Predicting Fire Severity Classes Within Historical Wildfire Burn Perimeters in California

Chanyoung Park
chp025@ucsd.edu

Jahnavi Naik
janaik@ucsd.edu

Reema Alsaeed
ralsaeed@ucsd.edu

Rushyendra Katabathuni
rkatabathuni@ucsd.edu

Ilyes Meftah
imeftah@cotality.com

Abstract

In recent years, California has experienced unprecedented levels of frequency and intensity of wildfires. Climate change and various environmental factors, such as droughts and high winds, have been linked to the severe burning of vegetation and destruction of property in regions of high risk. The more traditional wildfire severity assessment relies heavily on expensive and time-intensive satellite imagery analysis. This brings us to the focus of this paper, where we will construct a model to predict fire severity classes within historical burn perimeters using readily available geospatial, fuel, topographic, and weather data instead. This has significant implications for real-time severity assessment during active fire events, retrospective analysis where satellite data is unavailable, and cost-effective severity modeling for insurance risk assessment. We show how readily available data can be used to train logistic regression, random forest, and neural network models, evaluating their performance against ground truth wildfire severity labels.

Code: https://github.com/ChanyoungPark07/wildfire-risk

1 Introduction

Wildfire severity has become an increasingly important issue in California as recent fires have produced high-intensity burns and caused significant threats to nearby communities. Understanding and predicting wildfire severity is essential for informing fire response, supporting communities, and guiding insurance risk assessment. However, wildfire severity assessment traditionally relies on satellite imagery analysis, such as pre- and post-comparisons. This process is costly, computationally intensive, time-consuming, and often unavailable during active fire events, because it relies on obtaining post-fire satellite imagery.

In this project, we explore whether fire severity classes within historical burn perimeters can be predicted using readily available geospatial, fuel, topographic, and weather data. Our models indicate that environmental and geospatial variables can moderately predict burn severity classes, with Neural Networks outperforming Logistic Regression and Random Forest models. However, overall performance remained limited, which suggests that further investigation may be needed. For example, performance may improve by incorporating additional variables, such as features from wildland-urban interface (WUI) areas, or using more expressive model architectures. This approach nonetheless has important implications, including providing severity assessment during active fire events, supporting analysis where satellite data may not be available, and offering cost-effective severity modeling for insurance risk assessment.

Recent efforts to understand and predict wildfire behavior in California highlight the changing wildfire patterns and the potential for data-driven severity modeling. For example, Spatial and temporal pattern of wildfires in California from 2000 to 2019 explains the significant shift in California wildfires, including an increased frequency of small human caused wildfires and changing spatial patterns of wildfire occurrence (Li and Banerjee (2021)). Their multivariate analyses identified key environmental variables, such as temperature, vapor pressure deficit, and vegetation cover, that influence wildfire occurrence risk.

More recently, modeling wildfire burn severity in California using a spatial Super Learner approach used an ensemble of predictive models with a spatial Gaussian process to predict burn severity using pre-fire ecological and topographic data from four major California wildfires (Simafranca et al. (2023)). Their model outperformed standard linear regression models and proved important predictors of wildfires such as vegetation, weather conditions, and elevations.

These studies together demonstrate an increased need for robust and predictive fire-risk models and severity tools as a result of wildfire frequency and condition changes, and that environmental, spatial, and weather conditions are useful variables that can be used in predicting wildfire severity. These studies, along with others that have been done previously, often use satellite imagery models that tend to be time intensive and expensive, making it hard to access and produce. Our efforts use lighter machine learning models that use readily available public data. Creating this model would allow for real time, cost effective, severity assessment and analysis when satellite data is unavailable.

The datasets in this study provide high resolution spatial and environmental information necessary to predict the wildfire burn severity class at a pixel level for wildfires in California from 2018-2024. The historical fire perimeters and burn severity data from Monitoring Trends in Burn Severity (MTBS) provide ground truth data that can be used for statistical based modeling, a common wildfire modeling method that involves correlating and quantifying the relationships of historical wildfire observations and the various environmental conditions that led to the fire. Environmental data from LANDFIRE, including data such as fuel models and vegetation help capture conditions that influence wildfire behavior. Land cover information from the National Land Cover Database (NLCD) and weather data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) further help to understand the landscape and weather conditions around the fires. By integrating these diverse data sources, we are left with a pixel level dataset with comprehensive environmental, topographic, and meteorological variables, enabling the development of predictive models for wildfire burn severity.

2 Methods

For this study, we selected wildfire events in California occurring between 2018 and 2024, and integrated multiple spatial and environmental datasets to comprehensively capture the fires extent, severity, and conditions influencing fire behavior. The MTBS Fire Perimeters dataset provided polygon geometries that represent each individual wildfire, along with the burn dates and acres burned (MTBS (2025a)). Vegetation and topography data were provided by LANDFIRE, which includes fuel models, elevation, slope, aspect, and vegetation variables as 30 meter resolution raster layers (LANDFIRE (2025)). We also obtained land cover data from the National Land Cover Database, which contains classifications such as developed, forest, shrubland, etc., in a 30 meter resolution raster layer (Multi-Resolution Land Characteristics Consortium (2025)). Monthly weather variables for the fire's ignition month came from the PRISM dataset, including temperature, humidity, vapor pressure deficit, and windspeed, in a 800 meter resolution (PRISM Climate Group, Oregon State University (2025)). Finally, our target variable, burn severity classification, came from the MTBS Fire Severity, providing classifications unburned, low, moderate, and high for each pixel in the fire perimeter at a 30 meter resolution (MTBS (2025b)).

All spatial datasets were reprojected to a common metric coordinate reference system, EPSG:5070, to ensure accurate and consistent spatial calculations. For each fire perimeter, a bounding box extending 500 meters beyond the perimeter was created to capture the surrounding environmental conditions of the fire. Raster values for fuel type, slope, aspect, elevation, vegetation type, and land cover within the fire boundaries were extracted from the respective datasets to create input variables. Monthly weather statistics, including max temperature, total precipitation, humidity, etc., were extracted based each fire's ignition month. Wildfire burn severity classification, the target variable, was then included to provide the classification for each pixel within the fire. To prepare the data for modeling, multiple preprocessing steps were taken. Categorical variables such as fuel type, vegetation type, and land cover class were one-hot encoded, transforming them into numerical inputs suitable for machine learning models. The original categorical variables were then removed after the one-hot encoding to remove redundancy in the data. Additionally, with aspect being a circular variable (0°-360°), sine and cosine transformations were performed to preserve the data while allowing the model to correctly interpret the directions. Lastly, all continuous variables were standardized to ensure comparable scales across all features. The final dataset contained one observation per 30 meter pixel, with all the input features and the corresponding fire severity class, allowing pixel level classification of fire severity to be processed.

All models in the study used the same preprocessing and train, validation, and test split. To ensure spatial independence, splitting was performed on the fire event level, randomly partitioning unique Event_IDs into training (70%), validation (15%) and testing (15%). After preprocessing the data mentioned above, the non-predictive columns such as Event_ID and the spatial coordinates X and Y were removed.

First a multivariate logistic regression classifier was trained using the one-vs-rest method to predict burn severity. The model was trained using the liblinear solver with balanced class weights to address the class imbalance in the dataset, and the hyperparameters were optimized to obtain the best performing model. Since logistic regression provides a coefficient for each feature and class, feature importance were extracted by taking the top 20 maximum absolute coefficients across all classes and features. Then, a second logistic regression model was trained using only these features using the same hyperparameters, and the validation performance was assessed.

The next model trained was the Random Forest model. To tune the model, a range of different hyperparameter combinations was explored, focusing on maximum tree depth, number of trees, and the minimum number of samples required per leaf. Due to the size of the dataset, hyperparameter tuning was conducted on a randomly selected subset of the training dataset, which reduced computational cost while maintaining variability across fires.

Additionally, the class balancing parameter was held constant to address the imbalance across the severity classes and parallelization was used to speed up the training process. These two parameters were held constant across all parameter configurations and each parameter combination was trained on the subset and evaluated on the validation dataset using the accuracy metric. The configuration with the highest validation accuracy was selected and then used on the train dataset to create a Random Forest model with the best parameter configuration found.

To interpret the results from the Random Forest model, feature importance were extracted, all features were ranked in descending order, and the top 20 features were selected. Both the top 20 ranked features and the trained Random Forest model were saved to allow reproducibility.

The final model was the Neural Network where different model architectures, such as the number of layers and neurons, and parameters, such as the activation function and dropout rate, were tested on a validation set through k-fold cross validation. After the best model on the validation wildfire event set was selected, the final model was trained on the best parameters and evaluated on the test wildfire events.

3 Results

Logistic Regression with Top 20 Features

		precision	recall	f1-score	support
1	Unburned to Low	0.029401132666586335	0.009026005252839123	0.013811841956300237	27033.0
2	Low	0.3320864028702928	0.6031138172301448	0.428327311693361	187230.0
3	Moderate	0.41208614593378334	0.007136018524703316	0.014029097793317027	179652.0
4	High	0.14850220104801956	0.1903120380252839	0.16682744273079114	150847.0
5	macro avg	0.2305189706296705	0.20239696975824278	0.15574892354344236	544762.0
6	weighted avg	0.2926131959269688	0.26278448203068494	0.1987195744606439	544762.0
7	accuracy	0.26278448203068494			544762.0

Figure 1a: Classification report for Logistic Regression model using only the top 20 features showing precision, recall, F-1 scores for each class and overall accuracy

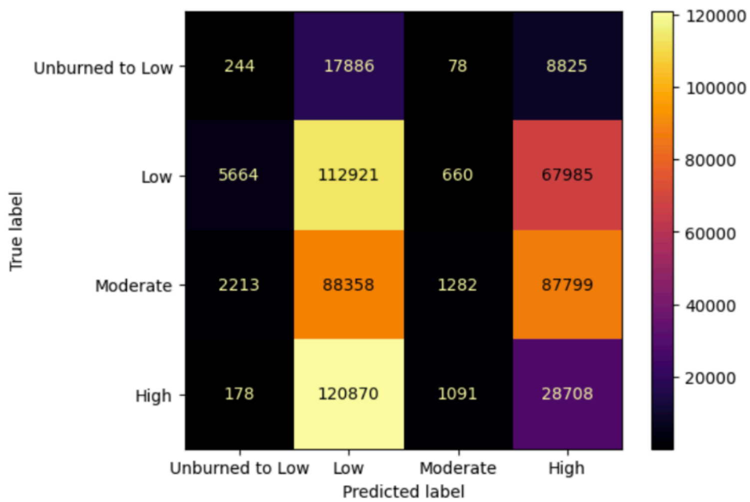

Figure 1b: Confusion Matrix for Logistic Regression model using only the top 20 features showing predicted vs. observed severity classes

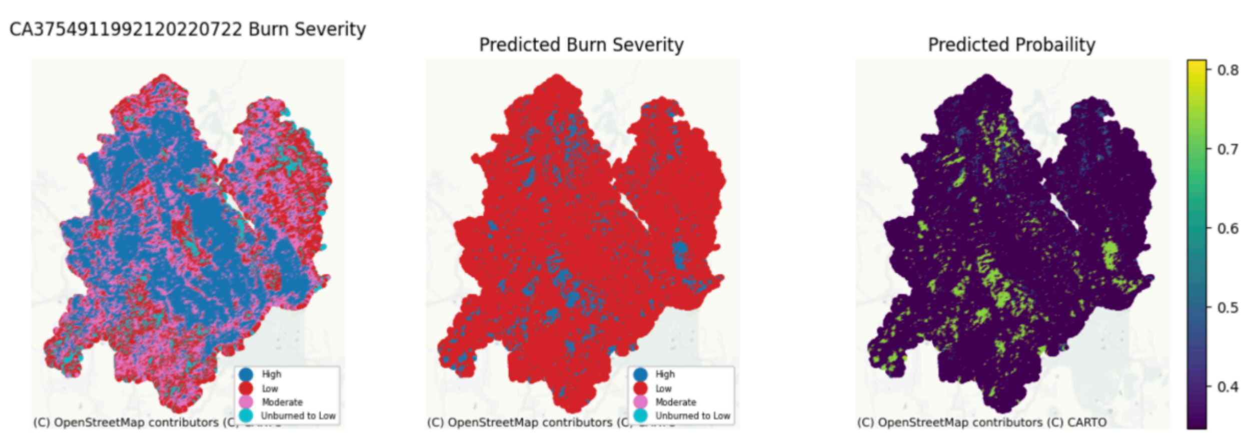

Figure 1c: Map showing true burn severity (left) vs. predicted burn severity (middle) for one fire using a Logistic Regression model with only the top 20 features. The predicted probability (right) represents the confidence of the model on the predicted class

Full Logistic Regression

		precision	recall	f1-score	support
1	Unburned to Low	0.08226824501898065	0.27096511670920725	0.12621585064314084	27033.0
2	Low	0.42311219617094753	0.35729850985419004	0.38743024929271985	187230.0
3	Moderate	0.2410137526502933	0.16261438781644513	0.19420005650374753	179652.0
4	High	0.30037867622049386	0.351269829694989	0.32383705473780067	150847.0
5	macro avg	0.2616932175151788	0.28553696101870785	0.25792080279435226	544762.0
6	weighted avg	0.31216031738566186	0.2871419078423238	0.29313504872321533	544762.0
7	accuracy	0.2871419078423238			544762.0

Figure 2a: Classification report for Logistic Regression model using all features showing precision, recall, F-1 scores for each class and overall accuracy

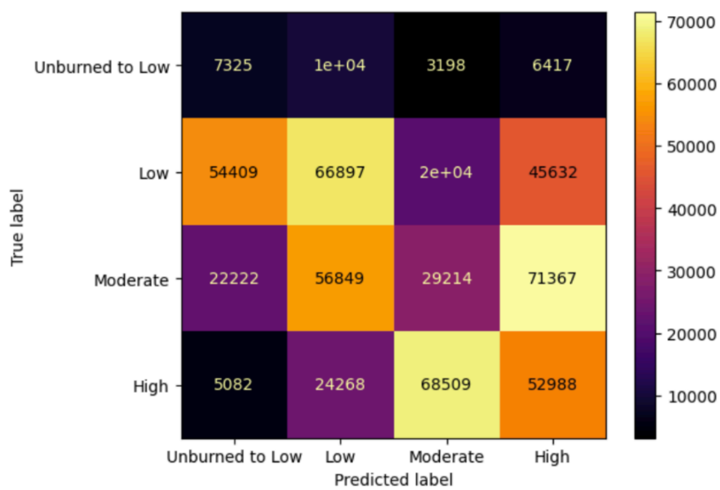

Figure 2b: Confusion Matrix for Logistic Regression model using all features showing predicted vs. observed severity classes

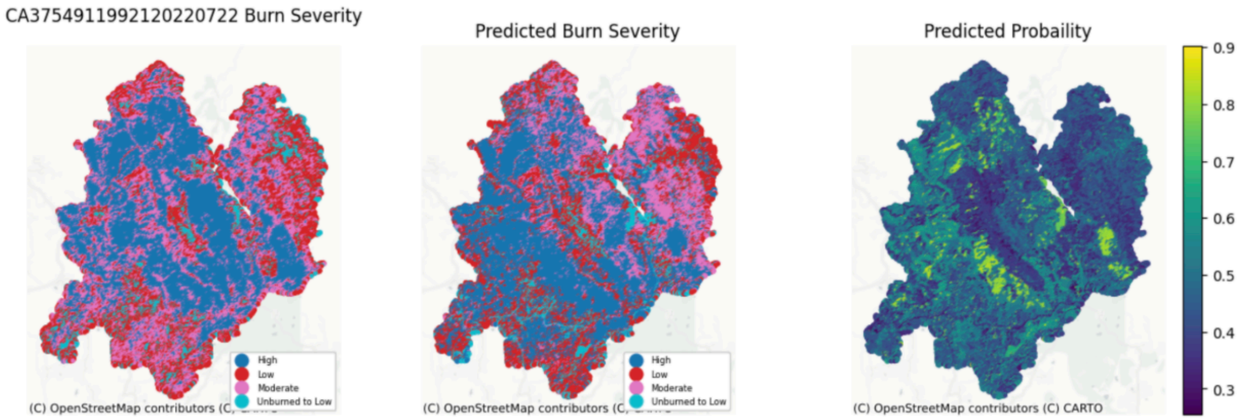

Figure 2c: Map showing true burn severity (left) vs. predicted burn severity (middle) for one fire using a Logistic Regression model with all features. The predicted probability (right) represents the confidence of the model on the predicted class

The logistic regression model with only the top 20 most influential features achieved an overall accuracy score of 26%. This indicates that although the top features captured meaningful predictors, important information was still lost during feature reduction. The feature importance analysis showed that vegetation variables such as EVT_7135, EVT_7137, and EVT_9810, had the strongest coefficients, indicating that vegetation contributed most to predicting burn severity. Both the confusion matrix and recall values show that the model struggles to predict Class 1 and 3 correctly. The logistic regression model with all features present achieved a slightly higher overall accuracy score of 29%, and performance varied across the four burn severity classes. Class 1 (Unburned to Low) and Class 3 (Moderate) had the lowest precision and recall, respectively. In contrast, Class 2 (Low) had both the highest precision and highest recall. Overall, the logistic regression model had a higher weighted than macro F1 score, indicating that it learned the larger classes better while struggling with the smaller classes. The burn severity maps for this model also showed a closer match than the model with only the top 20 features.

Random Forest

		precision	recall	f1-score	support
1	Unburned to Low	0.058776116028643466	0.44238523286353715	0.10376572668112798	27033.0
2	Low	0.35008587437002	0.3320514874752978	0.34083028384249553	187230.0
3	Moderate	0.29800954342673197	0.16025983568231916	0.20843182040302174	179652.0
4	High	0.6367904141641455	0.28325389301742826	0.3920971249759115	150847.0
5	macro avg	0.33591548699738527	0.3044876122596456	0.2612812389756392	544762.0
6	weighted avg	0.39784604488569186	0.2673607924194419	0.2995998644242556	544762.0
7	accuracy	0.2673607924194419			544762.0

Figure 3a: Classification report for the Random Forest model showing overall accuracy, precision, recall, and f1-score for each severity class

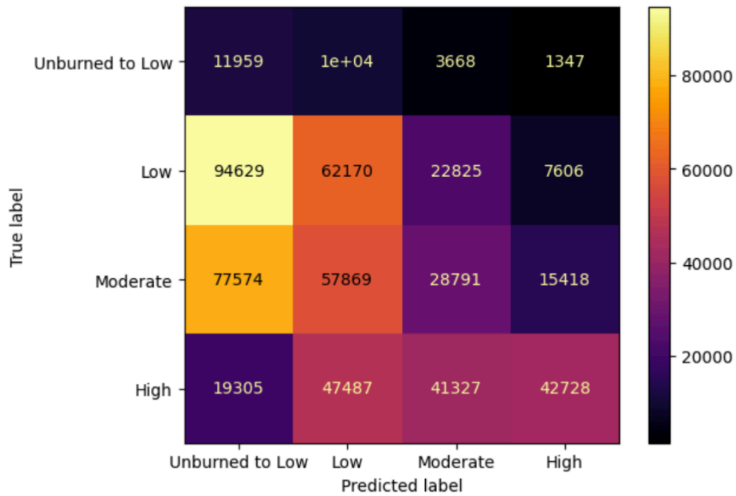

Figure 3b: Confusion matrix for the Random Forest model showing predicted vs. observed severity classes

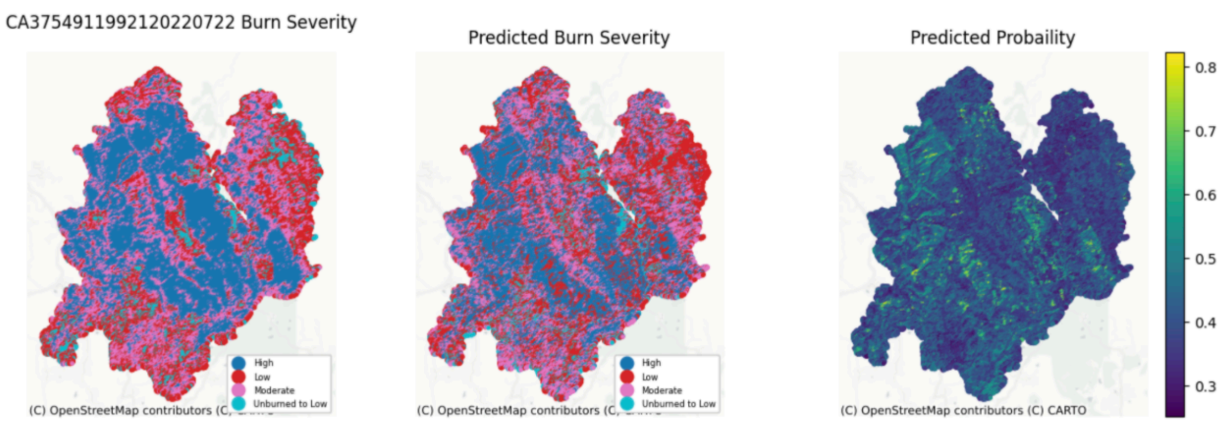

Figure 3c: Maps showing true burn severity (left), Random Forest predicted severity (middle), and predicted probability (right) for a specific fire

The Random Forest model achieved an overall accuracy of 26.7% on the full test dataset. More specifically, performance varied across the four burn severity classes. As shown in Figure 3a, Class 4 (High) had the highest precision at 0.63, while Class 1 (Unburned to Low) had the lowest precision at 0.05. For recall, the highest was Class 1 at 0.44, and the lowest was Class 3 (moderate) at 0.16. The F1-score was 0.1, 0.34, 0.2, and 0.39 for classes 1, 2, 3, and 4, respectively.

The confusion matrix in Figure 3b shows that predicted classes are concentrated in the lower severity classes. For instance, many true “Low” pixels are predicted as “Unburned to Low,” and many true “Moderate” pixels are predicted as “Unburned to Low” or “Low.” Overall, among the four classes, Class 2 (Low) had the highest number of correct predictions, with 62,170, while Class 1 (Unburned to Low) had the fewest correct predictions, with 11,959.

In Figure 3c, one specific fire is focused on and the model’s predictions for this fire are shown. The left panel displays the true burn severity classes for this fire, the middle panel shows the Random Forest's predicted burn severity classes, and the right panel displays the predicted probability for the predicted class. Based on the right panel, the model has moderate predicted probability for most of the fire area, with small regions showing higher predicted probability.

Lastly, a feature importance analysis was conducted to understand which variables contributed the most to the Random Forest model. The top-ranked features included elevation (LC20_Elev_220), existing vegetation height (LC24_EVH_250), and slope (LC20_SlpD_220) as the top three most important features for this model. Other influential features included aspect in both sine and cosine transformations (asp_sin and asp_cos), temperature and moisture variables (tdmean, tmax, vpdmax, and ppt), and fuel type categories (such as FBFM13_FBFM8, FBFM13_FBFM5, FBFM13_FBFM9). Overall, based on the feature importance analysis, the Random Forest model seems to heavily rely on topographic, vegetation, and fuel type variables, as they had some of the highest importance scores in this model.

Neural Network

		precision	recall	f1-score	support
1	Unburned to Low	0.2541958041958042	0.02689305663448378	0.04864014986786204	27033.0
2	Low	0.4228108087910172	0.7252256582812584	0.5341872944434828	187230.0
3	Moderate	0.29481076524477173	0.26475074032017454	0.27897333028335475	179652.0
4	High	0.5375618457810238	0.21175760870285787	0.3038298560415468	150847.0
5	macro avg	0.37734480600315423	0.30715676598469366	0.2914076576590616	544762.0
6	weighted avg	0.40400666392489043	0.3965346334729662	0.3621411099459868	544762.0
7	accuracy	0.3965346334729662			544762.0

Figure 4a: Classification report for the Neural Network model showing overall accuracy, precision, recall, and f1-score for each severity class

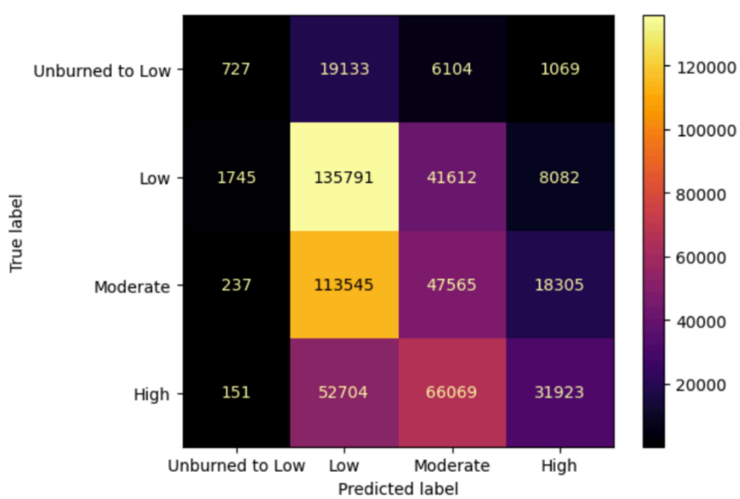

Figure 4b: Confusion matrix for the Neural Network model showing predicted vs. observed severity classes

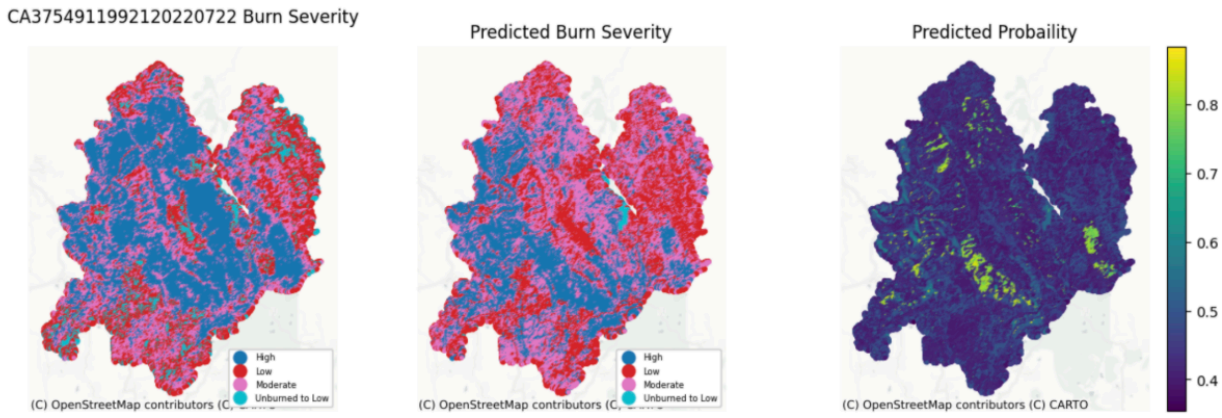

Figure 4c: Maps showing true burn severity (left), Network model predicted severity (middle), and predicted probability (right) for a specific fire

The neural network performed moderately well across the four metrics (Figure 4a) with some classes (Low and High) being more accurately predicted over the others (Figure 4b). Looking at the weighted averages, the neural network has 0.397 accuracy, 0.404 precision, 0.397 recall, 0.362 f1-score. The feature importance analysis showed that variables features, FBFM13_FBFM8, tmax, FBFM13_FBFM9, FBFM13_FBFM5 had the strongest coefficients, indicating that temperature and vegetation features contributed most to predicting burn severity.

4 Discussion

The three models had varying levels of success in predicting wildfire burn severity classes within California wildfire perimeters using the fuel, topography, weather, and land cover data. The Logistic Regression model trained using all of the preprocessed features achieved an overall accuracy of 28.7%, while the Logistic Regression model with the top 20 features performed slightly worse with an accuracy of 26.2%. Additionally, the Random Forest model with hyperparameter tuning was able to reach an accuracy of 26.7%, and the Neural Network was the best performing model with an accuracy of 39.6%. This implies that the relationship between environmental factors and burn severity might be non-linear and more complex, which benefits from the neural networks ability to learn complex relationships between features. These results suggest that although our models were able to identify some patterns in predicting burn severity, the problem of predicting burn severity with the data available was still challenging, as our target variable (satellite analysis burn severity classifications) might not interact well with ground-level environmental predictors like the ones used to build the models in this paper.

Although the data used in building our models was comprehensive, the results of the models highlights that burn severity is a complex outcome that can be influenced by a variety of factors and a plethora of ways those factors could interact with each other to amplify or modify wildfire effects. The scope of something as dynamic as burn severity may not be simply predictable using pre-fire and post-fire conditions alone. Wildfire behavior is inherently chaotic and difficult to predict, and this is further compounded by factors such as wind patterns and fire dynamics that weren't able to be predicted extremely effectively from the current set of features.

An important observation across all models is the consistent pattern in class-level performance: all models performed worse at predicting Class 1 (Unburned to Low) and Class 3 (Moderate) burn severity pixels, while performing far better at predicting Class 2 (Low) and Class 4 (High) burn severity pixels. This held true across all of the four models, suggesting that there is a fundamental pattern in how environmental features relate to burn severity.

The Top 20 Feature Logistic Regression model achieved F1-scores of 0.428 and 0.167 for Class 2 and 4 burn severities, respectively. While struggling significantly with Class 1 and 3 pixels, achieving F1-scores of 0.0140 and 0.0138, respectively. The Full Logistic Regression model performed similarly, with its highest F1-score of 0.387 for Class 2 pixels, followed by 0.324 for Class 4, 0.194 for Class 3, and 0.126 for Class 1 burn severity. The Random Forest model continued this pattern with F1-scores of 0.341 for Class 2, 0.392 for Class 4, 0.208 for Class 3, and 0.104 for Class 1 burn severity. Finally, the Neural Network performed the best out of all of the models, with F1-scores of 0.534 for Class 2, 0.304 for Class 4, 0.279 for Class 3, and 0.049 for Class 1.

When observing precision across models, another interesting pattern emerges: all of the models consistently achieved their highest precision values for Classes 2, 3, and 4, while having significantly lower precision for Class 1. For instance, the Random Forest model achieved precision values of 0.058, 0.350, 0.298, and 0.637 for Classes 1, 2, 3, and 4 respectively. This indicates that the models tend to be more accurate in their predictions for burned areas, compared to predicting unburned or minimally burned pixels. This means that when the models confidently predict that a pixel belongs to a burned class, they are often right. However, they struggled to identify unburned pixels accurately. This further suggests that fuel type, land cover and vegetation features are strong predictors of burned pixels, but are not sufficient for identifying what makes areas stay unburned.

Interestingly, the models each found different features to be their strongest predictors of burn severity. For instance, the Logistic Regression model found EVT_7135, EVT_7137, EVT_9810, and EVT_9301 (see Appendix Data Table) to be its strongest predictors. On the other hand, the Random Forest model found LC20_Elev_220, LC24_EVH_250, and LC20_SlpD_220 (see Appendix Data Table) to be its strongest predictors. Additionally, the Neural Network model found FBFM13_FBFM8, tmax, FBFM13_FBFM9, FBFM13_FBFM5 (see Appendix Data Table) to be its strongest predictors. This variance suggests that different models better capture different aspects of burn severity, suggesting that ensemble models that combine different model approaches together could prove valuable.

These models all share one similar pattern: fuel and vegetation data are essential factors in predicting burn severity. These findings corroborate the research done by Li and Banerjee (2021), who identified vegetation cover as key factors in wildfire occurrences. The presence of fuel factors in the group of strongest predictors highlights that fuel and vegetation characteristics directly influence the intensity and spread of a wildfire. Vegetation types like shrubs and grass, which are more susceptible to easy and prolonged burning are far more likely to initiate wildfires, which are what our models found as some of the strongest predictors.

It's important to note that although the research done by Simafranca et al. (2023) found that weather data such as TDMEAN, VPDMAX and VPDMIN were some of their strongest predictors, this may have not translated well into pixel-by-pixel data. The paper suggested that local AM (morning) soil moisture was their strongest predictor of wildfire burn severity, which the weather data used in our models did not contain. Additionally, the weather data was aggregated at a much coarser 800 meter resolution compared to the fuel and vegetation data, which was at 30 meter resolution. This broad 800 meter resolution likely doesn't properly capture smaller scale moisture trends, which could explain why our models didn't identify the weather variables as their strongest predictors. However, the Neural Network identified tmax as one of its strongest predictors, highlighting the potential impact that more fine-grained weather data could have on our models' predictive abilities.

Further expansions could improve the performances and applications of our models. Instead of general severity based on environmental predictors, our models could be expanded to predict structural damage and loss patterns in urban and WUI (Wildland-Urban Interface) areas. By incorporating CalFire Damage Inspection Reports, which contain ground-truth structural damage assessments, our model could identify and predict structural damage patterns and burn severities. Furthermore, by incorporating WUI Boundary Data, the potentially improved model could be used to understand the most common patterns and combinations of features that are likely to initiate wildfires where structures interface with wildland vegetation.

5 Conclusion

This report investigated whether burn severity classes within fire perimeters could be predicted using fuel, topographic, weather, and geospatial data as an alternative to traditional satellite-image analysis. We trained and evaluated the effectiveness of logistic regression, random forest, and neural network models on a comprehensive dataset containing a set of California wildfires from 2018-2024 at 30 meter pixel level observations.

Our findings demonstrate that environmental and geospatial features can be useful for predicting wildfire burn severity classification, although with only moderate accuracies ranging from 26.7% to 39.6%. The important feature analysis identified fuel and vegetation features as some of our models' strongest predictors, corroborating prior research. However, all of our models struggled with class imbalances, generally predicting Class 1 (Unburned-Low) and Class 3 (Moderate) burn severity classes worse than the other burn severity classes.

The key takeaway of our work is that readily available environmental feature data is a valid and complementary tool to traditional satellite-image based burn severity analysis. This alternative method offers many distinct advantages, such as minimal data collection costs, low computational requirements, and potential predictive ability for future fires. These characteristics make this approach better suited for simpler severity assessments and analysis where satellite data is unavailable or expensive.

In conclusion, this project demonstrates the potential and limitations of using publicly available environmental data for wildfire severity prediction. Although high prediction accuracy was difficult to attain due to the chaotic and complex nature of wildfire behavior, our approach provides actionable severity estimates at a fraction of the cost and time as satellite imagery. As California continues to face increasing wildfire frequency and intensity due to climate change, cost-effective and accurate assessment tools will become increasingly important for managing fire-prone areas and protecting communities.

References

LANDFIRE. 2025. "LANDFIRE WCS/WMS." https://landfire.gov/data/lf_wcs_wms

Li, S., and T. Banerjee. 2021. "Spatial and temporal pattern of wildfires in California from 2000 to 2019." *Scientific Reports* 11, p. 8779

MTBS. 2025a. "Direct Download MTBS." https://www.mtbs.gov/direct-download

MTBS. 2025b. "Frequently Asked Questions (FAQ) MTBS." https://mtbs.gov/index.php/faqs

Multi-Resolution Land Characteristics Consortium. 2025. "MRLC Data." https://www.mrlc.gov/data

PRISM Climate Group, Oregon State University. 2025. "PRISM Weather Data." https://prism.oregonstate.edu/

Simafranca, N. et al. 2023. "Modeling wildland fire burn severity in California using a spatial super learner approach." *Environmental and Ecological Statistics* 31 (2): 387–408

Appendices

A.1 Training Details

For Logistic Regression, to optimize the strength of L2 regularization, a grid search over the inverse regularization parameter was conducted:

$$C \in \{0.01,\ 0.1,\ 1,\ 10\}$$

This hyperparameter tuning was done using GridSearchCV, with 3-fold cross validation on the training dataset, and the macro-F1 score was used as the optimization metric so that performance was balanced across all severity classes. The grid search returned the best performing regularization parameter, C, and the tuned model was evaluated on the validation set, producing a classification report including metrics such as accuracy, F1, precision, and recall.

For the Random Forest model, when performing hyperparameter tuning, the search included parameter configurations such as:

$$\{n\_estimators : 150, max\_depth : 15, min\_samples\_leaf\ : 1\}$$

$$\{n\_estimators : 100, max\_depth : 20, min\_samples\_leaf\ : 1\}$$

Additionally, since the severity classes were imbalanced, class_weight=”balanced” was set, and n_jobs=-1 was used to enable full parallel processing and speed up the training process. To extract the most important features, feature_importances_ was used, and after ordering them in descending order, the top 20 features were selected. Lastly, both the top 20 ranked features and the trained Random Forest model were saved in .pkl files using joblib to allow reproducibility.

The final trained neural network was selected based on the best architecture and hyperparameters chosen from the 5-fold cross validation on the combined training and validation sets. The number of units for each fully connected layer was n_units = [256, 128, 64, 32, 16, 8], activation was set to “relu”, dropout_rate to 0.3, use_batchnorm to “False” and learning_rate to 0.0001. Early stopping of 10 epochs was used on the validation set with the best epoch step with the best validation loss being extracted as the final model. In order to get the feature importances, each feature was permuted randomly and passed through the model to get an F1 score. The difference between the baseline F1 score and permuted F1 score was taken for each feature and the top 5 features with the largest difference were taken.

A.2 Additional Tables

Data Table

Below is a table including the important variables and descriptions that were used in the model training process and mentioned in the paper:

Variable	Description
FBFM13_FBFM5	Low intensity fires, young, green shrubs with little dead material, fuels consist of litter from understory
FBFM13_FBFM8	Slow, ground burning fires, closed canopy stands with short needle conifers or hardwoods, litter consist mainly of needles and leaves, with little undergrowth, occasional flares with concentrated fuels
FBFM13_FBFM9	Longer flames, quicker surface fires, closed canopy stands of long-needles or hardwoods, rolling leaves in fall can cause spotting, dead-down material can cause occasional crowning
tmax	4km pixel-level temperature max per month
NLCD_52	Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions
EVT_7098	California Montane Woodland and Chaparral Woodland and Shrubland
EVT_7135	Inter-Mountain Basins Semi-Desert Grassland
EVT_7137	Mediterranean California Subalpine Meadow
EVT_9810	North American Warm Desert Ruderal & Planted Grassland
EVT_9301	California Ruderal Grassland and Meadow
LC20_ELEV_220	30 m pixel-level elevation data in meters above sea level
LC24_EVH_250	30 m pixel-level vegetation height
LC20_SlpD_220	30 m pixel-level slope in degrees