Risk Assessment of Potential Wildfire Outbreak Using Satellite Data

Members: Nathan Ress, Lily Smith, Ben Mellin, Uchswas Paul
nmress,lcsmit25,bmellin,upaul@ncsu.edu
North Carolina State University
Raleigh, NC, USA

ACM Reference Format:
Members: Nathan Ress, Lily Smith, Ben Mellin, Uchswas Paul. 2025. Risk Assessment of Potential Wildfire Outbreak Using Satellite Data. In . ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction and Background

1.1 Problem

Wildfires in the United States have resulted in significant loss of life, with over 500 fatalities recorded between 1990 and 2023, including 130 deaths in 2023 alone, marking the deadliest year on record [9]. Even in the last year, wildfires resulted in several deaths and the destruction of 4,552 structures in the United States [4]. Proper warning and initiation of evacuation processes in areas where wildfires may occur is crucial to mitigate the lives lost and economic damage in these acts of nature. Our project uses satellite data to forecast the risk of wildfire outbreak at a given point of reference using the NASA DAV data source [3]. This is important because satellite data does not require extensive ground sensors to measure meteorological conditions, meaning places across the world without access to localized weather data can still determine risks of wildfire outbreak in their community.

1.2 Related Work

Alonso-Betanzos et al. [2] demonstrated a very early use of a neural network to predict wildfire risk for the region of Galicia in Spain. Alonso-Betanzos et al. used temperature, daily humidity, daily rainfall, and fire history as features and were able to achieve an accuracy of .789 in wildfire prediction. Sakr et al. [6] provide another early example of applying machine learning to assess wildfire risk. Using minimal and maximal temperatures, average humidity, average solar radiation, average windspeed, and cumulative precipitation level measurements from the Lebanese Agricultural Research Institute, they found that a support vector machine could predict wildfire risk with about 85% accuracy. The authors also found that prediction accuracy varied greatly depending on the month of the instance they were attempting to predict.

Sayad, Mousannif, and Moatassime [7] used remote sensing data from Terra, Aqua, Landsat, and Aster satellites to obtain parameters related to vegetation coverage, land surface temperature, and if a fire is present. They ultimately found that a neural network not only had the greatest accuracy of machine learning models tested but also minimized the false positive rate (i.e., predicting fire when there was no fire). Pham et al. [5] compared the performances of the Bayes network, naïve Bayes, decision tree, and multivariate logistic regression models on the prediction and mapping of fire susceptibility in the Pu Mat National Park in Vietnam. Their model was based on data from 57 historical fires and nine features. Using AUC as a performance metric, they found that the Bayes network bested the other models with an AUC value of .96.

Sharma and Khanal [8] compared the performance of Decision Tree, Random Forest, Logistic Regression, Artificial Neural Network, Support Vector Machine, and convolutional Neural Network Nodels on predicting forest fires in South Carolina. Features used included meteorology, terrain, vegetation, and infrastructure. Sharma and Khanal found that the decision tree had the greatest accuracy of all methods tested. The authors also found a strong overlap between fire incidence and "carbon hotspots," or places with dense vegetation, soil, or forest, underscoring the need for better fire management in these regions.

2 Method

In this section, we will present the novelty, methodological approaches, and rationale of the approaches.

2.1 Novel Aspects

Our approach introduces five key innovations: analyzing feature importance, robust and balanced data collection, cleaning noisy wildfire records, and using risk-based evaluation to enable probabilistic fire forecasting.

2.1.1 Feature Analysis and Discovery: A key novelty of our work might be to uncover which features our models rely on most to predict wildfires, particularly those that may have been overlooked or underemphasized in prior research. For instance, Pham et al. [5] mention 9 features. Can we find some features that are equal to or more important than previous features identified? To do this, we will extract the important features from our models using permutation feature importance methods. An analysis of the feature importance of satellite data has not been conducted in the literature.

2.1.2 Dataset Scope and Scale: Our dataset is significantly more comprehensive in scale, scope, and geographic diversity compared to prior wildfire prediction studies. While earlier works—such as those by Alonso-Betanzos et al. [2], Sakr et al. [6], Sayad et al. [7], and Pham et al. [5]—focused on narrow regions like Galicia, Lebanon, or Vietnam and often used limited local meteorological

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

or remote sensing features, our dataset spans the entire continental United States. It includes over 29,000 instances (14,950 positive and 15,000 negative), each with rich satellite-derived weather features obtained via the NASA DAV API. This approach contrasts with past work that often used smaller, imbalanced, or region-specific datasets, thereby limiting generalizability.

2.1.3 Complex Data Handling: Another piece of novelty from our approach is the method in which we collect positive and negative instances of fire outbreaks. To get positive instances, we used the National United State Forest Service database of wildfire outbreaks. This data source provides the latitude, longitude, size, and date of fire outbreaks. This data was extremely messy, including many typos and incorrectly and nonsensically specified times, dates, latitudes, and longitudes. Cleaning a dataset this messy was part of the novelty. We wrote algorithms to ensure times and dates were correctly specified and latitudes and longitudes existed and were within the bounds of the United States.

For negative instances of fire outbreaks, since we would like our model to be robust, we want the full set of data instances to cover a large variation of possible locations represented by our fire presence data. To accomplish this, we developed an algorithm that randomly samples latitudes, longitudes, and date pairings within the United States that are not present within the fire outbreak data from the United States Forest Service dataset. By random sampling over 15,000 locations and date pairings, we have a large dataset with wide variation, allowing the model to handle data instances in a diverse range of target locations within the United States.

2.1.4 Contribution of a new Dataset: We also find novelty in our approach by obtaining balanced class data within the data collection process. Many state-of-the-art solutions exhibit one of the following issues: 1. The data is balanced but limited in scale, or 2. The data is unbalanced but large-scale. By utilizing the data collection approaches above, we obtained a class balance split of nearly 50/50, meaning we have a large set of balanced data on which to fit models. This dataset may serve as a valuable asset for future researchers.

2.1.5 Model Design and Evaluation: We also bring novelty to our solution by addressing risk assessment via ROC curves for our models. While not inherently exclusive to our approach, many research papers in this domain use accuracy, precision, recall, and F1 scores as the primary metrics for evaluating model success. By outputting 'confidence' of predicted labels by our model, we can effectively project the probability of a fire outbreak at a given location. In other words, we are able to forecast the risk of a fire outbreak rather than simply a binary response as output. By computing risk rather than binary output, our models can effectively be used to demonstrate fire-safety precautions for individuals living in areas where a fire outbreak may occur on a given day.

2.2 Approach

Our approach utilized a variety of machine learning classifiers to predict the binary label of wildfire outbreaks for each given location within our dataset. We selected 5 Machine learning algorithms: **random forests, decision trees, logistic regression, KNN, and multilayer perceptrons**. We conducted a variety of data engineering approaches to improve the evaluation metrics of our models, including standardization, feature selection, and hyperparameter tuning. We will use 5-fold **cross-validation** to evaluate our models on the evaluation metrics and for hyperparameter tuning, and use a **grid-search** approach. After hyperparameter tuning, we evaluate the model's performance using accuracy, recall, precision, F1 score, and ROC curve. We consider **F1 score and ROC–AUC** as our primary metrics. Finally, we used **permutation-importance** to measure the feature importance.

2.3 Rationale

2.3.1 Model Selection: We selected a diverse range of models to evaluate different facets of wildfire prediction comprehensively. We selected logistic regression as a transparent baseline to determine whether wildfire risk could be captured through simple additive relationships among meteorological variables. We included k--nearest neighbors to serve as a non-parametric, data-driven reference that would reveal the importance of local pattern sensitivity without imposing any distributional assumptions. We chose decision trees for their intuitive, rule-based structure and ability to model non-linear feature interactions, then adopted random forests to aggregate multiple trees for enhanced robustness and built-in feature-importance insights. We selected MLPs as they are well-known for their ability to handle complex data sets. Finally, our baseline classifier was an untuned random forest classifier against which we compared our best trained model.

2.3.2 Hyperparameter tuning: We'll use cross-validation because it provides a robust, unbiased estimate of model performance across multiple data splits, guiding us to select hyperparameters that generalize well rather than overfit to a single train–test split. Moreover, we aimed to maximize the F1 score and AUC-ROC on the validation set. That's why we chose grid search to determine the optimal combination of hyperparameters, though computationally heavy.

2.3.3 Evaluation Metrics: We chose accuracy, recall, precision, F1 score, and ROC-AUC to capture complementary aspects of classifier efficacy: accuracy for overall correctness, recall for detecting true fire events, precision for limiting false alarms, and F1 score to balance recall and precision given the rarity of fires. The ROC curve and its AUC further quantify the model's ability to discriminate between fire and no-fire across all decision thresholds. **Although false negatives (missed fire outbreaks) can be more catastrophic than false positives (false alarms), we regarded both error types as critically important—excessive false alarms can undermine trust, induce alert fatigue, and impose economic costs through unnecessary preparations**. By prioritizing F1 and ROC–AUC, we ensured our evaluation emphasized both the sensitivity–specificity trade-off and the model's intrinsic separability independent of any single threshold.

2.3.4 Feature Importance: We use permutation importance because it is an intuitive methdod that can be applied across all models. Other methods like normalizing coefficients or impurity scores can be confusing, since they only fit specific models and ignore how features work together. Permutation importance gives an easy, post-training list of what the model truly depends on.

3 Plan and Experiment

In this section, we present our experimental setup, which includes dataset construction, research question formulation, feature selection, hyperparameter tuning, model training and evaluation, and feature importance analysis.

3.1 Dataset

3.1.1 Dataset Overview. Our dataset comprises 29,950 instances—14,950 confirmed wildfire events (`FIRE=1`) from the U.S. Forest Service and 15,000 randomly sampled non-fire points (`FIRE=0`) across the continental United States. For each location and date, we retrieved meteorological and surface variables via NASA's DAV API: `WS2M` (wind speed at 2 m), `T2M` (near-surface air temperature at 2 m) with its daily maximum (`T2M_MAX`) and minimum (`T2M_MIN`), `T2MDEW` (dew point temperature at 2 m), `RH2M` (relative humidity at 2 m), `PS` (surface pressure), `ALLSKY_SFC_LW_DWN` (all-sky surface longwave downward irradiance), `ALLSKY_SFC_SW_DWN` (all-sky surface shortwave downward irradiance), and `ALLSKY_SFC_SW_DIFF` (all-sky surface shortwave diffuse downward irradiance). The temporal span ranges from January 1, 1980, to February 15, 2024, providing comprehensive coverage of seasonal and climatic variability for robust wildfire-prediction modeling.

3.1.2 Dataset Construction. The first step in our approach was the collection of confirmed wildfire locations and dates. This data comes from the United States Forest Service, which gave the latitude, longitude, and the date on which wildfires were present. After cleaning the dataset, as described in the novelty section, we then utilized the NASA DAV API, querying the database with the latitude, longitude, and date for the confirmed fire outbreaks and extracting the satellite weather data for that location on the given date. We assigned these queries with a 'FIRE' response of 1, the 'positive' class label for our approach. We utilized 14,950 instances of fire data during this process, extracting data for 14,950 wildfires and the weather data present from the API to collect positive class instance data.

For the next step in our approach, we collected data for the 'negative' class label, or instances in which wildfires were not present. We sought to collect data that fully encompassed the United States, matching the domain of our positive label dataset. Our logic here was that by providing data over the same domain as we used for our positive class, we could appropriately capture the widespread variation that comes from having such a large number of data samples. Our algorithm for collecting negative class data selected a random latitude within the continental United States (25N to 50N), a random longitude within the continental United States (65W to 126W), and a random date within the valid range of the APIs available data (January 1st, 1980 to February 15th, 2024). We passed these randomly generated values into an API query, collecting satellite weather data, assigning a 'FIRE' value of 0, and adding them to our dataset. If the query results in failure (given if point is over a body of water), missing data, or overlap with positive dataset, we threw out the query and tried again. We completed this process 15,000 times to collect 15,000 negative label instances.

3.2 Hypotheses

After constructing our dataset, we focused on three research questions that are interesting and relevant to the problem description. Those are:

- **RQ1** Can we predict the occurrence of wildfires with meaningful accuracy and uncertainty measures based on the features in the NASA DAV dataset?
- **RQ2** Does inference provide insight into which features are most important in predicting the likelihood of a wildfire occurring?
- **RQ3** What are the implications of using different models to predict wildfire on our dataset?

3.3 Experimental Design

3.3.1 Data Splitting and Feature Selection. We first split the full dataset into training and test sets using an 80/20 ratio, keeping the test set completely unseen during model development. We then partitioned the training set again into 5 subsets to be used for 5 fold cross validation in hyperparameter tuning. To ensure reproducibility, we fixed the random seed to 1234. After splitting, we isolated the target variable `FIRE` and dropped the `Date` attribute, since its high variability made it act like an identifier rather than a predictive feature. Moreover, We use *StandardScaler* to center each feature to zero mean and scale it to unit variance, which speeds up convergence and ensures all features contribute equally.

3.3.2 Hyperparameter Tuning: We identified key hyperparameters from the literature for each model and defined initial ranges tailored to our dataset. We then performed a grid search with 5-fold cross-validation to tune these hyperparameters and select the optimal settings. Table 1 summarizes each hyperparameter, its initial range, and the best value found.

3.3.3 Model Training, Evaluation, Feature Importance. Each tuned classifier was retrained on the full training set (training + validation) and then evaluated on the held-out test set. We computed accuracy, recall, precision, and F1-score, and plotted ROC curves to quantify both overall correctness and class separability. For each trained model, we applied permutation importance (15 repeats, `scoring='roc_auc'`) to obtain a true post-hoc ranking of input features. Finally, we inspected misclassification for the best-performing model (**Random Forest**) by grouping test examples by their (true, predicted) labels and comparing mean feature values across confusion-matrix cells to diagnose systematic errors.

4 Results

This section presents a comparative analysis of the model, structured around our research questions.

4.1 RQ1: Firebreak Detection Ability

Table 2 shows the accuracy, recall, precision, F1, and AUC-ROC scores of the five models we implemented (Random Forest, Decision Tree, Logistic Regression, K-NN, and Multilayer Perceptron) before and after hyperparameter tuning. For every model, we see a slight improvement in the primary evaluation metric after hyperparameter tuning. The best-performing model in terms of recall and

Model	Selected Hyperparameters	Initial Range/Values	Best Value
Decision Tree	`max_depth`, `ccp_alpha`	[10,30], [0.0,0.05]	11, .00038
MLP	`hidden_layer_sizes`, `activation`, `alpha`, `learning_rate_init`	{(32,),(64,),(32,16),(64,32)}, {`relu`, `tanh`}, $\{e^{-5},e^{-4},e^{-3}\}$, $\{e^{-3},e^{-2}\}$	(64, 32), relu, e^{-4}, e^{-2}
k-Nearest Neighbors	`n_neighbors`, `weights`	{1:20}, {uniform,distance}	17, distance
Random Forest	`max_depth`, `min_samples_leaf`	{[10, 30]}, {[1, 10]}	22, 2
Logistic Regression	`tol`, `c`	{0.01,0.001,0.0001,0.00001}, {0.01,0.2,0.4,0.6,0.8,1}	0.0008, 0.0775

Table 1: Hyperparameter settings for each model

	Before Tuning					After Tuning				
Model	Accuracy	Recall	Precision	F1	AUC–ROC	Accuracy	Recall	Precision	F1	AUC–ROC
Random Forest	0.973	0.976	0.969	0.973	0.994	0.973	0.978	0.968	0.973	0.994
Decision Tree	0.960	0.958	0.961	0.960	0.959	0.966	0.970	0.962	0.966	0.987
Logistic Regression	0.930	0.948	0.913	0.930	0.975	0.946	0.955	0.935	0.945	0.985
KNN	0.920	0.933	0.907	0.920	0.967	0.924	0.941	0.907	0.924	0.978
Multilayer Perceptron	0.964	0.979	0.949	0.964	0.992	0.964	0.966	0.961	0.964	0.992

Table 2: Model Evaluation Metrics Before and After Hyperparameter Tuning

ROC-AUC is Random Forest. While the Random Forest is good at maximizing our primary metric in AUC-ROC, the model also has the highest values for all other metrics. Generally, the tree based models (random forest and decision tree) showcased improved F1 scores compared to other models, indicating their ability to balance recall and precision a bit more successfully than other models. Overall, each model achieved a high F1 score and ROC-AUC, at least 92%. This implies that Machine learning models can be employed to detect firebreaks.

4.1.1 Baseline Confusion Matrices and ROC Curves. Figures 3 and 4 showcase the confusion matrix and ROC curve for the Random Forest baseline, which was the best-performing baseline model among all classifiers given our primary and secondary evaluation metrics.

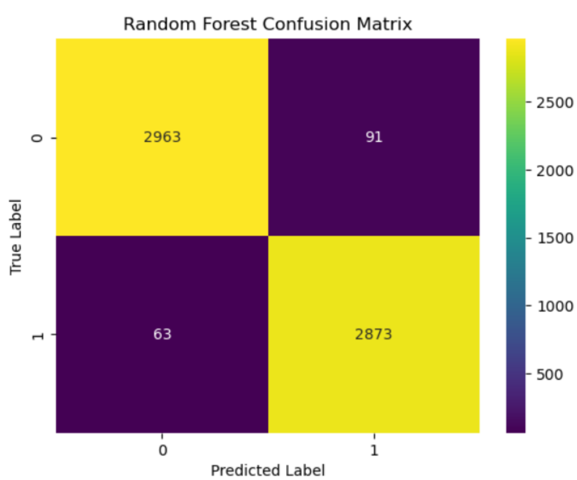

Figure 1: Confusion matrix for baseline random forest

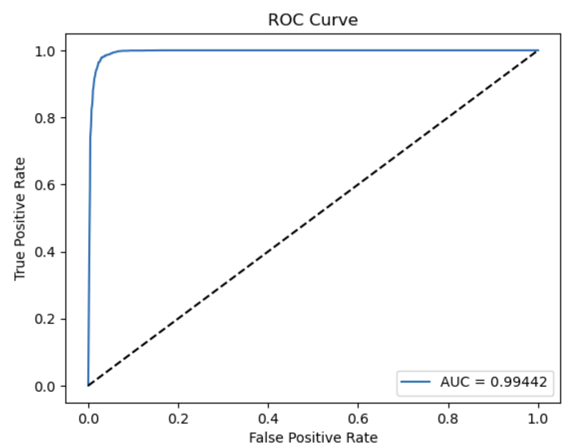

Figure 2: ROC Curve for baseline random forest

4.1.2 Tuned Confusion Matrices and ROC Curves. Figures 3 and 4 showcase the confusion matrix and ROC curve for the tuned Random Forest, which was the best-performing tuned model among all classifiers given our primary and secondary evaluation metrics.

4.2 RQ2: Feature Importance

Table 3 summarizes feature-importance scores for each predictor across the models used in this experiment, as well as a "Sum" column that aggregates each feature's total importance across all five classifiers. It highlights which meteorological and spatial variables consistently contribute most to wildfire prediction.

All but one of the models identified surface pressure (PS) as the strongest predictor of wildfire occurrence, assigning importance values of 0.355, 0.103, 0.352, 0.127, and 0.392 for each model respectively. For the best model, the Random Forest, geographic

Feature	Decision Tree	MLP	KNN	RF	LR	Sum
PS	0.355130	0.102687	0.352358	0.127	0.392	1.329175
Longitude	0.048066	0.155766	x	0.040	0.020	0.263832
Latitude	0.056772	0.044972	x	0.062	0.004	0.167744
WS2M	0.008211	0.002800	0.006052	0.002	0.000	0.019063
T2M	0.000462	0.000591	0.008598	0.002	0.001	0.012651
T2M_MAX	0.004581	0.000775	0.009286	0.002	0.002	0.018642
T2M_MIN	0.001165	0.006894	0.007471	0.000	0.001	0.016530
T2MDEW	0.000630	0.000614	0.012250	0.001	0.002	0.016494
RH2M	0.000153	0.004502	0.023393	0.002	0.006	0.036048
ALLSKY_SFC_LW_DWN	0.000000	0.002483	0.005989	0.000	0.000	0.008472
ALLSKY_SFC_SW_DWN	0.000000	0.000131	0.000146	0.000	0.000	0.000277
ALLSKY_SFC_SW_DIFF	0.000000	0.000468	0.002108	0.000	0.000	0.002576

Table 3: Feature importances across models and sum per feature

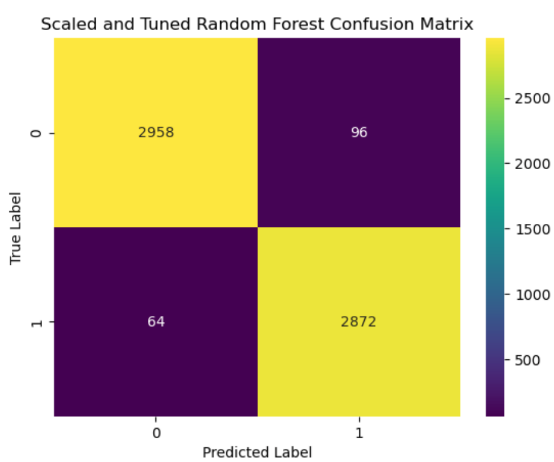

Figure 3: Confusion matrix for tuned Random Forest

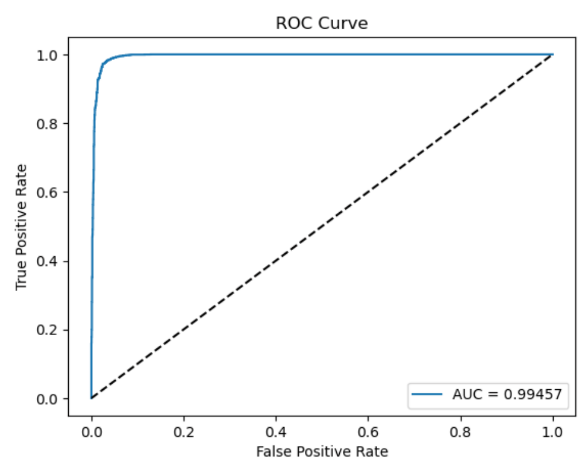

Figure 4: ROC Curve for tuned Random Forest

features—longitude (0.04) and latitude (0.062)—and meteorological variables such as wind speed (WS2M), dew point, and longwave irradiance each receive modest importance, reflecting the model's diverse split strategy. The only model that does not have surface pressure to be the most important feature, the MLP, has an importance distribution that shows longitude (0.155), surface pressure (0.103) and latitude (0.045) to be the most important features, respectively. All other features are near zero—suggesting a focused, pressure-driven decision boundary. Results are largely the same for the remaining models, with latitude and longitude being the second and third most important features for the model, with much lower degrees of significance compared to surface pressure. An exception to this pattern is in the logistic regression model, which actually showcases a low importance in the latitude feature, and a heavier emphasis on relative humidity (RH2M).

4.3 RQ3: Implications of Models and Features

The best model, Random Forest, found surface pressure to be the most important feature, followed by longitude and latitude as we can see from both Figure 5 and in Table 3. This was reflected in misclassified instances. In particular, the average surface pressure for false positives was much closer to the average surface pressure for true negatives, and the average surface pressure for false negatives was much higher than the average surface pressure for true positives, as seen in Table 4. This table was constructed using predictions from the tuned Random Forest.

Case	Avg. Long.	Avg. Lat.	Avg. Surf. Pres.
False Neg.	-106.89	44.41	87.06
False Pos.	-106.77	41.41	81.33
True Neg.	-95.17	37.17	96.40
True Pos.	-106.62	40.03	77.73

Table 4: Average values for most significant features in each classification group

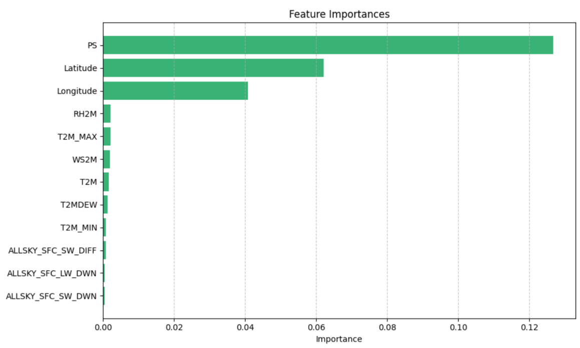

Figure 5: Feature importance in the Tuned Random Forest Model

5 Discussion

Classifiers: The results we show above indicate high performance of our models. We believe that this high performance is due in large part to the large amount of data we extracted from the NASA DAV API, which provides a much greater depth of data compared to other state-of-the-art solutions in this domain. Additionally, there is likely a significant degree of correlation between the features in our data and our output for this problem.

Feature Importance: Surface pressure was consistently found to be the most important feature in models where feature importance was evaluated. This makes sense–wildfires are most likely at times during transitions from high pressure fronts to low pressure fronts, which can create strong winds, produce thunderstorms, and enhance the behavior of the fire [1]. Latitude and longitude were also important in the decisions for our models. This also make sense–geographic location contains important information regarding fire likelihood. This may be because latitude and longitude have a correlation with other features, such as temperature, humidity, and surface pressure.

Selecting a model: When deciding on a model for risk assessment of wildfires using this dataset, it is important to identify which metric is most important to prioritize. We argue that maximizing AUC-ROC is the most important metric for this problem context. Since AUC-ROC well models probabilities, and probabilities can be associated with risks for each class (fire or no fire), this metric directly solves the problem statement. The next most important metric for these models is the recall score. When predicting wildfire risk, it would be much more costly to have a false negative rather than a false positive. This means recall scores are more important given the problem statement. With these aspects of the evaluation metrics in mind, the random forest would be the strongest candidate to deploy for predicting wildfire risk assessments. Not only does this model have the highest AUC-ROC score compared to the other models, but it also has the highest recall score. However, we note that all of these models showcase very high evaluation metric scores across all metrics, and the problem is well modeled with all of our tested machine learning models.

Model Limitations: While our models showcase high performance in wildfire risk and classification over the testing data, it is important to mention potential downsides with our approach. The most notable is the lack of data outside of the United States used in model training. While the NASA DAV API is a robust datasource that can be used internationally, we only used data from the continental United States, given that this was the only location we could find ground truth data for. Since our model leverages latitude and longitude and ground truth for United States wildfire data only, this model is unlikely to be as high performing outside of the United States. Going forward, using a machine learning model to predict wildfire outbreak risk outside of the United States would require the collection of new data from other regions of the world to ensure accurate fire predictions in areas outside of the United States.

Additionally, we note that our class data for no fire presence leveraged a random sampling of dates and latitude and longitude pairs in the United States. While this does provide some advantages, such as a more robust and diverse dataset to pull from, it may be more important to extract non-fire data from regions in which wildfire outbreak is more prevalent. The KNN model exemplifies a potential pitfall of our approach. When using latitude and longitude, it predicted the dataset perfectly. We hypothesize the reason for this is due to high correlation among fire locations, and the KNN model is able to exploit this better than any other model. For this reason, we left latitude and longitude out of the KNN model. Future work should consider the implications of this for other models.

It should be noted that the logistic regression model placed the least importance on the latitude and longitude features, indicating that it may be the best performer at forecasting wildfire risk in outlier locations in which there is a lack of ground truth data available for training. This may also make the model more robust internationally where data could be more sparse in general, however we did not investigate the availability of wildfire data outside the United States extensively.

Misclassifications: Using the Random Forest matrix and predictions, we plotted the average latitude and longitude pairs for each classification group (true positive, true negatives, false positives, and false negatives). In doing this, we were able to plot the coordinates on Google Earth to provide insight as to which locations are being misclassified at higher rates compared to correctly classified locations. Figure 6 plots the classification groups on Google Earth. From this map, we note that true negatives are much further east than any other class. To address this going forward, we can use sampling methods to boost the sampling of the FP and FN regions within the dataset. This would influence the model parameters and loss calculation to converge on values that represent these misclassified groups more successfully.

Figure 6: Classification groups using the Random Forest predictions plotted on Google Earth

6 Conclusions

Machine learning as a technique is increasingly being used to help solve complex problem statements in a variety of fields. Many modern state of the art machine learning models have shown high performance in classifying wildfires using daily weather data, however these models suffer from being non-robust, and utilizing binary classifications as the primary metric. Our proposed solution utilizes a more robust data-source in the NASA DAV API, as well as shifting the primary metric to AUC-ROC to evaluate model performance. This more closely aligns with the problem context of wildfires, which generally occur due to human behavior when weather conditions allow for a high risk of wildfire outbreak. Additionally, our methodology for the data extraction step allows for more balanced data to be used in training the machine learning models we employed for the problem statement.

Our approach of leveraging the NASA DAV API data-source allows for the construction of machine learning models that provide state of the art results. Moreover, our models are more applicable to diverse locations due to the leveraged dataset being more representative of under-served regions of the United States, where localized weather data may not be as easily accessible. The use of these models can help protect vulnerable communities that can still be greatly harmed both in terms of property and personal damages in the case of wildfire outbreak. It is our goal that these under-served communities will be able to utilize the models built through our approach to properly prepare for possible wildfire outbreaks, potentially reducing the cost burden of these disasters, and potentially even saving their lives if an evacuation is necessary.

While our results are promising, it is important to note the limitations our models present. Our models were only trained and tested on wildfire data originating from the continental United States. We only utilized data from the continental United States due to the availability of ground truth data within the country. In the future, we hope that these models can be expanded upon to leverage new data-sources that contain wildfire information from other regions of the world, particularly those that are under-served in the context of wildfires. This would allow the model to become even more robust, and continue to help reduce the economic and human cost that sudden wildfire outbreaks pose.

Furthermore, we can use the results from this experiment to show that with enough generalized data, even niche and highly specific phenomena can be well classified using a machine learning model. This proves that selecting a robust data-source with large class balance can be the key to identifying a potential high performing machine learning model. Many proposed state of the art machine learning models fall into the trap of utilizing data that is too specific to be generalized outside of the instances it was trained over. The methodology in this paper strives to showcase the importance of avoiding this pitfall to make a successful machine learning models.

References

[1] National Wildfire Coordinating Group (NWCG) [n. d.]. *Critical Fire Weather*. National Wildfire Coordinating Group (NWCG). https://www.nwcg.gov/publications/pms437/weather/critical-fire-weather Accessed: April 17, 2024.
[2] Amparo Alonso-Betanzos, Oscar Fontenla-Romero, Bertha Guijarro-Berdiñas, Elena Hernández-Pereira, María Inmaculada Paz Andrade, Eulogio Jiménez, Jose Luis Legido Soto, and Tarsy Carballas. 2003. An Intelligent System for Forest Fire Risk Prediction and Fire Fighting Management in Galicia. *Expert Systems with Applications* 25, 4 (2003), 545–554. https://doi.org/10.1016/S0957-4174(03)00007-X
[3] NASA. 2024. NASA DAV Dataset Methodology. https://power.larc.nasa.gov/docs/methodology/
[4] National Interagency Coordination Center. 2024. Wildland Fire Summary and Statistics Annual Report 2024. https://bit.ly/3XtNQ5W
[5] Binh Thai Pham, Abolfazl Jaafari, Mohammadtaghi Avand, Nadhir Al-Ansari, Tran Dinh Du, Hoang Phan Hai Yen, Tran Van Phong, et al. 2020. Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction. *Symmetry* 12, 6 (2020), 1022. https://doi.org/10.3390/sym12061022
[6] George E. Sakr, Imad H. Elhajj, George Mitri, and Uchechukwu C. Wejinya. 2010. Artificial Intelligence for Forest Fire Prediction. In *2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*. IEEE, 1311–1316. https://doi.org/10.1109/AIM.2010.5695921
[7] Younes Oulad Sayad, Hajar Mousannif, and Hassan Al Moatassime. 2019. Predictive Modeling of Wildfires: A New Dataset and Machine Learning Approach. *Fire Safety Journal* 104 (2019), 130–146. https://doi.org/10.1016/j.firesaf.2019.01.007
[8] Sanjeev Sharma and Puskar Khanal. 2024. Forest Fire Prediction: A Spatial Machine Learning and Neural Network Approach. *Fire* 7, 6 (2024), 205. https://doi.org/10.3390/fire7060205
[9] Statista. 2024. Number of Deaths Due to Wildfires in the U.S. (1990-2023). https://www.statista.com/statistics/1422130/usa-number-of-deaths-due-to-wildfires/

6.1 Meeting Attendance

- Tuesday, April 8th 6pm-7pm . Attendees: Nathan Ress, Lily Smith, Ben Mellin, Uchswas Paul
- Tuesday, April 15th 6pm-7pm . Attendees: Nathan Ress, Lily Smith, Ben Mellin, Uchswas Paul
- Thursday, April 17th 6pm-7pm . Attendees: Nathan Ress, Lily Smith, Ben Mellin, Uchswas Paul
- Tuesday, April 22nd 6pm-7pm . Attendees: Nathan Ress, Lily Smith, Ben Mellin, Uchswas Paul