“Predictive Analytics on the Recent California's Wildfire Risk Assessment and Damage Estimation Using Climate and Environmental Data”

Student’s Name

Date

**Table of Contents**

[1. Introduction 3](#_Toc191257265)

[1.1 Overview and Purpose of the Project 3](#_Toc191257266)

[1.2 Research Question, Aim and Objectives 4](#_Toc191257267)

[1.2.1 Research Question 4](#_Toc191257268)

[1.2.2 Aim 4](#_Toc191257269)

[1.2.3 Objectives 4](#_Toc191257270)

[2. Background 4](#_Toc191257271)

[2.1 Introduction to California’s Wildfire 4](#_Toc191257272)

[2.2 Selection Criteria for Papers 5](#_Toc191257273)

[2.3 Critical Analysis of Key Papers 6](#_Toc191257274)

[2.3.1 Paper 1: “Data-driven wildfire risk prediction in northern California”. 6](#_Toc191257275)

[2.3.2 Paper 2: “From static to dynamic prediction: Wildfire risk assessment based on multiple environmental factors” 6](#_Toc191257276)

[2.3.3 Paper 3: "Where there is Smoke, there is Fire: Wildfire Risk Predictive Modelling via Historical Climate Data" 7](#_Toc191257277)

[2.3.4 Paper 4: “Predicting electricity infrastructure induced wildfire risk in California”. 8](#_Toc191257278)

[2.4 Summary of Literature Review 8](#_Toc191257279)

[References 10](#_Toc191257280)

# 1. Introduction

## 1.1 Overview and Purpose of the Project

Wildfires may be devastating, yet paleoenvironmental records show that fire has been part of Earth's ecology for hundreds of millions of years. Some desert ecosystems have adapted to wildfires. Wildfires have become a threat as humans have overrun the earth. Unintentional, purposeful, or malicious human behaviour may be increasing wildfire frequency. Due to complex dynamics and atmospheric feedback, fire behaviour is problematic to predict systems, fire behaviour is hard to predict, and fire behaviour is challenging to predict. Wildfires may spread quickly, whether started intentionally or spontaneously. Wildfires may spread at 10-20 km/h and cover tens to hundreds or even thousands of square kilometres (Hincks *et al.*, 2011).

Embers may ignite vulnerable structures within seconds hundreds of meters or miles ahead of the fire front. A large blaze might reach several hundred degrees Celsius, making survival or escape difficult or impossible. Consequences include agricultural and structural destruction, environmental damage, economic losses, social instability, and smoke emissions that increase carbon and greenhouse gas emissions and lower visibility and human health (Hu, Tanchak and Wang, 2024). Wildfires are uncontrolled conflagrations occurring naturally, often igniting in grasslands, woodlands, shrublands, forests, or peatlands. These include fires initiated by natural phenomena, accidental or intentional ignition, and controlled burns that become uncontrollable (e.g., for land clearing or forest management). Wildfires transpire on every continent except Antarctica, anywhere there exists a fuel supply and conducive circumstances for burning (Hincks *et al.*, 2011).

The frequency, timing (e.g., seasonality), extent, and severity of fire in a region are collectively termed the fire regime, which may be characterised qualitatively or numerically (Bhowmik *et al.*, 2023). The risk of wildfires has been more pronounced in recent years. Wildfires are a costly natural hazard, with rising expenses, mainly due to heightened susceptibility. However, climate change may also be a contributing factor. In January and February 2009, wildfires in Victoria, Australia, were estimated to have inflicted damages of at least AU$4 billion, including AU$1.2 billion in insurance claims, and resulting in 173 fatalities and the loss of 2,133 homes (Hincks *et al.*, 2011). Significant wildfires periodically transpire in California; the fall 2007 firestorms are estimated to have incurred US$1.8 billion in property damage, while the 2008 fire season caused losses above US$2 billion (Hincks *et al.*, 2011). The expenses associated with fire department suppression and preparation actions are rising. The purpose of this study is to understand the risks and causes of wildfires that occur due to climate change with the help of evaluating the results to make data-driven decisions.

## 1.2 Research Question, Aim and Objectives

### 1.2.1 Research Question

* How do machine learning models use climate, environmental, and historical wildfire data to accurately predict wildfire risk and estimate potential damage in high-risk Californian regions?

### 1.2.2 Aim

The study aims to analyse by predicting California's wildfire risk assessment.

### 1.2.3 Objectives

* To create machine learning models utilising climate and environmental data to predict areas in California at heightened risk for wildfires.
* To utilise predictive models to estimate potential damages associated with predicted wildfire events, including property loss and economic impact.
* To provide insights to policymakers and emergency management agencies to inform resource allocation and develop effective wildfire mitigation strategies.

# 2. Background

## 2.1 Introduction to California’s Wildfire

Hundreds of wildfires occur yearly in California, making them a regular event. The vast majority are insignificant, do little harm to the environment, and provide benefits to the environment. People and wildfires are incompatible (Lesser and Feinstein, 2020). In the years 2017 and 2018, California was struck by several deadly wildfires, the most notable of which was the Camp Fire of 2018, which claimed the lives of 85 people and was the worst wildfire in the state's history. Most wildfires are not started by electric utility infrastructure, including fallen power lines, nor do they involve the infrastructure of electric utilities. There were 3,470 events in 2017, according to data supplied by the California Department of Forestry and Fire Protection (CalFire). Of those incidents, 408 (12%) were traced to electric power equipment. 2017 was a very harsh year for wildfires (Lesser and Feinstein, 2020). Vehicles were responsible for 309 events, which is 9% of the total. Burning debris was responsible for 437 occurrences, 13%, and arson was responsible for 222 incidents, 6%. It is yet unknown what caused 871 wildfires, which accounts for 25% of the total (Yao *et al.*, 2022).

Additionally, wildfires related to electric utility equipment are often comparatively modest. Since 2014, electric companies have been required to report any wildfires that are caused by electric operations to the California Public Companies Commission (CPUC). Throughout the period from 2014 to 2018, the three investor-owned electric utilities in the state of California, namely Pacific Gas & Electric (PG&E), San Diego Gas & Electric (SDG&E), and Southern California Edison (SCE), recorded a total of 2,583 fire incidents (Lesser and Feinstein, 2020). Out of that total, 506 were distributed within a radius of less than three meters (10 feet) from the site of their ignition. Furthermore, 1,538 were classified as being smaller than one-quarter of an acre (Qiu *et al.*, 2022). Because of this, seventy-five percent of the flames had a size that was slightly more than one-fourth of an acre.

On the other hand, 22 flames were more significant than 100 acres, while 10 fires were more significant than 1,000 acres. Consequently, fewer than 0.4% of the wildfires reported over five years were more significant than 1,000 acres. Seventy-seven per cent of the wildfires were caused by the service region of PG&E, eighteen per cent by SCE, and five per cent by SDG&E (Lesser and Feinstein, 2020). Between 2009 and 2018, 45 critical wildfires over 300 acres were ascribed to failing electrical equipment and fallen power lines. All of these wildfires, except one, occurred inside the service area of PG&E. This accounts for fewer than 8% of the 654 significant wildfires in the state. In 2017, twenty wildfires were caused by electric activities, while in 2018, there were six wildfires caused by electric activities (Jiang *et al.*, 2021).

## 2.2 Selection Criteria for Papers

The selection criteria for papers related to predictive analytics on California wildfire risk assessment and damage estimation revolve around several critical factors. It is essential to start with relevance to the specific topic of wildfires, especially the ones that have been affecting California. Such studies using machine learning models, climate, and environmental data are also included, as well as studies that use historical wildfire data to predict wildfire risk and associated damages. The methodological rigour of the studies is second, and papers with robust analytical frameworks or with innovation in the data analysis are preferred. Thirdly, the recency of research ensures the findings are conducted on current trends and wildfire management challenges.

## 2.3 Critical Analysis of Key Papers

### 2.3.1 Paper 1: “Data-driven wildfire risk prediction in northern California”.

The study by Malik, Rao, *et al.* (2021) titled "Data-driven wildfire risk prediction in Northern California" presents a substantial contribution to wildfire risk assessment using machine learning techniques. In the Monticello and Winters regions, Malik, Rao, *et al.* (2021) used two random forest models to predict wildfire risk, achieving a perfection accuracy of 92% (Malik, Rao, *et al.*, 2021). The ability of data-driven approaches to increase the precision level by combining various datasets, which potential vegetation indices, climate variables and geographical features could improve wildfire prediction capability. Nevertheless, the study has its limitations. Secondly, on the positive side, the model is entirely accurate. However, they should always be mindful of the risk of overfitting, given that the model they used was a relatively complex random forest. Past fire data may not be sufficient for future fire dynamics affected by climate change and human activity. Second, the main geographic focus described in this study generally limits the generalisability of the findings to other areas with different environmental conditions or fire regimes.

Additionally, Malik, Rao, *et al.* (2021) did not discuss how variable selection and preprocessing techniques affect model performance. Another potential contribution is performing a more complete analysis of the parameters of the data most influencing predictions. Finally, while the study demonstrates the value of predictive analytics in policymakers, the study fails to dwell on how those predictions can be turned into actionable strategies for managing and mitigating wildfire (Malik, Rao, *et al.*, 2021).

### 2.3.2 Paper 2: “From static to dynamic prediction: Wildfire risk assessment based on multiple environmental factors."

Jiang et al.'s (2021) study "From static to dynamic prediction: Wildfire risk assessment based on multiple environmental factors" offers a complete wildfire risk assessment methodology that transitions from traditional static models towards dynamic prediction methods. Using a variety of environmental factors, such as days at population density, NDVI, or PDSI, their predictions were more accurate in California (Jiang *et al.*, 2021). Their methodology, which is based on dividing the state into a grid system for analysis, generates a finer, more nuanced understanding of wildfire risks that incorporate the spatial and temporal location of fires. Some limitations can be found in this research as well. More specifically, although the study considers dynamic factors, it may not fully reflect the quick changes in environmental conditions caused by variability in climate and human activity. Potential inaccuracies in risk assessments could result from the fact that these could not be continuously updated. Second, the results can be generalised to only California due to the chaining of Eq. 2 to Eq. 3 and the attempt to generalise their model to Washington without fine-tuning, which might miss regional differences in fire behaviour and environmental conditions (Jiang *et al.*, 2021).

Furthermore, the use of specific environmental variables may ignore other vital components determining wildfire risks, e.g. socio-economic dynamics or management practices. While the study has predictive capabilities, they are not discussed well enough in a real-world insurance operationalisation within applied forest management, thus lacking practical applicability for policymakers and emergency responders (Jiang *et al.*, 2021).

### 2.3.3 Paper 3: “Where there is Smoke, there is Fire: Wildfire Risk Predictive Modelling via Historical Climate Data”

The paper by Gholami *et al.* (2021), "Where There is Smoke, there is Fire: Wildfire Risk Predictive Modeling via Historical Climate Data," provides a data analytics solution that takes a look at the complexities in predictive modelling of wildfire risk from historical climate data and machine learning techniques (Gholami *et al.*, 2021). Gholami *et al.* (2021) underline the importance of a more complex take on predictive modelling, sensitive to the fact that most previous studies ignore temporal dynamics and lean on static data. This study's comprehensive evaluation of many predictive models on various landscapes in India dramatically strengthens the robustness of their findings. By successfully demonstrating they can improve predictive accuracy by incorporating historical burned areas, climate and geospatial data, Gholami *et al.* (2021) provide valuable information to decision-makers in wildfire management (Gholami *et al.*, 2021).

However, several limitations are present. However, because the study focuses on specific geographical regions, it may restrict the generalisation of its findings to other areas with different ecological conditions (Gholami *et al.*, 2021). Moreover, while Gholami *et al.* (2021) treat temporal aspects of data to some extent, it has not thoroughly discussed how rapidly changing climate conditions can specifically affect long-term prediction. Additionally, historical data might be vulnerable to biases rooted in historical records that might not completely capture the latter factors ultimately responsible for improved future wildfire risks with climate change. Finally, the paper does not discuss in depth the operationisation of these predictive insights to actual wildfire prevention strategies for practical application (Gholami *et al.*, 2021).

### 2.3.4 Paper 4: “Predicting electricity infrastructure induced wildfire risk in California”.

The study by Yao *et al.* (2022), "Predicting electricity infrastructure induced wildfire risk in California," critically examines the intersection between electrical infrastructure and wildfire risks, particularly in the context of California's increasing wildfire incidents. Yao *et al.* (2022) use advanced predictive modelling techniques to determine when and where power lines can ignite a wildfire (Yao *et al.*, 2022). This research is intense due to its emphasis on data, which acts upon giving results and creating a more reliable predictive model. This integration of different environmental factors, such as weather patterns and vegetation type, makes it possible to gain a complete understanding of weather patterns and conditions conducive for power line-induced wildfires. Since the history of catastrophic fires has historically been connected to electrical infrastructure, the timing of this approach is particularly relevant in light of the 2018 Camp Fire (Yao *et al.*, 2022).

However, several limitations are apparent, such as the fact that the study primarily concerns historical data, which would not adequately reflect future climate change impacts on wildfire dynamics of significant magnitude. Moreover, Yao *et al.* (2022) lack a discussion about the prospects of overfitting in their models, and thereby, they neglect the possibility of making misleading predictions in other contexts. Additionally, their findings regarding policy and utility management issues have not been explored entirely. An added discussion on how utilities may implement these predictive insights into an actionable wildfire mitigation strategy would further increase the practicality of the study (Yao *et al.*, 2022).

## 2.4 Summary of Literature Review

The review of predictive analytics in wildfire risk assessment literature indicates that more and more research is being conducted on machine learning and environmental data to improve prediction determinacy (Malik, Jalin *et al.*, 2021). While most studies highlight the need for different types of datasets, climate variables, historical wildfires, or geographical features that would enable robust predictive models, few studies compare and evaluate different repeatable model construction approaches. For example, Malik *et al.* (2021) developed an accurate data-driven prediction of wildfire risks in Northern California, and Jiang *et al.* (2021) show how static but dynamic models considering the temporal changes in environmental conditions are preferred. Nevertheless, the limitations common to these studies include the limited generalisability of the findings to other areas and the risk of overfitting both numeric and complex models. Furthermore, many papers fail to explain how their predictive insights could be operationalised to develop effective wildfire management strategies (Masoudvaziri *et al.*, 2022). It is also frequently noted that continuous model updates are often needed to reflect changes in climate conditions.

# References

Bhowmik, R.T., Jung, Y.S., Aguilera, J.A., Prunicki, M. and Nadeau, K. (2023) ‘A multi-modal wildfire prediction and early-warning system based on a novel machine learning framework’, *Journal of environmental management*, 341, p. 117908.

Gholami, S., Kodandapani, N., Wang, J. and Ferres, J.L. (2021) 'Where there is Smoke, there is Fire: Wildfire Risk Predictive Modeling via Historical Climate Data', in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 15309–15315. Available at: <https://ojs.aaai.org/index.php/AAAI/article/view/17797> (Accessed: 24 February 2025).

Hincks, T.K., Malamud, B.D., Sparks, R.S., Wooster, M.J. and Lynham, T.J. (2011) ‘Risk assessment and management of wildfires’, in *Risk and uncertainty assessment for natural hazards*. Cambridge University Press, pp. 398–444. Available at: <https://research-information.bris.ac.uk/en/publications/risk-assessment-and-management-of-wildfires> (Accessed: 24 February 2025).

Hu, P., Tanchak, R. and Wang, Q. (2024) ‘Developing risk assessment framework for wildfire in the United States–a deep learning approach to safety and sustainability’, *Journal of Safety and Sustainability*, 1(1), pp. 26–41.

Jiang, T., Bendre, S.K., Lyu, H. and Luo, J. (2021) ‘From static to dynamic prediction: Wildfire risk assessment based on multiple environmental factors’, in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 4877–4886. Available at: <https://ieeexplore.ieee.org/abstract/document/9672044/> (Accessed: 24 February 2025).

Lesser, J.A. and Feinstein, C.D. (2020) *PLAYING WITH FIRE: California’s Approach to Managing Wildfire Risks*. Available at: <https://media4.manhattan-institute.org/sites/default/files/californias-approach-to-managing-wildfire-risks-JL.pdf>.

Malik, A., Jalin, N., Rani, S., Singhal, P., Jain, S. and Gao, J. (2021) ‘Wildfire risk prediction and detection using machine learning in San Diego, California’, in *2021 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, internet of people and smart city innovation (smartworld/scalcom/uic/atc/iop/sci)*. IEEE, pp. 622–629. Available at: <https://ieeexplore.ieee.org/abstract/document/9604370/> (Accessed: 24 February 2025).

Malik, A., Rao, M.R., Puppala, N., Koouri, P., Thota, V.A.K., Liu, Q., Chiao, S. and Gao, J. (2021) ‘Data-driven wildfire risk prediction in northern California’, *Atmosphere*, 12(1), p. 109.

Masoudvaziri, N., Ganguly, P., Mukherjee, S. and Sun, K. (2022) ‘Impact of geophysical and anthropogenic factors on wildfire size: a spatiotemporal data-driven risk assessment approach using statistical learning’, *Stochastic Environmental Research and Risk Assessment*, 36(4), pp. 1103–1129. Available at: <https://doi.org/10.1007/s00477-021-02087-w>.

Mitchell, J.W. (2023) ‘Analysis of utility wildfire risk assessments and mitigations in California’, *Fire safety journal*, 140, p. 103879.

Qiu, L., Chen, J., Fan, L., Sun, L. and Zheng, C. (2022) ‘High-resolution mapping of wildfire drivers in California based on machine learning’, *Science of The Total Environment*, 833, p. 155155.

Yao, M., Bharadwaj, M., Zhang, Z., Jin, B. and Callaway, D.S. (2022) ‘Predicting electricity infrastructure induced wildfire risk in California’, *Environmental Research Letters*, 17(9), p. 094035.