EasyChair Preprint

№ 6334

A Comprehensive Review on Prediction and Detection of Forest Fires Using Machine Learning Algorithms

Anshul Hooda, Antara Asthana and Dr.Vidya Patkar

August 21, 2021

A Comprehensive Review on Prediction and Detection of Forest Fires using Machine Learning Algorithms

Anshul Hooda, Antara Asthana, Vidya Patkar

Symbiosis Institute of Geoinformatics, Symbiosis International (Deemed University)

drvidyapatkar@gmail.com

ABSTRACT

Forests play a very important role in sustaining the environment. Wildfires or forest fires are not only responsible for the destruction of the natural environment, but also affects the ecological balance. A large number of fires are considered under man made causes and climate change. Although other factors like drought, wind, topography, plants, etc., have an important influence on fire appearance and its spreading. The prediction and detection of fire movement is important for fire prevention, organization of preventive measures and optimal storage of firefighting resources. An important tool for the prediction and detection of forest fires is modeling the relations between the fire threat and the influence factors and it can be done with the help of Data Mining and the Algorithms. Data mining is extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from data in large databases. It is also known as knowledge discovery in databases (KDD). The knowledge discovered from the data by applying various supervised, unsupervised and semi-supervised algorithms helps in predicting the outcomes and taking decisions. In the current study the comprehensive review of the research done by different reviewers is done and analyzed to identify the most common factors for the forest fires. The study concludes that temperature, relative humidity, precipitation, wind speed, month, heat, smoke and atmospheric gases affect the prediction of forest fires. The study also reveals that different supervised, unsupervised and semi-supervised algorithms play important role for prediction of forest fire. Keywords: forest fire, algorithms, supervised, unsupervised, semi-supervised, machine learning, data mining

Introduction

Forest fires, also known as bushfires or wildfires, became an increasing threat for the environment. There can be varied reasons for the occurrence of forest fires. It can be caused by natural calamity or due to human activity. Earlier forest fires were majorly caused periodically by natural calamities such as lightning and there was no human involvement but they have a low intensity of fires. Somehow, it benefits the ecosystem of the forests by clearing out the debris of flora and fauna and giving a path to young plants and trees to begin their life. In other words, the fires help in circulating the life cycle of forests in a healthy manner.

But in the present time, the wildfires occur due to rise in global warming and the rapid climate change. These fires can be traced all over the world, once started end up with high intensity fires making an irreversible change in the ecological balance, harming animals and plants, coercing humans to migrate and many more.

The core issue behind this is that these fires increase in an uncontrolled manner. To deal with this problem in today's world, Data Mining and its Techniques proved to be very helpful. Various Data Mining algorithms assisted in identifying and analyzing the patterns of the fires with the help of the data. These patterns were analyzed by using tools such as WEKA, MATLAB, R-programming, SQL, based on the researcher's approach. The data mining approach proceeded with the segmentation of data and evaluating the probability of future events, that is prediction of forest fires.

Methodology

To start the review, we defined the research questions to be answered in the research papers. Once research questions were formulated, the databases were selected with the similar studies. Different databases that were utilized for the review are Elsevier, Springer Link, IEEE, IOP Science, ResearchGate and Google Scholar. When the layout was formed for the review the research papers with appropriate relevance were chosen.

Research Strategy

With the help of this we were able to extract 45 research articles from the mentioned databases of different journals and which matched with our research questions.

Following are the research questions (RQs) which we defined for our study:

RQ1. Which data mining techniques or algorithms were used in the research for forest fire prediction?

RQ2. What are the different tools used for detection of forest fires?

RQ3. What are the major parameters or factors that affect the increase or decrease in forest fires?

RQ4. Is there any innovation made to improve the data collection as well as forest fire prediction?

Elimination Strategy

For our literature review we also defined some conditions or criteria to exclude research papers or articles that are not relevant. Data Mining is a broad domain with numerous techniques, it simply suggests that there are many research papers and articles published in different databases that will not be the perfect fit for our review. We started our search with basic inputs that were "data mining techniques" and "forest fires prediction". Research papers and articles were retrieved from different databases and their abstracts were read to identify the objective of the research.

The conditions made to eliminate irrelevant studies are as follows:

1. The research study is not related to forest fires prediction and detection.
2. The research study is not written in English.
3. Whether the research study is already identified and analyzed from another database.
4. The research study is not completed.

Comprehensive Review

The solution for **RQ1.** i.e., The data mining techniques or algorithms that were used in the research for forest fire prediction along with methodology and analysis in brief are:

[1] The whole data collected from organisation and surveys was used to create the courter map of weather and the fire location map. After data pre-processing, the database was divided into three categories - Weather database, Forest Fire Database and Terrain Database based on the basis of their attributes. An association algorithm **Apriori** was used in which the desired data was extracted from the database using **SQL**. Four kinds of fire number were formed i.e., often, sometime, rare and none.

By analyzing the association rules, it was found that forest fires are affected by two main factors i.e., Temperature and Moisture. They also found wind speed and rainfall as elements affecting wildfire.

[2] Kaggle is used to extract the whole dataset and it got divided into training and test set. Two classification algorithms **K-Nearest Neighbors classification** and **Multilayer Perceptron** was used to make detection models. After choosing the model with highest accuracy, data preprocessing takes place. Following which K Nearest Neighbor Classification Algorithm is used after which the Artificial Neural Network Algorithm Multi-Layer Perceptron is implied.

After generating fire detection and classification models of both the algorithms and comparing them with each other, we conclude that Multilayer Perceptron algorithm gave with high accuracy

[3] The data was extracted from different organization's databases. An **interpolation technique (Topo to Raster)** to make Digital Elevation Model (DEM). **ArcGIS** software was used to extract different layers. **ENVI** was used to process images from Remote Sensing. **Apriori** algorithm (available on Weka) is used to determine spatial rules between spatial variables and wildlife occurrence. Also, **Fuzzy c-means (FCM)** clustering is used to make the map of wildfires.

Association rules were generated and it was found that 14 rules showed strong relationship between occurrence of forest fires and the eight variables (distance from settlement, population density, distance from road, slope, standing, dead oak wood, temperature, land cover and distance from farm land).

[4] The collected meteorological data was subjected to three distinct classification algorithms. **1) Discriminant Analysis Algorithm, 2) K-Nearest Neighbor Algorithm and 3) Naïve Bayesian Algorithm.** The resulting reports of all the three algorithms from the dataset were generated using **Weka** tool.

On comparing the reported results of the three algorithms it was found that the overall error rate of K-Nearest Neighbor Algorithm is 0%. On the other hand, the error rate of other two algorithms are more than or equal to 50%. After that the whole dataset is updated in **R software** to plot the graph and get the accurate results. The results predicted that there are chances of forest fires.

[5] The spatial and hotspot data is collected from previous research. The tools used in this study were **ArcGIS** and **MATLAB**. Data preprocessing helped to relate attributes and target objects by using topological and metric operations. The **Adaptive Neuro Fuzzy Interference System (ANFIS)** algorithm is used to generate rules in building the model.

The research concludes that this algorithm can be suggested for prediction of hotspots since it gave low error for training as well as testing results. It helped to recognize most influential factor for forest fires by the probability of true and false alarm of the attribute NEAR_ROAD.

[6] The data of hotspots is taken from the database of NASA. For clustering, there were three attributes: longitude, latitude, and date of hotspot occurrence in the dataset. The density-based cluster algorithm, **Ordering Points to Identify the Clustering Structure (OPTICS)** is used which is available in **Weka** to generate clusters and to identify outliers on the dataset of hotspot.

The results of **Density-Based Spatial Clustering Algorithm with Noise (DBSCAN)** algorithm taken from another research were compared with the results of OPTICS. It was found that percentage of outliers was less when OPTICS was applied. The focus was to find number of objects in cluster in OPTICS instead of number of clusters in DBSCAN.

[7] Data preprocessing gave the number of daily and monthly hotspot frequency. **K-Means** algorithm, based on clustering method was implied on the dataset of hotspots to detect outliers. The R statistical tool has a k-means function that is used to create clusters. The detected outliers were categorized as global and collective by calculating the Outlier Score.

In the time period of 12 years, it was observed that global and collective outliers are found in February, March, June, July, and August. Lowest and Highest frequency of daily and monthly hotspots was also analyzed that occurs in specific clusters.

[8] **K-means** algorithm is applied on the dataset to cluster all the hotspots. **Shiny framework** and **R** language is also used with the algorithm to detect outliers and make a web-based application on it.

All the major functions can be easily used to visualize the clustering results, detecting global and collective outliers, and visualizing them on the Riau province map.

[9] The research was divided into three parts: area, tools, and stages of research. The source of hotspot data is FIRM MODIS. **R application, SpatialEpi, Quantum GIS** are the tools used in this research. After preprocessing of the extracted data **Kulldorff Scan Statistic Method** is used with **Poisson models** to decide the likelihood ration function. Cluster Validation was done with the **Monte Carlo** approach and cluster analysis is based on the physical parameters.

Through the study of hotspot distribution in the selected area, it was found that there is an increase in the hotspot occurrence in that time interval. The increase or decrease of hotspots also depends upon maturity and thickness of peat as well as the land cover.

[10] The research was conducted in two steps: 1) the secure transmission of the sensitive collected data (images of forest) from sensor nodes to base station. 2) analyzing the data of forest images with the help of **Direct Visualization Cluster Model** algorithm to predict the occurrence of fire.

The fire severity is predicted with the help of **Spatial Association rule mining**, applied to the pixel values in the images of four different clusters.

[11] **Agglomerative Hierarchical** Clustering, which is a "bottom-up" approach is an Image mining technique implied on the satellite images to plot the clusters and **RGB** values are extracted from them to predict the forest fires.

The list of hotspots obtained based on clusters and their image processing, the prediction of forest fires and its spreading direction can be analyzed with the help of wind direction.

[12] The whole method is divided into: a) Image Segmentation, b) Binarization, c) Noise Processing. All three were used to extract physical features of the flames. After extraction, **Back Propagation (BP)** neural network algorithm is implied on data integration to get the description of the images.

The algorithm used to identify fire has high sensitivity and low false alarm rate because of its good ability of adapting. It has been discovered that it is capable of recognizing the complex circumstance of non-fire and that its anti-interference capability is extremely strong.

[13] In this research, **Image processing** is done with the help of **YCbCr** colour model which is used to classify flame pixels based on statistical information from the fire image, such as mean and standard deviation since the relation between pixels is compared to other colour models. Three distinct set of images were considered to do the research, 1) contain fire, 2) contain fire like regions & 3) contain fire center like regions. Four rules are imposed for segmentation. The fire flame region is segmented using two rules, and the high temperature fire centre zone is segmented using two rules.

It is analyzed that the method applied, with the use of statistical features of fire image in YCbCr colour space such as mean and standard deviation, it not only separates fire flame pixels but also separates elevated temperature fire centre pixels. This study found that there was a higher rate of fire detection and a lower percentage of false alarms.

[14] In this research, an Unmanned Aircraft System, integrated with different sensors and devices is used and its work is divided into three steps: fire search, fire confirmation, and fire observation. The image processing algorithms used for tracking and detecting fire are image collection, image preprocessing (including noise reduction, color model conversion), **threshold segmentation**, **morphological operations,** and **blob counter** to detect and track the sequence of forest fires.

The results of this research showed better results with the help of Lab color model in detection of forest fire images on comparing to another previous research.

[15] This study was supported by Hunan Provincial Natura. In this study, researchers used **YCbCr** colour space which made it possible to separate chrominance from luminance/illumination. Segmentation problem is illustrated, the problem adopted the **CIE L*a*b*, YCbCr** colour space, and **K-means clustering** algorithm to isolate the fire. The colour YCbCr is defined by the concepts of brightness and chromatic aberration. The CIE created it as a uniform colour space approximation with three primary locations, such as L*a*b*. When employing heuristics like Lloyd's algorithm, k-means clustering is simple to design and deploy even on enormous data sets.

This research analyzes and deals with real-time data. the linear conversion between RGB and YCbCr color spaces. YCbCr colour space to model fire pixels is used. The segmentation algorithm has shown that the flame pixel identification method is faster and has high recognition rate. But the results are affected significantly by the threshold and in some cases, it is easier to produce many grains.

[16] Images from three different digital videos (forest fire, flying red flag and moving red car) are used as data to conduct the experimental research. The area of fire is taken from image by taking the help of threshold values in HSV (hue, saturation, value) color space since it gives more familiar results of describing colours comparing to RGB color model. Once segmentation is done, their physical features were extracted for fire regions. Dynamic characteristics are defined, i.e., average and mean square deviation of area, roundness, and contour. Now these characteristics were taken as input for the **Back Propagation** Neural Network algorithm to get the results.

Based on experimental research, it is analyzed that forest fires can be predicted with the help of limits of training and testing samples to a certain output value.

[17] The data of images is extracted from 6 videos taken from internet where 4 videos were of actual fire, and two were of fire-coloured objects. 1) receiving the input video from the input device, 2) using background subtraction algorithm **Movement Containing Region Detection Based (MRDB)** to detect movement contained regions implemented using **MATLAB,** 3)

converting the input image sequence from **RGB** to **YCbCr** color space and 4) putting the fire detection principles into action, as well as temporal variation.

The algorithm yields good detection rates, according to this study. The F-score is used to analyze the outcomes. Furthermore, the results of this technology are precise and can be employed in automatic forest fire alert systems.

[18] The research was conducted using **MATLAB**. The input data is extracted from spatial data of the selected forest region. Image Processing and Artificial Intelligence techniques are used for the detection of fire from the images of spatial data. A fuzzy rule base is formed from the spatial data to detect the fires. To detect fire zones, the digital images from the spatial data are transformed to **YCbCr** colour space and then segmented using **anisotropic diffusion**. The colour space values of the fire regions are then used to generate a fuzzy set. Fuzzy rules are also derived using **fuzzy logic reasoning**.

The result shows that this model can be used to detect fire in a particular forest area using the algorithms and y following the procedure.

[19] In this research a wide range of IoT sensors are deployed in the forest terrain to get large volumes of data about Wildfire Consequent Attributes (WCAs - Temperature, Relative Humidity, Precipitation, Wind Speed, Month) and Wildfire Influent Attributes (WIAs - Heat, Smoke, Atmospheric gases). **Fuzzy K-nearest-neighbor (FK-NN)** classifier algorithm method is used to identify wildfires and estimate the present level of wildfire proneness in a forest landscape. This algorithm is used in the Fog layer. In cloud layer, 1) temporal mining-based monitoring, 2) emergency alert generation, 3) wildfire proneness forecasting component and 4) cloud storage are the four components.

The suggested technique was tested using a dataset from the Forest Divisions of Hoshiarpur and Pathankot in Punjab, India. The accuracy, specificity, sensitivity, and precision of the woodland landscape are found to be high using this approach.

[20] In this research, a prediction algorithm, the risk of forest fires is calculated using a **Fuzzy Inference Technique** that is implemented using rechargeable wireless sensor networks. According to the idea for studying the causes of forest fires, the fuzzy system is used to analyze

the causality from uncertain components. The parameters taken for the implementation of this algorithm are 1) Temperature and 2) Humidity.

The study's analysis relies on a large number of sensor nodes and a monitoring system. It was shown that around 80% of forest fires happened in the spring or autumn seasons. Fire dangers on a large scale can be addressed as a strategic planning tool. This technique is built on fuzzy variables that can be changed to different landmarks of the variables to prevent forest fires.

[21] The Decision tree classifier and a smart sensor node was found suitable to predict forest fires based on characteristics such as temperature, wind speed, and humidity for fully automated detection. The data was collected from meteorological department. Using a binary decision tree, and three qualities from two regions of Algeria, a forecast of 'fire' or 'no fire' is created using IF ELSE.

The decision tree was found appropriate because it provides significant performance and can be converted to a rule-based system, making its hardware implementation very straightforward using fewer resources.

[22] To understand how common vegetation ignites so that detection of the fire risk of a powerline vegetation failure and issue the appropriate warning in real time the use of a Hybrid Step XGBoost (HSXG), a new time series model for recognizing the warning pattern of distinct vegetation species based on their current data in the ignition process. The k-nearest neighbor (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Random Forest (RF) data mining techniques were employed to test the method's practicality.

After employing these algorithms, it was found that the suggested HSXG model was able to determine the safe and dangerous periods and issue real-time bushfire warnings. It could be integrated into other IoT (Internet of Things) systems to help prevent bushfires caused by vegetation problems on powerlines.

[23] For predicting the occurrence of forest fires: Neural Networks and Decision Trees Both were simple to use and inexpensive, especially in nations like Lebanon.

These approaches that can identify high fire risks, allowing for fire prevention or at the very least limiting the consequences of fires.

[24] The Five alternative DM approaches, such as Support Vector Machines (SVM) and Random Forests, as well as four different feature selection settings (using spatial, temporal, FWI components, and weather variables) were used. The meteorological data was obtained in Portugal's northeast region. The local sensors in weather stations detect forest fires.

As the DM approaches were applied after the data was obtained, this was based on off-line learning. However, this approach paves the way for the creation of automated firefighting equipment.

[25] For forest fire monitoring and detection, a novel technique based on wireless sensor networks was presented. Based on monitoring and combining real data from various sensors (temperature, humidity, light, and smoke), as well as employing the Nave Bayes (NB) classifier to detect fires.

Applying Data Mining techniques minimized the amount of data, eliminates redundancy, increases WSN performance, and lowers network trac to extend the network's life-time and provide quick decision-making and early fire detection.

[26] To employ Image Processing and Artificial Intelligence techniques to extract useful information from photos in forest spatial data and use it to determine places at danger of fire.

The images in the spatial data generated from remote sensing were used by the presented method for forest fire detection.

[27] Two data-mining models – logistic regression (LR) and an artificial neural network (ANN) were applied to the data gathered from the database to forecast fire occurrence in Brazil's Federal District, which is located within the Brazilian Carrado. The data was divided into training and testing data and after defining an optimal decision threshold, the models were trained for the best area under the receiver operating characteristic curve (AUROC, or simply AUC), and then compared and verified based on their accuracy.

It was found that Both models performed well, with identical accuracy and AUC values. When examining purely non-burned areas, the ANN model had a higher AUC value and greater accuracy.

[28] Clustering was chosen as the data mining method because it can convert hotspot data into information that can be utilized to inform areas prone to hotspots. The k-means technique is used in this clustering, which groups data based on comparable properties.

The k-means algorithm clustering approach can be used to categorize data into groups, such as hotspot-prone areas and non-hotspot-prone areas.

[29] To create a predictive model based on climatic variables the model was built for predicting the area burned during forest fires using Linear Regression, Random Forest, Bagging, and Logistic Regression. This type should be used in any areas where there is a higher chance of catching fire.

The accuracy of the random forest and boosting algorithm can be increased by modifying the parameters and adding some other factors like forest vegetation, forest cover, type of trees in the forest, and Buildup Index.

[30] Python (3.7) was used to write the code and using neural network technology, a prototype system for detecting spots of fire in Quadro copter images. It was determined to design a software solution to handle the problem of forest fire detection. The software's goal is to locate forest fires using data captured by Quadro copters in the infrared spectrum. During training, the neural network had a 97 percent accuracy rate and a loss function of 0.092.

It was found that as the training was conducted on a PC with a NVidia GeForce GTX 1050 graphics card, and satisfactory results were obtained when the system was tested as a whole.

[31] To detect forest fires, an artificial neural network was utilized, and the generated system automatically detects the fire and displays the value as ‘1'. It will show as ‘0' if there is no forest fire. Drone Remote Sensing gives information on forest fires and can be used to report the occurrence to the appropriate authorities when the forest fire breaks out. The authors discovered that, in addition to the neural network technique, remote sensing drone data can be combined to determine the specific location of a forest fire.

Small substance with planned expenses, easy monitoring of spatial and sequential decision, high-intensity data gathering, and lack of hazards to crews are all advantages of drone inaccessible sensing.

[32] Evaluation of wildfire likelihood across two geographical regions comprising most of Italy's territory: Alpine and subalpine region and Insular and peninsular region, using an Artificial Neural Network model and fire data (2004–2012). Using a comprehensive set of biophysical and human factors, we used an ANN model to assess the chance of wildfire occurrence in Italy.

For understanding the probability of wildfire occurrence, the usage of ANNs is efficient and statistically sound.

[33] New approach for detecting early forest fires that incorporates Data Mining techniques into sensor nodes. The focus on measuring and combining real data from several sensors, as well as employing an Artificial Neural Network classifier to identify fires.

Applying DM techniques decreases data size, eliminates redundancy, increases WSN speed, and extends the network's life-time to ensure fire detection as soon as possible.

[34] Fog-IoT-centric system for detecting wildfires in real time was employed. A cost-effective real-time solution for reducing the damage caused by wildfires. K-means Clustering is initially utilized to detect the wildfire outbreak at the fog layer, followed by the generation of real-time alerts to the administration and community. A cloud layer-based Adaptive Neuro Fuzzy Inference System is also utilized to analyze a forest block's sensitivity to forest fires and classify it into one of five danger zones based on the Forest Fire Vulnerability Index.

The suggested framework's implementation results demonstrate its efficacy in detecting and predicting wildfires, and real-time warning production further boosts the system's efficacy.

[35] On the MNP dataset, several Soft Computing techniques such as the cascade correlation network (CCN), multilayer perceptron neural network (MPNN), polynomial neural network (PNN), radial basis function (RBF), and support vector machine (SVM) were used to find the most effective predictor that could give more accurate results for forest fires. When compared to alternative predictors with low estimation error, the SVM algorithm gives more precise predictions.

In comparison to previous methods, the collected findings show that the SVM enhances prediction accuracy and is suitable for forest fire prediction.

[36] To create a forest fire risk map using an integrated strategy that included ANP and fuzzy logic. Using the frequency ratio approach, fuzzy membership functions were calculated for each of the criteria used in the development of a forest fire risk map, and a fuzzy map was created for each of them. Following that, in a GIS context, fuzzy maps were applied to the ANP with the derived coefficients to produce the forest fire risk map.

It was analyzed that when paired with fuzzy logic and the ANP model, remote sensing data with its spatial information allows fire managers and people to anticipate "when and where" forest fires will most likely occur.

[37] For fire detection, the method employs spatial data mining, image processing, and artificial intelligence approaches. Forest spatial data with and without fires were given into the proposed system as input.

With the help of the developed fuzzy rules, the proposed system was able to detect fires in the data successfully.

[38] Horus, a new and efficient fire detection system, was proposed in this paper. It comprises of background reduction, the Fuzzy Fire Search, and the Motion Dynamic Test.

The Bo-Ho Cho algorithm is excessively sensitive to fire-colored objects, resulting in false positives, whereas the Celik approach is too conservative, resulting in a high proportion of false negatives.

[39] Prediction of natural disasters and disaster management based on fuzzy logic and hybrid PSO is described in this study to forecast forthcoming natural disasters. Prediction approaches include preprocessing spatial data, cluster analysis using hybrid PSO, and event prediction utilizing fuzzy logic methods.

It was found that strategy is effective, and correct results are obtained based on data from natural events and information available on the knowledge base.

[40] The Intelligent Soft Computing Multivariable Analysis system (ISOCOMA) is proposed in this paper for determining appropriate wildfire risk indices. Collecting twelve original fire risk

indices into four classes of partial risk indices and then combining them into a single overall one, the Conceptual Risk Index (CRI), for each fire incidence and study region. The Actually Burned Surfaces Index (ACBUS) was then calculated based on the burned area.

As a consequence of a comparison analysis between the produced indices and the ACBUS, this modelling research endeavor has yielded high rates of accurate classifications.

[41] MATLAB was used to eliminate false fire alarms in fire alarm systems, which divert emergency personnel away from actual crises that could result in loss of life and property. A fuzzy logic methodology was used to conduct the investigation (FLA). The approach contains 125 rules because there are three variables: fire, smoke, and gas, as well as five language variables.

The fuzzy logic approach reduces the risk of a false alert, which could generate unneeded panic and/or bring resources (such as firefighters) to a location where they are not needed.

[42] A genetic algorithm to predict the progress of a forest fire was based on four parameters: dryness factor, temperature, relative humidity, and wind, a genetic algorithm and geometric semantic genetic algorithm-based fire spread prediction system is built.

The resulting result depicts the ranges of values for input parameters that contribute to fire start and spread. The system can forecast the spread category of the impacted area with good accuracy using the ranges.

[43] Using a fused dataset, the research seeks to understand the reliability of three forecasting approaches. The accuracy of three alternative data mining algorithms, KNN, SVM, and decision tree models, is tested on a small database in this paper. The goal of this study is to develop a model that accurately predicts in a short dataset using both binary and multi-class classifiers.

On the basis of supervised learning of structured data, the model's validity will be validated.

[44] The forest fire region is identified using geographic data in three phases: preprocessing, training, and detection. Forest fire detection utilizing the ANFIS system has been implemented in the working environment MATLAB (version 7.8) in this proposed study, which makes use of publicly available geographical data.

The fire zones are recognized and noted in the image tested in the created ANFIS system.

[45] Forest fire prevention using Spatio-temporal Data Mining and Knowledge Discovery (STDMKD). Spatiotemporal data mining techniques and tasks include spatiotemporal forecasting and trend analysis, spatiotemporal association rule mining, spatiotemporal sequential pattern mining, spatiotemporal clustering and classification, and more.

The high prediction accuracy of ISTIFF (enhanced spatio-temporal forecasting framework) is demonstrated by a comparison of ISTIFF with ARIMA and STIFF (a spatio-temporal integrated framework).

In all these research distinct types of algorithms were used that can be classified into:

1. Supervised Learning: When the machine is supervised while it is "learning", the training type is called supervised learning. It means that we provide the machine with the information about a case and provide it with the case outcome. The outcome is called the labelled data while the rest of the information is used as input features. Supervised learning is divided into two techniques, classification, and regression.

The Supervised Learning Algorithms used in different studies are:

a. K-Nearest Neighbor
b. Multilayer Perceptron
c. Naïve Bayesian
d. Discriminant Analysis
e. Adaptive Neuro Fuzzy Inference System
f. Ordering Points to Identify the Clustering Structure
g. Back Propagation
h. Support Vector Machine
i. Artificial Neural Networks
j. Binary Decision Tree
k. Random Forest
l. Logistic Regression
m. Linear Regression
n. Cascade Correlation Network

2. Unsupervised Learning: In case of unsupervised learning, there is no help from the user for the computer to learn. In the lack of labelled training sets, the machine identifies patterns in the data that is not so obvious to the normal human brain. So, unsupervised learning is extremely useful to recognize patterns in data and help us make decisions. Unsupervised learning is divided into two techniques, Clustering and Association rule mining.

The Unsupervised Learning Algorithms used in the above-mentioned studies are:

a. Apriori
b. K-Means
c. Direct Visualization Cluster Model
d. Agglomerative Hierarchical Clustering
e. Spatio-temporal Association Rule Mining

3. Semi-Supervised Learning: A combination of labeled and unlabeled data is used for the purpose of machine learning. A combination of all four techniques of supervised and unsupervised learning are used in semi-supervised learning.

The Semi-Supervised Learning Algorithms used in the above-mentioned studies are:

a. Fuzzy c-means
b. Fuzzy Logic
c. Geometric Semantic Genetic

The solution for **RQ2.** i.e., Researchers used very distinctive approaches for the detection of forest fires using different tools. Distinct types of satellite sensors were used to track forest fires on time and the data was collected for the research. In a study an unmanned aerial vehicle (UAV) was used with latest technology and sensors to detect forest fires in less period. IoT framework was applied with the sensors and other required tools and equipment were installed in the forest sending the real time observation to the base monitoring station in less time.

The solution for **RQ3.** i.e., The major parameters or factors that affect the increase or decrease in forest fires are:

1. Temperature
2. Relative Humidity
3. Precipitation
4. Wind Speed
5. Month
6. Heat
7. Smoke
8. Atmospheric Gases

The solution for **RQ4.** i.e., Any innovation made to improve the data collection as well as forest fire prediction can be seen in some studies like a mobile application was made using Shiny framework and R where the data can be collected simply through the mobile app of the area. This can help researchers to directly get the data access without any delay.

Conclusion

The purpose of this comprehensive review of 45 research papers and articles was to know how much work has been done till now and on what grounds we can proceed further. The key role of an algorithm for any research under any random domain is that the algorithm must be of such kind where it is easy to interpret the output along with consuming less time for the calculation and giving the highest accuracy. Majority of the studies mentioned above aimed at defining the proneness level of wildfires to the forest with the help of Machine Learning. The research conducted with combination of diverse algorithms provided higher accuracy with minimum research gaps in them. As per our study, we conclude that combination of Machine Learning algorithms in further studies will improve the accuracy of results as well as covering more research gaps.

References

[1] Hu, L., Zhuo, G., & Qiu, Y. (2009, August). Application of apriori algorithm to the data mining of the wildfire. In *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery* (Vol. 2, pp. 426-429). IEEE.

[2] Kumar, N., & Kumar, A. (2020, July). Australian Bushfire Detection Using Machine Learning and Neural Networks. In *2020 7th International Conference on Smart Structures and Systems (ICSSS)* (pp. 1-7). IEEE.

[3] Jafarzadeh, A. A., Mahdavi, A., & Jafarzadeh, H. (2017). Evaluation of forest fire risk using the Apriori algorithm and fuzzy c-means clustering. *Journal of forest Science*, *63*(8), 370-380.

[4] Kamalakannan, J., Chakrabortty, A., Bothra, G., Pare, P., & Kumar, C. P. (2019). Forest fire prediction to prevent environmental hazards using data mining approach. In *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology* (pp. 615-622). Springer, Singapore.

[5] Wijayanto, A. K., Sani, O., Kartika, N. D., & Herdiyeni, Y. (2017). Classification model for forest fire hotspot occurrences prediction using ANFIS algorithm. In *IOP Conference Series: Earth and Environmental Science* (Vol. 54, No. 1, p. 012059). IOP Publishing.

[6] Febriana, N. L., & Sitanggang, I. S. (2017, March). Outlier detection on hotspot data in riau province using OPTICS algorithm. In *IOP Conference Series: Earth and Environmental Science* (Vol. 58, No. 1, p. 012004). IOP Publishing.

[7] Sitanggang, I. S., & Baehaki, D. A. M. (2015, July). Global and collective outliers detection on hotspot data as forest fires indicator in Riau Province, Indonesia. In *2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)* (pp. 66-70). IEEE.

[8] Suci, A. M. Y. A., & Sitanggang, I. S. (2016). Web-based application for outliers detection on hotspot data using k-means algorithm and shiny framework. In *IOP Conference Series: Earth and Environmental Science* (Vol. 31, No. 1, p. 012003). IOP Publishing.

[9] Hudjimartsu, S. A., Djatna, T., & Ambarwari, A. (2017). Spatial temporal clustering for hotspot using kulldorff scan statistic method (KSS): A case in Riau Province. In *IOP Conference Series: Earth and Environmental Science* (Vol. 54, No. 1, p. 012056). IOP Publishing.

[10] Divya, T. L., Manjuprasad, B., Vijayalakshmi, M. N., & Dharani, A. (2014, April). An efficient and optimal clustering algorithm for real-time forest fire prediction with. In *2014 International conference on communication and signal processing* (pp. 312-316). IEEE.

[11] Divya, T. L., & Vijayalakshmi, M. N. (2015, March). Analysis of wild fire behaviour in wild conservation area using image data mining. In *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-3). IEEE.

[12] Xiao, J., Li, J., & Zhang, J. (2009, October). The identification of forest fire based on digital image processing. In *2009 2nd International Congress on Image and Signal Processing* (pp. 1-5). IEEE.

[13] Premal, C. E., & Vinsley, S. S. (2014, March). Image processing based forest fire detection using YCbCr colour model. In *2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014]* (pp. 1229-1237). IEEE.

[14] Yuan, C., Liu, Z., & Zhang, Y. (2015, June). UAV-based forest fire detection and tracking using image processing techniques. In *2015 International Conference on Unmanned Aircraft Systems (ICUAS)* (pp. 639-643). IEEE.

[15] Chen, R., Luo, Y., & Alsharif, M. R. (2013). Forest Fire Detection Algorithm Based on Digital Image. *J. Softw.*, *8*(8), 1897-1905.

[16] Zhang, D., Han, S., Zhao, J., Zhang, Z., Qu, C., Ke, Y., & Chen, X. (2009, April). Image based forest fire detection using dynamic characteristics with artificial neural networks. In *2009 International Joint Conference on Artificial Intelligence* (pp. 290-293). IEEE.

[17] Mahmoud, M. A., & Ren, H. (2018). Forest fire detection using a rule-based image processing algorithm and temporal variation. *Mathematical Problems in Engineering*, *2018*.

[18] Angayarkkani, K., & Radhakrishnan, N. (2009). Efficient forest fire detection system: a spatial data mining and image processing based approach. *International Journal of Computer Science and Network Security*, *9*(3), 100-107.

[19] Kaur, H., & Sood, S. K. (2020). Soft-computing-centric framework for wildfire monitoring, prediction and forecasting. *Soft Computing*, *24*(13), 9651-9661.

[20] Lin, H., Liu, X., Wang, X., & Liu, Y. (2018). A fuzzy inference and big data analysis algorithm for the prediction of forest fire based on rechargeable wireless sensor networks. *Sustainable Computing: Informatics and Systems*, *18*, 101-111.

[21] Abid, F., & Izeboudjen, N. (2019, July). Predicting Forest Fire in Algeria Using Data Mining Techniques: Case Study of the Decision Tree Algorithm. In *International Conference on Advanced Intelligent Systems for Sustainable Development* (pp. 363-370). Springer, Cham.

[22] Ma, J., Cheng, J. C., Jiang, F., Gan, V. J., Wang, M., & Zhai, C. (2020). Real-time detection of wildfire risk caused by powerline vegetation faults using advanced machine learning techniques. *Advanced Engineering Informatics*, *44*, 101070.

[23] Karouni, A., Daya, B., & Chauvet, P. (2014). Applying decision tree algorithm and neural networks to predict forest fires in Lebanon. *Journal of Theoretical and Applied Information Technology*, *63*, 282-291.

[24] Cortez, P., & Morais, A. D. J. R. (2007). A data mining approach to predict forest fires using meteorological data.

[25] Saoudi, M., Bounceur, A., Euler, R., & Kechadi, T. (2016, March). Data mining techniques applied to wireless sensor networks for early forest fire detection. In *Proceedings of the International Conference on Internet of things and Cloud Computing* (pp. 1-7).

[26] Angayarkkani, K., & Radhakrishnan, N. (2010). An intelligent system for effective forest fire detection using spatial data. *arXiv preprint arXiv:1002.2199*.

[27] de Bem, P. P., de Carvalho Júnior, O. A., Matricardi, E. A. T., Guimarães, R. F., & Gomes, R. A. T. (2018). Predicting wildfire vulnerability using logistic regression and artificial neural networks: A case study in Brazil's Federal District. *International journal of wildland fire*, *28*(1), 35-45.

[28] Khairani, N. A., & Sutoyo, E. (2020). Application of k-means clustering algorithm for determination of fire-prone areas utilizing hotspots in West Kalimantan Province. *Int. J. Adv. Data Inf. Syst*, *1*(1), 9-16.

[29] Poluru Eswaraiah, R. L. Prediction of Fire Area in Forests with Regression Techniques.

[30] Golodov, V., Buraya, A., & Bessonov, V. (2019, October). Detection of Forest Fires Based on Aerial Survey Data Using Neural Network Technologies. In *2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)* (pp. 1-6). IEEE.

[31] Rao, G. N., Rao, P. J., & Duvvuru, R. (2016, September). A drone remote sensing for virtual reality simulation system for forest fires: semantic neural network approach. In *IOP Conference Series: Materials Science and Engineering* (Vol. 149, No. 1, p. 012011). IOP Publishing.

[32] Elia, M., D'Este, M., Ascoli, D., Giannico, V., Spano, G., Ganga, A., ... & Sanesi, G. (2020). Estimating the probability of wildfire occurrence in Mediterranean landscapes using Artificial Neural Networks. *Environmental Impact Assessment Review*, *85*, 106474.

[33] Saoudi, M., Bounceur, A., Euler, R., Kechadi, T., & Cuzzocrea, A. (2016, July). Energy-efficient data mining techniques for emergency detection in wireless sensor networks. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)* (pp. 766-771). IEEE.

[34] Kaur, H., & Sood, S. K. (2019). Adaptive neuro fuzzy inference system (ANFIS) based wildfire risk assessment. *Journal of Experimental & Theoretical Artificial Intelligence*, *31*(4), 599-619.

[35] Al_Janabi, S., Al_Shourbaji, I., & Salman, M. A. (2018). Assessing the suitability of soft computing approaches for forest fires prediction. *Applied computing and informatics*, *14*(2), 214-224.

[36] Abedi Gheshlaghi, H., Feizizadeh, B., & Blaschke, T. (2020). GIS-based forest fire risk mapping using the analytical network process and fuzzy logic. *Journal of Environmental Planning and Management*, *63*(3), 481-499.

[37] Angayarkkani, K., & Radhakrishnan, N. (2009). Efficient forest fire detection system: a spatial data mining and image processing based approach. *International Journal of Computer Science and Network Security*, *9*(3), 100-107.

[38] A. Buemi, D. Giacalone, F. Naccari and G. Spampinato, "Efficient fire detection using fuzzy logic," *2016 IEEE 6th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, 2016, pp. 237-240, doi: 10.1109/ICCE-Berlin.2016.7684764.

[39] Ravikumar, K., & Kannan, A. R. (2018). Spatial Data Mining for Prediction of natural events and disaster management based on fuzzy logic using hybrid PSO. *TAGA J*, *14*, 858-878.

[40] Anezakis, V. D., Demertzis, K., Iliadis, L., & Spartalis, S. (2016, September). A hybrid soft computing approach producing robust forest fire risk indices. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 191-203). Springer, Cham.

[41] Syafitri, N., Labellapansa, A., Kadir, E. A., Saian, R., Zahari, N. N. A., Anwar, N. H. K., & Shaharuddin, N. E. M. (2019). Early detection of fire hazard using fuzzy logic approach. International Journal of Advanced Computer Research, 9(43), 252-259.

[42] Amol, D. R. (2020). Prediction of Fire Propagation in Forest Areas using Genetic Algorithm. In *ITM Web of Conferences* (Vol. 32, p. 03046). EDP Sciences.

[43] Naganathan, H., Seshasayee, S. P., Kim, J., Chong, W. K., & Chou, J. S. (2016). Wildfire predictions: Determining reliable models using fused dataset. *Global Journal of Computer Science and Technology*.

[44] Angayarkkani, K., & Radhakrishnan, N. (2011, April). An effective technique to detect forest fire region through ANFIS with spatial data. In *2011 3rd International Conference on Electronics Computer Technology* (Vol. 3, pp. 24-30). IEEE.

[45] Wang, T. C. J. (2006). Applications of spatio-temporal data mining and knowledge discovery (STDMKD) for forest fire prevention.