Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

# Wildfire Predictions: Determining Reliable Models using Fused Dataset

Hariharan Naganathan[1], Sudarshan P Seshasayee[2], Jonghoon Kim[3], Jui-Sheng Chou[4] and Wai K Chong[5]

[1] Arizona State University

*Received: 14 December 2015 Accepted: 31 December 2015 Published: 15 January 2016*

---

## Abstract

Wildfires are a major environmental hazard that causes fatalities greater than structural fire and other disasters. Computerized models have increased the possibilities of predictions that enhanced the firefighting capabilities in U.S. While predictive models are faster and accurate, it is still important to identify the right model for the data type analyzed. The paper aims at understanding the reliability of three predictive methods using fused dataset. Performances of these methods (Support Vector Machine, K-Nearest Neighbors, and decision tree models) are evaluated using binary and multiclass classifications that predict wildfire occurrence and its severity. Data extracted from meteorological database, and U.S fire database are utilized to understand the accuracy of these models that enhances the discussion on using right model for dataset based on their size. The findings of the paper include SVM as the best optimum models for binary and multiclass classifications on the selected fused dataset.

---

***Index terms***— support vector machines, k-nearest neighbor, k-fold cross-validation, decision tree stumps, forest fire, binary and multiclass classifiers.

# 1 I. Introduction

ildfires are a major environmental hazard and a real world problem that affects human, wildlife and create damages to the economy. According to United States Department of Agriculture (USDA), fatalities caused by the wildfires are greater than structural fire and other disasters. Over 90% of the wildfires were caused by humans while others by a volcanic eruption and lightning. Data mining techniques have increased the possibilities of predicting forest fires that enhanced the firefighting capabilities in U.S. The National Interagency Fire Center (NIFC) provides daily information on wildfire events using various intelligence and predictive methods.

# 2 W

research to develop a generic model for predicting forest fires. Hierarchical information is a significant tool that connects factors and helps understand the start and growth of a forest fire. Such information helps fire managers make critical short and long-term decisions before the beginning of and during a wildfire. In addition to prediction, firefighting and fire restoration are also a part of wildfire mitigation. According to Burned Area Emergency Response (BAER), proper restoration and adaptation procedures after forest fires are a necessary and handy system to have. The active fire mapping program by the National Interagency Fire Center (NIFC) includes the location, severity, the type, burnt area, and the contaminant status of the wildfire region. It also specifies the causes of the fire that helps fire managers make a decision. The Wildfire Assessment System (WFAS) is a mapping tool that provides information on fuel and fire hazards. Also, the Federal government has a comprehensive fire prevention and prediction system that predicts, forecasts and contains information on forest fires through a national database on wildfires.

Predictive models integrating meteorological data from different weather stations (local sensors) and fire database still need improvement since it can possess lower predictive accuracy for larger fires. The accuracy also depends on the size of the database and its features. The motivation of this paper is to enhance the predictability of forest fires using predictive analytics to manage it effectively. The primary focus of this article is to develop prediction forecast models from spatial data, identify the areas prone to wildfires from previous meteorological and fire data using both binary and multi-class classifiers. While this is not a new approach, the applications have yet been fully tested to predict forest fire.

# 3 III. Research Objective

The objective is to understand the reliability of three techniques (that uses a dimensionality-reduced dataset) in predicting forest fires using USDA data. These techniques have been proven to provide insights for decision makers and computer scientists. The paper proposes a comparative study of the three techniques to analyze and predict forest fires using data from California, Idaho, Oregon, Nevada, and New Mexico. These states were selected due to the severity and frequency of occurrences between 2004 and 2014. The authors used three different predictive techniques in this paper to identify which one has greater accuracy with small-scale data.

Also, the data collection process involves feature extraction, and dimensionality reduction, to make the dataset more comprehensive. The paper is organized into sections that include objectives, a review of various fire predictions using support vector machine (SVM), K-nearest neighbors (KNN) and decision tree, addressing the gaps, research methodology, and discovery.

# 4 IV. Relevant Work

The section details on various models developed from previous studies, data mining techniques used in the models and finally addressing the gaps.

Climatic change is portrayed to be one of the reasons for wildfires at tropical regions (Over peck, Rind, & Goldberg, 1990). It is still a debate because fire is a set of complex set of interactions. According to National Oceanic and Atmospheric Administration (NOAA), 32 groups of scientists from around the world investigate 28 individual extreme events in 2014 and broke out various factors that led to the extreme events, including the degree to which natural variability and human-induced climate change played a role. The report added that the overall probability of California wildfires has increased (2,500 acres) due to humaninduced climate change (EPA, 2014). Hence, fires not only impact carbon sequestration by forests but emit greenhouse gasses and releases most carbon as CO2, which potentially affect the climate. It has some potential positive feedback since greenhouse-gas-driven climate warming may increase fire activity.

Machine-learning models were frequently used to predict forest fires in different countries and states (Alonso-Betanzos et al Service & Mountain, 2002). Most of them relied on general models for both large and small database predictions.

After a preliminary review of related work on predictive systems used (on forest fire), regression models such as SVM with other metrics are found to be the most frequently used models (Cortez & Morais, 2007). Similarly, Cortez and Morais (2007) subsequently used a k-fold cross validation on the model with Root Main Square Error (RMSE). The neural network is an alternative model utilized on large data sets (Breiman, 2001). Breiman (2001) also utilized back propagation with controlled layers of data that serve the purpose of predictions. Additionally, the use of data mining techniques was used to extract through sensor networks (Safi & Bouroumi, 2013). Iyer et. al. (2011) utilized Waikato Environment for Knowledge Analysis (WEKA) as an interface to implement decision tree analysis and study the behavior of algorithms conditions.

SVM is an effective classification technique that supports kernel mapping of the data points to a higher dimensional space for small dataset (Cortez & Morais, 2007). SVM could be used with convex optimization method to determine the decision boundary to split dataset **??** (Cortez & Morais, 2007). The time dependence of the forest area burned in a given year is inherently chaotic, and the predictions become less accurate as time increases (Malarz, Kaczanowska, & Kulakowski, 2002). The features extracted from the predicted class through data mining allows machine learning algorithms to perform the function of data transformation (Iyer et al., 2011). Viegas et al., (1999) examined five different methods of forest fire prediction and determined that the Canadian and modified Nesterov methods yielded the best overall performance. The K-Nearest Neighbor (KNN) method had also proven to be timely, costefficient, and accurate when applied in the Nordic countries and the United States (Finley, Ek, Bai, & Bauer, 2005). KNN is a non-parametric method used in regression analysis and the classification of data. The principle behind KNN is to determine, amongst the training data set, the points closest to the new point and predict the labels (Service & Mountain, 2002) Two of the features of the decision tree are that it neglects the linearity of parameters or is independent of the meteorological, temporal and spatial data. It is not affected by missing values or outliers, as it splits the data on ranges rather than absolute values. It does not require the transformation or scaling of parameters like regression analysis. Also, the decision trees implicitly perform feature selection. Decision tree modeling has its origins in artificial intelligence research where the aim was to produce a system that could identify existing patterns and recognize similar class membership (Ofren & Harvey, 1996). Sensor nodes collect measured data and send to their respective cluster nodes that collaboratively process the data by constructing a neural network (Yu, Wang, & Meng, 2005). This process is expensive when compared to other methods since it involves installation of sensor systems. Service & Mountain (2002) included linear models (LMs), generalized additive models (GAMs), classification and regression trees (CARTs), multivariate adaptive regression splines (MARS), and artificial neural networks (ANNs) to identify which suits better for predicting forest fires. The comparative study concluded that the model's accuracy changes with the real time and assumed datasets.

.

Though there were different techniques and models developed, the paper compares three different techniques with same datasets for both binary and multiclass classification to determine the accuracy percentage of each technique. The following section in this paper explains the research methods and results obtained from the analysis.

# 5 V. Research Methodology

This paper utilizes three different data mining techniques, KNN, SVM, and decision tree models to identify the accuracy of each technique on a small database. The data collected (feature extracted) for this research are from two different reliable sources: 1) the US meteorological department (climate data such as maximum and minimum temperature, humidity, precipitation and snowfall); and 2) the US forest fire database (Burnt area, severity, latitude, longitude). The feature extraction is a prime factor that contributes towards machine learning. The data collected are fused using Python programming language and is cleaned, processed, and integrated into the models.

The primary intention of this paper is to utilize data fusion technique and identify the regions of severity using three different prediction methods.These results are compared with UCI repository data set to prove that the models in this paper perform better. The UCI dataset consists of Fire weather index, which serves as the core parameters towards detection of forest fires. The paper utilizes this information to derive the probability of occurrence of a forest fire and plot a performance curve. While predominantly, most machine-learning problems involve feature extraction as its defining factor, the model is assumed to behave like a black box. This paper aims at modifying the model at its root and fit them according to the dataset and its characteristics.

# 6 a) Feature extraction

The primary task of feature extraction is to understand the interpretations of the dataset. The output label needs to be clearly stated that helps in correlating and analyzing the data features. It can be done using the Fisher's information that provides a way of measuring the extent of how much one feature is dependent on another within the dataset. It provides the amount of information a feature has towards the prediction of the output label. The dataset is analyzed for its ability to undergo dimensionality reduction that helps to understand the output visually. The paper tests the hypothesis of predicting forest fires using meteorological data (interchangeably used with Climate Data) and fire data from the Monitoring Trends in Burnt Severity (MTBS) data source.

The algorithm and data extraction are learned at the University of California, Irvine machine learning repository that has data sets of forest fires from Portugal. The 517 samples from the UCI repository contains features from the Fire Weather Index such as FFMC and DMC. These serve as major contributing factors, which are derived from Fisher's information for predicting forest fires.

31 Year 2016

# 7 ( ) C b) Data Fusion

The feature extracted data need to be fused together with specific date and region for all ten years. It is validated through the online metadata for US climate and MTBS data. In the Geospatial domain, we obtain localized points which on daily cycle records meteorological data. Additionally, the MTBS department also records the occurrence of forest fires. Using 'Beautiful Soup' library, a Python script is written that extracts data over a span of 10 years from 2004-2014. It is then fused with metadata that maps the occurrence of forest fire on a particular day with its respective climate data. It provides features such as Precipitation, Temperature (maximum and minimum), Burnt Area, Latitude, and Longitude of fire occurrence. If there is a date match with an occurrence of a fire, the dataset is integrated with its own forest fire affiliated data. If there is no burnt area, it is marked with a zero. It results in a wide separation between burnt severities and magnifies the confidence of prediction. While both datasets provide a binary label that allows us to predict if a forest fire occurred on a particular date given the meteorological data, the fused data also provides us with the liability to provide for the severity of the fire.

# 8 c) Data Preprocessing

Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, redundant information, noisy and unreliable data. While the dataset includes 21,000 samples from five states and seven different features with a small dimensionality, there is a need to look for false positives in the data and has to be omitted. Another python script is written that checks for such anomalies. It occurs because of the dataset during extraction, parses data at (0,0) latitude and longitude when there is no fire data against that date. Thus, it needs to be cleaned up or omitted to analyze in certain models.

Furthermore, this simplifies the search space a level further by consolidating valid samples. The first part is to infer the occurrence of forest fire whereas the second part is to identify the severity of the occurrences using MTBS reference table. It is performed using binary and multi-class classification while the former predicts the occurrence, the latter identifies the severity. The burnt severity is branched into five categories, namely: Very Small, Small, Medium, Large, and Very Large. Subsequently, these modes are separately passed through 3 models used for the classification of the data to derive meaningful results from the output.

# 9 d) Binary Classifiers

The process of Binary classification includes training, testing and validating data to determine the occurrence of wildfires from 21000 samples. These classification procedures are implemented in all three models respectively. Initially, a set of data is used to train the machine when the expected output is given to learning the pattern. Later, the data is tested to study the behavior of the machine and finally, the accuracy percentage is determined from each of the techniques by validation.

# 10 e) Multiclass classifiers

After training the machine to learn the prediction of burnt area from the sample provided by various features, the process of training and testing repeats with three different models for multi-class classifiers. The training includes severity data initially and then at the Year 2016 ( ) testing instance, the models are run to predict the right severity and validated later with real-time data to determine the accuracy percentage.

# 11 VI. Model Validation

The section validates three different models and explains the varied approaches used by the authors to improve the accuracy of prediction models. Support vector machines, K-Nearest Neighbors, and Decision tree stumps are trained and tested with modified algorithms to improve the accuracy.

# 12 VII. Svm Validation

Support vector machines (SVM) are learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. A set of training samples, each marked as belonging to one of two categories (0 or 1); an SVM training algorithm builds a model and make a not-probabilistic classifier. The samples are mapped so that the samples of the separate categories are divided by a clear gap that is as wide as possible. New samples are then mappedinto that same space and predicted to belong to a category based on which side of the decision boundary they fall on in the domain space. The principle behind this model is to maximize the distance between the two classes that are positive and negative classes.

# 13 a) Modified Approach

The open source machine-learning library LIBSVM implements the algorithm for kernel SVM. SVM requires data to be represented as a vector of real numbers. The most trivial approach is to define simply the training and testing data and pass it to the SVM model. It provides the desired output regarding the input data. However, this paper aims at modifying the black box SVM model and analyzing it on the fused dataset. The first step was transforming the data into numerical data and then to the format for the LIBSVM package. While choosing a model for the SVM, several parameters are taken into accounts such as the penalty parameter, C, and the kernel parameters. We found that the model worked best when the soft margin constant C was kept at 100. The smaller value of C will tend to ignore the points close to the boundary and causes false results. Kernel parameters also have a significant effect on the SVM model. As our feature set is small, we chose the RBF kernel as it non-linearly maps data into a higher dimensional space and handles non-linear relationships between class labels and features. The degree of the polynomial controls the flexibility of the classifier. We found that the 5-degree polynomial works best as it has a greater curvature. The nu-SVM model sets a lower and upper boundary on the number of data points that lie on the wrong side of the hyperplane and is advantageous for controlling the number of support vectors.

# 14 b) Results

The Receiver Operating Characteristic (ROC) curve plots the true positive rate against the false positive rate. Figure 2 shows the area under the curve for the ROC on the SVM model. The true positive rate resembles the burnt area in the spatial domain, whereas the false positive rate identifies the non-burnt area in the spatial domain.

using the SVM model with an RBF kernel over the given data set. The Mean square error obtained by implementing a Support Vector Regression model after taking a log(x+1) on the data set gives us 2.3117. It turns out to map onto the burnt area in a given spatial domain given its coordinates.

# 15 VIII. KNN Validation

Initially, a random set of points k is chosen. This k is the same number as neighbors and finds all the points in the training set that are closest. The weighted average of these points then moves k to a new place to balance the centroid in a spatial domain. Figure 3 shows the cells that depict the neighbors. distance. It is repeated several times to obtain a weighted average to test the validity of the code and the model. To elucidate further on this, we run a KNN model with up to 50 neighbors. With each new neighbor, an expected error is obtained on that models' neighbors' index. The test set is then applied to our trained model. The true error obtained here is compared to the expected error, and its accuracy is validated.

The second approach verifies the trained model and runs the k-fold cross validation on it. By this, the cross validation losses are obtained from each incremented neighbor. The index of which is then matched with the model that provides the least error. It provides us with an expected error per epoch. This, in turn, returns a minimum error of these neighbors. If the error obtained through cross-validation is lower than the expected error, the index at which the KNN flags optimum is incorrect and vice versa. This way the KNN model is used with both binary and multi-class classifiers.

# 16 b) Results

The KNN models are trained with UCI data primarily and then trained with the fused dataset. It is done to compare the accuracy and also to make machine optimize the pattern of output required.

Initially, k = 2, there would be {xj,yj} values where j ? size (D) closer to one of the k points. As we add another point to accommodate this phenomenon, the accuracy is accounted by the correctness of predicting the sample point in its respective polygon. The forest fire data occurs close to one another according to the feature space. Additionally, the features are localized to a spatial domain. Thus, if a model needs to predict the occurrence of forest fire based on meteorological data over a constrained area of land, its confidence is magnified if predicted correctly within the neighborhood of the previous occurrence with similar data. KNN does this exactly.

# 17 a) Modified Approach

Again manually altering the black box model, the author not only defined the model behavior but also increased the confidence by repeating the experiment several times. Each time the experiment is repeated, the number of neighbors is altered, and the behavior change of the pattern is observed and recorded.

Two different approaches tackle the model. First, the data set is separated into training and testing modules. The MATLAB code then produces an expected error from the training set. It is then matched against its test error or exact error, and the percentage of accuracy is derived using squared Euclidean The accuracy percentage for UCI dataset is 53 % for binary and 40% for multiclass whereas the accuracy percentage of the fused dataset is close to 55% in binary and 44% in multi-class.

# 18 IX. Decision Tree Validation

After the Nearest Neighbor approach to classification/regression, perhaps the second most intuitive model is Decision Trees. There are many possible trees can be used to organize (i.e., classify) the dataset. It is also feasible to get the same classifier with two very different trees. Tree classification becomes complex with lots of features. A tree that splits the data into all pure leaves is considered consistent with the data. It is always possible when no two samples have different outcomes but identical features. The hierarchy of the architecture leafs out in a manner where every level is a feature. The decision is made on a binary basis. Intuitively, the complexity of the tree increases the variance on the classification boundary.

# 19 a) Modified Approach

The data is separated into testing and training. Using the C4.5 Decision Tree classifier, WEKA produced results that proved that the fused dataset had more accuracy than the 517 sample set. It can be reasoned merely due to some instances (21421 instances of data) than the 517 dataset. The smaller data set could overfit the model. The other reason is due to our better feature selection of spatial data (latitude) and meteorological data; the output has a higher attribute ranking.

Based on the C4.5 classifier model, the UCI 517 dataset could predict correctly at 46.15 % while the

# 20 b) Results

The classifier is developed using WEKA tool that serves best on controlling attributes, enhance visualization and preprocessing data, and availability of a variety of decision tree algorithms. Open-source workbench called WEKA is a useful tool to quantify and validate results, which can be edited and modified. WEKA can handle numeric attributes well, so we use the same values for the weather data from the UCI repository datasets. The class variable has to be a nominal one, to allow WEKA. As WEKA uses kappa stats for evaluating the training sets, a standard score of > 60 % means training set is correlated, using C4.5 simulations. C4.5 is the popular decision tree algorithm, and the WEKA employs the J48 that is an open-source Java implementation of C4.5. The C4.5 or J48 is an improved version of original ID3 that has additional support to handle continuous features in the data and a better bottom-up pruning methodology. The C4.5 automatically handles the pruning (to manage the overfitting) by default.

fused dataset could achieve 50%. With reduced error pruning, the rate could be increased roughly by 1%. The classifier is right in predicting the small fires. It achieves good accuracy with Prediction, Recall and ROC area. From the output file, it predicts better based on the features for a lower severity. Particularly, the area under ROC curve outputs the fused dataset at a value of 0.77 in most classes and with a weighted average of 0.636. In contrast, the weighted ROC curve area for UCI dataset is 0.569. The class attribute of the burnt area that needs to be classified under supervised learning is a multiclass attribute that is based on the size of the burnt area. The accuracy percentage from binary classifiers is close to 57 % and percentage from multi-class classifiers is around 42 %. We employed the different algorithms for the Decision trees that could better suit the meteorological, spatial, and temporal data that are continuous.

# 21 X. K-Means Clustering

K-means clustering approach failed to deliver any useful results in this paper. The segregated dataset into five different classes to see the clustering based on the states were chosen and their burnt severity type respectively. This model changes its center after every iteration due to the highly localized data. Thus, it is unable to draw a conclusion on a stable centroid that distinctly separates the classes. Figure 8 depicts the clustering of burnt severity of five classes. Due to this unlikely occurrence of overlapping data, no classifier can accurately suggest a stable or correct output. Hence, the clustering is omitted for this small-scale dataset.

# 22 XI. Conclusion

There are many research on forest fire predictions. There have been very fewer approaches to identify the accuracy of these models for both binary and multi-classifiers. The data fused is used to predict the occurrence by training the machine using latitude, longitude, temperature, humidity, burnt area, burnt area severity, precipitation, and snowfall. The purpose of this paper is to arrive at a model that predicts accurately in a small dataset on both binary classifiers and multi-class classifiers.

The validity of the model will be tested based on supervised learning of structured data. The research is chosen, as there is a need to have different models for different sizes of data. The actual experiment results will tell the suitable method and throw some light on the nature of the problem. Table 3 details on accuracy percentages of both binary and multiclass classifiers of three predictive techniques. From the table 3, it is evident that many parameters come into play while considering models on a small database. With respect to the database, SVM behaves as the optimal model to implement a binary classification and KNN for multiclass classification. The future focus is to improve the algorithms and add satellite images to extract more features and improve the accuracy of machine learning models. The research team also focuses on visualizing data and study of hypothesis over such small dimensionality using Inference and graphical models. [1] [2]

Figure 1:

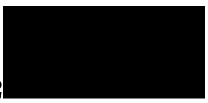
2 

Figure 2: Fig. 2 :

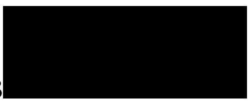
3 

Figure 3: Fig. 3 :

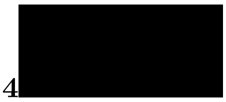
4


Figure 4: Fig. 4 :

6


Figure 5: Fig. 5 and 6 :

7


Figure 6: Fig. 7 :

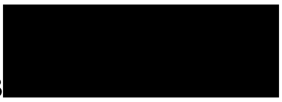
8


Figure 7: Fig. 8 :

9


Figure 8: Fig. 9 :

have been applied to identify the best model for
predicting fire occurrence and spread
Year 2016
30
Volume XVI Issue IV Version I
)
(
Global Journal of Computer Science and Technology

Figure 9:

**1**

| State | Date | Latitude | Longitude | Burnt Area |
|---|---|---|---|---|
| Nevada | 04-25-2007 | 36.647 | -116.435 | 330706 |
| Idaho | 06-15-2004 | 44.154 | -115.566 | 9862 |
| Oregon | 07-20-2010 | 38.469 | -112.473 | 42956 |
| New Mexico | 04-19-2008 | 37.623 | -78.422 | 807 |
| California | 07-13-2010 | 36.215 | -121.447 | 934 |

The above table randomly picks up tuples from each state of the test data and validates it against the MTBS metadata. It checks if the given forest fire occurred. It also crosses checks against its respective meteorological dataset.
Additionally, on analyzing the output as derived from MATLAB provides us with an accuracy of 75.67%

Figure 10: Table 1 :

**2**

| State | Date | Latitude | Longitude | Burnt Area |
|---|---|---|---|---|
| Nevada | 04-02-2007 | 39.014 | -116.867 | 6662 |
| Idaho | 06-13-2004 | 45.153 | -114.903 | 538167 |
| Oregon | 01-11-2010 | 28.903 | -82.194 | 450 |
| New Mexico | 07-23-2009 | 65.625 | -143.671 | 42649 |
| California | 10-21-2007 | 33.181 | -116.430 | 197990 |

Figure 11: Table 2 :

**3**

| Model | Accuracy |
|---|---|
| SVM | Binary: 65% Multiclass: 42% |
| Decision Tree | Binary: 57% Multiclass: 42% |
| KNN | Binary: 55% Multiclass: 44% |

Figure 12: Table 3 :

[Dale et al. ()] '723-734 ST -Climate change and forest disturbances'. V H Dale , L Joyce , S Mcnulty , R P Neilson , M P Ayres , M D Flannigan , P , M . 10.1641/0006-3568(2001)051[0723:CCAFD]2.0.CO;2. *Bioscience* 2001. 51 (9) . (Climate change and forest disturbances)

[De Groot et al. ()] 'A comparison of Canadian and Russian boreal forest fire regimes'. W J De Groot , A S Cantin , M D Flannigan , A J Soja , L M Gowman , A Newbery . 10.1016/j.foreco.2012.07.033. *Forest Ecology and Management* 2013. 294 p. .

[Cortez and Morais ()] 'A Data Mining Approach to Predict Forest Fires using Meteorological Data'. P Cortez , A Morais . *New Trends in Artificial Intelligence*, 2007. p. .

[Ofren and Harvey ()] *A Multivariate Decision Tree Analysis of Biophysical Factors in Tropical Forest Fire Occurence*, R S Ofren , E Harvey . 1996. p. . (Integrated Tools Proceedings)

[Alonso-Betanzos et al. ()] 'An intelligent system for forest fire risk prediction and firefighting management in Galicia'. A Alonso-Betanzos , O Fontenla-Romero , B Guijarro-Berdiñas , E Hernández-Pereira , M I Paz Andrade , E Jiménez , T Carballas . doi:10.1016/ S0957-4174(03)00095-2. *Expert Systems with Applications* 2003. 25 (4) p. .

[Bisquert et al. ()] 'Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data'. M Bisquert , E Caselles , J M Snchez , V &caselles . doi:10. 1071/WF11105. *International Journal of Wildland Fire* 2012. 21 (8) p. .

[Malarz et al. ()] *Are Forest Fires Predictable?*, K Malarz , S Kaczanowska , K Kulakowski . doi:10. 1142/S0129183102003760. 2002. p. 13.

[Flannigan et al. ()] *Climate Change and Forest Fires. The Science of the Total Environment*, M D Flannigan , B J J Stocks , B M Wotton . doi:10.1016/ S0048-9697(00)00524-6. 2000.

[Viegas et al. ()] 'Comparative study of various methods of fire danger evaluation in southern Europe'. D X Viegas , G Bovio , A Ferreira , A Nosenzo , B Sol . 10.1071/WF00015. *International Journal of Wildland Fire* 1999. 9 (4) p. 235.

[Service and Mountain ()] 'Comparing Five Modelling Techniques'. U S F Service , R Mountain . *Ecological Modelling* 2002.

[Mcroberts et al. ()] 'Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery'. R E Mcroberts , E O Tomppo , A O Finley , J Heikkinen . doi: 10. 1016/j.rse.2007.04.002. *Remote Sensing of Environment* 2007. 111 (4) p. .

[Franco-Lopez et al. ()] *Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbor's method. Remote Sensing of Environment*, H Franco-Lopez , A R Ek , M E Bauer . 10.1016/S0034-4257(01)00209-7. 2001. 77 p. .

[Park et al. ()] 'Fire Monitoring Handbook. Program, 285'. N Park , S National , I Fire . http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Fire+Monitoring+Handbook#1 *Wildfire Predictions: Determining Reliable Models using Fused Dataset 22*, 2003.

[Martins Fernandes ()] 'Fire spread prediction in shrub fuels in Portugal'. P A Martins Fernandes . doi:10.1016/ S0378-1127(00. *Forest Ecology and Management* 2001. 144 (1-3) p. .

[Gavin et al. ()] 'Forest fire and climate change in western North America: Insights from sediment charcoal records'. D G Gavin , D J Hallett , S H Feng , K P Lertzman , S J Prichard , K J Brown , D L Peterson . 10.1890/1540-9295(2007)5[499:FFACCI]2.0.CO;2. *Frontiers in Ecology and the Environment* 2007.

[Schoennagel et al. ()] *Implementation of National Fire Plan treatments near the wildland-urban interface in the western United States*, T Schoennagel , C R Nelson , D M Theobald , G C Carnwath , T B Chapman . 10.1073/pnas.090099110. 2009. Proceedings of the National Academy of Sciences of the United States of America

[Schoennagel et al. ()] 'Implementation of National Fire Plan treatments near the wildland-urban interface in the western United States'. T Schoennagel , C R Nelson , D M Theobald , G C Carnwath , T B Chapman . 10.1073/pnas.0900991106. *Proceedings of the National Academy of Sciences of the United States of America* 2009. 106 (26) p. .

[Cheng and Wang ()] 'Integrated spatiotemporal data mining for forest fire prediction'. T Cheng , J Wang . doi:10.1111/j. 1467-9671.2008.01117.x. *Transactions in GIS* 2008. 12 (5) p. .

[Finley et al. ()] *K-Nearest Neighbor Estimation of Forest Attributes: Improving Mapping Efficiency BT - Proceedings of the fifth annual forest inventory and analysis symposium*, A Finley , A Ek , Y Bai , M Bauer . 2005. p. .

[Chang and Lin ()] 'LIBSVM: A Library for Support Vector Machines'. C.-C Chang , C.-J Lin . 10.1145/1961189.1961199. *ACM Transactions on Intelligent Systems and Technology* 2011. 2 p. 27.

[Iyer et al. ()] *Machine Learning and Dataming Algorithms for Predicting Accidental Small Forest Fires. Weather*, V Iyer , T , S S I Paramesh , N Murthy , GR , U , MB S . 2011. p. .

[Mcroberts et al. ()] *Parametric, bootstrap, and jackknife variance estimators for the k-Nearest Neighbors technique with illustrations using forest inventory and satellite image data. Remote Sensing of Environment*, R E Mcroberts , S Magnussen , E O Tomppo , G Chirici . doi:10.1016/ j.rse.2011.07.002. 2011. 115 p. .

[Safi and Bouroumi ()] *Prediction of Forest Fires Using Artificial Neural Networks Description of the proposed method Artificial neural networks*, Y Safi , A Bouroumi . 2013. 7 p. .

[Breiman ()] 'Random Forrest'. L Breiman . 10.1023/A:1010933404324. *Machine Learning*, 2001. p. .

[Schoennagel and Nelson ()] 'Restoration relevance of recent National Fire Plan treatments in forests of the western United States'. T Schoennagel , C R Nelson . 10.1890/090199. *Frontiers in Ecology and the Environment* 2011. 9 (5) p. .

[Mcroberts et al. ()] *Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. Remote Sensing of Environment*, R Mcroberts , M Nelson , D Wendt . 10.1016/S0034-4257(02)00064-0. 2002. 82 p. .

[French et al. ()] 'Using Landsat data to assess fire and burn severity in the North American boreal forest region: an overview and summary of results'. N French , E Kasischke , R Hall , K Murphy , D Verbyla , E Hoy , J Allen . 10.1071/WF0800. *International Journal of Wildland Fire* 2008. 17 (4) p. .

[Mcroberts ()] *Using satellite imagery and the k-nearest neighbors technique as a bridge between strategic and management forest inventories. Remote Sensing of Environment*, R E Mcroberts . 10.1016/j.rse.2007.07.025. 2008. 112 p. .