Predicting the impact of wildfire using Machine Learning techniques to assist effective deployment of resources

Nikesh Kumar Pahuja
Information Systems Engineering and Management
Harrisburg University of Science and Technology
Harrisburg, PA 1701, USA
npahuja@my.harrisburgu.edu

Maria H. Rivero
Information Systems Engineering and Management
Harrisburg University of Science and Technology
Harrisburg, PA, 1701, USA
mrivero@my.harrisburgu.edu

***Abstract*—Wildfire incidents occur every year due to human-caused factors, droughts, vegetation, and climate change. It is complex to predict severe wildfires and even more difficult to extinct them. Structural damage may occur during wildfires incidents. Therefore, understanding the intensity of these damages could help mitigate the risk for the region and local communities. This study builds a predictive model using various Machine Learning techniques as a real-time decision-support tool. The model predicts the severity of damage to the structures affected during the wildfires in the state of California, United States. The study used The California Department of Forestry and Fire Protection (CAL FIRE)'s historical data from 2013 until 2019. The study trained several supervised machine learning algorithms for classification, such as Logistic Regression (LR), Decision Trees (DT), Random Forest (RF), Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), Naïve Bayes (NB), and XGBoost (XGB), to predict the structural damage. Predictions were classified in the following categories of expected Fire damage: Inaccessible, Destroyed (>50%), Major (26-50%), Minor (10-25%), Affected (1-9%), and No Damage. The best performing algorithm was the Decision Tree model. This study will allow fire and forest departments, other departments, corporations, and the public to forecast the expected fire damage when the wildfire is first reported. It will also enable the departments, corporations, and public to prioritize actions and deploy their resources effectively to reduce the impact of wildfires in the region.**

Keywords—Machine Learning, Artificial Intelligence, Wildfires, Decision-Support Tool, Environmental Science, Sustainability

I. INTRODUCTION

In 2018, over 1.6 million acres of land had burned and caused large sums of environmental damage throughout California (Malik, 2021). Wildfires are hazardous events that are complex to predict and can create a state of emergency. When wildfires occur, they pose health risks to the communities and cause damage to the land and the economy in the affected region. In many cases, it is difficult to quickly reduce the fire's severity, and it can cause irreparable damages to structures in the affected area. In 2018, The Internal Journal of Wildland Fire published that the effects of climate change and global warming will increase the possibility of wildfires turning the United States into "a more combustible country" (Pierre-Louis, Popovich, 2020). Since the beginning of the year 2020, wildfires incidents in California have increased, burning over 3.2 million acres of land, an area equivalent to that of the state of Connecticut (Silverman, 2020). Wildfires in the United States have affected the Pacific Northwest of the country, such as California, Oregon, and Washington, causing more than half a million people to flee from these states (Evans, 2020). Wildfires have also expanded to other U.S. states. For instance, the wildfire season affected Colorado state during its third driest year on record, which led to more than 175,000 acres burned during 2020 (Sakas, 2020). Indeed climate change, extensive droughts, and human activity have caused wildfires incidents to increase and become more hazardous events in the country.

Predicting the intensity and impact of wildfires throughout California, United States is complex. Fortunately, Artificial Intelligence (A.I.) predictive models can help fire and forest management departments and land agencies estimate the expected impact of these hazardous events and provide an exceptional service when these massive or local wildfires occur in real-time. This research uses historical data and builds a predictive model using Machine Learning techniques to predict expected damage on structures affected by wildfires in California. The study uses the damage inspection datasets provided by The California Department of Forestry and Fire Protection (CAL FIRE) collected from the historical period of 2013 until 2019. The study implements various supervised Machine Learning techniques, using the Fire Damage to the structures as the target variable. The selected Machine Learning algorithms are the Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), and

979-8-3503-2028-2/22/$31.00 ©2022 IEEE
DOI 10.1109/CSCI58124.2022.00140

XGBoost (XGB). The dataset is split into a train set (90%) and a test set (10%). The grid search is used to find the best parameters. Algorithms are then trained and validated on the training dataset by applying ten-fold cross-validation. Finally, algorithms are tested on a sample of the testing data set to find model accuracy and robustness. The best performing model predicts expected damage to structures during wildfires in California with 92% accuracy. It predicts and categorizes the expected impact to the structure in six categories: Inaccessible, Destroyed (>50%), Major (26-50%), Minor (10-25%), Affected (1-9%), and No Damage. Fire and forest management departments can use this model as a decision-support tool to predict expected fire damage when the wildfire is first reported and allocate resources efficiently for wildfire prevention initiatives, thereby reducing the risk and impact of these incidents on the affected regions.

II. Literature Review

A. *Structural damages resulting from wildfires*

Damage resulting from wildfires reaches significant levels across different land structures, mountains, urban areas, and vegetation. Land managers and fire departments develop wildfires frameworks to assess the damage intensity post-wildfire incidents. For instance, Mickler et al. (2013) presented a survey from the Joint Fire Science Program that showed a 10-year retrospective of fire research themes to understand historical wildland fire patterns and develop a comprehensive land management framework. Among the themes discussed, the post-wildfire erosion and mitigation assessment incorporate the burn severity ratings classified as low, medium, and severe, divided into vegetation and soil burn groups. Its relevance assesses the post-fire treatments and indicators for the land management department to mitigate the risk of wildfire incidents and understand the economic feasibility of resources.

However, recent wildfires aggravated the structural damage and have presented severe consequences on the west coast of the United States. Syphard et al. (2019) published a study analyzing 40,000 wildfire-exposed structures from the period of 2013 until 2018 within the state of California. This study explored the causes behind the wildfire incident and how they impact the structures. The study compared the homes that survived wildfires versus the homes destroyed during wildfires. This study emphasized that structural characteristics were significant to prevent the damage caused by wildfires more than defensible space distance. Features such as enclosed eaves, vent screens, and multi-pane windows mostly prevented the structures from collapsing during a wildfire. Other factors mentioned in this study acknowledged that active firefighting contributed to reducing the structure's collapse in the state's Northern and interior areas.

B. *Machine Learning approaches*

A.I. approaches can be significant for predicting the root cause of fire incidents, susceptibility mapping, the area of damaged land, and impact on the health of the local communities. However, there is an opportunity to explore studies related to structural damage caused by wildfires that could support effective resource deployment depending on the expected impact. Previous studies have addressed the early detection of wildfires through adaptable frameworks using Machine Learning techniques. For instance, Khan et al. (2017) proposed an early-fire detection framework using fine-tuned convolutional neural networks (CNN) for CCTV surveillance cameras to detect indoor and outdoor fire incidents. The proposed algorithm addressed the prioritization of cameras, and an autonomous response based on cognitive radio networks to ensure proper data dissemination. The model's experimentation resulted in mitigating disaster management and detecting fire in the early stages while reducing potential false fire alarms.

Another study proposing a framework for wildfire surveillance included a machine learning approach to measure hazardous events, including wildfires, and their impact on mountainous areas. Yousefi et al. (2020) designed a multi-hazard risk map for a mountainous region located in Southwestern Iran to model probabilities of natural disasters using machine learning models. The features included in this study as input were climatic, topographic, geological, social, and morphological factors. The resulting generalized linear model predicted wildfire among other natural disasters mapping with acceptable accuracy. The predictive maps obtained from the multi-hazard machine learning model provided a tool to manage future human interaction with wildfires and other extreme natural disasters.

Wildfire management tools must provide real-time monitoring of these incidents. For instance, fire authorities usually do not know the damage level until they approach the area. However, Machine Learning can solve this problem by using predictive analytics. Rishickesh et al. (2019) proposed a model to assess the occurrence of wildfire using historical data from the University of California Irvine (UCI) Center for Machine Learning and Intelligent Systems' Machine Learning repository. The study provides insights into the physical and climatic conditions that impact wildfires in Montesinho park in Portugal. This study predicts the occurrence of wildfires using supervised Machine Learning classification techniques. The target feature 'fire' was generated from the 'area' variable, where fire equals one if the value of the area is greater than zero. Otherwise, the value of the 'fire' variable remains equal to zero. A classification approach trained algorithms with and without Principal Component Analysis (PCA). The best performing algorithm with PCA was Logistic Regression with 68% accuracy.

Furthermore, Rodrigues et al. (2020) studied fire behavior. They predicted fire activity for government agencies to mitigate the structural impact of fire in a Eucalyptus plantation land located in Viçosa, Minas Gerais, Brazil using Machine Learning techniques. The study demonstrates the performance of two Machine Learning algorithms, the Lasso Elastic-Net Regularized Generalized Linear Model and Random Forest in predicting three different response variables: the Fire intensity, the Height of flames, and the Rate of fire spreading. Explanatory variables such as Days without rain, Relative Humidity, Live fuel moisture, Class 1 fuel moisture etc were included in the models according to the response variables addressed. The results highlighted the Random Forest algorithm as the best performing model, noting that climatic conditions and moisture of combustible materials significantly affected the fire behavior intensity. This study’s conclusion suggests the application of Random Forest algorithms for fire behavior prediction and susceptibility mapping.

Classification algorithms have undertaken categories or indices to measure damage intensity such as the Normalized Burn Ratio (NBR). Collins et al. (2018) expanded the significance of fire severity mapping using Random Forest classifiers as a robust method to scale and monitor severe wildfires across landscapes. Fire severity mapping is an essential resource for Scientists and land management agencies. This study collected point-based fire severity training data from sixteen large wildfires that occurred from 2006 and 2016 in south-eastern Australia using Landsat imagery. The Random Forest classifier algorithm considers the derived pre-and post-fire difference severity indices and classifications categories such as “Crown consumption”, “Crown scorch”, “Partial crown scorch”, “Crown unburnt”, and “Unburnt”. The results from the bootstrapping validation of the algorithm demonstrated that the Random Forest classifier reached a higher classification accuracy (>95%) in the severity classes of “Unburnt”, “Crown scorch”, and “Crown consumption”, also leading to increased accuracy between 6% and 21% over NBR. Also, the algorithm showed high classification accuracy (>74%) for low severity classes such as “Crown unburnt”, and “Partial crown scorch”.

Fire severity also may have an impact during fire occurrences and its spread in real-time. Zhai et al. (2020) proposed a learning-based wildfire spread model to predict the short-term wildfire spread. An index determined in this study is the Real-time rate of Spread (RoS), calculating the normal movements of fire fronts and enhancing its prediction in unburnt areas using Machine Learning techniques. This study methodology designed a level-set measure to simulate the fire front movement. With the multilayer perception neural network model and current RoS, the learning-based model could predict short-term fire spread in shrubland fires which has a significant impact on modeling the spread of natural wildfire.

Finally, wildfire research initiatives play an important role in studying the prevalence of wildfires, their causes, and how government agencies respond to them locally. New and effective methods are constantly assessed and deployed to increase the efficiency of management and resource allocation to mitigate wildfires. However, there is an opportunity to forecast the expected damage to the structure when a fire is reported and provides government agencies a decision support tool to allocate resources in real-time when the incident starts. This study's research objectives will emphasize creating a Machine Learning model that proposes a real-time decision support tool to predict the level of damage to the structures when a wildfire incident is reported.

III. METHODOLOGY

This study analyzes the damage inspection data from The California Department of Forestry and Fire Protection (CAL FIRE) between 2013 and 2019 in California, USA. This study trains several Machine Learning models to find the best model. The data comprises eleven predictors and one target variable influencing the fire incidents published on these reports. These variables include the month of occurrence, cause, latitude, longitude, and acres burned during a wildfire. Table I. Presents a descriptive list of the variables. Fire Damage is the target variable. All other parameters are used as predictors in the model.

TABLE I. DATA SOURCES AND PARAMETERS

Parameters	**The California Department of Forestry and Fire Protection (CAL FIRE)**		
	Units	***Source***	***Remarks***
Fire Damage	-	Damage Inspection Dataset (2013 – 2019)	Target Variable Categories: Inaccessible Destroyed (>50%) Major (26-50%) Minot (10-25%) Affected (1-9%) No Damage
Street type	-		Categorical data
City	-		Categorical data
CAL FIRE Unit	-		Categorical data
County	-		Categorical data
Structure type	-		Categorical data
Structure category	-		Categorical data
Year-built parcel	-		Numeric data
Latitude	degree		Numeric data
Longitude	degree		Numeric data
Incident Start year	-		Numeric data
Incident Start month	-		Numeric data

The study predicts the expected Fire Damage caused by a wildfire using the Supervised Machine Learning approaches. Target variable is the Fire Damage to the structures. The study trained the following algorithms: Logistic Regression (L.R.), Decision Tree (D.T.), Random Forest (R.F.), Naïve Bayes (N.B.), Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), and X.G. Boost (XGBoost). Eleven variables were selected as predictors to the model: Fire Damage, Street type, City, CAL FIRE unit, County, Structure type, Structure category, Year built parcel, Latitude, Longitude, Incident start year, and Incident start month. The model trained and predicted the expected Fire Damage into six target categories: Inaccessible, Destroyed (>50%), Major (26-50%), Minor (10-25%), Affected (1-9%), and No Damage. The dataset was split into a train set (90%), and a test set (10%). The grid search was used to find the best parameters. The supervised learning models were trained and validated on the Train dataset. These algorithms were evaluated for their accuracy and performance on the Test set using Accuracy metrics and confusion matrix. The best performing model will provide a prediction of expected damage to the structures during wildfires in California.

IV. RESULTS

The Methodology presented in this study built a predictive model using Machine Learning algorithms that used the selected features from the CAL FIRE's damage inspection dataset and predicted the expected damage to the structures during a wildfire using Supervised Learning technique. The performance of various models indicated in the methodology of this study is compared with each other. The Decision Tree recorded the best performance with 92% classification accuracy among all the different algorithms trained and evaluated in this study. Table 2. presents the performance metrics of all the Machine Learning algorithms applied in the study across ten-fold cross-validation. It also highlights the mean score and standard deviation of each of the models.

TABLE II. MACHINE LEARNING MODEL RESULTS

Method	Cross-Validation				
	Score (Fold 1)	***Score (Fold 2)***	***Score (Fold 3)***	***Score (Fold 4)***	***Score (Fold 5)***
Logistic Regression	0.89	0.89	0.89	0.89	0.89
Decision Tree	0.92	0.92	0.92	0.92	0.92
Random Forest	0.91	0.92	0.92	0.92	0.92
Naïve Bayes	0.64	0.66	0.66	0.66	0.65
SVM	0.81	0.81	0.81	0.81	0.81
SGD	0.81	0.81	0.81	0.81	0.81
XGBoost	0.92	0.92	0.92	0.92	0.92
Method	***Score (Fold 6)***	***Score (Fold 7)***	***Score (Fold 8)***	***Score (Fold 9)***	***Score (Fold 10)***
Logistic Regression	0.89	0.88	0.89	0.89	0.89
Decision Tree	0.92	0.92	0.92	0.92	0.92
Random Forest	0.92	0.92	0.92	0.92	0.92
Naïve Bayes	0.66	0.65	0.63	0.65	0.62
SVM	0.81	0.81	0.81	0.81	0.81
SGD	0.40	0.81	0.81	0.40	0.81
XGBoost	0.92	0.92	0.93	0.92	0.93

Fig 1. shows the Confusion Matrix of Decision Tree model. Fig. 2. shows the Actual vs. Predicted Fire damage of the structure by Decision Tree (D.T) model on a sample of the testing data set. The actual Fire Damage is plotted on the y-axis in blue color, while the predicted Fire Damage is plotted on the y-axis in orange color. Observations are plotted on the x-axis. These results recognize the accuracy of the Decision Tree model. These also offer the fire and forest management departments, other departments, corporations, and public a data-driven model to efficiently allocate resources in these susceptible areas.

V. CONCLUSION

Wildfires are hazardous events and are increasing in the Pacific Northwest region in the United States. These natural disasters are often complex to predict and can create a state of emergency within a region and local communities.

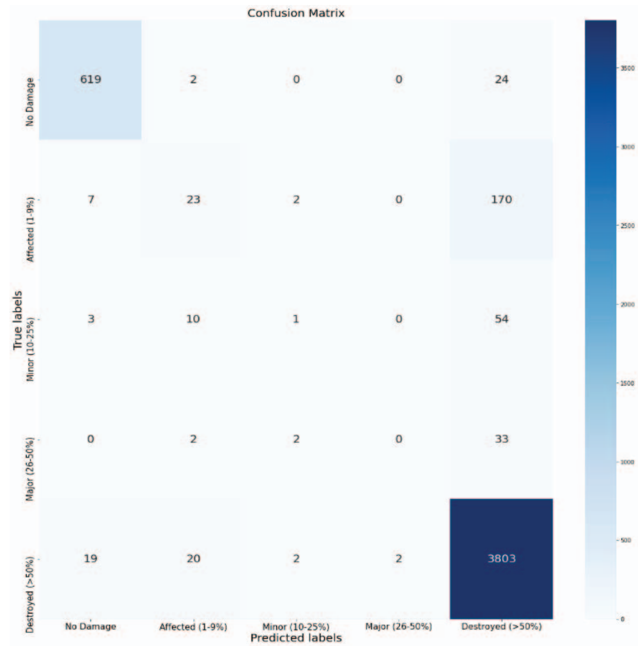

Fig. 1. Confusion Matrix Results

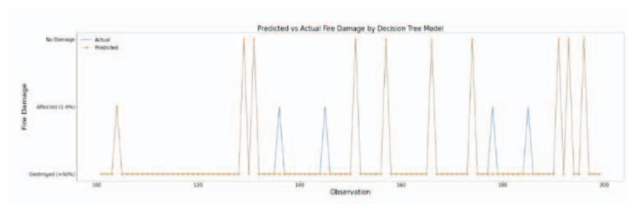

Fig. 2. Actual vs. predicted fire damage by Decition Tree model on a sample of Test data

Other consequences of these events pose severe health risk, land damage, vegetation extinction, and financial damage in the affected region. The wildfire damages caused to structures could be complex to understand in real-time. Severe wildfires can cause irreparable damage to structures in the affected area. A.I approaches such as Machine Learning can provide predictive analytics tools to forecast the expected damage in real-time and assist effective deployment of resources from fire and forest departments. Previous and recent studies have built comprehensive structural damage models for mitigation, multi-hazardous frameworks using machine learning, and other classification models to group the fire severity and analyze indices responding to the level of wildfire damage impact. This study presents a Machine Learning model that can be used as a decision-support tool for pre-fire assessment of structural damage within the state of California. The model used CAL FIRE's historical dataset from 2013 until 2019 and predicted expected impact to the structure. The best performing model was the Decision Tree model with an accuracy of 92% and standard deviation of .001. Therefore, this study acknowledges that Artificial Intelligence and Machine Learning techniques provide a robust model to analyze the historical datasets to unveil insights that can support strategic goals. The fire and forest management departments, other departments, corporations, and public could use this study's model to prioritize and deploy their resources and mitigate impact to the structures during wildfires. This study innovates the current resource deployment frameworks by introducing advanced analytics for rapid and transparent decision-making.

VI. Future Direction

This research study's future direction suggests improving the current model. The wildfires also impact communities and the nation from a social and economic perspective. This research will also build a new predictive model that can estimate the financial loss and economic impact due to severe wildfires. This research will continue to evolve and optimize the existing Machine Learning and Deep Learning models by including new parameters to assess wildfires' financial impact. Such a model will allow fire and forest management departments, other departments, corporations, and public to forecast the financial impact of wildfires and formulate strategic objectives and initiatives to efficiently allocate their budget to improve the management of the wildfire incidents within the United States and globally.

Acknowledgment

We want to thank The California Department of Forestry and Fire Protection (CAL FIRE) for supporting the data retrieval of the Damage Inspection (DINS) dataset for this research.

References

[1] K. Pierre-Louis and N. Popovich, "Climate Change is Fueling Wildfires Nationwide, New Report Warns," *The New York Times*, p.1, November 27, 2018. [Online]. Available: The New York Times. https://www.nytimes.com/interactive/2018/11/27/climate/wildfire-global-warming.html [Accessed September 01, 2020].

[2] H. Silverman, "California wildfires have burned an area almost the size of Connecticut," *CNN*, p.1, September 14, 2020 [Online]. Available CNN https://www.cnn.com/2020/09/14/us/california-wildfires-monday/index.html [Accessed September 14, 2020].

[3] C. Evans, "Oregon braces for a "mass fatality incident" as wildfires rage in western states," *CBS News*, p.1, September 12, 2020 [Online]. Available: CBS News https://www.cbsnews.com/news/oregon-wildfires-state-braces-for-mass-fatality-incident/ [Accessed September 12, 2020].

[4] E. Sakas, "Colorado Wildfires Are Climate Change 'In The Here And Now' – And A Sign Of Summers To Come," *CPR News*, p.1, August 20, 2020.[Online].Available:CPRNewshttps://www.cpr.org/2020/08/20/colorado-wildfires-climate-change-drought-snowpack/ [Accessed September 1, 2020].

[5] R. A. Mickler, J. H. Cissel, J. Laurence. "Advances in understanding and managing wildland fire: communicating wildland fire research to land-management practitioners.", *International Journal of Wildland Fire*, vol. 22, 2013. [Online] Available: http://dx.doi.org/10.1071/WFv22n1_FO

[6] A. D. Syphard, J. E. Keeley, "Factors Associated with Structure Loss in the 2013-2018 California Wildfires," Fire, vol. 2, pp. 49, September 2 2019. [Online] Available: doi:10.3390/fire2030049

[7] L. Collins, P. Griffioen, G. Newell, A. Mellor, "The utility of Random Forests for wildfire severity mapping," *ELSEVIER Remote Sensing of Environment*, vol 216, July 17, 2018. [Online] Available: https://doi.org/10.1016/j.rse.2018.07.005 [Accessed August 20, 2020]

[8] K. Muhammad, J. Ahmad, S. W. Baik, "Early fire detection using convolutional neural networks during surveillance for effective disaster management," *ELSEVIER Neurocomputing*, vol 288, December 29, 2017 [Online] Available: https://doi.org/10.1016/j.neucom.2017.04.083 [Accessed August 20, 2020]

[9] R. Rishickesh, A. Shahina, A. Nayeemulla Khan, "Predicting Forest Fires using Supervised and Ensemble Machine Learning Algorithms" *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, pp. 2277 - 3878, July 2019. [Online] Available: doi: 10.35940/ijrte.B2878.078219

[10] V. B. Rodrigues, F. T. P. Torres, "Fire Behavior Prediction using Machine Learning Algorithms," in *Revista Bras. Biom.*, 2020, Brazil, 2020, Lavras [Press]

[11] S. Yousefi, H. R. Pourghasemi, S. N. Emami, S. Pouyan, S. Eskandari & J. P. Tiefenbacher, "A machine learning framework for multi-hazards modeling and mapping in a mountainous area," *Scientific Reports NatureResearch*, July 22 2020. [Online] Available: https://doi.org/10.1038/s41598-020-69233-2

[12] C. Zhai, S. Zhang, Z. Cao, X. Wang, "Learning-based prediction of wildfire spread with real-time rate spread of measurement," *ELSEVIER Combustion and Flame*, vol 215, February 28, 2020 [Online] Available: https://doi.org/10.1016/j.combustflame.2020.02.007 [Accessed August 20, 2020]