to classify these images in the augmented dataset as fire or not fire and the model achieved an accuracy of 93% on 10 folds cross-validation. Pourghasemi, et al. [98] assessed forest fire susceptibility in the Fars Province Iran using Landsat-8 OLI and MODIS satellite images from 358 locations. They have identified ten most relevant factors (elevation, slope, topographical wetness index, aspect, distance from urban areas, annual mean temperature, land use, distance from the road, annual mean rainfall, and distance from the river) from the dataset and trained three machine learning algorithms, Boosted regression tree (BRT), General linear model (GLM) and Mixture discriminant analysis (MDA). Forest fire susceptibility map is generated by these three models with four risk levels, very high, high, moderate and low. The area under the ROC curve (AUC) for BRT, GLM, and MDA was found to be 0.89, 0.86, 0.866 respectively. Identification of high-risk areas can help the government to make an effective fire fighting plan.
**Table 2:** Satellite-based monitoring of forest fire.

<table>
<thead>
<tr>
<th>Reference, Year</th>
<th>Purpose</th>
<th>Data</th>
<th>Classifier</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sifakis, et al. [85]</td>
<td>Hot spot detection</td>
<td>MSG-SEVIRI geostationary data</td>
<td>Image Processing</td>
<td>85% accuracy</td>
</tr>
<tr>
<td>Giglio, et al. [87]</td>
<td>Hotspot detection</td>
<td>36 spectral bands (MODIS)</td>
<td>Image Processing</td>
<td>False Alarm Rates less than 2%</td>
</tr>
<tr>
<td>Mallinis and Koutsias [91]</td>
<td>Burned Area mapping</td>
<td>Landsat Thematic Mapper (TM) imagery</td>
<td>SVM and 9 others</td>
<td>More than 93% Accuracy</td>
</tr>
<tr>
<td>Mohler and Goodin [92]</td>
<td>Burned Area mapping</td>
<td>spectral bands (MODIS)</td>
<td>Supervised minimum distance classification</td>
<td>90% Accuracy</td>
</tr>
<tr>
<td>Pourghasemi, et al. [98]</td>
<td>Forest fire susceptibility map</td>
<td>OLI and MODIS satellite images</td>
<td>Boosted regression tree</td>
<td>AUC is 0.89</td>
</tr>
<tr>
<td>Gibson, et al. [99]</td>
<td>Fire Severity Mapping</td>
<td>Sentinel2 satellite imagery</td>
<td>Random Forest</td>
<td>98% Accuracy</td>
</tr>
</tbody>
</table>

Table 2 provides some of the machine learning algorithms’ useful applications to detect the burned area or fire severity mapping [85,87,91,92,98,99].

**Wireless sensors network-based monitoring and prediction**

In satellite-based monitoring of forest fires, it is difficult to locate the fire accurately, and satellite images are of low spatial and temporal resolution. Wireless sensor networks are being used in many real-life applications successfully including animal tracking and health monitoring [100,101], precision agriculture [102], environmental monitoring, security and surveillance [103], smart buildings [104,105] and smart cities [106], health care [107], monitoring water quality [108] etc.

A wireless sensor network can be deployed in larger areas where there is a history of forest fires at ground level. Wildfire monitoring using wireless sensors to measure temperature, relative humidity, and barometric pressure, and whenever there is a wildfire starts, these sensors can issue warnings to the base station [109]. Early warning about the fire started can initiate the early fire extinguishing process. Wireless sensor networks can also be used to sense the wind speed to monitor the local spread of fire [110]. Lin, et al. [111] used the fuzzy rule-based system to predict the forest fire using rechargeable wireless sensor networks. Bui, et al. [7] introduced a new machine learning algorithm called DFP-MnPbAn algorithm based on Differential Flower Pollination (DFP) and mini-match backpropagation (MBps) for spatial modeling of forest fire danger. Ten features, namely land slope, aspect, elevation, land use, Normalized Differential Vegetation Index (NDVI), distance to the road, distance to the residential area, temperature, wind speed, and rainfall are used to train the proposed DFP-MnPbAn algorithm. The data consists of 540 historical fire locations provided by the Department of Forest Protection (Ministry of Agriculture and Rural Development of Vietnam). The proposed algorithm is compared with Random Forest (RF), SVM, Least Square SVM, BpAnn, and Particle Swarm Optimized Neural Fuzzy model (PSO-NF) algorithms. Classification accuracy of 70% training and 30% testing datasets is the best for the proposed algorithm DFP-MnPbAn (Accuracy 89% and AUC - 0.94) compared to other algorithms, which shows classification accuracy more than 85% also. Airborne sensor platforms on drones or small planes can also be used for airborne fire sensing. Allison, et al. [112] reviewed variously manned, and unmanned aerial platforms used to detect the fire. Electro-optical sensors, including videos and thermal infrared, can be useful remote sensing technologies for fire detection using automatic image processing algorithms. Heat, smoke, infrared, and light detection can be used in a combination or separate fire detection during daytime and night time. Lee, et al. [113] applied deep convolutional neural networks on the aerial images captured by unmanned airborne vehicles and achieved an accuracy of 94% with a classification time of about 7 seconds on a small prototype fire experiment. Bosch, et al. [114] presented an idea of a multi-sensor wireless network system for automatic detection of forest fires. Airborne sensing systems can provide a larger coverage area with a quicker response at the cost of flight expenses. Spectroscopic analysis to map the wild-land fire effects can be used through remotely sensed imagery by unmanned aircraft systems. The class means spectral reflectance from black ash, and white ash helps the classifier to differentiate between low severity of burned area and high severity burned areas [115]. Moreover, the separation between black ash and vegetative class is also possible through the mean reflectance of visible and near-infrared spectrum above 350 nm. Arkin, et al. [116] emphasized the importance of higher spatial resolution fire severity maps. In this regard, they have proposed an integrated framework consisting of high spatial resolution post-fire imagery and digital aerial photogrammetric point clouds acquired from an unmanned aerial vehicle (UAV). The correlation-based feature selection method is used to select the most relevant features from the suite of spectral, structural, and textural features extracted from the data acquired by UAV. Random forest classifier is used to produce a 5-meter and 1-meter resolution map with a high accuracy of 89.5% for 5-m resolution and 85.4% for 1-m resolution maps. Table 3 summarize few applications of wireless sensors network to monitor and predict the forest fire and issues alerts [7,111,113,116-118].

**Prediction of fire risks**

Fire risk analysis is critical in wildfire and forest fire management. Ignition of fire in the forest is a highly stochastic
process, and fire risk can be predicted with a certain probability only. Three components, the likelihood of fire event, the intensity of the fire, and the effect of fire in terms of the burn area, are essential in the fire risk analysis. A good review of the wildfire risk analysis can be found in [119,120]. Logistic regression analysis is used in many fire risk analysis, and occurrence of fire probability in temporal and spatial domains [120-122]. Methods to predict the Fire danger risk can be divided into statistical methods and dynamical methods [123]. Statistical methods can forecast long-term fire risk using topological parameters, frequency of occurrences and vegetation types etc. Dynamical methods are based on data that changes daily, and they can predict short term fire risk indices. A big list of factors that can affect the occurrence of wildfires are given in [124] that can be used in the prediction of wildfire and forest fire risks. To predict the probability of forest fire occurrences, different fire prediction indices based on mathematical modeling are proposed between 1960 and 1970 including Fire Weather Index (FWI), Angstrom index, KBDI index, Nesterov index, and Baumgartner index based on meteorological parameters [125]. Human factors are also associated with forest fire risks, including socio-economic transformation in rural areas, human presence in the urban area, traditional activities linked with fire, accidental or negligent events, intentional fires, and forest policy [126]. Logistic regression models can be used to estimate the human-caused fires [127,128]. A good review of different models including logistic regression, Autoregressive regression, Classification and Regression Trees (CART), Poisson regression, etc. for short term and long term prediction of the Human-caused fire occurrence events can be found in [129]. Lightning with high winds can start the fire, and winds can drive high fire growth [130]. Canadian Forest Fire Danger Rating System (CFFDRS) [131] rate the risk of fire as the Fire Weather Index (FWI) based on many components. Fine Fuel Moisture Code (FFMC) is calculated based on daily rainfall, relative humidity, Temperature, and Wind. Duff Moisture Code (DMC) is based on rainfall, relative humidity, and temperature. It indicates fuel consumption in the moderate duff layer. Drought Code (DC) represents the average moisture content of the deep organic layer and depends on daily rainfall and temperature. Initial Spread Index (ISI) is the predicted fire spread rate and depends on wind speed and FFMC, whereas Buildup Index (BUI) is the measure for available fuel for combustion and depends on DMC and DC values. Finally, FWI is calculated using numerical values of ISI and BUI. Safi and Bouroumi [132] optimized neural network architecture to predict total burned area based on 12 parameters including (location coordinates, day and month, FFMC, DMC, DC, ISI, temperature, relative humidity, wind speed, daily rainfall) on dataset available at machine learning repository [133]. The error rate of the best neural network was 9%. Sakr, et al. [134] predicted fire danger index on a scale from one to four based on temperature, average humidity, solar radiation, wind speed, and cumulative precipitation level using neural networks (NN) and support vector machine (SVM). Prediction accuracy of fire/no fire binary decision by NN and SVM is found to be more than 90% for all months in summer on the data of Lebanese Agricultural Research Institute (LARI), covers the Lebanese territory and spans the nine years between 2000 and 2008. Performance on the four-level scale was not found to be accurate. Spatial patterns of fire occurrences in Mediterranean Europe was studied by Oliveira, et al. [135]. The machine learning algorithms estimated the number of fires occurring per square area (km²). Data for training and prediction by the models were obtained from the European Fire Database of the European Forest Fire Information System. Due to irregular patterns present in the region (space and time), human factors were also considered in contributing to the fire occurrences. Hence non-parametric models, random forest, and multiple regression methods were selected for the prediction of the fire occurrences. Random forest performed better than the linear regression model, and 96% of the variance is successfully explained by this model using twelve important features. They have found total precipitation in the off-season as the most important feature and, interestingly, unemployment as the second important feature in the list. Hence human-made fires may be linked to the dissatisfaction of the rural population due to employment. Cortez and Marias [136] used multiple regression, SVM, RF, and NN algorithms to predict the burn area by small fires in the northeast region of Portugal. The SVM classifier performed the best with only four direct weather inputs consisting of temperature, relative humidity, rain, and wind speed (Error of 12.7%).

Alonso-Betanzos, et al. [137] predicted forest fire risks (four categories, low, medium, high, extreme) using neural networks so that firefighters can focus on high-risk areas. Four inputs are used, namely, the temperature of the day, daily humidity, daily rainfall, and fire history. Map of Galicia, Spain, was divided into zones, and for every zone, the risk cat-

### Table 3: Wireless sensors network based monitoring and prediction.

<table>
<thead>
<tr>
<th>Reference, Year</th>
<th>Purpose</th>
<th>Data</th>
<th>Classifier/Predictor</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin, et al. [111]</td>
<td>Forest fire prediction</td>
<td>Weather, location and time history data</td>
<td>Fuzzy inference system</td>
<td>75% prediction accuracy</td>
</tr>
<tr>
<td>Lee, et al. [113]</td>
<td>Fire detection</td>
<td>UAV imagery</td>
<td>Deep convolutional NN</td>
<td>&gt; 95% accuracy</td>
</tr>
<tr>
<td>Arkin, et al. [116]</td>
<td>Fire severity mapping</td>
<td>UAV imagery</td>
<td>Random Forest classifier</td>
<td>89% accuracy</td>
</tr>
<tr>
<td>Jiao, et al. [117]</td>
<td>Forest fire detection</td>
<td>UAV-based aerial images</td>
<td>Convolution neural network (CNN)</td>
<td>83% Accuracy</td>
</tr>
<tr>
<td>Barmpoutis, et al. [118]</td>
<td>Early fire detection</td>
<td>Aerial 360-degree imagery</td>
<td>Convolution neural network (CNN)</td>
<td>94% Accuracy</td>
</tr>
</tbody>
</table>
category is predicted by the neural network with a prediction accuracy of 79%.

Arpaci, et al. [138] have conducted a study on a dataset of fire presence only in Tyrol province, Austria, from 1993 to 2011, including the topography data vegetation, climate, and socio-economic parameters. Their study was to predict the driving parameters of the spatial fire distribution and locate the areas with fire danger. Maximum Entropy (MaxEnt) and Random forest algorithms are applied to determine the main governing parameters of fire distribution. Authors have analyzed a comprehensive list of parameters consisting of Socio-economics (n = 6), infrastructure (n = 19), Forest vegetation (n = 4), Topology (n = 3), and FWI as fire risk days per year. They have found that population density and climate are two significant factors contributing to spatial fire susceptibility. Random forest predicted fire ignition distribution (fire or no fire in spatial location) with a classification accuracy of 75 to 78%. Further classification of fire susceptibility levels into six classes starting from very low to extremely high was also compared between MaxEnt and RF algorithms. Prediction of both algorithms was comparable except in the extremely high fire susceptibility class where MaxEnt found 28 spatial locations compared to two found by the RF algorithm. Liang, et al. [139] compared three types of neural networks, namely backpropagation NN (BPNN), recurrent NN (RNN), and long short-term memory (LSTM) to predict different scales of wildfire in the forest of Alberta, Canada. The scale of wildfire (6 levels according to the severity of fire) was calculated from the wildfire data (Location, duration of fire, and burned area), whereas weather data (features from Temperature, Rain, Snow, Wind speed) after normalization was used to estimate the wildfire scale. The LSTM model obtained the best predictive accuracy of 90.9% out of all three algorithms on the test data set. Different machine learning algorithms are used in predicting the human-caused wildfire occurrences. Rodrigues and Riva [140] covered the entire peninsula of Spain, except some cities to evaluate human-induced wildfires in Spain. This study used one of the oldest databases in Europe, which is Spanish EGIF, from the year 1988 to 2007. It contained 308,893 fire during this period consists of approximately 93,537 samples, and it was divided into 60% for training, 40% for testing, and 37,429 samples for validation. Different machine learning algorithms, Linear Regression (LR), SVM, RF, and Boosting Regression Trees (BRT), are compared on the data set. The results indicated that Random Forest (RF) and BRT had shown the best results with AUC = 0.746 and 0.730, respectively. Tien Bui, et al. [141] proposed a machine learning methodology that analyzes and forecasts the spatial patterns of the tropical forest fires in Lao Cai state, Vietnam. The database used in this study was created for the Geographic Information System (GIS). The study area was about 2253 Km² with a 257 fire location. Ten features were recorded: Temperature, slope, road distance, height, vegetation index, humidity, wind speed, land use, rainfall, and aspect. The suitability of these features with forest fires was determined and evaluated by the algorithm for mutual information. A new hybrid AI model called MARS-DFP based on Multivariate Adaptive Regression Splines and Differential Flower Pollination fire locations is proposed. They graded the danger rating of fire to five levels: Very low, low, moderate, high, and very high. The MARS-DFP algorithm achieved high performance (Accuracy = 86.57% and AUC = 0.91) compared to other machine learning models (BPANN, ANFIS, RBFANN, RE, and MARS). Jaafari, et al. [142]. Ferreira, et al. [143] emphasized that the prediction of fire occurrences on global and seasonal scales can be very beneficial for fire management and control. Global models are used for long-time forecasting (decades to centuries), whereas seasonal models are short-range forecasting only specific to seasonal time scales. Authors have used different time series forecasting methods for seasonal fire forecasting as time series forecasting is fast and straightforward. Several forecasting methods (naïve forecasting, autoregressive integrated moving average models (ARIMA), Exponential smoothing, Short term load forecasting, Generalized linear model, NN) are compared for seasonal fire forecasting on time series of

<table>
<thead>
<tr>
<th>Reference, Year</th>
<th>Purpose</th>
<th>Data</th>
<th>Classifier/Predictor</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safi and Bouroumi [132]</td>
<td>Fire weather index prediction</td>
<td>Weather observations</td>
<td>NN</td>
<td>9% error rate</td>
</tr>
<tr>
<td>Oliveira, et al. [135]</td>
<td>Fire density</td>
<td>Environmental, demographic, infrastructure, socio-economic</td>
<td>RF</td>
<td>96% variance explained</td>
</tr>
<tr>
<td>Cortez and Moarias [136]</td>
<td>Fire weather index prediction</td>
<td>Weather observations</td>
<td>SVM</td>
<td>12.7% Error</td>
</tr>
<tr>
<td>Arpaci, et al. [138]</td>
<td>Fire prediction</td>
<td>Weather, topology, infrastructure, socio-economic</td>
<td>RF</td>
<td>78% Error</td>
</tr>
<tr>
<td>Liang, et al. [139]</td>
<td>Wildfire scale Prediction</td>
<td>Weather and wildfire data</td>
<td>LSTM</td>
<td>90.9% Accuracy</td>
</tr>
<tr>
<td>Rodrigues and Riva [140]</td>
<td>Human caused wildfire occurrences</td>
<td>socioeconomic and economic activity, Fire causing possibilities</td>
<td>LR, SVM, RF</td>
<td>AUC=0.746</td>
</tr>
<tr>
<td>Tien Bui, et al. [141]</td>
<td>Spatial Pattern of forest fires</td>
<td>Weather, vegetation and infrastructure</td>
<td>MARS-DFP</td>
<td>86.5% Accuracy</td>
</tr>
<tr>
<td>Qu, et al. [144]</td>
<td>Fire occurrence forecasting</td>
<td>Weather data</td>
<td>Auto-sklearn framework</td>
<td>87% Accuracy</td>
</tr>
</tbody>
</table>
fire counts. Fire season severity (FSS) for the area divided into cells defined as the sum of fire counts in a season. They claimed that their method could predict the FSS accurately in many regions globally. Table 4 presents some applications of machine learning algorithms in forest fire forecasting using a different type of data collected from weather, environmental, socio-economics, and infrastructure details of the region [132,135,136,138-141,144].

Prediction of propagation of fire

The fire propagation rate depends on fuel characteristics (types of vegetation), terrain, and weather conditions. In the weather factors, wind speed plays a vital role in spreading forest fires to larger areas. Monitoring the wind speed and direction can be used to predict the spread of forest fires. Prediction of wildfire spread models can be grouped into four major groups [145]. In data-driven models or empirical models, correlations between data recorded from different wildfire events are obtained [146] and are statistical. In quasi-physical or semi-physical models, only physical equations are used to model the fire spread behavior, whereas physical models include physics and chemistry of the fire spread. Quasi-empirical models use the physical framework on which statistical modeling is based. A good survey on these types of models from 1990 to 2007 can be found in [146]. Guelpaa, et al. [147] proposed a model reduction technique to the physical model named Proper Orthogonal Decomposition (POD) to reduce the computational cost of the model. Sanjuan, et al. [148] suggested that high-resolution wind speed and direction are necessary for the prediction of fire propagation, whereas wind speed and direction provided by meteorological stations are of low resolution. Hence they proposed a map partitioning methodology to compute partial wind field maps to be integrated with the forest fire propagation prediction framework. Thus, a deep analysis to optimize the parameters such that wind field differences remain within a reasonable limit and execution time should also be minimized. Carrillo, et al. [149] mentioned that forest fire spread simulation requires a diverse set of parameters that may not accurately measurable and contain high uncertainty. To solve this problem, they used a two-stage methodology to calibrate the input parameters, adjustment stage (calibration is done), and the prediction stage (to improve the quality of prediction). Hence, different error functions are proposed, and their effect on the prediction of forest fire has been analyzed on an example of a fire in Greece during the summer season of 2011. Hajian, et al. [150] predicted the fire travel time distribution by using Monte Carlo simulation on the landscape that is considered as graph network, and fire propagation time is modeled as stochastic shortest path problem. The network reduction technique is also applied to reduce the computational cost of the simulation. The fire propagation rate was calculated on Montague Plains Wildlife Management Area (MPWMA) in West-Central Massachusetts and showed good accuracy. Zhou, et al. [151] integrated ensemble Kalman filter to the FARSIGHT prediction model to import the forecast of the wildfires by updating dynamically evolving fireline position. Cruz, et al. [152] provided a comprehensive summary of the Australian fire rate of spread models depending on fuel type and wind speeds. Authors have compared these models on different fuel types like grasslands, temperate shrublands, semi-arid shrublands, dry eucalypt forests, and conifer forests.

Hong, et al. [153] utilized the genetic algorithm (GA) to optimize the number of forest fire-related variables out of thirteen most common forest fire-related variables (elevation, slope angle, aspect, curvature, land use, soil cover, heat load index, normalized difference vegetation index, mean annual temperature, mean annual wind speed, mean annual rainfall, distance to river network and distance to road network) extracted from the data of forest fire events in Dayu county, China during 1980 to 2010. A fire susceptibility map was generated by RF and SVM using a full set of variables (Thirteen variables) and an optimized variables set (eight variables). RF classifier outperformed the SVM classifier for both sets of variables. AUC of cumulative forest fire susceptibility percentage was 0.816, 0.849 for original and optimized RF, respectively. Another paper by Bui, et al. [154] proposed the Particle Swarm Optimized Neural Fuzzy (PSO-NF) method to predict the spatial fire susceptibility map of forest fire in the Tropical Forest at the province of Lam Dong (Central Highland of Vietnam) in 2013. Ten fire ignition factors are used to train the classifier, and the overall classification accuracy of the proposed algorithm was found to be 89% with AUC = 0.932. Three fuzzy-metaheuristic algorithms, namely the adaptive neuro-fuzzy system (ANFIS) incorporated with GA, PSO, and differential evolution (DE), were used by Moayed, et al. [155] to find the fire susceptibility map of a fire-prone region of cities Gilakesh and Minudasht in Iran covering 1531 km2 forest area. Five classes of fire susceptibility (Very low, Low, Moderate, High, and Very high) are defined for each cell. GA-ANFIS performed the best with AUC equals 0.85.

Detection of Smoke spread after forest fires

Forest fire smoke can be one of the biggest concerns for the population living nearby by contributing to the poor air quality in a wider region. Wildfire smoke is harmful to human health and may cause death. If climate changes in the world due to human-made or natural reasons contribute to more forest fires and wildfires, smoke generated from these fires containing a mixture of pollutants may spread to wider regions polluting the fresh air resulting in significant concerns for public health. Wildfire smoke is composed of a blend of gases including acrolein, benzene, carbon mono oxide, polycyclic aromatic hydrocarbons, and microscopic particles from burned material known as the delicate particulate matter (PM) with aerodynamic diameter = 2.5 m (PM2.5), which is the most hazardous [156]. PM2.5 when inhaled results in the death of the person, global mortality caused by wildfire smoke is 339,000, including 157,000 in sub-Saharan Africa and 110,000 in Southeast Asia [157]. Landscape Fires are a global occurrence consisting of wildfires, Deforestation fires, peat fires, agriculture burning, and grass fires result in polluting the environment by release of 2 petagrams of carbon annually. The harmful nature of smoke is often neglected [158]. Forest fires smoke influences global warming and affects regional cloud formation and rainfall patterns, causing the escalation of the fire’s frequency and intensity, which enhances...
the spread of fire and elongates the fire sessions burning larger areas [157,159].

Smoke from larger fires can spread over thousands of square kilometers. Exposure to smoke is associated with the spread of various health hazards, including respiratory diseases and cardiovascular diseases [160-163]. In the last couple of decades, the availability of new satellite records helped researchers and decision-makers predict emissions accurately from forest fires. Yao, et al. [164] developed a random forest (RF) model to predict the height of the smoke layer observed by Cloud-Aerosol Lidar with Orthogonal Polarization (CALIPSO) satellite using fire activity and its geographic location and meteorological conditions. A total of twenty-two variables is identified from the CALIPSO data and NASA Modern-Era Retrospective-analysis for Research and Applications (MERRA) program data. In the final model of RF, thirteen variables are used, and the model explained 82% of data variance. RF classifier predicted the minimum height with great accuracy if the minimum height is above 100 m.

**Discussions, Future Directions and Challenges**

Forecasting forest fires is a complex phenomenon as it depends on many parameters that are difficult to predict. Fire ignition can be due to natural causes (self-ignition due to natural heat-generating process or striking of lightning etc.), or it can also be caused by human behaviors (negligence or intentional). Short term forecasting (within a year), both factors are stochastic. After fire ignition, its spread depends on various factors including weather conditions, terrain, type of fuels, etc. Prediction methods for forest fires are designed to consider multiple factors measurable by sensors in the spatial and temporal domain. These prediction models can not work as standalone applications. Still, they require forecasting in the required parameters, including weather forecasting, population forecasting, urban area development forecasting, socio-economical situation forecasting, seasonal drought predictions, etc. Hence strong integration of different modules in a greater framework is required. In the last decade, computational power and resources are increased exponentially, and new machine learning algorithms are getting places into various real-life applications. An intelligent automatic decision making and smart city concepts are developing, and in the coming decade, the concept of smart countries with automated intelligent governance may also evolve. Early warning systems for catastrophic events like big fire eruption will be in place in such frameworks. Wildfires not only pollute the climate, but the severity of wildfires may also contaminate the clean water supply to the population [165]. Wildfire forecasting can be divided into short term forecasting (in terms of days) and Long term forecasting (in terms of years).

Short term forecasting can be beneficial for allocating budget, mobilization of firefighters, moving the population to safer places, and, if possible, avoiding fire ignition. Long-term forecasting can help in country-level decision making for the next five or ten years by modifications in the factors (minimization of human-made changes in weather etc.) related to the fire events to secure the vegetation, forest, and wildlife, and healthy environment. Such studies help to understand the dependencies of fire events and spread forecasting on long term weather changes (human-made or natural), socio-economic conditions in the country, population growth, fuel demand forecasting, etc. [166,167]. Hearing the sound of the forest by putting various self-powered sensors in the forest can also provide lots of useful information in better forest care. Proactive preparedness and planning in terms of behavioral, cultural, structural and institutional play an important role in mitigating the hazardous effect of wildfires. Social media now can play its role in increasing the resilience of the people directly or indirectly exposed to wildfire threats [168]. Hence there is a strong need for community collaboration and its involvement in the decision-making process to develop effective wildfire mitigation schemes [169,170]. Adaptive governance at multiple scales with the help of aggressive data processing and involving all the stakeholders is the key to success in mitigating the effect and hazards of the wildfire [171]. In this regards, good papers the readers can read include [172-177]. Big data measurement and analysis related to fire occurrences and prediction of fire events require a more significant framework incorporating the government and communities living nearby fire-prone areas to increase wildfire resilience. More research is needed to study such types of framework constituted on social media interactions of communities and crowdsource sensing of the events. Crowdsource sensing of the events can help in the prediction of man-made fire events. Deep learning is a new field that has a promising future in wildfire/forest fires predictions as it is successfully applied in various similar events prediction [178-180]. Deep learning algorithms can be used for short-term and long-term fire occurrence events, fire spread prediction, and fire severity mapping. Already some researchers started using deep learning towards solving wildfire occurrence and spread forecasting.

There are not many standard data sets available for the application of machine learning algorithms in the wildfire. Few data sets available publicly are given in Table 5 [136,181-183]. Although lot of research has been done in forest fire and wildfire management and prediction using different machine learning algorithms, we could not find a well structured data sets available publicly. Researchers have to collect data from different sources and agencies.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>Purpose</th>
<th>Size</th>
<th>Number of Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortez and Morais [136]</td>
<td>2007</td>
<td>Predict burned Area</td>
<td>Small (517 instances)</td>
<td>13 (mostly meteorological)</td>
</tr>
<tr>
<td>Abid, et al. [181]</td>
<td>2019</td>
<td>Binary Classification (Fire, No fire)</td>
<td>Small (244 instances)</td>
<td>12 (mostly meteorological)</td>
</tr>
<tr>
<td>Sayad, et al. [182]</td>
<td>2019</td>
<td>Binary Classification (Fire, No fire)</td>
<td>Medium (804 instances)</td>
<td>3</td>
</tr>
<tr>
<td>Short and Karen [183]</td>
<td>2017</td>
<td>Fire Intensity Categories</td>
<td>Huge (1.8 million instances)</td>
<td>Many (can be selected)</td>
</tr>
</tbody>
</table>
References

8. List of wildfires.
open computer vision. 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), 287-292.


105. Killough SM, Mohammed MO, Pooran J, et al. (2019) Low-


125. Hamadeh N, Daya B, Hilal A, et al. (2015) An analytical review on the most widely used meteorological models in forest fire prediction. 2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECE), 239-244.


138. Alexander A, Bodo M, Oliver S, et al. (2014) Using multi variate data mining techniques for estimating fire susceptibility of Ty-
rolean forests. Applied Geography 53: 258-270.


Copyright: © 2021 Arif M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.36959/673/372