Projected Escalation of Wildfire Severity Due to Climate Change in Maryland, USA: A Machine Learning and Climate Simulation Approach.

Abstract

With the increase in climate change, wildfires are becoming more hazardous for humans as well as the ecosystem. They are growing in severity and spread causing economic, health, and ecological damages. Therefore predicting and projecting wildfires to take preventive measures, has become an essential task. Computational methods such as Machine Learning (ML) and Climate Simulations have been used to project future wildfire severity by researchers. This paper uses a combination of both for a variety of tasks. Firstly, this paper used a Random Forest (RF) ML model to conduct a feature importance test to better understand which climate factors most influence the severity of a wildfire. Secondly, to simulate potential future climate scenarios, the Climate Explorer Toolkit was used to acquire anticipated climate states for distinct counties or regions in the state of Maryland, USA. Using these forthcoming climate states as predictive inputs for the Machine Learning models, this study used RF, K Nearest Neighbors (KNN), and Support Vector Machine (SVM) models to project the wildfire severity of the state of Maryland, USA by approximately 10 years in the future. All models were trained using historical wildfire severity data correlated with specific climate variables of temperature, precipitation, wind speed, humidity, and remoteness (non-dimensional distance to the closest city). The results showed that remoteness, humidity, wind speed, temperature, and precipitation, in that order, were the most important factors in determining the severity of the wildfire. There was also agreement among the ML models that the severity of wildfires is projected to increase uniformly in Maryland. Except for highly urbanized counties, the most likely wildfire type to occur in 2030-2045 will be the two most severe wildfire categories possible. This projection emphasizes the concerning pattern of increasing wildfire severity in Maryland, a region that has historically experienced fewer instances of such hazardous wildfires. As climate change continues, wildfires are expanding beyond their previous limits, highlighting the importance of taking preventative measures to address their harmful effects.

1. Introduction

Wildfires, also known as forest fires or bushfires, are uncontrolled fires that spread rapidly through vegetation. They are a natural part of many ecosystems and have been occurring for millions of years. Wildfires regenerate their ecosystem by allowing nutrients to re-enter the soil and create new niches for different animals and plants. By removing dead organic matter, wildfires allow for new and different organisms to thrive. Wildfires also allow for nutrients stored in the burned organic matter to return to the soil more quickly. This allows for healthier soil as well. Therefore, wildfires allow for a healthier and more diverse ecosystem. Small wildfires are especially helpful to the ecosystem. They remove excessive shrubs and vegetation which may serve as fuels for larger wildfires. By doing so, small wildfires reduce the chances of a larger or mega fire [1,2].

However, wildfires have been getting severe with multiple megafires and a longer wildfire season. Wildfires have more than doubled since 1984 [3]. More severe wildfires also damage healthy and permanent trees which harm the forests instead of helping them. This results in no regrowth of trees and, consequently, the clearing of forests [4,5]. These huge wildfires severely harm the health and economy of the region. Wildfire smoke can cause various health issues ranging from asthma to premature death [6]. Wildfires also result in property damage. The cost of these property damages and the cost to put out the wildfires sum up to a considerable amount of money. In 2017, the Thomas Fire in Southern California burned 281,893 acres of land and cost the US government 177 million dollars to fight. In 2018, the Camps Fire in Northern California burned 153,336 acres of land and cost 16.5 billion dollars to fight. The economic costs of fighting wildfires in the USA alone keep increasing. In 2020, it cost 3.5 billion dollars and in 2021, the government spent another 4.4 billion dollars. Overall, wildfires take up a huge part of government expenses and this cost is also increasing due to the worsening climate, making wildfires more frequent and dangerous [5,7,8]. The effects of such severe wildfires can even be seen in places where wildfires are not prevalent. In 2023, the Canadian Wildfire's smoke spread to the East Coast of the USA. The smoke engulfed many cities such as New York and Cincinnati, reducing vision and worsening air conditions.

Wildfires have long been known to be affected by key environmental variables. Climate variables such as temperature, relative humidity, wind speed, precipitation, and drought determine the intensity and spread of wildfires and contribute to the formation of fire weather [9]. It can be seen that hotter temperatures, low humidity and precipitation, and prolonged periods of drought are directly correlated to the chances of a severe wildfire. These fire weather characteristics are also caused by climate change [10]. Climate change is a phenomenon caused by the emission of excessive greenhouse gasses, such as Carbon dioxide, Methane, and Nitrous Oxide, which blanket Earth and trap heat. This heat leads to a hotter climate and ideal fire weather conditions. Wildfires also release greenhouse gasses such as carbon, methane, and nitrous oxide stored in trees into the atmosphere. As these gasses contribute to climate change, a vicious cycle of climate-fire feedback is also starting to appear [11]. Therefore, predicting and preventing wildfires under changing environmental conditions has become very important.

There are multiple approaches to predicting wildfire trends in different geographical regions. One approach is the use of fire weather indices such as the Keetch-Byram Drought Index (KBDI) and Fire Weather Index (FWI) [12]. These indices use atmospheric and soil quality to give a unitless measurement of the probability of a wildfire in that region. These models can be coupled with climate simulation models such as the General Circulation Model (GCM) to determine the wildfire trends of the future. GCMs provide simulated future climates that reflect the responses of the global climate system to GHG emissions scenarios [13]. Different ML models can also be trained on previous wildfire-climate data to categorize future wildfires or quantify the probability of wildfires occurring in a region. These models can then be used to understand how wildfire behavior and intensity will change under changing climatic conditions [14]. Future climate properties and ML models or fire weather indices can then be used together to understand the wildfire properties of the future.

This paper aims to predict the most likely category of wildfires to occur based on future climate using both ML and climate simulation models. The categories are ordered from A to G with A being the least severe wildfire (less than 0.25 acre) and G being the most severe (5000+ acre). A table detailing the characteristics of each type of wildfire is provided in Appendix A as Table 1. This paper proposes a two-part research plan. The first part is the creation and training of a classifier model on past wildfire-climate data. The second part will be to use climate simulation models to predict the future states of climate variables. The classifier model will then be used to get the most likely category of wildfire to occur under the inputted climate scenario. Due to data and time limitations, this paper will focus only on predicting the future climate change-induced wildfire category in one of the states on the East Coast of the USA. Wildfires are already affecting the East Coast as seen in the case of the 2023 Canadian wildfires. This study aims to quantify the effect of climate change on wildfires on the East Coast of the USA.

2. Related Work

As climate change increases, our environment becomes more wildfire-prone. While there is no standard way of predicting wildfires [15] to reduce their impact and harm, researchers are exploring multiple avenues. To accurately predict and assess wildfire risk, scientists and policymakers use a variety of methods including but not limited to climate modeling, vegetation modeling, machine learning, drought and fire weather indexing, and data mining. These methods help researchers explore and understand the hidden causes and subtle effects of wildfires. They also inform policymakers of any new understanding such that they are able to revise or create better policies. Among these methods, the use of Machine learning (ML) to improve our understanding of wildfire has been met with success and excitement [16,17]. As this paper also implements Machine Learning algorithms, ML will be one of the primary focuses.

2.1 ML Models

Many different ML algorithms have been explored in the context of wildfire science. An environmental review paper by Jain et al. provides an extensive review and reports that ML models have been used in many different problem settings: fuel characterization, fire-climate interaction, fire risk assessment, fire behavior prediction, fire effects, and fire management [12]. Multiple different ML models were also used for these tasks. Jain et al. report that the most frequently used models were Random Forests (RF), Maximum Entropy (MaxEnt), Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machines (SVM), and genetic algorithms [12]. As wildfires are more extensively documented and simulated, the available data also increases. Monte Carlo Simulations are receiving increasing attention as a method of acquiring more data [18]. This helps ML models train and project more accurate results.

As wildfire science is such a huge field with many factors contributing to the formation of one, ML data science and models have grown to include all relevant variables. These variables include topographical variables: slope inclination, altitude, plane curvature, topographic position index (TPI), topographic roughness index (TRI), topographic wetness index (TWI); vegetation variables: types of fuels, amount of vegetation, soil type; atmospheric variables: mean temperature, precipitation, humidity, drought duration, wind speed and direction; and human-environment variables: proximity of settlement, type of land use and human settlement densities [10,15,19,20,21,22]. However, many issues such as the risk of overfitting and increasing space and time complexity arise with high dimensional data. That is why dimensionality reduction, the systematic removal of features, has become very important. Dimensionality reduction also increases the accuracy of a model by removing potential noise, or insignificant features from the data set. Researchers are able to do so by conducting feature selection using different models. Feature selection also reveals which predictive variables have a significant impact on the target variable, a variable related to wildfires in this case [23].

2.2 Feature Importance and Selection

There are multiple ways of conducting feature selection such as forward selection, backward elimination, correlation tests, and other ML models such as decision trees and logistic regression [24]. One study done by Pourtaghi et al. used three different models: boosted regression tree (BRT), generalized additive model (GAM), and random forest (RF), to perform feature selection in order to reveal the variables that contribute more to forest fire occurrence. Their results showed that annual rainfall, slope degree, distance to roads, land use, and annual temperature were the most effective factors in forest fire occurrence [25]. More specific studies have also been conducted to reveal how a domain of variables such as topographical and atmospheric contribute to wildfires. One such study was conducted by Rodrigues et al. to model human-caused wildfires. This study also compared different models, namely BRT, RF, and SVM to another more traditional model called the Logistic Regression (LR) model by comparing their AUC (Area under the Curve) score. This score gives a measure of how well a model is able to classify its target variable using [26]. All models return a better AUC score than LR with the best AUC score of 0.746 and 0.730 for RF and BRT respectively. The aggregate decisions reached by its models identified human density and population as making the most significant impact on wildfire ignition risk [21]. Model comparison studies are also very common in this growing field of ML wildfire data science. A study by Jaafari et al. compared different types of Tree-based models, Alternating Decision Tree (ADT), Classification and Regression tree (CART), Functional tree (FT), Logistic model tree (LMT), and Naïve Bayes tree (NBT), in their ability to assess wildfire risk in spatial intervals [15]. Another study compared different types of Artificial Neural Networks in their ability to predict wildfire on a global scale [20]. The main threshold of comparison for both studies was the AUC score of the different models.

2.3 Climate Simulation Models

ML models and climate simulations have been used to project wildfire risk under climate change. Park et al. used past wildfire data to train a BRT model to predict wildfire risk and severity as a function of climatic variables in the western USA. The climatic projection from General Circulation Models (GCM), a type of Climate Model, was then used as inputs to the model to get the future wildfire states in different temporal intervals [27]. Moritz et al. used the Maximum Entropy (MaxEnt) model to project future fire probability globally in spatial intervals as well as temporal. By incorporating climate change's effect on different biomes, Moritz et al. used the MaxEnt model to project wildfire probabilities across different biomes [28]. Another paper by Stralberg et al. used a fire simulation model called Burn-P3 in combination with future vegetation projections to understand the future wildfire scenarios that are likely to occur [29]. There is a recurrent pattern of implementing a two-part method being observed in such studies. The first plan is to train a model on past data and the second is to project future climatic or environmental states by the use of climate models or statistical models to serve as inputs to the trained wildfire models. This paper adopts this method to project the wildfire severity in one state on the East Coast, USA.

There are also many different ways of forecasting the future climatic and environmental states. Time series projection by means of statistical models has been explored extensively. One of the most popular models is the Autoregressive Moving Average (ARIMA) model. While the ARIMA model is not accurate in projecting long-term climatic states, it has high accuracy in the short term and it also portrays the seasonal variations of variables such as temperature and precipitation [30]. Other statistical models such as ETS (Error Trend Seasonality Model) and

Holt-Winters (HW) were also compared to ARIMA models, and as expected they were found to have higher accuracy in the long term [31]. Akaike Information Criteria (AIC) and Root Mean Square Error (RMSE) were used to compare the results of these models [31]. ML models can also be used to forecast future climate states. Exponential models, Decision Tree Regression Models, and Long-Short Term Memory models are a few examples of ML models that can effectively forecast climatic variables such as temperature, precipitation, and humidity [32,33,34]. In addition to these models, climate and atmospheric modeling using physical, derived formulas is also another method of forecasting climate. This method has an edge over statistical and ML models as it is not as heavily dependent on past data [35]. Therefore, climate models such as the GCM have received increasing attention in the wake of climate change. This study uses one such model called the Climate Explorer Toolkit in order to project the future climate variable states. Using these projected climate states, the wildfire severity is projected.

3. Wildfire-Climate ML Modeling Methods

3.1 Data

The data was acquired from a Machine Learning community website called Kaggle [36]. As noted on the website, this data was aggregated from multiple different sources [37,38,39,40]. The given data has many features describing the location of each wildfire, the category, size, and area burned of each wildfire, the temperature, wind, humidity, precipitation, and a measure of remoteness at the time of the wildfire, among other features. The dataset contains 50,000 wildfire occurrences in the USA for the years 1992 to 2015. This data set was chosen because it gave the climate states such as temperature, precipitation, wind speed, and humidity at the times of wildfire. This helped the models understand the relationship between climate variables and wildfire.

3.2 Data Preprocessing

First, the filtration of unnecessary features was conducted. The features which were removed gave logistical information about the wildfire such as how long it took to put out the fire, and which station the fire was discovered in. These data points are irrelevant to this research which is more focused on modeling wildfire-climate relations. The remaining features are shown in Table 2 in Appendix A.

There were also numerous occurrences of missing data. These data points were marked as -1 and 0s. For example, Figure 1 demonstrates how each case of missing data may be represented in the data set. Data point 3 has a 0.000 value at the 'Temp_cont' feature even though the temperature observed over the last 30, 15, and 7-day intervals was around 18 degrees C. This unnatural drop is characteristic of missing data on the day of the fire. Data point 4 is filled with -1 values in all features associated with temperature. This is also characteristic of missing data. These rows were filtered out and a new data set of the remaining 7284 wildfire occurrences was used.

	fire_size	fire_size_class	stat_cause_descr	Vegetation	Temp_pre_30	Temp_pre_15	Temp_pre_7	Temp_cont
0	10.0	1	8	12	24.480974	24.716923	24.902597	24.527961
1	3.0	0	0	15	7.553433	7.010000	0.343529	10.448298
2	60.0	1	0	16	4.971930	5.782766	5.558750	13.696600
3	1.0	0	3	0	16.275967	18.996181	18.142564	0.000000
4	2.0	0	7	12	-1.000000	-1.000000	-1.000000	-1.000000

Figure 1: Example of missing data shown in rows 3 and 4

3.3 Data Normalization

In order to reduce the time complexity and help our models converge faster, each feature was normalized using the min-max normalization method [41]. Min-max normalization is a statistical technique that rescales numerical values into a range between 0 and 1. The minimum value of each feature is subtracted from each data point and then divided by the range of the feature. This technique preserves the relative order and distance of the data points and also helps ML models achieve convergence quickly. A formulaic description of the min-max normalization method is shown in Appendix A, Figure 1. A snapshot of the dataset prior to normalization is shown in Figure 2.

Temp_pre_30	Temp_pre_15	Temp_pre_7	Temp_cont
4.971930	5.782766	5.558750	13.696600
16.851939	16.997783	20.434783	11.985560
8.410983	9.007193	8.209000	14.328047
5.181473	6.003609	7.773430	11.375329
24.566175	25.548871	25.232870	25.610425
...	...	...	...
21.443127	22.495833	22.608523	17.337247
3.390566	3.966190	5.167179	8.710361
4.010314	4.852619	6.844324	8.686747
15.546194	12.890633	10.734328	12.713333
23.350785	23.093421	24.573373	24.465102

After min-max normalization →

Temp_pre_30	Temp_pre_15	Temp_pre_7	Temp_cont
0.303748	0.269907	0.323068	0.524050
0.598885	0.560861	0.678577	0.484270
0.389185	0.353559	0.386403	0.538731
0.308954	0.275636	0.375994	0.470082
0.790531	0.782703	0.793242	0.801040
...	...	...	...
0.712945	0.703498	0.730525	0.608693
0.264462	0.222779	0.313710	0.408123
0.279859	0.245776	0.353790	0.407574
0.566446	0.454308	0.446754	0.501190
0.760337	0.719001	0.777481	0.774412

Figure 2: A snapshot of the dataset prior to and after normalization

3.4 Data Exploration and Climate Feature Selection

The initial data exploration consisted of understanding how the classes of wildfires differed in temperature, wind speed, precipitation, humidity, remoteness, and vegetation type. How wildfires caused by different factors (lightning, arson, debris burning, etc.) differed in the above-mentioned variables was also explored. This was done by picking two variables from the set and creating 2-dimensional scatter plots. Example graphs are shown below in Figure 3. Feature exploration such as collinearity tests and importance was also conducted.

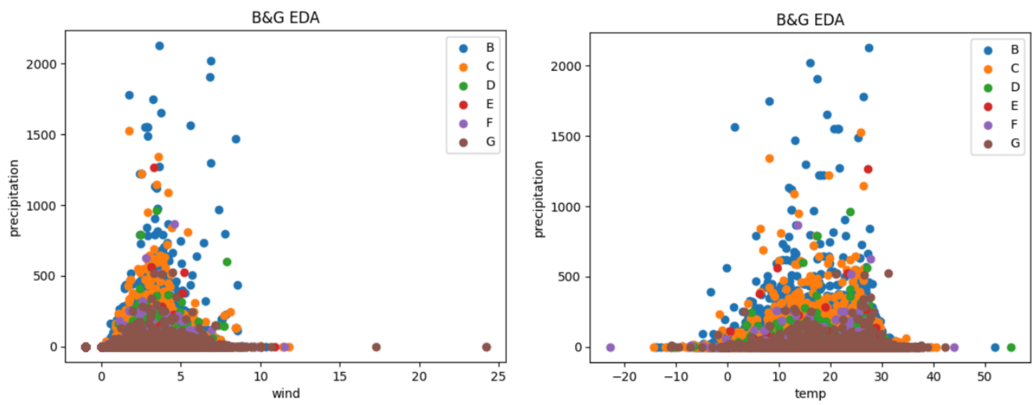


Figure 3; Example scatter plots used for initial exploration: On the left, wind speed over precipitation was explored for each category of wildfire; on the right, temperature over precipitation was explored for each category of wildfire.

3.4.1 Feature Selection

Feature selection is an essential part of any Machine Learning or data-oriented research [42,43,44]. Selecting important features with high correlation to the target variable has become an essential task in acquiring more accurate models. By selecting important features and removing redundant or unnecessary features, one can reduce the noise of a data set. Noise is meaningless data which makes it harder for models to understand the correlation between the predictive variables and the target variables.

3.4.2 Pearson's Correlation Coefficient

Pearson's Correlation Coefficient is a number between -1 and 1 which measures the strength of correlation between two variables. It is the most common way of measuring linear correlation. Values closer to 0 signify weaker correlation while values closer to -1 or 1 signify stronger correlation. Positive values signify direct correlation while negative values signify the opposite [45]. Pearson's Correlation Coefficient was used to understand which features were correlated with each other and can be removed to reduce the noise of the data set. By removing features with high correlation, we remove potentially redundant data from the data set.

3.4.3 Measuring Feature Importance

Feature importances were explored through the use of Random Forests (RF). RF is an ensemble tree model and the nuances of the model are described in section 3.5. RF, in addition to serving as a data mining model, can also measure feature importance. It does so by comparing features on their ability to decrease the impurity, or diversity, of a sample. Therefore, a feature that clearly distinguishes one category of the target variable from another will have a higher importance measure. RF is able to conduct this experimental research on its own through the use of multiple decision trees. RF has proven to be a very useful approach in doing so [46] and was therefore used in this study.

3.4.4 Climate Feature Selection

Climate feature selection was done mainly through the use of the Pearson Correlation Matrix and Random Forest Generated feature importance graphs.

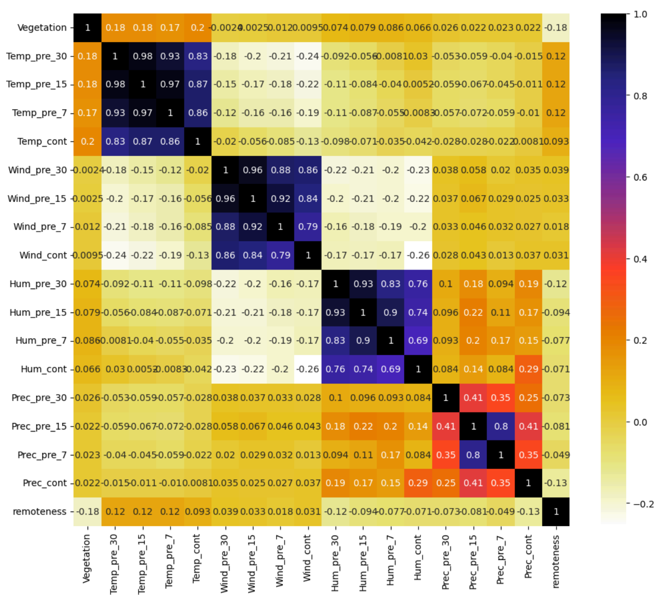


Figure 4: This Pearson Correlation Matrix spanning across all the climate features, shows how climate variables measured at different temporal intervals before the wildfire are very correlated with each other. This demonstrates the redundancy of using every temporal measurement.

As demonstrated by Figure 4, different temporal measurements of the same climate variable proved to be too correlated with each other. This signifies potential redundancy in the data which can hamper model training and forecasting. Therefore, different temporal measurements of the same climate variable were filtered out and the climate states on the day of the wildfire were kept. A Random Forest feature importance graph was then generated on the remaining features and is shown in Figure 5.

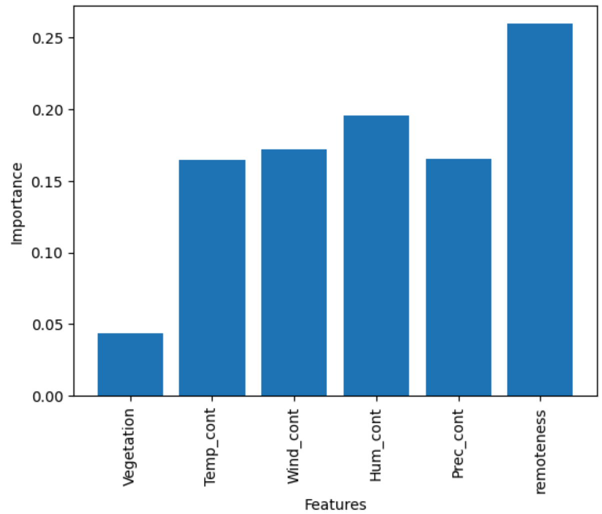


Figure 5: This graph ranks the remaining features on their ability to distinguish one category of wildfire from another. The Random Forest classifiers rank remoteness as the most important and vegetation as the least important.

Several other feature importance tests were run on specific caused wildfires and their feature importance is shown below. In all of their feature important graphs, vegetation remained the lowest ranked. Therefore, Vegetation was determined to be an unnecessary feature and was filtered out. The feature importance graphs are shown in Figure 6.

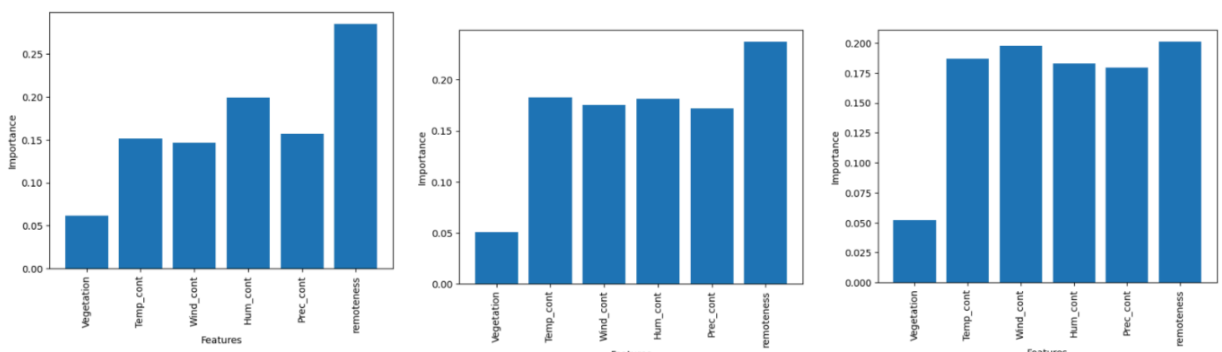


Figure 6: Feature importance graphs of only lightning, arson, and debris burning caused wildfire data, from left to right.

The final list of features used in this study is shown in Table 1. Apart from the features shown, 'fire_size_class' (category of wildfire) and 'stats_cause_descr' (cause of the wildfire) were also used for modeling. The 'fire_size_class' was chosen to be the target variable of all models. The other features, except for 'stats_cause_descr' were chosen to be the predictive variable. The data in the 'stats_cause_descr' attribute provides the reason behind each instance of a wildfire. While this attribute is not employed as a predictive factor, it was utilized to divide the wildfires dataset according to their causes. This categorization helps in comprehending the varying degrees of predictive influence exhibited by different variables, based on the wildfire's cause.

Feature Name	Feature Description
Temp_cont	temperature in deg C at the location of fire up to the day the fire occurred
Hum_cont	humidity in % at the location of fire up to the day the fire occurred
Wind_cont	wind in m/s at the location of fire up to the day the fire was contained
Prec_cont	precipitation in mm at the location of fire up to the day the fire was contained
remoteness	non-dimensional distance to closest city A higher remoteness factor signifies a closer distance while a lower number signifies a greater distance.

Table 1: Final list of features used for modeling

3.5 Model Overview

Three different models were used in this study: *k-nearest neighbor* (*KNN*), *support vector machines* (*SVM*), and *random forest* (*RF*). All these models were structured to model wildfire-climate relations. By understanding all of wildfire-climate relations based on past data, these models were able to predict the most likely wildfire category based on climate variables of temperature, precipitation, humidity, wind speed, and remoteness.

$$F(temperature,\ wind\ speed,\ humidity,\ precipitation,\ remoteness) \ = \ MLWC$$
$$MLWC\text{:}\ Most\ Likely\ Wildfire\ Category\ to\ Occur$$

Figure 7: A formulaic description of the structure of each model

K-nearest neighbor (KNN): KNN is a relatively simple ML algorithm that classifies a data point based on how its neighbors are classified. It is a supervised algorithm that uses the training data sets and their corresponding labels as a reference to new, test data. It can be used for both classification and regression problems [47]. Many different parameters can determine the performance of a *KNN* model.

SVM: SVM or *Support Vector Machine* creates a decision boundary through the training data to classify new data. It does so by creating a hyperplane or a set of hyperplanes to divide the classes into distinct planes or clusters. The most optimal decision boundaries are created by maximizing the distance to data points. This reduces the chances of misclassification. *SVMs* also use a technique called the kernel trick which transforms low-dimensional data to a higher dimension to create a better decision hyperplane [48]. *SVMs* are especially good at binary classification problems [49].

RF: A *random forest* is an ensemble model based on Decision Trees. A Decision Tree is a model that creates divisions or decisions on a given data set to reduce the impurity of the data. In other words, each decision that the tree makes aims to divide the data sets into their labeled categories. A *random forest* has a number of these decision trees and it aggregates their predictions to give its prediction [50].

3.6 Model Validation

K-Fold cross-validation score: K-fold cross-validation is a technique for evaluating predictive models. It involves splitting a dataset into k subsets, or fold, where k can be any integer. The model is trained and evaluated k times, using a different data fold as the validation set each time. Performance metrics from each fold are averaged to estimate the model's generalization performance. This technique can also be used to find the best parameters for the given model. The K-fold cross-validation scores were calculated for each model and are displayed in the Results section.

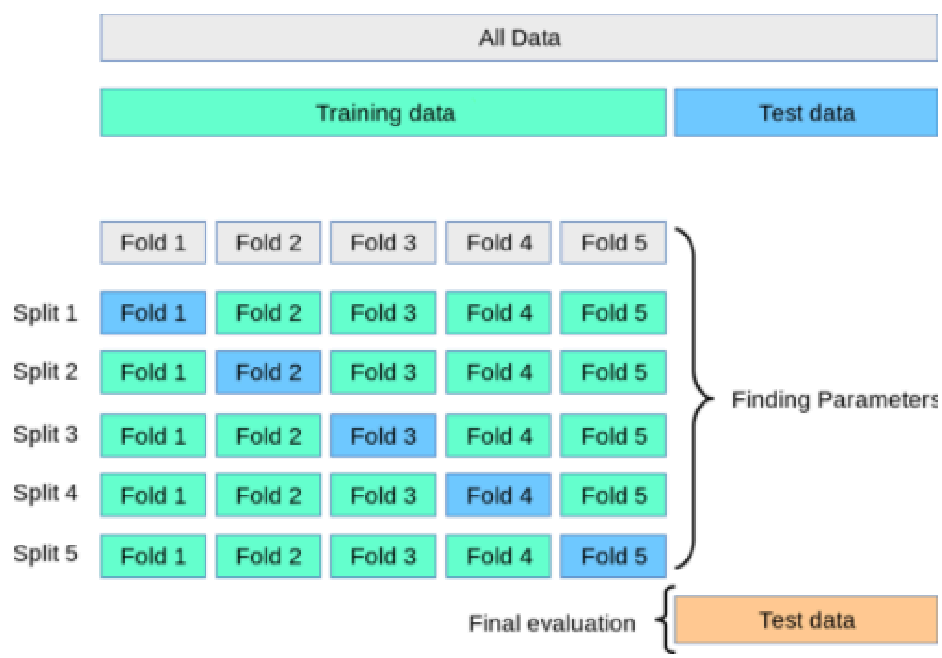


Figure 8: K-fold cross-validation description [51]

Testing data: The data set was split into training and testing sets in a 70 to 30 ratio. The testing data was also used to measure the performance of each model. The accuracies of each model are displayed in the results section.

4. Wildfire Projections Methods

4.1 Study Area

In order to best observe how the severity of wildfires will change on the East Coast of the USA, states with historically low wildfire frequency and risk were made into a subgroup. A region with historically low wildfire frequency would be the most ideal study area since the impact of climate change on wildfire severity will be the most prominent. The state of Maryland in the United States of America was randomly chosen from among these states as the study area. The latitude and longitude of Maryland are 39° 2 '44.72"N and 76° 38' 28.58"W, or 39.045755 and -76.641271. Maryland is located in the southeastern part of the United States. Surface wildfires have occurred in Maryland and Maryland Forest Services responded to approximately 123 wildfires which burn 1,780 acres of land annually [52]. Nonetheless, wildfire occurrences in Maryland are comparatively infrequent compared to states like California, known as one of the most wildfire-prone regions in the USA. For instance, in 2020, California faced a staggering 10,000 wildfires that consumed over 4 million acres of land. Given this context, this study seeks to predict how the intensity of wildfires in Maryland might alter in the coming years due to the impacts of climate change.

4.2 Data Acquisition and Creation

The data for the predictive variables (temperature, wind speed, humidity, precipitation, and remoteness) was collected using the Climate Explorer Toolkit built by a partnership led by the National Oceanic and Atmospheric Administration (NOAA) [53]. The Climate Explorer Toolkit is a mapping and graphing tool that provides historical and projected climate data for every county in the contiguous United States. Future climate states of all 23 counties of Maryland were recorded using the Climate Explorer toolkit. The daily maximum temperature and the daily precipitation for July averaged over time ranges of 2010 - 2040 were used as input data for temperature and precipitation predictive features. The projections using the higher emissions pathway were used to represent the worst-case scenario.

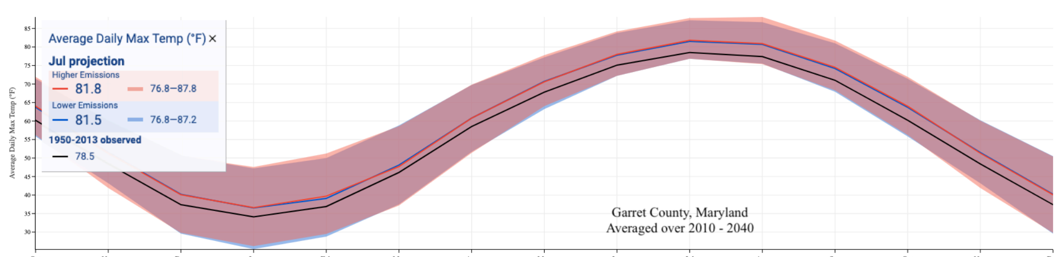


Figure 9: Daily Maximum Temperature averaged over the 2010 - 2040 time period graph for Garret County, one of the 23 counties of Maryland [53].

As wind speed and humidity forecasts were not available in the Climate Explorer Toolkit, wind speed, and humidity were held constant across all counties of Maryland. The average summer wind speed of 9 miles per hour was used as the wind speed constant and the average summer humidity percentage of 77% was used as the humidity constant [54,55,56].

The remoteness feature for each county was determined using the population density 2010 census [57]. As population density increased in urban areas, counties with a higher population density were given a higher remoteness number (i.e. 1 remoteness factor being cities and 0 remoteness factor being rural areas). The thresholds used to create the remoteness factor data for each county are summarized in Table 2. In this study, it was assumed that there would be minimal population migration within the state of Maryland. This assumption was made to maintain a consistent level of remoteness in the future, akin to the present circumstances.

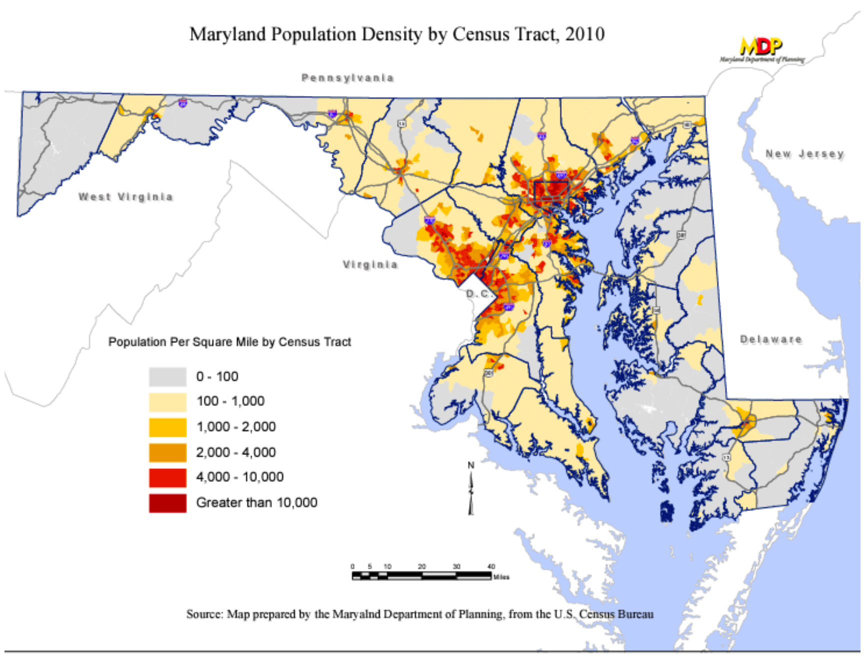


Figure 10: Maryland 2010 Census of population density [57]

Population per square mile	Remoteness factor
< 1000	0.09
1000 - 2000	0.2
2000 - 4000	0.4

4000 - 10,000	0.7
>10,000	0.99

Table 2: Population thresholds used to create the remoteness factor.

By this method, a data set containing future temperature, and precipitation states for each county in the state of Maryland was acquired. The current wind speed and humidity levels were constant across all counties. The current remoteness of each county was also held constant.

4.3 Model-Specific Data Transformation

The temperature data acquired from the Climate Explorer toolkit was in Fahrenheit. However, the training data used to create the model was in Celsius. Therefore, a scalar transformation on the temperature feature was applied to change the Fahrenheit data points to Celsius. The precipitation data acquired from the Climate Explorer toolkit was in inches. However, the training data used to create the model was in millimeters. Therefore, a scalar transformation on the precipitation feature was applied to change the inches of data points to millimeters. The min-max normalization was applied to all features of this projected data set. The min-max values acquired from the past wildfire occurrences data were used to create the models.

5. Results

5.1 Model Validation

The K-fold cross-validation score (k-score) and the testing accuracy score were recorded for each model. The data used were also varied by the cause of the wildfire and the scores obtained for each model were also recorded. These scores are recorded in Tables 3 and 4.

All wildfires		
	K-Fold Cross Validation Score	Testing Data Score
RF	69%±2	70.03%
KNN	67%±1	69.03%
SVM	66%±0	65.87%

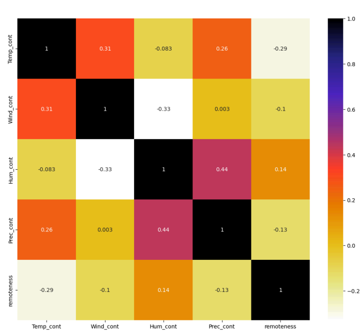
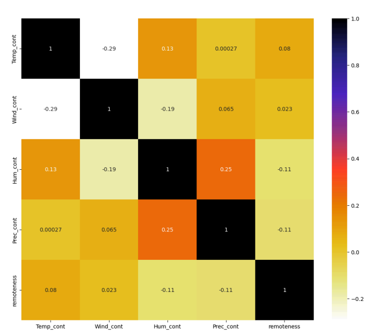
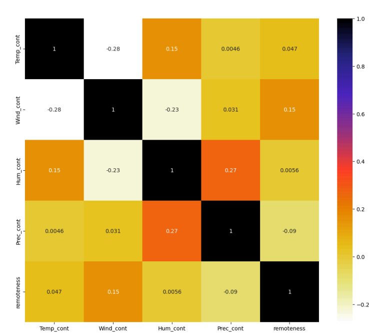
Table 3: Accuracy score of each type of model. The data used to train the model was 70% of the 7284 complete wildfire data, while 30% of the data was used for testing the model.

Different wildfire data set by cause						
	Lightning		Arson		Debris Burning	
	k-score	testing	k-score	testing	k-score	testing
RF	57%±5	57.14%	61%±5	61.48%	77%±3	76.07%
KNN	56%±3	55.10%	62%±4	60.12%	77%±3	75.92%
SVM	48%±3	47.81%	65%±1	65.03%	78%±0	78.22%

Table 4: The accuracy of each model trained on different subsets of the data set

It can be seen from Table 4 that the accuracy of each model increases when the data set is changed from Lightning only wildfire to Arson only wildfire to Debris Burning only wildfire. The accuracy obtained by all models when trained on the Debris Burning only wildfire data set even exceeds the 70% threshold accuracy of the models when trained on the complete data set.

The Pearson Correlation Matrix and the feature importance graphs for each type of wildfire are also displayed in Table 5. It can be seen from matrices that the Lightning data set consists of very correlated data while the Arson and Debris Burning data sets have relatively low correlation. It also be seen that the feature importance test on the Lightning data set has a noticeable difference between 'remoteness' and the other attributes. The difference between 'remoteness' and the other attributes is less noticeable in the Arson data set and is nonexistent in the Debris Burning data set.

Lightning	Arson	Debris Burning
		

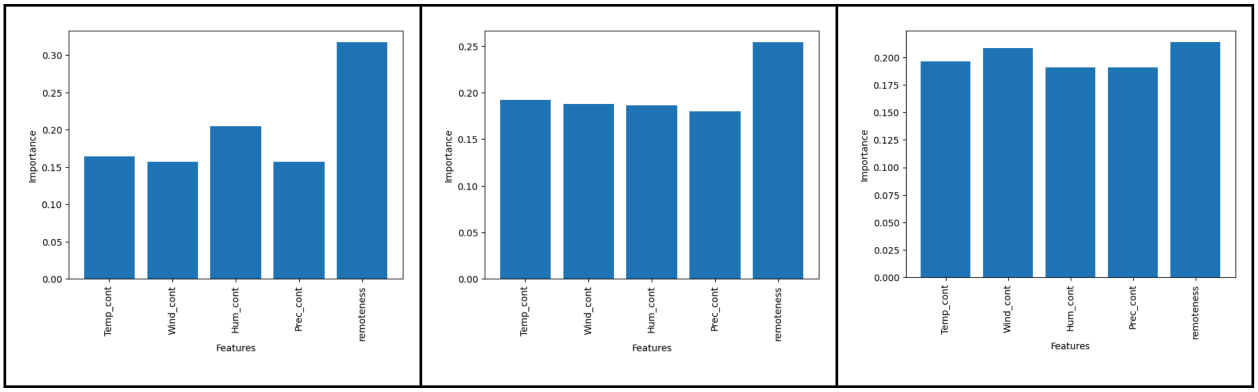


Table 5: This table shows the correlation among predictive features and the importance of each feature based on the data subset chosen.

5.2 Wildfire Projections

The Most Likely Wildfire Category (MLWC) to occur forecast is shown in Table 6. Here due to the absence of category A wildfire after the filtration of missing data, 0 signifies a B category wildfire while 5 signifies a G category wildfire.

Maryland MLWC Forecast (2010 - 2040)							
County Name	*KNN*	*SVM* Wildfire	*Random Forest*	County Name	*KNN*	*SVM*	*Random Forest*
Allegany	5	5	5	Howard	5	5	5
Anne Arundel	0	5	5	Kent	5	5	4
Baltimore	0	5	5	Montgomery	0	5	5
Calvert	5	5	5	Prince George's	0	5	5
Caroline	5	5	4	Queen Anne's	5	5	4
Carroll	5	5	5	Somerset	5	5	5
Cecil	5	5	4	St. Mary's	5	5	5
Charles	5	5	5	Talbot	5	5	4
Dorchester	5	5	4	Washington	5	5	5
Frederick	5	5	5	Wicomico	5	5	5

Garrett	5	5	4	Worcester	5	5	4
Harford	0	5	5				

Table 6: Model MLWC forecast for each Maryland county

6. Discussion

Wildfires are becoming more severe and widespread with the escalation of climate change. Wildfires are also contributing to climate change by releasing greenhouse gasses stored in vegetation. This is causing a wildfire-climate feedback loop to occur. Therefore, understanding, predicting, and preventing wildfires has become an essential task. This study attempts to project the Most Likely Wildfire Category (MLWC) to occur in the state of Maryland, USA. It uses Random Forest, K Nearest Neighbors, and Support Vector Machines to first understand the relation between wildfire category and climate variables (temperature, precipitation, humidity, wind speed, and remoteness). The classifier ML models will then employ the predicted climate scenario to determine the most likely wildfire category. The decision to focus on an East Coast state in this research is driven by the intention to forecast the effects of climate change on wildfire patterns in a region where they are infrequent. Apart from forecasting the type of wildfires, this study also endeavors to comprehend which climate factor holds a more significant influence in determining the wildfire category.

The findings indicated that wildfires in Maryland are expected to intensify in severity over the forthcoming decades. KNN, SVM, and RF models primarily projected F or G-category wildfires for most counties in Maryland. However, for specific counties, there were significant divergences in the model predictions. Notably, in Anne Arundel, Baltimore, Montgomery, Prince George, and Harford counties, KNN forecasted B-category wildfires, the least severe, while Random Forest and SVM anticipated G-category wildfires, the most severe. This discrepancy could potentially be attributed to KNN's heavy reliance on the remoteness factor, as these counties were assigned a remoteness factor of 0.99, signifying heightened urbanization. In contrast, RF and SVM made no such distinction, leading to their projection of G-category wildfires. In this context, KNN's projections appear more plausible, considering that wildfires are improbable in urban areas. The disagreement among the models is very interesting and should be explored in future studies. Nonetheless, the findings suggest that Maryland's wildfires will experience increased severity under simulated climate conditions. These results align with prior studies that indicate a projected rise in wildfires, both within the USA and worldwide, attributed to the effects of climate change [14,28,58].

Among the climate factors examined, it was observed that remoteness had the most notable impact in determining the wildfire category [Figure 5]. Subsequently, wind speed, humidity, temperature, and precipitation exhibited roughly equivalent influence over the category.

Interestingly, it was observed that these influences shifted based on the data subset utilized. When considering only wildfires induced by lightning, the gap between remoteness and the other atmospheric variables widened. Conversely, for wildfires resulting from debris burning, no significant discrepancies were evident among the variables. The outcome of the feature importance analysis for lightning-induced wildfires was particularly intriguing. Despite these wildfires being solely ignited and influenced by climatic conditions, the prominence of atmospheric variables in categorizing them was unexpected. As lightning-induced wildfires are primarily contingent upon the climate state for their ignition and behavior, it was anticipated that atmospheric variables would carry more influence over their categorization. Surprisingly, this was not the case, as remoteness continued to emerge as the predominant influencing factor [Table 5]. Our discovery of the significant impact of remoteness corroborates findings from previous research. Another study conducted by Jaafari et al. similarly identified remoteness as the foremost crucial feature influencing their model's capability to predict wildfires [15]. Climate change is also leading to rising sea levels which has a definitive effect on population movement moving them inwards. This can change the locations of wildfire-prone areas. Therefore, changing remoteness and population movements can also be an interesting factor to focus more on in future studies.

The KNN, SVM, and RF models exhibited similar accuracy levels, achieving an approximate 70% accuracy score. Nevertheless, variations in accuracy were evident when the models were trained on different subsets of the dataset. As illustrated in Table 4, when focusing solely on wildfires initiated by lightning, the accuracy decreased to 60% across all models. Conversely, when concentrating exclusively on wildfires arising from debris burning, the accuracy improved to 78% across all models. Additionally, an observation was made that the debris-burning dataset resulted in a scenario where all climate variables wielded roughly equal influence over the wildfire category. This equal distribution of influence was not observed in the case of the lightning-induced wildfire dataset, as depicted in Table 5. Consequently, it is conceivable that a positive correlation might exist between a higher accuracy score and a more balanced distribution of feature importance across variables. Understanding this correlation more extensively as well as exploring how different subsets of the data impact the accuracy of the models is a topic for future studies. Improving the accuracy of the models with the complete dataset should also be explored in future studies.

6.1 Limitations and Future Work

While this study provides quantitative projections of wildfire severity, there were several limitations. The data used for training the model had many missing and misleading data points. They were all filtered out but this greatly reduced the size of the dataset. The dataset also included many overlapping features which reduced the predictive power of the attributes. Figure 3 demonstrates this characteristic with two different attributes. The features of the dataset were also greatly correlated with each other. This increased the probability of feeding potential noise to the models. The model wildfire category forecasts were also not validated against real data. Overall, this study has had multiple limitations and there is room for improvement. However, the findings of this study align with previous work and demonstrate a reason for reducing climate change and preventing wildfires.

7. Conclusion

Due to climate change, wildfires are becoming more severe and widespread. Therefore, understanding, predicting, and preventing wildfires has become an essential task. By using Random Forest, SVM, and KNN machine learning models to understand the relation between wildfire severity and climate behavior, this study attempts to project wildfire severity in Maryland, USA with respect to its projected climate state. The results showed a general increase in Maryland's fire severity in the upcoming decades. All models projected wildfire categories of F or G based on the inputted future climate states of Maryland. This shows how climate change is increasing the severity of wildfires in regions where wildfires are quite uncommon. All the model's k-cross validation scores were approximately 70%, but their scores changed significantly depending on the data subset used. Feature importance graphs also demonstrated 'remoteness' to be the most important criterion for determining the wildfire category. Overall, this study demonstrates the huge impact of climate change on wildfire severity and spread. Without preemptive measures taken by policymakers and the general public to reduce climate change and prevent wildfires, wildfires in the state of Maryland are projected to increase in the coming decades.

8. Acknowledgments

I am very thankful to my mentor Dr. Sabie for guiding me through this research. I am thankful to the Lumiere team for providing me with resources. I am also very thankful to my parents for supporting me and giving me an environment suitable for research.

References

[1] wfca_teila, “Are Wildfires Good for the Environment?,” *WFCA*, Aug. 03, 2022. https://wfca.com/articles/are-wildfires-good-for-the-environment/ (accessed Aug. 21, 2023).

[2] National Geographic Society, “The Ecological Benefits of Fire | National Geographic Society,” *education.nationalgeographic.org*, Jul. 15, 2022. https://education.nationalgeographic.org/resource/ecological-benefits-fire/ (accessed Aug. 21, 2023).

[3] J. T. Abatzoglou and A. P. Williams, “Impact of anthropogenic climate change on wildfire across western US forests,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 42, pp. 11770–11775, Oct. 2016, doi: https://doi.org/10.1073/pnas.1607171113.

[4] L. Supriya, “Ecosystems could once bounce back from wildfires. Now, they’re being wiped out for good,” *www.science.org*, Dec. 19, 2017. https://www.science.org/content/article/ecosystems-could-once-bounce-back-wildfires-now-they-re-being-wiped-out-good

[5] C. S. Stevens-Rumann *et al.*, “Evidence for declining forest resilience to wildfires under climate change,” *Ecology Letters*, vol. 21, no. 2, pp. 243–252, Dec. 2017, doi: https://doi.org/10.1111/ele.12889.

[6] O. US EPA, “Wildland Fire Research: Health Effects Research,” *US EPA*, Mar. 30, 2017. https://www.epa.gov/air-research/wildland-fire-research-health-effects-research#:~:text=The%20effects%20of%20smoke%20from

[7] J. Jones, “Forest Service Ready for Western Wildfire Season,” *US Forest Service*, Jul. 06, 2016. https://www.fs.usda.gov/features/forest-service-ready-western-wildfire-season (accessed Aug. 22, 2023).

[8] National Interagency Fire Center, “Wildfire Mitigation | National Interagency Fire Center,” *Nifc.gov*, 2021. https://www.nifc.gov/fire-information/fire-prevention-education-mitigation/wildfire-mitigation

[9] Y. Liu, S. L. Goodrick, and J. A. Stanturf, “Future U.S. wildfire potential trends projected using a dynamically downscaled climate change scenario,” *Forest Ecology and Management*, vol. 294, pp. 120–135, Apr. 2013, doi: https://doi.org/10.1016/j.foreco.2012.06.049.

[10] National Oceanic and Atmospheric Administration, “Wildfire climate connection,” *www.noaa.gov*, Aug. 08, 2022. https://www.noaa.gov/noaa-wildfire/wildfire-climate-connection

[11] F. Thomson, “Record-breaking amount of carbon dioxide emissions from wildfires in 2021,” *Open Access Government*, Mar. 07, 2023. https://www.openaccessgovernment.org/record-breaking-amount-carbon-dioxide-emissions-wildfires-2021/154310/

[12] P. Jain, S. C. P. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, “A review of machine learning applications in wildfire science and management,” *Environmental Reviews*, vol. 28, no. 4, pp. 478–505, Dec. 2020, doi: https://doi.org/10.1139/er-2020-0019.

[13] V. Wagner, *Development and structure of the Canadian forest fire weather index system*, vol. 35. 1987.
[14] H. An, J. Gan, and S. Cho, “Assessing Climate Change Impacts on Wildfire Risk in the United States,” *Forests*, vol. 6, no. 12, pp. 3197–3211, Sep. 2015, doi: https://doi.org/10.3390/f6093197.
[15] A. Jaafari, E. K. Zenner, and B. T. Pham, “Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers,” *Ecological Informatics*, vol. 43, pp. 200–211, Jan. 2018, doi: https://doi.org/10.1016/j.ecoinf.2017.12.006.
[16] S. Kondylatos *et al.*, “Wildfire Danger Prediction and Understanding With Deep Learning,” *Geophysical Research Letters*, vol. 49, no. 17, Sep. 2022, doi: https://doi.org/10.1029/2022gl099368.
[17] L. Surya, “Fighting Fire with AI: Using Deep Learning to Help Predict Wildfires in the US,” *Social Science Research Network*, Oct. 08, 2020. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3786539 (accessed Aug. 22, 2023).
[18] H. Xue, F. Gu, and X. Hu, “Data assimilation using sequential monte carlo methods in wildfire spread simulation,” *ACM Trans. Model. Comput. Simul.*, vol. 22, Art. no. 4, 2012, doi: https://doi.org/10.1145/2379810.2379816.
[19] Z. Makhaya, J. Odindi, and O. Mutanga, “The influence of bioclimatic and topographic variables on grassland fire occurrence within an urbanized landscape,” *Scientific African*, vol. 15, p. e01127, Mar. 2022, doi: https://doi.org/10.1016/j.sciaf.2022.e01127.
[20] G. Zhang, M. Wang, and K. Liu, “Deep neural networks for global wildfire susceptibility modelling,” *Ecological Indicators*, vol. 127, p. 107735, Aug. 2021, doi: https://doi.org/10.1016/j.ecolind.2021.107735.
[21] M. Rodrigues and J. de la Riva, “An insight into machine-learning algorithms to model human-caused wildfire occurrence,” *Environmental Modelling & Software*, vol. 57, pp. 192–201, Jul. 2014, doi: https://doi.org/10.1016/j.envsoft.2014.03.003.
[22] J. A. Cardille, S. J. Ventura, and M. G. Turner, “ENVIRONMENTAL AND SOCIAL FACTORS INFLUENCING WILDFIRES IN THE UPPER MIDWEST, UNITED STATES,” *Ecological Applications*, vol. 11, Art. no. 1, 2001, doi: https://doi.org/10.1890/1051-0761(2001)011%5B0111:EASFIW%5D2.0.CO;2.
[23] J. Li *et al.*, “Feature Selection,” *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, Jan. 2018, doi: https://doi.org/10.1145/3136625.
[24] M. McCombe, “Intro to Feature Selection Methods for Data Science,” *Medium*, Jun. 07, 2019. https://towardsdatascience.com/intro-to-feature-selection-methods-for-data-science-4cae2178a00a
[25] Z. S. Pourtaghi, H. R. Pourghasemi, R. Aretano, and T. Semeraro, “Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques,” *Ecological Indicators*, vol. 64, pp. 72–84, May 2016, doi: https://doi.org/10.1016/j.ecolind.2015.12.030.

[26] Geeks For Geeks, “AUC-ROC Curve,” *GeeksforGeeks*, Nov. 25, 2020. https://www.geeksforgeeks.org/auc-roc-curve/
[27] S. A. Parks, C. Miller, J. T. Abatzoglou, L. M. Holsinger, M.-A. Parisien, and S. Z. Dobrowski, “How will climate change affect wildland fire severity in the western US?,” *Environmental Research Letters*, vol. 11, no. 3, p. 035002, Feb. 2016, doi: https://doi.org/10.1088/1748-9326/11/3/035002.
[28] M. A. Moritz *et al.*, “Climate change and disruptions to global fire activity,” *Ecosphere*, vol. 3, no. 6, p. art49, Jun. 2012, doi: https://doi.org/10.1890/es11-00345.1.
[29] D. Stralberg *et al.*, “Wildfire-mediated vegetation change in boreal forests of Alberta, Canada,” *Ecosphere*, vol. 9, no. 3, p. e02156, Mar. 2018, doi: https://doi.org/10.1002/ecs2.2156.
[30] Y. Lai and D. A. Dzombak, “Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation,” *Weather and Forecasting*, vol. 35, no. 3, pp. 959–976, Apr. 2020, doi: https://doi.org/10.1175/WAF-D-19-0158.1.
[31] L. Gowri, K. R. Manjula, K. Sasireka, and D. Deepa, “Assessment of Statistical Models for Rainfall Forecasting Using Machine Learning Technique,” *Journal of Soft Computing in Civil Engineering*, vol. 6, no. 2, pp. 51–67, Apr. 2022, doi: https://doi.org/10.22115/scce.2022.304260.1363.
[32] D. Endalie, G. Haile, and W. Taye, “Deep learning model for daily rainfall prediction: case study of Jimma, Ethiopia,” *Water Supply*, Nov. 2021, doi: https://doi.org/10.2166/ws.2021.391.
[33] M. Raval, P. Sivashanmugam, V. Pham, H. Gohel, A. Kaushik, and Y. Wan, “Automated predictive analytics tool for rainfall forecasting,” *Scientific Reports*, vol. 11, no. 1, p. 17704, Sep. 2021, doi: https://doi.org/10.1038/s41598-021-95735-8.
[34] W. Cai, R. Wei, L. Xu, and X. Ding, “A method for modelling greenhouse temperature using gradient boost decision tree,” *Information Processing in Agriculture*, Sep. 2021, doi: https://doi.org/10.1016/j.inpa.2021.08.004.
[35] Science Direct, “General Circulation Model - an overview | ScienceDirect Topics,” *Sciencedirect.com*, 2014. https://www.sciencedirect.com/topics/earth-and-planetary-sciences/general-circulation-model
[36] Kaggle, “U.S. Wildfire data (plus other attributes),” *Kaggle*. https://www.kaggle.com/datasets/capcloudcoder/us-wildfire-data-plus-other-attributes
[37] K. C. Short, “Spatial wildfire occurrence data for the United States, 1992-2015 [FPA_FOD_20170508] (4th Edition),” *Forest Service Research Data Archive*, doi: https://doi.org/10.2737/rds-2013-0009.4.
[38] NOAA National Centers for Environmental Information, “Integrated Surface Hourly [1992-2015],” 2001.
[39] P. Meiyappan and A. K. Jain, “Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years,” *Frontiers of Earth Science*, vol. 6, no. 2, pp. 122–139, Jun. 2012, doi: https://doi.org/10.1007/s11707-012-0314-2.
[40] Simple Maps, “World Cities Database,” *Simplemaps.com*, 2019.

https://simplemaps.com/data/world-cities
[41] "Max Normalization - an overview | ScienceDirect Topics," *www.sciencedirect.com*. https://www.sciencedirect.com/topics/computer-science/max-normalization
[42] V. Kumar, "Feature Selection: A literature Review," *The Smart Computing Review*, vol. 4, no. 3, Jun. 2014, doi: https://doi.org/10.6029/smartcr.2014.03.007.
[43] K. Kira and L. A. Rendell, "A Practical Approach to Feature Selection," *ScienceDirect*, Jan. 01, 1992. https://www.sciencedirect.com/science/article/pii/B9781558602472500371
[44] I. Guyon and A. De, "An Introduction to Variable and Feature Selection André Elisseeff," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003, Available: https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf?ref=driverlayer.com/web
[45] W. Kirch, "Pearson's Correlation Coefficient," *Encyclopedia of Public Health*, pp. 1090–1091, 2019, doi: https://doi.org/10.1007/978-1-4020-5614-7_2569.
[46] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *Journal of Big Data*, vol. 7, no. 1, Jul. 2020, doi: https://doi.org/10.1186/s40537-020-00327-4.
[47] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN Model-Based Approach in Classification," *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, vol. 2888, pp. 986–996, 2003, doi: https://doi.org/10.1007/978-3-540-39964-3_62.
[48] Science Direct, "Support Vector Machine - an overview | ScienceDirect Topics," *Science Direct*. https://www.sciencedirect.com/topics/computer-science/support-vector-machine
[49] W. Dudzik, J. Nalepa, and M. Kawulok, "Evolving data-adaptive support vector machines for binary classification," *Knowledge-Based Systems*, vol. 227, p. 107221, Sep. 2021, doi: https://doi.org/10.1016/j.knosys.2021.107221.
[50] IBM, "What is Random Forest? | IBM," *IBM*, 2023. https://www.ibm.com/topics/random-forest
[51] SciKit-Learn, "3.1. Cross-validation: evaluating estimator performance — scikit-learn 0.21.3 documentation," *Scikit-learn.org*, 2009. https://scikit-learn.org/stable/modules/cross_validation.html
[52] Maryland, Department of Natural Resources, "Wildland Fire in Maryland," *dnr.maryland.gov*. https://dnr.maryland.gov/forests/Pages/wfm.aspx
[53] FernLeaf Interactive and National Environmental Modeling and Analysis Center (NEMAC), "Climate Explorer," *Climate Explorer Toolkit*. https://crt-climate-explorer.nemac.org
[54] University of Maryland, "Maryland State Climatologist Office," *Maryland's Climate*, Nov. 06, 1976. http://www2.atmos.umd.edu/~climate/ClimateInfo/mdClimate.php (accessed Aug. 22, 2023).
[55] National Oceanic and Atmospheric Administration (NOAA), "Climate Report National Weather Service Baltimore MD/WASHINGTON DC," *Weather.gov*, 2023. https://w1.weather.gov/data/LWX/CLIBWI (accessed Aug. 22, 2023).
[56] "Most Humid States 2023," *Wisevoter*. https://wisevoter.com/state-rankings/most-humid-states/ (accessed Aug. 22, 2023).

[57] Maryland State Data Center, “Census 2010,” *Census 2010*.
https://planning.maryland.gov/MSDC/Pages/census/Census2010/maps.aspx (accessed Aug. 22, 2023).
[58] S. Mansoor *et al.*, “Elevation in wildfire frequencies with respect to the climate change,” *Journal of Environmental Management*, vol. 301, p. 113769, Jan. 2022, doi:
https://doi.org/10.1016/j.jenvman.2021.113769.
[59] National Weather Service, “Parts of a Wildland Fire ,” *National Weather Service*.
https://training.weather.gov/pds/firewx/documents/s130_m1_Module2.docx (accessed Sep. 06, 2023).

Appendix A: Reference Materials

Fire Class	**Description**
Class A	0 to 0.2 acres (0 to 0.1 ha)
Class B	0.3 to 9.9 acres (0.2 to 4 ha)
Class C	10 to 99.9 acres (4.1 to 40.4 ha)
Class D	100 to 299.9 acres (40.5 to 121.4 ha)
Class E	00 to 999.9 acres (121.5 to 404 ha)
Class F	1,000 to 4,999.9 acres (405 to 2,024 ha)
Class G	5,000 acres (2,025 ha) or more

Table 1; Description of each type of wildfire based on area spread or burned [59]

Feature Variable Name	Feature Description
fire_size_class	Class of Fire Size (A-G)
stat_cause_descr	Cause of Fire
Vegetation	Dominant vegetation in the areas 1:Tropical Evergreen Broadleaf Forest 2:Tropical Deciduous Broadleaf Forest 3:Temperate Evergreen Broadleaf Forest 4:Temperate Evergreen Needleleaf Forest 5:Temperate Deciduous Broadleaf Forest 6:Boreal Evergreen Needleleaf Forest 7:Boreal Deciduous Needleleaf Forest 8:Savanna 9:C3 Grassland/Steppe 10:C4 Grassland/Steppe 11:Dense Shrubland 12:Open Shrubland 13:Tundra Tundra 14:Desert 15:Polar Desert/Rock/Ice 16:Secondary Tropical Evergreen Broadleaf Forest 17:Secondary Tropical Deciduous Broadleaf Forest 18:Secondary Temperate Evergreen Broadleaf Forest 19:Secondary Temperate Evergreen Needleleaf Forest 20: Secondary Temperate Deciduous Broadleaf Forest 21:Secondary Boreal Evergreen Needleleaf Forest 22:Secondary Boreal Deciduous Needleleaf Forest 23:Water/Rivers Water 24:C3 Cropland

	25:C4 Cropland 26:C3 Pastureland 27:C4 Pastureland 28:Urban land
Temp_pre_30	temperature in deg C at the location of fire up to 30 days prior
Temp_pre_15	temperature in deg C at the location of fire up to 15 days prior
Temp_pre_7	temperature in deg C at the location of fire up to 7 days prior
Temp_cont	temperature in deg C at the location of fire up to the day the fire was
Wind_pre_30	wind in m/s at the location of fire up to 30 days prior
Wind_pre_15	wind in m/s at the location of fire up to 15 days prior
Wind_pre_7	wind in m/s at the location of fire up to 7 days prior
Wind_cont	wind in m/s at the location of fire up to the day the fire was contained
Hum_pre_30	humidity in % at the location of fire up to 30 days prior
Hum_pre_15	humidity in % at the location of fire up to 15 days prior
Hum_pre_7	humidity in % at the location of fire up to 7 days prior
Hum_cont	humidity in % at the location of fire up to the day the fire was contained
Prec_pre_30	precipitation in mm at the location of fire up to 30 days prior
Prec_pre_15	precipitation in mm at the location of fire up to 15 days prior
Prec_pre_7	precipitation in mm at the location of fire up to 7 days prior
Prec_cont	precipitation in mm at the location of fire up to the day the fire was contained
remoteness	non-dimensional distance to closest city

Table 2: List of features relevant to wildfire-climate modeling

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Figure 1: Min-max normalization formula

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Figure 2: Formula for calculating the Pearson Coefficient of two variables.