Determination of Optimal Set of Spatio-temporal Features for Predicting Burn Probability in the State of California, USA

Javier Pastorino
U.S. Geological Survey
Denver, Colorado, USA
jpastorinogonzalez@contractor.usgs.gov

Joseph W. Director
U.S. Geological Survey
Denver, Colorado, USA
jdirector@contractor.usgs.gov

Ashis Kumer Biswas
University of Colorado, Denver
Denver, Colorado, USA
ashis.biswas@ucdenver.edu

Todd J. Hawbaker
U.S. Geological Survey
Denver, Colorado, USA
tjhawbaker@usgs.gov

ABSTRACT

Wildfires play a critical role in determining ecosystem structure and function and pose serious risks to human life, property and ecosystem services. Burn probability (BP) models the likelihood that a location could burn. Simulation models are typically used to predict BP but are computationally intensive. Machine learning (ML) pipelines can predict BP and reduce computational intensity. In this work, we tested approaches to reduce the set of input features used in an ML model to estimate BP for the state of California, USA, without loss of predictive performance. We used Principal Component Analysis (PCA) to determine the optimal set of features to use in our ML pipeline. Then, we mapped BP and compared model performance when using the reduced set and when using the whole set of features. Models using optimized input achieved similar prediction performance while using less than 50% of the input features.

CCS CONCEPTS

• **Applied computing** → **Environmental sciences**.

KEYWORDS

Wildland Fire, Burn Probability, Feature Selection

ACM Reference Format:
Javier Pastorino, Joseph W. Director, Ashis Kumer Biswas, and Todd J. Hawbaker. 2022. Determination of Optimal Set of Spatio-temporal Features for Predicting Burn Probability in the State of California, USA. In *2022 ACM Southeast Conference (ACMSE 2022), April 18–20, 2022, Virtual Event, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3476883.3520228

1 INTRODUCTION AND BACKGROUND

Wildfires play a critical role in determining ecosystem structure and function and pose serious risks to human life, property, and ecosystem services [3][11][13]. Thus land management agencies are keenly interested in assessing wildfire potential and risk [1][28].

Burn probability (BP) indicates the likelihood of fire for a given region and time interval [9], and has been predicted using simulation models for more than two decades [19]. BP modeling uses weather data, topography, availability of vegetation/fuels, and ignition patterns as basic inputs to estimate fire spread [19], although other input data can also be used. After simulating spread for thousands of individual fires, burn probabilities are then calculated as the mean number of times a pixel burned across all simulations. Modeled BP can be used to assess vulnerability for a given community and estimate hazards and risks from wildfires [28].

Simulating wildfire spread and growth can assist incident command teams to control wildfires [2], and modeling several wildfires in a region can depict landscape-scale patterns of wildfire behavior [19]. However, predicting BP is extremely challenging, because fire behavior varies strongly among landscapes and with changing weather conditions and wildfire spread simulations are computationally intensive and require integration of data with large spatial and temporal variability [1].

Using machine learning (ML) models to predict BP is an alternative approach to simulation models; however these approaches can also be data intensive. Reducing the data volume or number of features these models use can increase computational efficiency. Currently available solutions to estimate BP rely on manual analysis to determine the features to use [21], use all available data [19] or are constrained to a limited time span with limited number of features [17]. The goal of this work was to determine an approach to effectively reduce the number of input features in a BP ML model without reducing model performance. To address that goal, we applied two well known techniques to determine the optimal set of spatio-temporal features for predicting BP while reducing input data requirements.

Feature selection and feature extraction are two techniques used to reduce the size of the inputs in an ML model; thereby reducing its complexity. Feature selection is the process of selecting the best features among all the features that are useful to discriminate classes [15]. This can be achieved, by determining which features are highly correlated to the target class and removing those with low correlation, i.e., keeping those features that provide more information to the classification task. Feature extraction transforms the original features into new, more informative features [15]. Hybrid models that combine both feature selection and feature extraction can also be used to develop an ML pipeline.

ACMSE 2022, April 18–20, 2022, Virtual Event, USA

ACM ISBN 978-1-4503-8697-5/22/04…$15.00
https://doi.org/10.1145/3476883.3520228

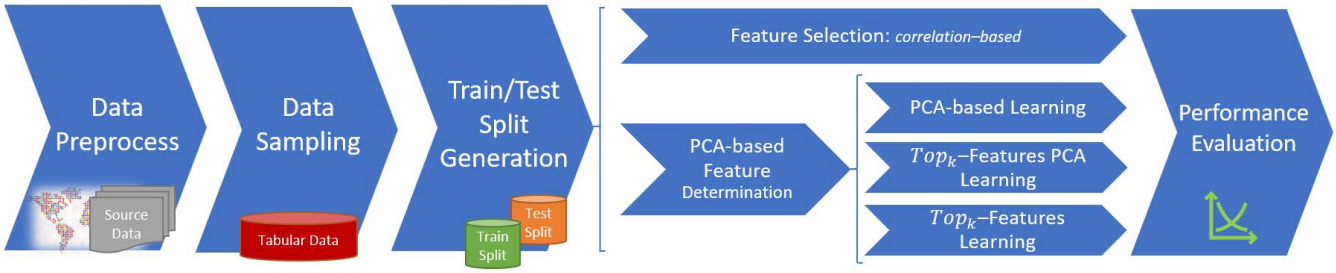

Figure 1: Overall Methodologic Workflow Using Feature Selection by Correlation and Principal Component Analysis (PCA)

1.1 Feature Selection

Identifying which set of features are relevant for a classification task is an important step in extracting knowledge from data [23]. Feature selection improves model performance and efficiency when modeling complex systems with a large number of predictors known to be important but highly correlated. Correlation-based feature selection (CFS) places features in subsets according to the degree of redundancy among the features [15]. Different metrics have been used to determine the significance of the features to select, including chi-squared, Euclidean distance and information gain. In the context of information gain, metrics such as mutual information, are used to analyze the dependence between two random features, and therefore quantifies the amount of information provided about one of the features by knowing the value of the other [7]. Total correlation, introduced by Watanabe [33], extends the idea of mutual information, measuring the amount of information shared by a set of features and as a result measures the interactions among a set of features [29]. The information obtained by these metrics can help identify which features are most relevant to predicting the values of another random variable.

1.2 Feature Extraction

For ML models, it is important that the data are represented in a manner that facilitates analysis. Principle Component Analysis (PCA) is an unsupervised technique used to transform the input features into a set of more meaningful features. It uses orthogonal transformations to convert samples belonging to correlated features into samples of linearly uncorrelated features, or principal components (PC) [15].

1.3 Contributions

We tested a suite of different feature extraction and selection routines. A large number of features related to weather, vegetation, human activity, topography, and disturbance history have been shown to be important to predict wildfire occurrence [12][18][20][22][26]. However, efficient ML models must be able to identify and make use of only the most informative predictors. In the following sections we present the data preprocessing conducted, focusing on the suite of feature extraction and selection routines we tested. Then we present the model implementation, training, and evaluation, and finally we discuss the performance of these techniques towards the goal of predicting BP.

2 METHODOLOGY

The overall workflow used in this study is depicted in Figure 1. We used bagged decision tree classifiers to predict the probability a location burning each month in a wildfire [5] and evaluate different approaches to feature extraction and selection. Given the temporal nature of the data, we used data from 1984 through 2014 as our training dataset and from 2015 through 2019 as the test split. The train split was used for feature selection and model fitting while the test split was used for the final model performance evaluation.

For each approach, we tuned hyperparameters for a bagging classifier using cross validation with the training data split, including the number and depths of trees. Specifically, we used 5–fold time series cross–validation to maintain the temporal order of the data for the training and validation splits within each cross–validation fold as shown in Figure 2.

We tested two general approaches. The first one transformed the data using PCA and used the PC as inputs for the model (PCA-base Learning). The second approach, used these PCA results to select features. Then, we used the set of selected features to train another PCA and used the new PC as model inputs (Top_k-Features PCA Learning) or used the raw features as model inputs (Top_k-Features Learning).

We built a pipeline for each technique, extracting and selecting features, then fully training the bagging classifier, and selecting hyperparameters using the train data split. We conducted a final evaluation using the test data split with area under the receiver operating characteristic curve (ROC-AUC) and F1 scores.

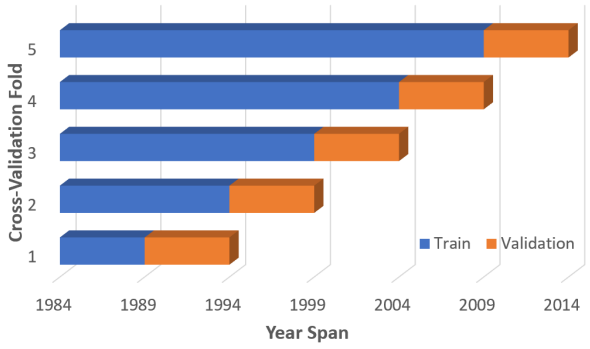

Figure 2: Years Spanned in Each Time Series Cross–validation Split

2.1 Study Area

We selected the state of California, USA (CA) as a study area to test our feature extraction and selection routines. The terrestrial extent of California is 423, 970 km^2, of which 5% is developed land, 11% is cropland, 5% is barren, and the remaining 79% is wildland vegetation (55% grass/shrub, 23% forest, and 1% wetland; [6]). Wildfires occur most years in California's wildland vegetation. Between 1984 and 2019, 1, 687 large fires ($\geq$ 4 km^2) occurred and burned 72, 000 km^2 [8]. These large fires are of concern because of the severe risks they present to human lives, property, and ecosystem services [4][16]. In places like California, where wildfires are common, predictive models of the potential for wildfire occurrence can help communities and government agencies assess and evaluate wildfire risks and target actions to mitigate risk. However, building such predictive models is challenging as there are many areas with wildland vegetation that can potentially burn in wildfires but only a small portion of those places actually burn each year.

2.2 Datasets

We acquired geospatial data layers for past wildfire occurrence and all predictive features considered in this study. A spatial grid within the state's boundaries was constructed using 1 km^2 pixels. We used the Monitoring Trends in Burn Severity (MTBS) burned area dataset for observed burned locations [8]. The MTBS data included fire perimeters for large fires ($\geq$ 4 km^2) from 1984 through 2019. We selected a suite of features to represent weather and landscape patterns that influence BP. We averaged weather-related features such as daily precipitation, daily minimum and maximum relative humidity, daily minimum and maximum temperature, and daily wind speed, for each month from the GridMET dataset [8]. We also calculated 1−, 2−, 3−, 6−, 9−, 12-, 18−, and 24−month lagged means of each monthly weather-related predictor except wind speed.

We also included land cover from the Land Change Monitoring, Assessment, and Projection (LCMAP) primary land cover data for 1985–2019 [6]. Because LCMAP data were not available for 1984, we used the 1985 LCMAP data for that year. We assigned the nearest National Forest Inventory and Analysis (FIA) forest type group [27] to pixels classified as forest in the LCMAP data to represent variability in forest types. We excluded water, snow/ice, or barren land cover in the LCMAP primary land cover data [6] from all analyses. Additionally, we simplified the LCMAP land cover data to an additional layer representing wildland vegetation (grass/shrub, forest, or wetland). To represent fuel connectivity and potential for large fire spread, we calculated the density of wildland vegetation within 2, 4, 6, 8, 10, and 12 km of each pixel.

We added several features to represent potential human influences on ignitions, including a wildland-urban interface (WUI) category, distance to edges of protected areas, wilderness areas, and developed areas, and distance from powerlines, railroads, and roads [6][30–32]. Other landscape-related features such as topography, including slope, aspect, and elevation, and previously burned areas were also included in the input data. Finally, we included indicators of burning within the previous 5−, 10−, and 15−years using MTBS perimeters. In total, we considered 155 features.

We standardized or scaled continuous non-weather features to z-scores by subtracting their mean and dividing them by their standard deviation. We scaled weather-related predictors to z-scores based on the mean and standard deviation of each pixel's time series. We converted all categorical features to binary indicators. We resampled all data to 1-km resolution using nearest-neighbor resampling for categorical variables and bilinear interpolation for continuous features. Given the imbalance in the data towards non-events (i.e., 2, 820 non-events for every fire occurrence), we applied an under–sampling strategy. Thus, for any given month m, we determined ne_m, the number of events in that month, and included all events in the sample. In addition, we randomly sampled $max(ne_m, C)$ non-events for every month m, with C a constant that was determined based on the monthly average of events. Hence every month contained data samples whether events were observed or not.

The sampled dataset contained 155, 038 samples of which 71, 603 were events and 83, 435 were non-events. Of these, 116, 926 (75%) were used for training and the remaining 38, 112 were used for testing.

2.3 Proposed Methodology

2.3.1 ***Feature Selection by Correlation.*** For the first pipeline, we used correlation among features to determine which were redundant and could be removed from the set of input features prior to fitting a bagging classifier.

Figure 3 shows the pipeline used to determine the most relevant features using correlation and the bagging classifier (*CLF*). Algorithm 1 depicts the correlation analysis algorithm to select the features including the final set of features with low correlations.

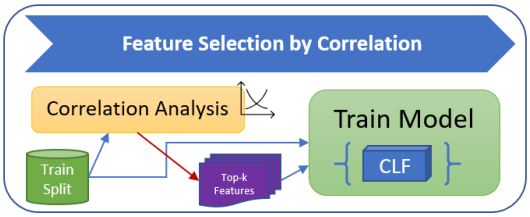

Figure 3: Feature Selection by Correlation

Features were combined into three groups based on their similarities. Monthly weather features were placed in the first group and a second group was formed by features not related to weather (elevation, distance to roads, etc.). Using the training data split, we analyzed the CLF prediction score for each feature in each group with respect to the other features in the same group. We retained the feature with the lower logloss score, and removed those that were highly correlated to those retained ($correlation > 0.7$), thus reducing the size of features to fit the bagging classifier.

Once the initial groups of features were obtained, we considered lagged weather features. For each weather feature (e.g., temperature, humidity, etc.), a new group with the 1−, 2−, 3−, 6−, 9−, 12−, 18− and 24−month lags was generated and analyzed for correlations.

2.3.2 ***PCA-based Traditional Feature Extraction Pipeline.*** The second pipeline incorporated a PCA to transform the original data to reduce dimensionality while maximizing the variance

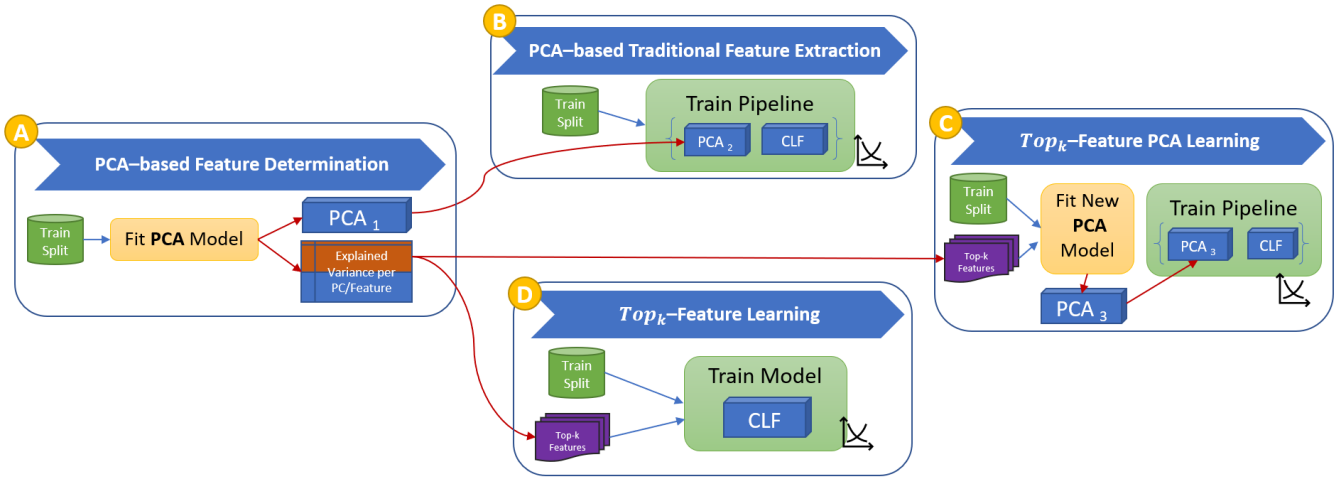

Figure 4: Feature Determination by Principal Component Analysis (PCA)

Algorithm 1: Correlation Analysis

```
Data: List of feature sets, threshold
Result: Reduced dimension feature set
begin
    for a_set in list_feature_sets:
        for var in a_set:
            compute_CV_prediction_score(var)
            for var2 in list_feature_sets:
                if var != var2:
                    compute_CV_prediction_score(var, var2)
        for var in a_set:
            corr_preds =
              get_corr_features_for(var,threshold)
            a_set.remove(corr_preds)
    return list_feature_sets
```

explained by each PC. Using all samples and features, except the response (class) in the training data split, we fit a PCA model, PCA_1, as shown in Figure 4 step **A**. We determined t, the number of components required to explain 95% of the data variance using the PC obtained by PCA_1. Then, we incorporated, as the first stage in the pipeline, PCA_2 model to transform the dimensional space of the data using t principal components, and the bagging classifier (CLF) model as the second stage (Figure 4 step **B**.).

After fitting and hyper–parameter tuning for the pipeline, summarized by steps **A** and **B** in Figure 4, we evaluated the pipeline using the test data split.

2.3.3 Top_k***–Feature PCA Learning.*** The third pipeline we tested used the same fitted PCA model as in Section 2.3.2, PCA_1. The PC were ordered based on the variance they explained. For each PC, we then determined which feature contributed most to the transformation by analyzing the PC eigenvectors and producing a mapping between the PC and the input features. Algorithm 2 depicts the method used to this end.

Using the mapping between PC and input features, and the order obtained before for the PC, we selected the top-k features required to explain 95% of the cumulative variance. We projected the train split with these top-k features and using this projection we fitted a new PCA model, PCA_3. We included PCA_3 as the first step in the pipeline to obtain the PC and as second step the same bagging classifier described in Sections 2.3.1 and 2.3.2. After fitting and tuning hyper–parameter for the pipeline summarized by steps **A** and **C** in Figure 4, we evaluated it using the test data split.

Algorithm 2: Most Significant Features for Principal Components

```
Data: PCA model, original features
Result: List of most significant features
begin
    variance_csum = np.cumsum(pca.explained_var_ratio_)
    pca_pcs = pca.components_
    pcs = range(pca_pcs.shape[0])
    main_featur = [np.abs(pca_pcs[i]).argmax() for i in pcs]
    main_names = [features[main_featur[i]] for i in pcs]
    items = list()
    for i in pcs::
        items.append(i,main_names[i], variance_csum[i] )
    return items
```

2.3.4 Top_k***–Feature Learning.*** For the fourth, and final pipeline, we used the same top-k features as determined in Section 2.3.3. Then, we used the untransformed data for these top-k features from the training split as input to train and tune the hyper–parameters for the bagging classifier. The pipeline is summarized by steps **A** and **D** in Figure 4. Finally, we evaluated the performance using the test data split.

2.4 Burn Probability Spatial Mapping

After the best performing pipeline was identified, BP was predicted using the selected pipeline and mapped for each month from 2015 – 2019. Additionally, as a baseline for BP, we trained the same bagging classifier using all available features. Burn probability was calculated as the average of the inverse logit of the sum of the correction factor (ϵ) and the predicted probability (p) of every estimator in the bagging classifier as shown in Equation 1 [12][14].

$$BP = mean\left(\frac{1}{1 + e^{\epsilon + \ln\left(\frac{p}{1-p}\right)}}\right) \quad (1)$$

The correction ϵ was computed as $\ln(\alpha / \beta)$, the log ratio between non-events sample rate $\alpha = NE_{sample}/NE_{population}$ and events sample rate $\beta = E_{sample}/E_{population}$. With NE_i, E_i the number of *non-events* and *events* in sample i respectively.

We mapped the BP of each pixel in the map using both the baseline and the selected pipeline with reduced input. Then, we computed the difference between the prediction of both models for visual comparisons.

3 EXPERIMENTAL RESULTS

3.1 Feature Determination

Feature selection by correlation (pipeline 1) selected a total of 28 features of which 16 were related to weather. Elevation and terrain aspect were included but not slope. Selected features related to human development included housing density, and distance to power lines and roads. Indicators of fire occurrence within the past 5–, 10–, and 15–years were also included.

Figure 5 shows the cumulative variance explained by the fitted PCA_1 model components described in Section 2.3.2. From this, we determined that 69 PC explained 95% of the data variance. Thus, PCA_2 in pipeline 2 (Section 2.3.2) was set to use 69 PC.

Using the information provided by the first 69 PC of PCA_1 we constructed the *PC – Feature* mapping based on Algorithm 2. Both pipeline 3 (Section 2.3.3) and pipeline 4 (Section 2.3.4) were trained with the top–54 features extracted from those 69 PC. The top-20 features are shown in Table 1. A comprehensive list is available in our source code repository [25].

The top-10 features were related to vapor pressure deficit, on a 6– and 1–month lag, type of vegetation, temperature, relative humidity and other one-hot features identifying ecoregions. Additionally, the top-20 included weather features like wind speed and precipitation (on a 2–month lag), features describing the terrain like slope, and human development such as distance to power lines.

3.2 Pipeline Performance

Based on the CV-test AUC values, pipeline 1 (feature selection by correlation AUC 0.72 ± 0.03) had the best training performance, followed by pipeline 4 (Top_k-feature learning; AUC 0.61 ± 0.07), then pipeline 3 (top_k-features PCA learning; AUC 0.57 ± 0.10). All 3 pipelines performed better than a bagging classifier fit using PCA transformed space (pipeline 2; AUC 0.55 ± 0.11).

CV-test AUC also varied depending on the number of years included in the CV-training fold (Figure 6a). CV-test AUC varied little for pipeline 1 (feature selection by correlation) until the 5th CV fold when the greatest number of years were included in the training fold (number of years = 24). In contrast, CV-test AUC was greatest for other pipelines for the 3rd CV fold, potentially in response to an increase in the proportion of events in the sample.

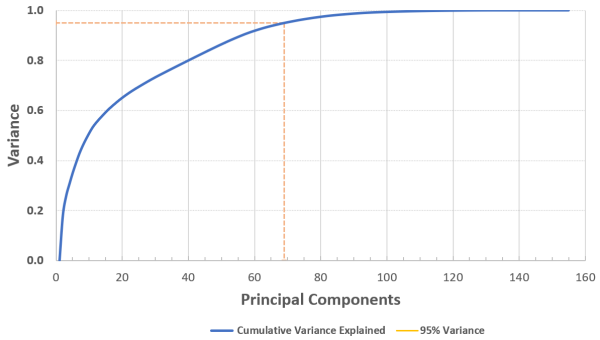

Figure 5: Cumulative Sum of Explained Variance by Principal Component for PCA_1 Model

Although the feature selection pipeline using correlation produced the better results on the training data, it underperformed when evaluating the model with the test data. The Top_k feature learning (pipeline 4), however, maintained a performance in the test data split comparable to the one achieved during training. Figure 6b shows the evaluation performance on the test data for all four pipelines. Based on this, we concluded that the Top_k feature learning selected features were the most informative for the model compared to the other three methods. However based on the variability in CV-test AUC across CV folds, we recognize that the test-AUC values may be dependent on fire activity in the years used to train the pipeline (1984 – 2014) and the years used to evaluate it (2015 – 2019).

3.3 Performance and Input Size Trade-off

After determining that the Top_k feature learning pipeline performed best with the test evaluation data, we further analyzed the performance of the model as we reduced the cumulative explained variance in the PCA model.

We conducted 17 experiments using between 5 and 69 PC with a cumulative explained variance between 39% and 95% respectively (Figure 7). Based on these experiments, by selecting 20 PC (an explained variance of 66%), the Top_k feature learning model (pipeline 4) achieved an AUC of 0.73 ± 0.03 and an F1–score of 0.72 ± 0.05 respectively. This model's performance was comparable to the model trained with 69 PC (an explained variance of 95%) but required 34 less input features.

3.4 Burn Probability

We mapped predicted burn probabilities from the baseline and the Top_k feature selection pipeline (pipeline 4). Figure 8 shows the mapped burn probability estimates for both models for July through October 2019. These months often have higher rates of wildfires in the state, and 2019 is the most recent period for which data were available. Predicted BP for both models during the 2015 – 2019 period is available at our data repository [24].

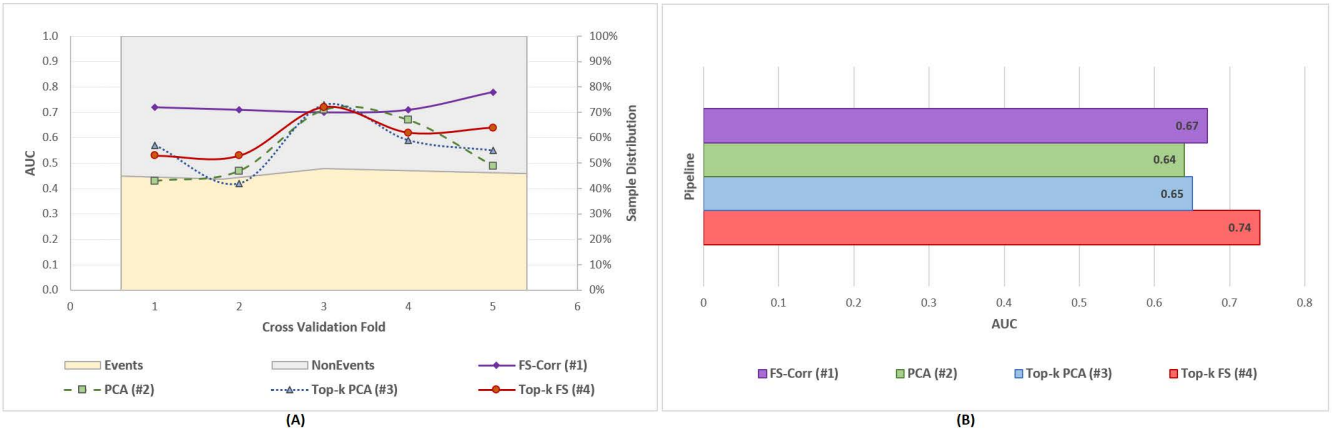

Figure 6: A) Validation AUC for Each CV–fold vs. Event/Non-Event Distribution B) AUC Performance on Test Data

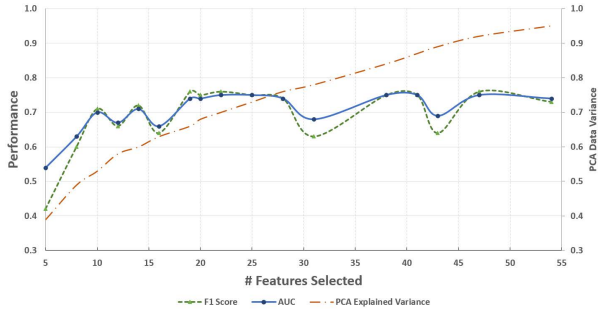

Figure 7: Test Area Under the Curve (AUC) vs. Input Size. FS: Feature Selection, PCA: Principal Component Analysis

The baseline predictions are shown on Figure $8a - d$ (top row), and the proposed pipeline on Figure $8e - h$ (bottom row). Figure $8i - l$ (middle row) show the hotspots where the baseline and the proposed pipeline predictions are different ($\Delta(BP) \geq 1 \times 10^{-3}$). The statewide average of BP differences for July through October 2019 were between -1.37×10^{-5} and 1.6×10^{-5}.

We further analyzed the hotspots of the differences in the predictions. In most cases, the baseline model predicted a higher BP for areas where there was limited fuel for a fire to burn, or where the area was highly developed. Moreover, in other cases such as January 2019, shown on Figure 9, the baseline model predicted even higher BP for areas in, or close to the Death Valley National Park, Mojave National Preserve and Joshua Tree National Park. These areas did not experience extensive burned area in the past, but Joshua Tree did experience a large wildfire in 2020.

Considering these patterns, we concluded that the Top_k feature selection pipeline, predicted more stable and realistic burn probabilities compared to the model that used all the features.

4 CONCLUSIONS AND FUTURE WORK

Modeling wildfire burn probability (BP) is a challenging task due to large class imbalance, high year–to–year and month–to–month variability, and the large number of potential features. Here, we analyzed different methodologies to reduce spatio–temporal features needed as input to model BP for the state of California, USA.

Based on the analysis, we concluded the best pipeline was the Top_k feature learning, using a PCA to determine the most informative features to predict BP. Using this pipeline, the bagging classifier predicting burn probability achieved a 74% AUC based on the test data. We presented the spatial BP predictions and demonstrated that they contained less noise than predictions from a baseline model which used all available features.

There are several limitations to our approach that could be improved in future studies. First, multiple features may have contributed to each PC. However, our approach only considered the top feature from each individual PC. Second, predictive performance was moderate. Finally, in this study we relied on 1 km^2 pixel grid and monthly–lagged temporal data, which may limit predictive performance.

Future work may consider selecting features that have individual PC contributions above a threshold, or cumulative contributions, instead of the most important feature for each PC to ensure these interactions are retained. Furthermore, performance may be improved by considering other types of models such as deep neural networks. Fire potential and behavior varies temporally, over minutes, hours, days, and months, and predictive performance may also be improved by incorporating features that better represent the temporal range of weather important to fuel conditions and fire spread. In addition to these potential improvements, this study could be expanded to predict other components of fire occurrence such as fire size and burn severity, which would provide additional information about the impact of fires on ecosystems. For example, burn severity is critical for estimating the amount of biomass consumed in fires [10].

The pipelines considered here demonstrated that machine learning models can effectively predict burn probability with a reduced set of features. Even though the predictive performance was moderate ($AUC = 0.74$), the models captured well–known spatial and temporal patterns of fire occurrence in California. Our efficient modeling approach can provide information to communities and

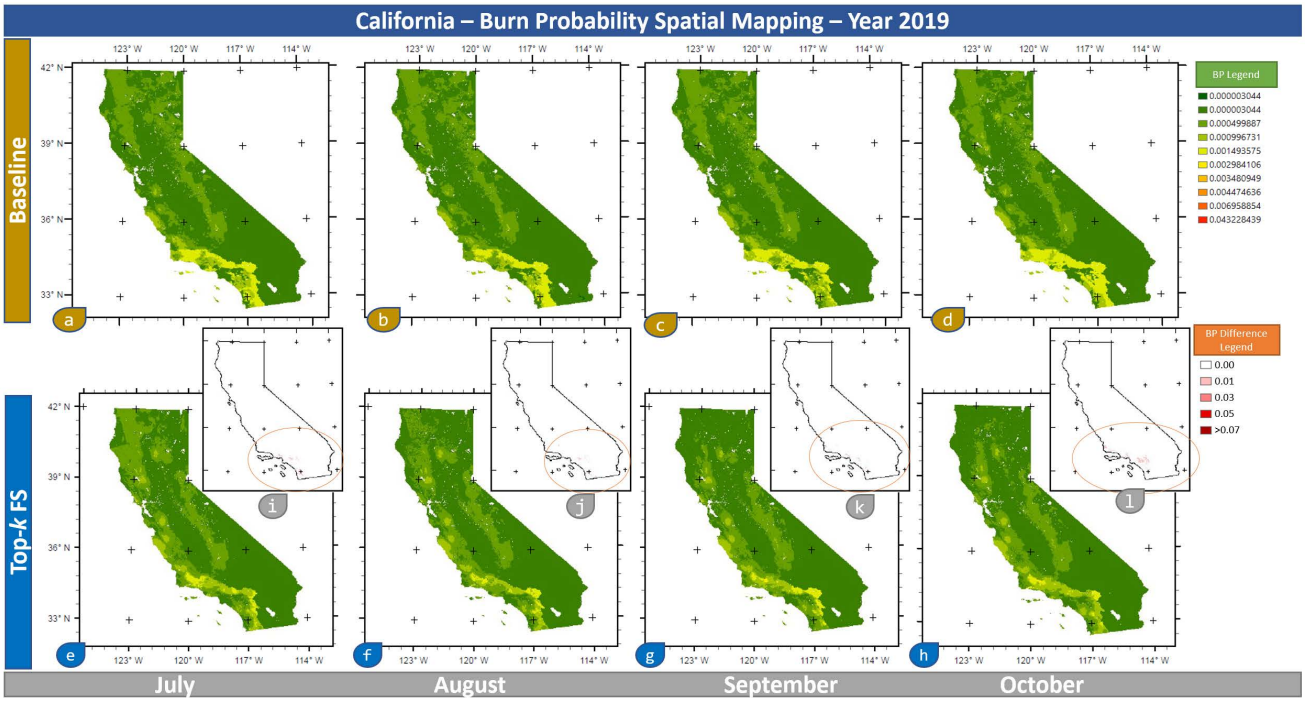

Figure 8: Burn Probability (BP) Spatial Mapping for Baseline Model (*a-d*) Top_k Feature Learning Model (*e-h*), and Difference Hotspots Between Them (*i-l*)

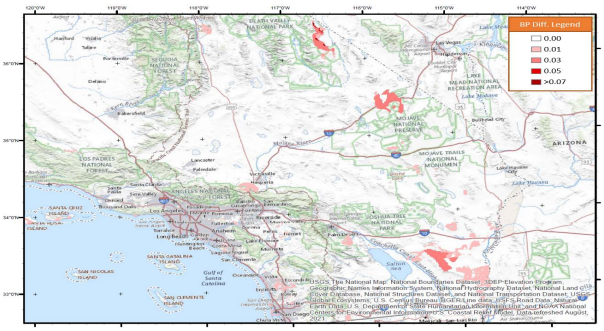

Figure 9: Burn Probability (BP) Difference for January 2019

agencies concerned about wildfire occurrence and the impacts on human and ecological systems.

Disclaimer: Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REFERENCES

[1] Bishrant Adhikari, Chen Xu, Paddington Hodza, and Thomas Minckley. 2021. Developing a Geospatial Data-driven Solution for Rapid Natural Wildfire Risk Assessment. *Applied Geography* 126 (2021), 102382. https://doi.org/10.1016/j.apgeog.2020.102382

[2] Bachisio Arca, Tiziano Ghisu, Marcello Casula, Michele Salis, and Pierpaolo Duce. 2019. A Web-based Wildfire Simulator for Operational Applications. *International Journal of Wildland Fire* 28 (2019), 99–112. Issue 2. https://doi.org/10.1071/WF18078

[3] David M.J.S. Bowman, Jennifer K. Balch, Paulo Artaxo, William J. Bond, Jean M. Carlson, Mark A. Cochrane, Carla M. D'Antonio, Ruth S. DeFries, John C. Doyle, Sandy P. Harrison, Fay H. Johnston, Jon E. Keeley, Meg A. Krawchuk, Christian A. Kull, J. Brad Marston, Max A. Moritz, I. Colin Prentice, Christopher I. Roos, Andrew C. Scott, Thomas W. Swetnam, Guido R. van der Werf, and Stephen J. Pyne. 2009. Fire in the Earth System. *Science* 324 (2009), 481–484. Issue 5926. https://doi.org/10.1126/science.1163886

[4] David M.J.S. Bowman, Grant J. Williamson, John T. Abatzoglou, Crystal A. Kolden, Mark A. Cochrane, and Alistair M.S. Smith. 2017. Human Exposure and Sensitivity to Globally Extreme Wildfire Events. *Nature Ecology and Evolution* 1 (2017), 0058. Issue 3. https://doi.org/10.1038/s41559-016-0058

[5] Leo Breiman. 1996. Bagging Predictors. *Machine Learning* 24 (1996), 123–140. Issue 2. https://doi.org/10.1007/BF00058655

[6] Jesslyn F. Brown, Heather J. Tollerud, Christopher P. Tollerud, Qiang Zhou, John L. Dwyer, James E. Vogelmann, Thomas R. Loveland, Curtis E. Woodcock, Stephen V. Stehman, Zhe Zhu, Bruce W. Pengra, Kelcy Smith, Josephine A. Horton, George Xian, Roger F. Auch, Terry L. Sohl, Kristi L. Sayler, Alisa L. Gallant, Daniel Zelenak, Ryan R. Reker, and Jennifer Rover. 2020. Lessons Learned Implementing an Operational Continuous United States National Land Change Monitoring Capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) Approach. *Remote Sensing of Environment* 238 (March 2020), 111356. https://doi.org/10.1016/j.rse.2019.111356

[7] Thomas M. Cover and Joy A. Thomas. 2006. *Elements of Information Theory*. Wiley. https://doi.org/10.1002/047174882X

[8] Jeff Eidenshink, Brian Schwind, Ken Brewer, Zhi-Liang Zhu, Brad Quayle, and Stephen Howard. 2007. A Project for Monitoring Trends in Burn Severity. *Fire Ecology* 3 (12 2007), 3–21. https://doi.org/10.4996/fireecology.0301003

[9] Mark A. Finney. 2005. The Challenge of Quantitative Risk Analysis for Wildland Fire. *Forest Ecology and Management* 211 (2005), 97–108. Issue 1. https://doi.org/10.1016/j.foreco.2005.02.010 Relative Risk Assessments for Decision–Making Related To Uncharacteristic Wildfire.

[10] Nancy H.F. French, William J. de Groot, Liza K. Jenkins, Brendan M. Rogers, Ernesto Alvarado, Brian Amiro, Bernardus de Jong, Scott Goetz, Elizabeth Hoy, Edward Hyer, Robert Keane, B.E. Law, Donald McKenzie, Steven G. McNulty, Roger Ottmar, Diego R. Pérez-Salicrup, James Randerson, Kevin M. Robertson, and Merritt Turetsky. 2011. Model Comparisons for Estimating Carbon Emissions from North American Wildland Fire. *Journal of Geophysical Research: Biogeosciences* 116 (2011). Issue G4. https://doi.org/10.1029/2010JG001469

[11] A. Malcolm Gill, Scott L. Stephens, and Geoffrey J. Cary. 2013. The Worldwide "Wildfire" Problem. *Ecological Applications* 23 (2013), 438–454. Issue 2. https://doi.org/10.1890/10-2213.1

Table 1: Top 20 Features with Most Impact on the Principal Components (PC)

Feature	Description	Explained Variance by PC	Explained Variance Cumulative Sum
VPDlag6	Vapor Pressure Deficit (6-month lag)	0.1797	0.1797
LCPRI_wildland_8	Wildland vegetation (8 km radius)	0.0778	0.2575
TMmnlag9	Minimum temperature (9-month lag)	0.0483	0.3058
LCPRI_3	Grass/shrub	0.0449	0.3507
VPDlag1	Vapor Pressure Deficit (1-month lag)	0.0395	0.3902
ECO_l1_10	Ecoregion: North American Deserts	0.0371	0.4273
ECO_l1_11	Ecoregion: Mediterranean California	0.0307	0.4580
RMaxlag24	Maximum relative humidity (24-month lag)	0.0270	0.4850
ECO_l1_7	Ecoregion: Marine West Coast Forest	0.0246	0.5095
TMmxlag0	Maximum temperature	0.0224	0.5319
ECO_l2_1001	Ecoregion: Cold Deserts	0.0175	0.5494
WUI_0	Wildland-Urban-Interface: Wildland vegetation	0.0145	0.5639
ECO_l3_110103	Ecoregion: Southern And Baja California Pine-Oak Mountains	0.0137	0.5776
VSlag0	Wind speed	0.0133	0.5909
LCPRIf_920	Western oak forest type group	0.0118	0.6027
PL_dist	Powerline distance	0.0113	0.6140
PRlag2	Precipitation (2-month lag)	0.0110	0.6250
LCPRI_10	Developed	0.0103	0.6353
slope	Slope	0.0097	0.6450
month_8	Month of Year:August	0.0092	0.6542

[12] Todd J. Hawbaker, Volker C. Radeloff, Susan I. Stewart, Roger B. Hammer, Nicholas S. Keuler, and Murray K. Clayton. 2013. Human and Biophysical Influences on Fire Occurrence in the United States. *Ecological Applications* 23 (4 2013), 565–582. Issue 3. https://doi.org/10.1890/12-1816.1

[13] Tianhua He, Byron B. Lamont, and Juli G. Pausas. 2019. Fire as a Key Driver of Earth's Biodiversity. *Biological Reviews* 94 (2019), 1983–2010. Issue 6. https://doi.org/10.1111/brv.12544

[14] David W. Hosmer Jr., Stanley Lemeshow, and Rodney X. Sturdivant. 2013. *Applied Logistic Regression* (3rd ed.). Wiley and Sons, Inc.

[15] Samina Khalid, Tehmina Khalil, and Shamila Nasreen. 2014. A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning. *2014 Science and Information Conference*, 372–378. https://doi.org/10.1109/SAI.2014.6918213

[16] H. Anu Kramer, Miranda H. Mockrin, Patricia M. Alexandre, Susan I. Stewart, and Volker C. Radeloff. 2018. Where Wildfires Destroy Buildings in the US Relative to the Wildland–urban Interface and National Fire Outreach Programs. *International Journal of Wildland Fire* 27 (2018), 329–341. Issue 5. https://doi.org/10.1071/WF17135

[17] Ashima Malik, Megha Rajam Rao, Nandini Puppala, Prathusha Koouri, Venkata Anil Kumar Thota, Qiao Liu, Sen Chiao, and Jerry Gao. 2021. Data-Driven Wildfire Risk Prediction in Northern California. *Atmosphere* 12 (2021). Issue 1. https://doi.org/10.3390/atmos12010109

[18] Nicolas Mansuy, Carol Miller, Marc-André Parisien, Sean A. Parks, Enric Batllori, and Max A. Moritz. 2019. Contrasting Human Influences and Macro-environmental Factors on Fire Activity Inside and Outside Protected Areas of North America. *Environmental Research Letters* 14 (5 2019), 64007. Issue 6. https://doi.org/10.1088/1748-9326/ab1bc5

[19] Marc-André Parisien, Denyse A. Dawe, Carol Miller, Christopher A. Stockdale, and O. Bradley Armitage. 2019. Applications of Simulation-based Burn Probability Modelling: A Review. *International Journal of Wildland Fire* 28 (2019), 913–926. Issue 12. https://doi.org/10.1071/WF19069

[20] Marc-André Parisien, Carol Miller, Sean A. Parks, Evan R. DeLancey, François-Nicolas Robinne, and Mike D. Flannigan. 2016. The Spatially Varying Influence of Humans on Fire Probability in North America. *Environmental Research Letters* 11 (7 2016), 75005. Issue 7. https://doi.org/10.1088/1748-9326/11/7/075005

[21] Sean A. Parks, Marc-André Parisien, and Carol Miller. 2011. Multi-scale Evaluation of the Environmental Controls on Burn Probability in a Southern Sierra Nevada Landscape. *International Journal of Wildland Fire* 20 (2011), 815–828. Issue 7. https://doi.org/10.1071/WF10051

[22] Sean A. Parks, Marc-André Parisien, Carol Miller, and Solomon Z. Dobrowski. 2014. Fire Activity and Severity in the Western US Vary Along Proxy Gradients Representing Fuel Amount and Fuel Moisture. *PLOS ONE* 9 (12 2014), 1–8. Issue 6. https://doi.org/10.1371/journal.pone.0099699

[23] Javier Pastorino and Ashis Kumer Biswas. 2020. Hey ML, What Can You Do for Me? *2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 116–119. https://doi.org/10.1109/AIKE48582.2020.00023

[24] Javier Pastorino, Joseph W. Director, Ashis K. Biswas, and Todd J. Hawbaker. 2022. Burn Probability Predictions for the State of California, USA Using an Optimal Set of Spatio-temporal Features. U.S. Geological Survey Data Release. https://doi.org/10.5066/P9GLB4VB

[25] Javier Pastorino, Joseph W. Director, Ashis K. Biswas, and Todd J. Hawbaker. 2022. Source Code for Determining the Optimal Set of Spatio-temporal Features for Predicting Burn Probability in the State of California, USA. U.S. Geological Survey Code Release. https://doi.org/10.5066/P96W3OBD

[26] Karin L. Riley, John T. Abatzoglou, Isaac C. Grenfell, Anna E. Klene, and Faith Ann Heinsch. 2013. The Relationship of Large Fire Occurrence with Drought and Fire Danger Indices in the Western USA, 1984–2008: The Role of Temporal Scale. *International Journal of Wildland Fire* 22 (12 2013), 894–909. https://doi.org/10.1071/WF12149

[27] B. Ruefenacht, M.V. Finco, M.D. Nelson, R. Czaplewski, E.H. Helmer, J.A. Blackard, G.R. Holden, A.J. Lister, D. Salajanu, D. Weyermann, and K. Winterberge. 2008. Conterminous US and Alaska Forest Type Mapping Using Forest Inventory and Analysis Data. *Photogrammetric Engineering & Remote Sensing* 74, 11 (2008), 1379–1388.

[28] Matthew P. Thompson, David E. Calkin, Mark A. Finney, Alan A. Ager, and Julie W. Gilbertson-Day. 2011. Integrated National-scale Assessment of Wildfire Risk to Human and Ecological Values. *Stochastic Environmental Research and Risk Assessment* 25 (2011), 761–780. Issue 6. https://doi.org/10.1007/s00477-011-0461-0

[29] Nicholas Timme, Wesley Alford, Benjamin Flecker, and John M. Beggs. 2014. Synergy, Redundancy, and Multivariate Information Measures: An Experimentalist's Perspective. *Journal of Computational Neuroscience* 36 (2014), 119–140. Issue 2. https://doi.org/10.1007/s10827-013-0458-4

[30] U.S. Department of Homeland Security. 2020. Homeland Infrastructure Foundation-Level Data (HIFLD). https://gii.dhs.gov/hifld/

[31] U.S. Geological Survey. 2020. The National Map, Supporting Themes, Transportation. https://www.usgs.gov/core-science-systems/national-geospatial-program/supporting-themes

[32] U.S. Geological Survey Gap Analysis Project. 2018. Protected Areas Database of the United States (PAD-US). https://usgs.gov/gapanalysis/PAD-US

[33] Satosi Watanabe. 2010. Information Theoretical Analysis of Multivariate Correlation. *IBM Journal of Research and Development* 4 (11 2010), 66–82. Issue 1. https://doi.org/10.1147/rd.41.0066