



Machine Learning for Forest Fire Prediction: A Case Study in North Algeria

Youcef Ghibeche^{1*}, Abdellah Sellam², Nabil Nouri¹, Ahmed Khaldi³, Amine Harrane³, Ismail Ghibeche⁴

¹ Telecommunications and Smart Systems Laboratory, Department of Computer Science, Faculty of Exact Sciences and Computer Science, Ziane Achour University of Djelfa, Djelfa 17000, Algeria

² Department of Computer Science, Institute of Sciences, University Center El Cherif Bouchoucha, Aflou 3001, Algeria

³ Department of Computer Science, Faculty of Exact Sciences and Computer Science, Ziane Achour University of Djelfa, Djelfa 17000, Algeria

⁴ Department of Hydraulics, Faculty of Sciences and Technology, Ziane Achour University of Djelfa, Djelfa 17000, Algeria

Corresponding Author Email: y.ghibeche@univ-djelfa.dz

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290133>

ABSTRACT

Received: 5 October 2023

Revised: 14 January 2024

Accepted: 23 January 2024

Available online: 27 February 2024

Keywords:

forest fire, fire prediction system, machine learning, decision tree, random forest

Wildland fires are the most common peril for forests due to climate change. Furthermore, it is an uncontrollable disaster and poses a great deal of threat to human health and ecosystems. In Algeria, almost 40,000 hectares are burned each year, approximately 1% of all existing woodlands of the country. In this work, the forest fire event prediction is highlighted using machine learning. The study utilized data sets from several sources, including fire data obtained from the fire information system for resource management by NASA (FIRMS) and climate data accessed from the NASA energy project API, derived from the MODIS satellite (NASA forecasting of energy resources around the world). Fire data from NASA provides real-time information, spanning from 2000 to 2020. The methodology process of creating the prediction system involved collecting the data, pre-processing the data, finding the best models, training and testing the models, and evaluating them for validation. The machine learning model was trained and validated using 70% and 30% of the set features with a performance accuracy of up to 86%. Upon completion, we deployed our selected machine learning model to create a Web platform enables different end users to check possible future forest fires by select a geographical area on a world map. The objective of our machine learning model is to analyze the weather data of the selecting area on the map in real time and predict whether a fire will occur or not. This prediction system will enhance early detection, allowing prompt response measures to be implemented, reducing the risk of uncontrolled wildfires and safeguarding ecosystems and communities.

1. INTRODUCTION

Forests are a major natural resource that plays a crucial role in maintaining environmental balance. Moreover, it is the most necessary resource part of our ecosystem. Human health and wealth are inextricably linked to forest health; from the fresh air we breathe to the natural products we rely on. Therefore, the health of the forest in any given area is a real indicator of the ecological condition prevailing in that area.

Fire has been closely associated with mankind from the beginning of civilization [1]. The discovery of fire and its uses have directly or indirectly permitted a man to live and survive in the temperate zone. However, fire also can be a danger whose potential for disaster is a source of growing concern all over the world (every year millions of hectares of land are destroyed by fire, which causes some damage to the natural environment [2], the forest fire also causes an increase in the proportion of CO₂ in the air, which causes suffocation and respiratory diseases for human.

Forest fires are considered to be a potential hazard with physical, biological, ecological, and environmental

consequences [3]. Forest fire results in partial or complete degradation of vegetation cover, thus modifying the radiation balance by increasing the surface albedo, and water runoff, and raising soil erosion [4].

Forest fires are the most common peril in forests. It is an uncontrollable event that occurs in nature that poses a great deal of threat to the wildlife and the people who live there. It is reported that each year in the last decade, a total number of 4 to 6 million wildfire events happened worldwide [5]. Therefore, the probability of occurrence of it depends on the ignition causes and environmental preconditions [6].

Algeria is the one considered the fire hotspot on the southern rim of the Mediterranean Basin (MB) [7], which has suffered from forest fires due to the presence of flammable fuels such as shrub lands and forests [2]. These recent forest fires that happened last year in "Algeria" causes huge damage that required international support to stop (Figure 1) and the death of many Algerians to throw the process of rescuing. Around 90 people, including 57 civilians and 33 soldiers (during rescue operations) [8], and according to study [9], the total area of vegetation cover affected by fires during the

summer of 2021, more than 100,000 hectares through 1,631 fire outbreaks recorded in 21 wilayas. Around 260,135 hectares of forests, 21,040 hectares of bushes, 16,415 hectares of scrub, 16,160 hectares of fruit trees, and 352 hectares of esparto were ravaged by the fires. Also, 19,178 farm animals burned in the fires [10], and 1,705 homes burned [11].

Indeed, the Algerian forests of high productivity and conservation value have suffered during the recent decades of degradation and fragmentation from repeated fires [12, 13]. According to study [14], fires lead to consumption of six times more than these forest ecosystems could produce.

To avoid these losses, predicting the forest fire before it happens has effectiveness and influence more than detecting it (At least, we gain some time to take action to avoid many of those losses). Prediction of wildfire occurrence also plays a major role in resource allocation, mitigation, and recovery efforts.

Nowadays, there are various technologies to predict forest fires depending on the availability of data collected from remote sensing satellites and location detection systems; Machine learning as a sub-branch of Artificial Intelligence (AI) is one of these technologies.

A variety of machine learning (ML) models with different architectures has been used in literature to predict forest fires in different zones around the world, and some of them used only meteorological data. These studies are capable to perform good results and obtaining high accuracy for predicting forest fires [15-19]. However, some challenges and limitations are encountered by many studies including [20]:

-The effectiveness of forest fire prediction relies on gathering a substantial volume of data from diverse sources. However, the presence of incomplete or inaccurate data introduces challenges that can impact the reliability and validity of prediction models.

-In the context of forest fire prediction, models must grapple with the complexity and uncertainty surrounding fire behavior and spread, including interactions and feedbacks among diverse factors and scales. Despite these challenges, accurately measuring, modeling, and validating certain aspects of this intricate system may prove challenging.

-For forest fire prediction to be effective across diverse regions, scenarios, and conditions, models must be adaptable. However, some prediction models rely on specific assumptions or parameters that may not be universally applicable. Regular testing, calibration, and updates are essential to ensure the accuracy and robustness of these models in various situations and environments.

In Algeria, a prediction study of fire behavior based on terrain, wind conditions, and fuel characteristics presented in [21]. In study [22], a predictive model based on the decision tree for forest fires prediction using data mining techniques presented three meteorological attributes namely: temperature, relative humidity (RH), and wind speed.

While study [23] applied artificial neural networks to predict forest fires in embedded devices using collected meteorological data from wireless sensor networks, nine machine learning algorithms were investigated and compared based on the obtained results, and they propose an embedded forest fire prediction model. The study [24] addresses wildfire prediction using a recent dataset from 2012, employing an artificial neural network (ANN) that outperforms other classifiers in accuracy, precision, and recall. Key features influencing predictions include relative humidity (RH), drought code (DC), and initial spread index (ISI).

The main objective of our work, is to create a forest fire prediction system to predict the fire early based on the data climate and build a complete system that it can be accessed from any device in the world at any time, and we focus on:

- Create the system with the lowest budget possible.
- Let the users interact with the system from any device.

A comparison between three algorithms was applied by examining the experimental results to obtain the best model, K-Nearest Neighbors (KNN), Decision Trees, and Random Forest, which are diverse machine learning models with distinct characteristics. KNN is an instance-based algorithm that classifies data points based on their proximity to neighboring instances. Decision Trees construct a tree-like structure to make decisions by splitting the dataset at different features. Random Forest, an ensemble method, leverages multiple decision trees to enhance predictive accuracy and reduce overfitting. While KNN relies on local relationships, Decision Trees focus on hierarchical decision-making, and Random Forest combines the strengths of multiple trees for robust predictions. Each model suits different scenarios, offering a range of approaches for various machine learning tasks.

Several classifiers other than the three mentioned in the comment were evaluated, however empirically KNN, decision trees and random forest showed the most promising results.

Furthermore, the random forest classifier yields the best experiment results in terms of prediction accuracy, which makes it a useful model for predicting fires in Algeria.

The practical application of our forest fire forecasting systems will bring great benefit to the state and individuals on the other hand, including that, firstly, a low-budget, easy to implement, and accessible system for forecasting forest fires. Secondly, predicting fires before they occur facilitates the process of rapid intervention to extinguish them and prevent their spread, and it also avoids losses and damages, that were previously mentioned in the introduction.

The rest of the paper is organized as follows: The methodology section discusses the method we have pursued to create the prediction system, including the system architecture, the data collection, and the algorithm used to build the forest fire prediction system. While the obtained results will be presented and discussed in the results and discussion section. Finally, we conclude our work with a conclusion and suggest some perspectives.

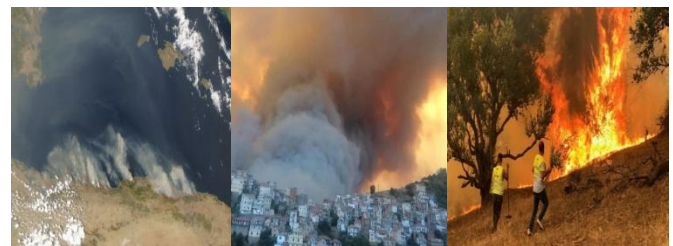


Figure 1. Wildfire in Algeria, 2021

2. METHODOLOGY

In this study Algeria country has been selected owing to:

- (1) The number of fires in Algeria has increased in recent years. Based on the availability of fire data, we chose a study period of 21 years from 2000 to 2020 (Figure 2 presents a map of the occurrence of fires number in northern Algeria of the

period 2001-2018).

- (2) Lack of equipment needed to deal with fires.
- (3) Lack of early detection systems for fires.
- (4) The difficulty of the region's terrain.

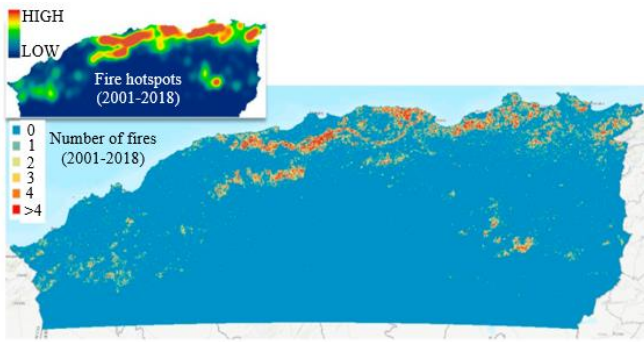


Figure 2. Number of fires in northern Algeria between 2001-2018, the enclosed figure (upper left) indicates the fire hotspots and their level of density [7]

In our work, we used machine learning algorithms to train an AI model that can be applied to the future climate dataset to predict forest fires before they start.

This will help the authorities to take adequate precautions and make necessary arrangements to reduce possible losses.

In order to achieve this goal, the designed system uses a trained machine learning model to analyze weather data in the area selected by the user to predict if a potential fire may occur.

To analyze the data provided by NASA for the climate change and previous fires, the system starts by normalizing this data, and then it applies one of the machine learning algorithms: K-Nearest Neighbor (KNN), decision tree (DT), and random forest (RF).

The website provides a simple interface that allows the user (local authorities, and civilians) to select a geographical area, then displays the possibility that a fire may start.

The following figure (Figure 3) summarizes the architecture of our system.

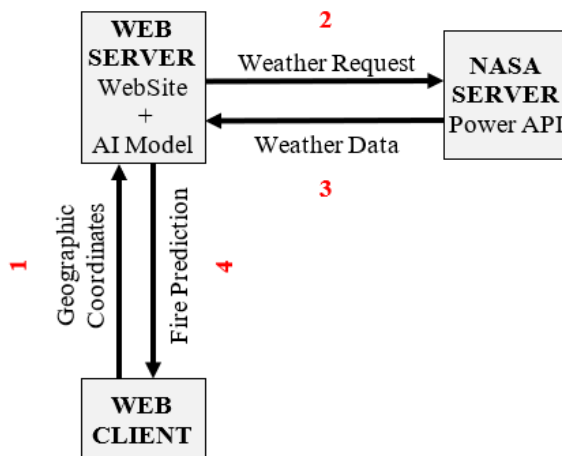


Figure 3. Number general system architecture

The process of creating the prediction system consists of the following steps:

2.1 Collecting the data

In this step, we collect two types of data from two different

sources:

-The geographical location and date of actual fires.

-The weather data corresponding to a specific date and geographical location.

The fires data. The fire data provided by FIRMS from NASA is divided according to country and year. The data in each file consist of: latitude, longitude, brightness, scan, track, acq_date, acq_time, satellite, instrument, confidence, version, bright y31, frp, daynight, and type, but in our case we need only to:

-Latitude and longitude: represent the geographical coordinates of the fire.

-acq_date: represents the date when the fire happens.

-Confidence: This confidence estimate, which ranges between 0% and 100%, is used to assign one of the three fire classes (low-confidence fire, nominal-confidence fire, or high-confidence fire).

For our study, we collected fire information in the period between 2000 and 2020 for the region of Algeria.

The climate data. The climate data is provided using the NASA POWER PROJECT API from the MODIS satellite.

The climate data that we collect consists of many features described in Table 1, in the following:

Table 1. Features description of climate data

Feature	Description	Measuring Unit
Latitude	Center of 1km fire pixel	Degree
Longitude	Center of 1km fire pixel	Degree
acq_date	Date of the fire	Yyyy/mm/dd
T2M_RANGE	Temperature at 2 Meters Range	Celsius
TS	Earth Skin Temperature	Celsius
T2MDEW	Dew/Frost Point at 2 Meters	Celsius
T2MWET	Wet Bulb Temperature at 2 Meters	Celsius
T2M_MAX	Temperature at 2 Meters Maximum	Celsius
T2M_MIN	Temperature at 2 Meters Minimum	Celsius
T2M	Temperature at 2 Meters	Celsius
QV2M	Specific Humidity at 2 Meters	g/kg
RH2M	Relative Humidity at 2 Meters	g/kg
PRECTOTCORR	Precipitation Corrected	mm/day
PS	Surface Pressure	kPa
WS10M	Wind Speed at 10 Meters	m/s
WS10M_MAX	Wind Speed at 10 Meters Maximum	m/s
WS10M_MIN	Wind Speed at 10 Meters Minimum	m/s
WS10M_RANGE	Wind Speed at 10 Meters Range	m/s
WS50M	Wind Speed at 50 Meters	m/s
WS50M_MAX	Wind Speed at 50 Meters Maximum	m/s
WS50M_MIN	Wind Speed at 50 Meters Minimum	m/s
WS50M_RANGE	Wind Speed at 50 Meters Range	m/s

The features that the table shows cover: coordinates, date, Temperature, Humidity, Wind, Surface Pressure and Precipitation.

Process of Collecting. The method to download the data is

shown in Figure 4.

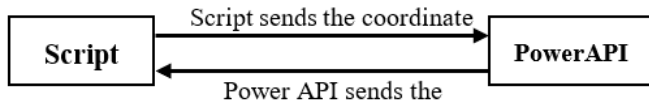


Figure 4. The process of collecting data

To create a machine learning model, we need positive and negative samples. To obtain negative samples, we create a Power API request using the same geographical coordinates of positive areas with a date decreased by 10 days.

Consequently, the data consist now of:

1. Fire data: the geographical coordinates and the date of the fires.
2. Positive climate data: the weather data of the day and the fire events consists of Temperature, Humidity, Precipitation, Surface Pressure, Wind and each feature has sub-features that are related to it, with a total of 23 features.

3. Negative climate data: data of normal days without any fire have the same features as the Positive climate data.

We organized the data into files corresponding to the different year of the studied period. Each data file has a different number of samples labeled as positive or negative:

- Positive: represents 50% of samples, indicated by 1 in the column 'fire', (meaning there was a fire in that location that day).
- Negative: represents 50% of samples, indicated by 0 in the column 'fire', (meaning there wasn't a fire in that location that day).

2.2 Pre-processing the data

Normalization alters raw datasets by creating new values and maintaining general distribution as well as a ratio in data. The most used type of normalization in machine learning is: Min-Max Scaling which subtract the minimum value from each column and divide by the range (max - min). Each new column will have a minimum value of 0 and a maximum value of 1.

There are many techniques in Normalization such: Min-Max, Z-score and more, but we focus on using the Min-Max approach.

2.3 Finding the best models

The process of choosing a machine learning model depends on many factors ranging from the type of problem at hand to the type of output you are looking for, some of these factors are:

- Size of the Training Data.
- Accuracy and/or Interpretability of the Output.
- Speed (Training Time).
- Number of Features.

In the following, we tested the three models described above:

1. K-Nearest Neighbor.
2. Random forest.
3. Decision tree.

2.4 Training and testing the models

We evaluated each model using the following train-test ratio:

- 70 percent of the data will be used for training and 30

percent of the data will be for testing.

- 80 percent of the data will be used for training and 20 percent of the data will be for testing.

- 90 percent of the data will be used for training and 10 percent of the data will be for testing.

2.5 Evaluation

We will evaluate the accuracy of the models on the test collection by calculating the Accuracy Score performance metric which is a scoring system in binary classification (i.e., determining if an answer or returned information is correct or not). It represents the ratio of correctly predicted outputs.

3. WEBSITE

The goal of the website is to allow any user that is interested in forest fire prediction to access the system from any device at any time.

The website is built using frontend and backend languages, and it consists of simple pages, the structure of the website is shown in Figure 5:

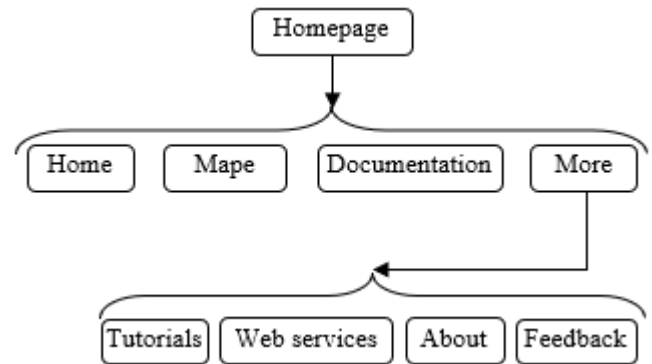


Figure 5. The website structure

The map page is where the user can test the system; it consists of input fields to specify the coordinates of a specific geographical area. Alternatively, he can select an area using the interactive map and check if there is a possibility that a forest fire may start.

The request of the user will be sent to the Webserver responsible for the prediction using our trained models. This server responds with the result that will be displayed to the user on the selected area.

4. RESULTS

We restrict our processing only to the data that is related to Algeria, where the confidence of fire is above 80%. The data is grouped by years as indicated in Figure 6.

Because the data already consists of clean vectors of measurements (features), there wasn't so much preprocessing needed. The only preprocessing method we applied to the data was the Min-Max normalization (see the equation below) to give the same importance (influence on training) to all the features of our data even if they have different ranges of values.

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

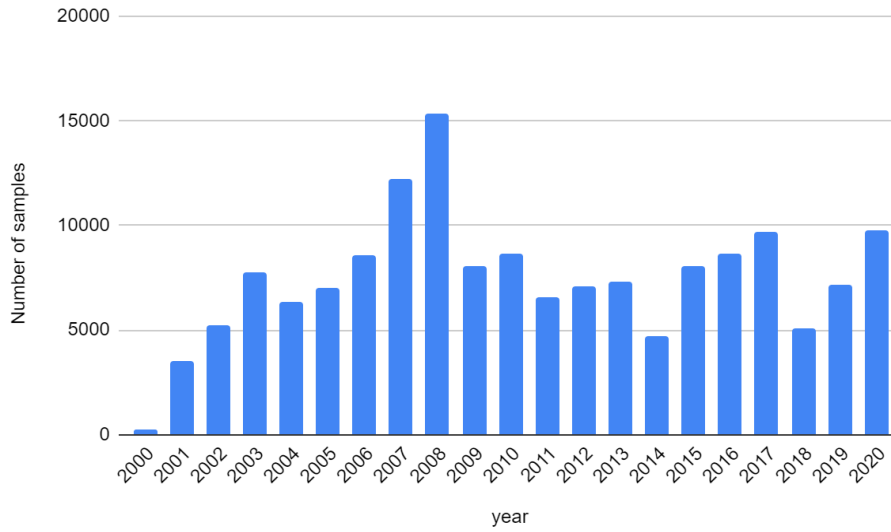


Figure 6. The distribution of samples with respect to each year

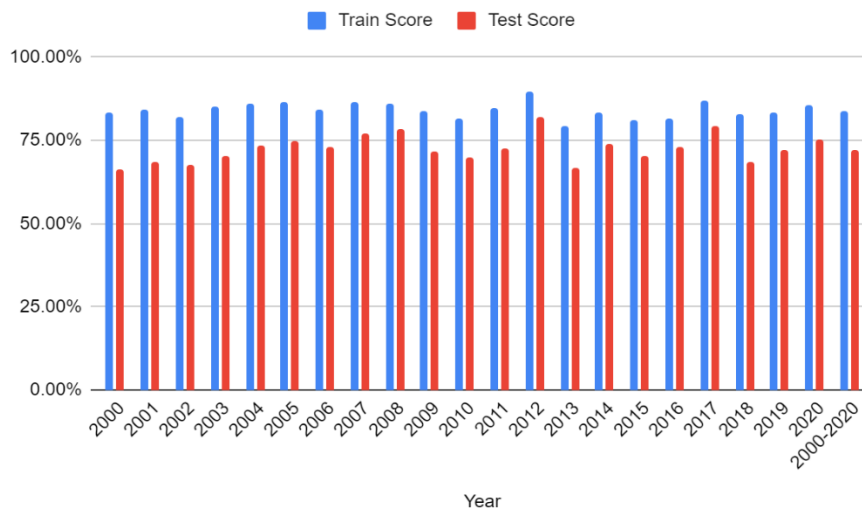


Figure 7. The accuracy of the KNN classifier for different years in the period 2000-2020

Table 2. The accuracy of KNN classifier

Year	Test Set Size: 30		Test Set Size: 20		Test Set Size: 10	
	Train Score	Test Score	Train Score	Test Score	Train Score	Test Score
2000	0.835443	0.661764	0.838888	0.739130	0.857142	0.739130
2001	0.843991	0.685660	0.846044	0.705128	0.851758	0.726495
2002	0.819218	0.674683	0.823794	0.691358	0.826261	0.696394
2003	0.849212	0.701709	0.849951	0.691025	0.847677	0.711538
2004	0.861185	0.735849	0.868710	0.759433	0.870370	0.757861
2005	0.867074	0.746077	0.871922	0.755349	0.871888	0.750356
2006	0.842245	0.731830	0.843002	0.740675	0.843482	0.724941
2007	0.865110	0.770277	0.860234	0.771335	0.859606	0.769795
2008	0.858538	0.783977	0.859039	0.790621	0.857328	0.792317
2009	0.836408	0.716059	0.840863	0.748603	0.845293	0.758064
2010	0.817415	0.696607	0.818102	0.727167	0.820460	0.731791
2011	0.847783	0.723892	0.850372	0.751718	0.845252	0.751145
2012	0.896869	0.820224	0.898525	0.822331	0.902465	0.823033
2013	0.794512	0.665910	0.802656	0.670524	0.801422	0.670748
2014	0.833282	0.737809	0.836382	0.755037	0.836397	0.737288
2015	0.813184	0.705276	0.810112	0.704390	0.809484	0.726823
2016	0.815109	0.730293	0.815915	0.746666	0.812008	0.755504
2017	0.871090	0.794792	0.873972	0.790339	0.873130	0.789311
2018	0.827993	0.686107	0.823211	0.695181	0.824918	0.707269
2019	0.835690	0.720854	0.843875	0.728223	0.848048	0.729805
2020	0.855907	0.752735	0.850820	0.751282	0.851442	0.754871
2000-2020	0.836670	0.719687	0.839329	0.730016	0.840485	0.734478

4.1 Model selection

The experimental protocol consists of two main parts:

- The performance metric: a function computed on the true (ground-truth) labels and the labels predicted by the model to assess how good its performance is. In our study, we focused mainly on the test accuracy (score) defined as the ratio of correctly predicted labels.

- The validation protocol: generally, two protocols are used to evaluate machine learning methods: K-Fold Cross Validation and Train-Test-Splitting. In our study, we used the second approach, because the first one is used on small datasets to increase the confidence of the results. The Train-Test protocol requires the definition of a Train-Test-Ratio value that indicates the percentages of samples in each of the two folds (train/test), in our study we used a Train-Test-Ratio of 70%-30%.

As stated previously, we trained and tested a specific model for each year in the studied period (2000-2020).

Since we ourselves collected the dataset, we thought we wanted to experiment with different test set sizes to set a benchmark for future researchers who may work on this data.

K Nearest Neighbors (KNN)

The first algorithm we experimented is the K Nearest

Neighbors (KNN) approach, the Figure 7 demonstrates and summarizes the obtained results in terms of the accuracy (score) performance metric.

As we can clearly see from Table 2, the best results in terms of the test accuracy (score) are obtained in the year 2012, with an accuracy of 82% (an error rate of 18%). We can also see that the accuracy of the classifier trained on all the data (2000-2020) is around 72%. The average test score for the individual year is around 73%.

The parameters value Number of neighbors= 3, weights ='uniform', algorithm='auto'.

Decision trees

The second algorithm we experimented is the decision tree model. The scikit-learn library's implementation of the decision tree algorithm provides a long list of hyper-parameters that can be set manually by the developer or left to their default values. The most important hyper-parameter is the max-depth parameter that limits the depth (number nested feature tests) of the tree.

In our experiment we set the max-depth hyper-parameter to a value of 21.

The Figures 8 and Table 3 illustrate and summarize the obtained results in terms of the accuracy performance metric for all years in the period 2000-2020.



Figure 8. The accuracy of the decision tree classifier for different years in the period 2000-2020



Figure 9. The accuracy of the random forest classifier for different years in the period 2000-2020

Table 3. The accuracy of decision tree classifier

Year	Test Set Size: 30		Test Set Size: 20		Test Set Size: 10	
	Train Score	Test Score	Train Score	Train Score	Test Score	Train Score
2000	0.968354	0.705882	0.972222	0.717391	0.970443	0.695652
2001	0.928716	0.678062	0.903064	0.678062	0.854292	0.669515
2002	0.887622	0.699367	0.877701	0.694207	0.871859	0.694497
2003	0.890069	0.703418	0.820134	0.682692	0.858364	0.710256
2004	0.938005	0.758909	0.934944	0.773584	0.934661	0.786163
2005	0.927217	0.764621	0.922761	0.767475	0.902172	0.787446
2006	0.867233	0.741158	0.868075	0.738344	0.854236	0.745920
2007	0.901516	0.772727	0.883614	0.777868	0.895181	0.776326
2008	0.899115	0.798523	0.886726	0.799413	0.887224	0.790364
2009	0.893718	0.739652	0.896134	0.749844	0.893734	0.766749
2010	0.862194	0.722821	0.867119	0.741618	0.869554	0.745664
2011	0.869840	0.723892	0.865086	0.741023	0.881943	0.766412
2012	0.940208	0.850187	0.936446	0.855337	0.935861	0.856741
2013	0.810468	0.672719	0.851183	0.695711	0.819433	0.676190
2014	0.886935	0.785865	0.883054	0.780487	0.876944	0.766949
2015	0.836161	0.709810	0.751662	0.666048	0.850171	0.733003
2016	0.835321	0.736862	0.865197	0.765797	0.862775	0.775202
2017	0.912347	0.809181	0.911742	0.813463	0.907845	0.811921
2018	0.876053	0.688073	0.861322	0.715830	0.881530	0.766208
2019	0.903804	0.758012	0.902770	0.772125	0.875464	0.749303
2020	0.902814	0.767783	0.893535	0.773333	0.891574	0.781538
2000-2020	0.779506	0.693731	0.789485	0.699767	0.789507	0.711195

Table 4. The accuracy of random forest classifier

Year	Test Set Size: 30		Test Set Size: 20		Test Set Size: 10	
	Train Score	Test Score	Train Score	Test Score	Train Score	Test Score
2000	0.968354	0.676470	0.972222	0.760869	0.970443	0.608695
2001	0.936863	0.701804	0.930862	0.712250	0.926512	0.749287
2002	0.903365	0.722784	0.897649	0.731244	0.892970	0.721062
2003	0.910406	0.735470	0.905097	0.720512	0.898831	0.746153
2004	0.942048	0.780922	0.938482	0.802672	0.936582	0.819182
2005	0.934964	0.783166	0.929896	0.791012	0.926906	0.791726
2006	0.907046	0.775359	0.904081	0.784382	0.900362	0.784382
2007	0.904550	0.782798	0.900765	0.783993	0.897086	0.782040
2008	0.901163	0.810247	0.897312	0.813090	0.894173	0.811848
2009	0.904364	0.762831	0.900481	0.780881	0.898150	0.787841
2010	0.880039	0.745181	0.874638	0.759537	0.873281	0.765317
2011	0.909150	0.753948	0.904070	0.771581	0.900628	0.8
2012	0.940208	0.858146	0.936446	0.869382	0.935861	0.863764
2013	0.862424	0.712210	0.860378	0.706603	0.855456	0.697959
2014	0.886935	0.773851	0.883320	0.779427	0.879773	0.802966
2015	0.868151	0.732893	0.865316	0.730364	0.860068	0.749072
2016	0.870775	0.771638	0.868386	0.772753	0.866512	0.780996
2017	0.912347	0.818773	0.911742	0.824768	0.910129	0.825282
2018	0.896571	0.725425	0.890828	0.747295	0.886338	0.754420
2019	0.907388	0.776590	0.905035	0.786062	0.902881	0.795264
2020	0.903400	0.786251	0.898794	0.794358	0.898415	0.797948
2000-2020	0.899495	0.763624	0.895470	0.771494	0.892516	0.780916

Table 5. The accuracy of the 2012 obtained model for different years of the study

Year	Test Score	Year	Test Score
2000	0.485294	2011	0.503821
2001	0.494777	2012	0.858146
2002	0.533544	2013	0.519292
2003	0.511111	2014	0.520141
2004	0.511006	2015	0.510717
2005	0.480266	2016	0.49459
2006	0.5274	2017	0.492634
2007	0.515787	2018	0.519004
2008	0.528875	2019	0.488621
2009	0.463162	2020	0.48119
2010	0.51542	2000-2020	0.52555

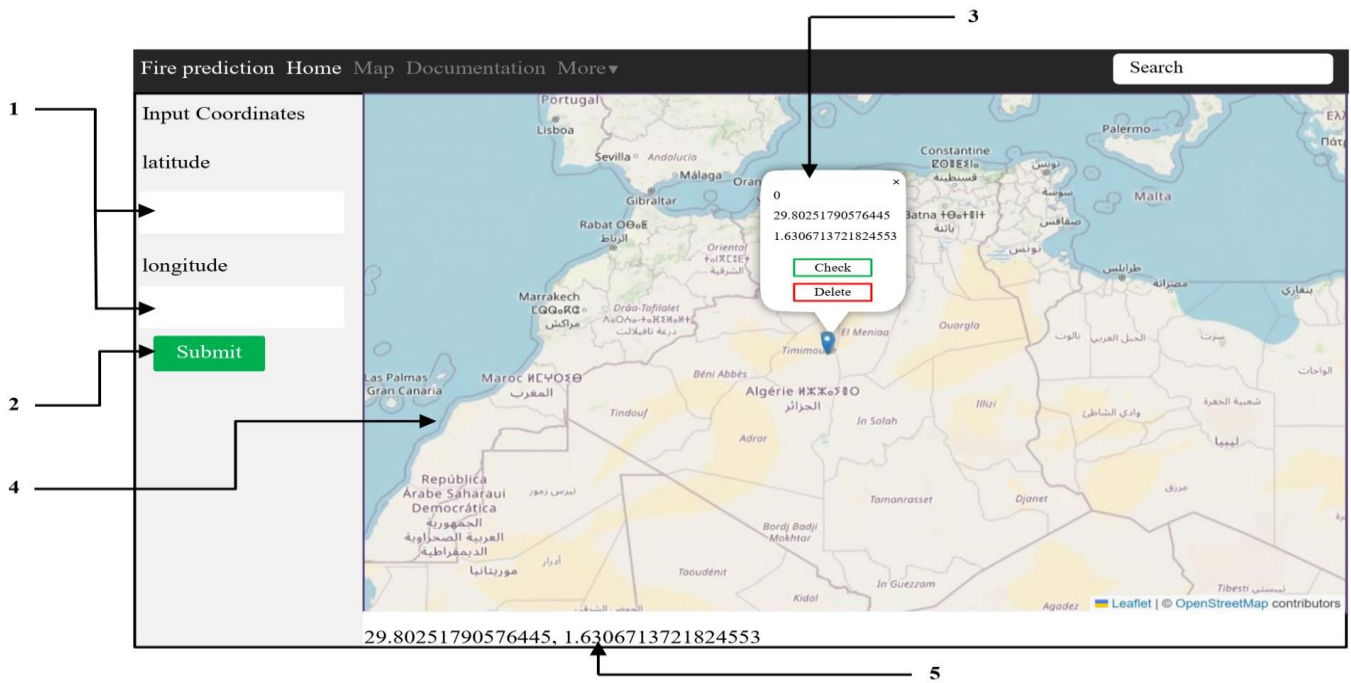


Figure 10. The main page of our website (map page)

We can see that the best results in terms of the test accuracy (score) are obtained in the year 2012, with an accuracy of 85% (an error rate of 15%). We can also see that the accuracy of the classifier when trained on all the data (2000-2020) is around 69%. The average test accuracy for the individual years is around 74%.

The parameters value: max depth=21, random state=33.

Random Forest

The last machine learning approach we tested was the Random Forest Classifier. The random forest classifier shares a similar hyper-parameter ‘max-depth’ with the decision tree classifier. Additionally, the random forest classifier has another hyper-parameter called ‘number of estimators’ ($n_estimators$ in the implementation of scikit-learn) that specifies how many decision trees the forest has.

In our experiment, we set the max-depth hyper-parameter to a value of 21 and the $n_estimator$ hyper-parameter to 600.

We can infer from Table 4 that the best results in terms of the test accuracy (score) are obtained in the year 2012, with an accuracy of 87% (an error rate of 14%). We can also see that the accuracy of the classifier when trained on all the data (2000-2020) is around 76% (see Figure 9). The average test accuracy for the individual years is around 76%.

The parameters value: random_state=42, number of jobs=-1, max depth=21, number of estimators=600, oob score=True, bootstrap=True.

Because random forest algorithm gives the best score we apply the model that we were created using the data of the year 2012 on all other years and the results presented in the following table (Table 5).

We also evaluated the models based on the processing time needed to fit each one of them. We run three experiments and record the processing time of each model then average the results of the 5 experiments. The obtained results are shown in Table 6.

By comparing the three types of models using the best performance (max accuracy) which is obtained in the year 2012 for all models, we can see that the random forest classifier outperforms all the other models with an accuracy of

86% followed by the Decision Tree classifier with an accuracy of 85% while the last place is occupied by the KNN classifier with an accuracy of 82%.

Table 6. Models evaluation based on the processing time

Model	Time in Seconds					Average Time
Experiment	1	2	3	4	5	
KNN	1.13	1.00	1.01	1.00	1.02	1.03
Decision Tree	1.03	1.11	1.06	1.07	1.04	1.06
Random Forest	3.58	3.63	3.59	3.61	3.56	3.59

A Similar conclusion can be deduced from the comparison using the average test accuracy on the individual years from 2000-2020. The random forest classifier has the highest score: 76% followed by the Decision Tree classifier with an accuracy of 74% while the last place is occupied by the KNN classifier with an accuracy of 73%.

However, the test score obtained when training on all the samples from all the years reveals a slightly different conclusion concerning the KNN and Decision Tree classifiers, in which KNN outperforms the Decision Tree classifier with an accuracy of 72% versus 69% for the Decision Tree classifier. Nevertheless, the Random Forest classifier still outperforms both other classifiers with an accuracy of 76%.

Clearly, we can conclude that the random forest classifier is the best choice for our solution; we attribute this superiority to its use of the ensemble learning paradigm, which largely reduces its degree of overfitting.

This conclusion leads us to invest more time in the future test on this model and other models of the ensemble learning family.

In order to enable different end users (authorities or civilians) to check for possible future forest fires, we deployed our selected machine learning model to a website we built.

The backend of the website was developed using the framework DJANGO available in the open-source repository of the programming language Python.

The main page displays a map of the entire world from

which the user can select the geographical area he wants to check for possible forest fires. The figure below (Figure 10) shows the different parts of this page.

The different parts (numbered in the figure) are explained in the following list:

(1) Two input fields that enable the user to manually type the geographical coordinates (latitude and longitude) of the area he wants to check for forest fires.

(2) The submit button that allows the user to send the data (latitude and longitude) to our server to check for forest fires.

(3) A container that displays the id of the selected coordinates, the latitude, and the longitude.

(4) A map that shows the different countries and regions of the world. It allows the user to select a geographical area to check for fires in a faster and more user-friendly manner by clicking on it with the mouse or touching it on mobile and tablet devices.

(5) Shows the latitude and longitude of the location where the mouse is located.

The website we built provides a simple interface to end users (local authorities or civilians) that allows them to test our models and predict potential forest fires.

5. CONCLUSIONS

Forest fire prediction systems often use a large number of monitored features that make them complicated and strenuous to implement in developing countries. This paper presented a prediction system based only on weather data using machine learning algorithms. Three different models were tested and compared in terms of prediction accuracy. The random forest classifier yields the best experimental results, which is selected to be deployed on our website. The website we built provides a simple interface to end users (local authorities or civilians) that allows them to test our model and predict potential forest fires. This work can save the forests from being burned by predicting the fire before it happens, thus saving people, animals, and all living beings. Added a notification system that alerts local authorities when a forest fire is likely to start near the area they are supervising, and improving the accuracy of our model for different fire seasons and study areas will be further investigated in the future.

REFERENCES

[1] Scott, A.C., Chaloner, W.G., Belcher, C.M., Roos, C.I. (2016). The interaction of fire and mankind: Introduction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1696): 20150162. <https://doi.org/10.1098/rstb.2015.0162>

[2] (2020). Global forest resources assessment 2020. Food and Agriculture Organization of the United Nations, pp. 184. <https://www.fao.org/documents/card/en?details=ca9825en>

[3] Jhariya, M.K., Raj, A. (2014). Effects of wildfires on flora, fauna and physico-chemical properties of soil-An overview. *Journal of Applied and Natural Science*, 6(2): 887-897. <https://doi.org/10.31018/jans.v6i2.550>

[4] Darmawan, M., Aniya, M., Tsuyuki, S. (2001). Forest fire hazard model using remote sensing and geographic information systems: Toward understanding of land and

forest degradation in lowland areas of East Kalimantan, Indonesia. In 22nd Asian Conference on Remote Sensing.

[5] Global forest watch fire in 2019. <https://fires.globalforestwatch.org/home/>, accessed on May 2020.

[6] Bachmann, A., Allgöwer, B. (1999). The need for a consistent wildfire risk terminology. In the Joint Fire Science Conference and Workshop: Crossing the Millennium: Integrating Spatial Technologies and Ecological Principles for a New Age in Fire Management, pp. 67-77.

[7] Curt, T., Aini, A., Dupire, S. (2020). Fire activity in Mediterranean forests (The Algerian case). *Fire*, 3(4): 58. <https://doi.org/10.3390/fire3040058>

[8] (2021). Algerian wildfires still raging, death toll hits 90 including 33 soldiers. *Africa News*. <https://www.africanews.com/2021/08/15/algerian-wildfires-still-raging-death-toll-hits-90-including-33-soldiers/>

[9] (2022). Feux de forêt: Plus 100.000 hectares ravagés dans 21 wilayas durant l'été 2021. *www.aps.dz* (in French). <https://www.aps.dz/economie/139524-feux-de-foret-plus-100-000-hectares-ravages-dans-21-wilayas-durant-l-ete-2021>

[10] (2021). Tizi-Ouzou/Incendies: Plus de 5.100 ha d'arbres fruitiers et plus de 19.100 animaux d'élevage brûlés. *www.aps.dz* (in French). 22 August 2021. <https://www.aps.dz/regions/126427-tizi-ouzou-incendies-plus-de-5-100-ha-d-arbres-fruitiers-et-plus-de-19-100-animaux-d-elevage-brules>

[11] "Incendies de Tizi-Ouzou: Plus de 1700 habitations brûlées expertisées par le CTC. *www.aps.dz* (in French). <https://www.aps.dz/regions/126787-incendies-de-tizi-ouzou-plus-de-1700-habitations-brulees-expertisees-par-le-ctc>

[12] Meddour-Sahar, O. (2014). Les feux de forêts en Algérie: Analyse du Risque, étude des Causes, évaluation du Dispositif de Défense et des Politiques de Gestion. Ph.D. Thesis, University of Tizi Ouzou, Tizi-Ouzou, Algeria. <https://dspace.ummto.dz/handle/ummto/21588>

[13] Chergui, B., Fahd, S., Santos, X., Pausas, J.G. (2018). Socioeconomic factors drive fire-regime variability in the Mediterranean Basin. *Ecosystems*, 21: 619-628. <https://doi.org/10.1007/s10021-017-0172-6>

[14] (2012). GFN Report of Mediterranean Ecological Footprint Trends. Plan Bleu Edition: Paris, France. <https://planbleu.org/en/publications/mediterranean-ecological-footprint-trends/>

[15] Özbayoğlu, A.M., Bozer, R. (2012). Estimation of the burned area in forest fires using computational intelligence techniques. *Procedia Computer Science*, 12: 282-287. <https://doi.org/10.1016/j.procs.2012.09.070>

[16] Satir, O., Berberoglu, S., Donmez, C. (2016). Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. *Geomatics, Natural Hazards and Risk*, 7(5): 1645-1658. <https://doi.org/10.1080/19475705.2015.1084541>

[17] Tonini, M., D'Andrea, M., Biondi, G., Degli Esposti, S., Trucchia, A., Fiorucci, P. (2020). A machine learning-based approach for wildfire susceptibility mapping. The case study of the Liguria region in Italy. *Geosciences*, 10(3): 105. <https://doi.org/10.3390/geosciences10030105>

[18] Liang, H., Zhang, M., Wang, H. (2019). A neural

- network model for wildfire scale prediction using meteorological factors. *IEEE Access*, 7: 176746-176755. <https://doi.org/10.1109/ACCESS.2019.2957837>
- [19] Joshi, J., Sukumar, R. (2021). Improving prediction and assessment of global fires using multilayer neural networks. *Scientific Reports*, 11(1): 3295. <https://doi.org/10.1038/s41598-021-81233-4>
- [20] de Dios, V.R., Nolan, R.H. (2021). Some challenges for forest fire risk predictions in the 21st century. *Forests*, 12(4): 469. <https://doi.org/10.3390/f12040469>
- [21] El-Bouhissi, M., Miloua, H., Bachir-Bouiadjra, S.E., Soummar, A. (2022). Fire analysis and prediction in the Zid-Elmoumen forestry (Northwest Algeria). *Ukrainian Journal of Ecology*, 12(3): 46-56. https://doi.org/10.15421/2022_353
- [22] Abid, F., Izeboudjen, N. (2020). Predicting forest fire in Algeria using data mining techniques: Case study of the decision tree algorithm. In *Advanced Intelligent Systems for Sustainable Development*, 4: 363-370. https://doi.org/10.1007/978-3-030-36674-2_37
- [23] Merabet, M., Kourtiche, A. (2022). Embedded ANN-based forest fire prediction case study of Algeria. *International Journal of Distributed Artificial Intelligence (I.J.D.A.I.)*, 14(1): 1-18. <https://doi.org/10.4018/IJDAI.291085>
- [24] Zaidi, A. (2023). Predicting wildfires in Algerian forests using machine learning models. *Heliyon*, 9(7): E18064. <https://doi.org/10.1016/j.heliyon.2023.e18064>