Data Analysis of the Land Fire in California

Wenhan Zheng[a]
Faculty of Science, University of Bristol, Bristol, U.K.

Keywords: Wildfire Prediction, Landfire Risk Assessment, Machine Deep Learning, Predictive Modeling, Climate Data Analysis.

Abstract: On a typical afternoon in Los Angeles, California, on January 7, 2025, an unexpected natural disaster - the "land fire" - broke out. This epidemic is slightly different from the past. The fire was not brought under timely control, but continued to spread and decay, losing control. This article will attempt to focus on several types of data collected so far. Firstly, it analyzes the linear relationship between firefighting costs and time, then summarizes the different causes and percentages of fires in California and analyzes data on volcanic coverage areas and total numbers before 2023. These data involve the analysis and prediction of wildfires that have occurred in California over the past five years, including factors such as high-frequency periods, scale, and causes, as well as attempts to use mathematical models to minimize and predict the possible consumption of human and financial resources. Finally, some suggestions and prospects for fires were provided.

1 INTRODUCTION

To date, in the case of the previously mentioned California mountain fires, strong winds continue to accelerate the spread of the blazes, which have affected more than 26,000 acres and forced the evacuation of approximately 150,000 people (Southern, Harding, Hurley, et al, 2025). So to better prepare people to protect themselves in the face of disaster, accurate fire prediction and the costs involved have long-term implications for human development. This research aims to optimise the performance of the models by making more effective predictions to provide more hilarious early warning scenarios, allocate resources more efficiently and reduce the likelihood of natural disasters threatening people's lives (Hino & Field, 2023).

According to the annual statistics of natural disasters in the United States, it is not difficult to find that California, as a high incidence of hill fires, has actually shown some problems in recent years, for example, by the growing impact of greenhouse gas emissions, the global warming has also made some changes in the temperature of California, and with the increase in human activities it is inevitable, because people need to make a reasonable allocation of limited resources. Rational deployment, so we should prevent or reduce the damage caused by disasters more efficiently (Abatzoglou & Williams, 2016). According to research findings, it is not difficult to see that as global temperatures rise, resulting in a slowdown in the accumulation of water vapour in the clouds, the most direct impact of reduced precipitation is that the water content of the vegetation will be reduced, which is more for the occurrence of fires to create conditions conducive to the occurrence of fires. Therefore, this has led to the outbreak of mountain fires in high frequency.

Since the 1970s, the number of such disasters has increased significantly, more than half of the increase being due to man-made climate change. The area covered by hill fires has also nearly doubled during this period, and over the years, the occurrence of hill fires has led to a very large economic expenditure, resource loss, and even casualties.

The U.S. California Mountain fire is not the first time. Still, in recent years, as the fire is more and more difficult to control, the California mountain fire that occurred in early 2025 brought unprecedented havoc to the local people. Therefore, when more and more people began to pay attention to this phenomenon, it is not difficult to find that early many scholars have noticed such a problem, hoping to have through the

[a] https://orcid.org/0009-0004-7902-2954

Zheng, W.
Data Analysis of the Land Fire in California.
DOI: 10.5220/0013688200004670

In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 292-298
ISBN: 978-989-758-765-8

study of mountain fires occurring in time or spatial patterns. Among the many findings, Westerling et al. (2006) used historical data to analyse seasonal patterns, turning a vague abstract perception into a more intuitive and rational mathematical model, he argued that the first half of the year, in fact, that is, spring and summer, California's mountainous and forested areas are usually high-frequency occurrence of mountain fires, summing up the occurrence of the California Mountain Fire The reason for the occurrence of the California mountain fires may not be because of the season, but there must be some kind of strong correlation between the two.

In subsequent analyses, some scholars have suggested that a mixture of more frequent daily human activities and insufficient management and protection of forest resources have indirectly contributed to the problem.

This paper focuses on the collected data on wildfires that have occurred in California over the past five years, including analysis of factors such as high-frequency periods, scale, and causes of wildfires, as well as attempts to use mathematical models to minimize and predict the potential consumption of human and financial resources.

2 RELATED WORKS

In recent years, with the continuous development of data science and the unremitting efforts of scholars, people have tried to believe that mathematical models can help people solve many life problems. Predictive modelling is the best solution to meet various natural geological disasters and other emergencies. In the paper of Abatzoglou & Kolden (2013), they took and analyzed various collected data and constructed a more perfect statistical model for the prediction of mountain fires, because in the whole process, they took into account as much as possible all kinds of variables that may affect the results, topographic factors, wind factors, climate conditions are all taken into account, which makes the prediction based on real data as a support to the model. This makes the prediction models supported by real data more accurate and more convincing.

Later on, with the continuous updating and iteration of observation technology, remote sensing technology can be more powerful through satellites outside the sky, better use of spatial distribution, real-time follow-up of various data, as well as UHF transmission of real-time monitoring data through satellites have become the best aid to assist in the judgement process, but some human factors will greatly affect the analysis process and results, which is an insurmountable problem for the time being. This is an insurmountable problem. But the emergence of remote sensing technology, to a large extent, in addition to human influence, completes the task of data collection. And some scholars have borrowed such remote sensing data combined with data models to analyze the high-risk areas of the California wildfires, and ultimately succeeded in predicting the outbreak of several areas.

The deep learning model of the machine has made the accuracy of the model in the prediction process much better than before. Many researchers use algorithms such as SVM (SVM: Support Vector Machine), RF (Random Forest), etc. to perform simulations and give the final results of the algorithm. This process is also due to the rise of machine self-learning technology, which helps research teams to continuously improve and optimize, which is very helpful to analyze and find out the potential factors affecting the prediction results, which are different from the normal logic of human thinking, and may be overlooked.

The use of mathematical models is not really out of reach for humans, on the contrary as early as in the paper by Green, Kaiser & Shenton (2020) the evolution of the regions around the mountain fires has been predicted through the analysis of various graphical data and modeled by deep convolutional neural networks with very successful results in the Sierra Nevada of the Western Sierras, so it is all the more important to try to find commonalities of successful cases in this process, which is very meaningful for the development of new Machine Deep Learning models. This process of finding commonalities in successful cases is very relevant for the development of new Machine Deep Learning models.

Overall, as of the end of 2019, the researchers have counted around 300 publications on the most commonly used machine learning methods across a wide range of problem domains, including Max Ent, Artificial Neural Networks, Decision Trees, Genetic Algorithms, and so on. This means that there are differences in the ability of each model to learn and produce results under different conditions and environments, which leaves the human race to think about progress and hopefully continue to develop or derive a completely new algorithm (Jain et al., 2020).

More advanced deep learning methods such as CNNs (Convolutional Neural Networks) and LSTMs (Long Short-Term Memory Networks) are being developed to optimize both temporal and spatial data, Hosseini et al. (2020). In this process, through the real occurrence of mountain fires after data collection, and the prediction of comparison, it is not difficult to find that the LSTM model for the California mountain fire time series prediction, in the long time dependence of the performance is pretty good. In recent years, with the continuous development of data science and the unremitting efforts of scholars, people have tried to believe that mathematical models can help people solve many life problems. Predictive modelling is the best solution to meet various natural geological disasters and other emergencies. In the paper of Abatzoglou & Kolden (2013), they took and analyzed various collected data and constructed a more perfect statistical model for the prediction of mountain fires, because in the whole process, they took into account as much as possible all kinds of variables that may affect the results, topographic factors, wind factors, climate conditions are all taken into account, which makes the prediction based on real data as a support to the model. This makes the prediction models supported by real data more accurate and more convincing.

Later on, with the continuous updating and iteration of observation technology, remote sensing technology can be more powerful through satellites outside the sky, better use of spatial distribution, real-time follow-up of various data, as well as UHF transmission of real-time monitoring data through satellites have become the best aid to assist in the judgement process, but some human factors will greatly affect the analysis process and results, which is an insurmountable problem for the time being. This is an insurmountable problem. But the emergence of remote sensing technology, to a large extent, in addition to human influence, completes the task of data collection. And some scholars have borrowed such remote sensing data combined with data models to analyze the high-risk areas of the California wildfires, and ultimately succeeded in predicting the outbreak of several areas.

The deep learning model of the machine has made the accuracy of the model in the prediction process much better than before. Many researchers use algorithms such as SVM (SVM: Support Vector Machine), RF (Random Forest), etc. to perform simulations and give the final results of the algorithm. This process is also due to the rise of machine self-learning technology, which helps research teams to continuously improve and optimize, which is very helpful to analyze and find out the potential factors affecting the prediction results, which are different from the normal logic of human thinking, and may be overlooked.

The use of mathematical models is not really out of reach for humans, on the contrary as early as in the paper by Green, Kaiser & Shenton (2020) the evolution of the regions around the mountain fires has been predicted through the analysis of various graphical data and modeled by deep convolutional neural networks with very successful results in the Sierra Nevada of the Western Sierras, so it is all the more important to try to find commonalities of successful cases in this process, which is very meaningful for the development of new Machine Deep Learning models. This process of finding commonalities in successful cases is very relevant for the development of new Machine Deep Learning models.

Overall, as of the end of 2019, the researchers have counted around 300 publications on the most commonly used machine learning methods across a wide range of problem domains, including Max Ent, Artificial Neural Networks, Decision Trees, Genetic Algorithms, and so on. This means that there are differences in the ability of each model to learn and produce results under different conditions and environments, which leaves the human race to think about progress and hopefully continue to develop or derive a completely new algorithm (Jain et al., 2020).

More advanced deep learning methods such as CNNs (Convolutional Neural Networks) and LSTMs (Long Short-Term Memory Networks) are being developed to optimize both temporal and spatial data, Hosseini et al. (2020). In this process, through the real occurrence of mountain fires after data collection, and the prediction of comparison, it is not difficult to find that the LSTM model for the California mountain fire time series prediction, in the long time dependence of the performance is pretty good.

3 RESEARCH METHODOLOGY

3.1 Data Overview and Collection

The dataset used in this analysis contains information such as the time of the hill fire, the exact location, the cause of the fire, the duration, the estimated spread area, the number of casualties, the expected damage to property, casualty statistics, etc. Before analyzing the acquired data, the analysis focused on the fires that occurred in Los Angeles, California, USA, because of the random nature of the fires.

1. The time of the occurrence
2. Duration of the disaster
3. The scene of the incident
4. Types of incentives

5. Casualty statistics

3.2 Statistic Analysis

In comparison with the 5-Year Average, the 2025 statistics show a much higher number of fires (164 vs. 76) and acres burned (40,695 vs. 30) compared to the 5-year average. This further confirms the severity of the current fire season in comparison to previous years.

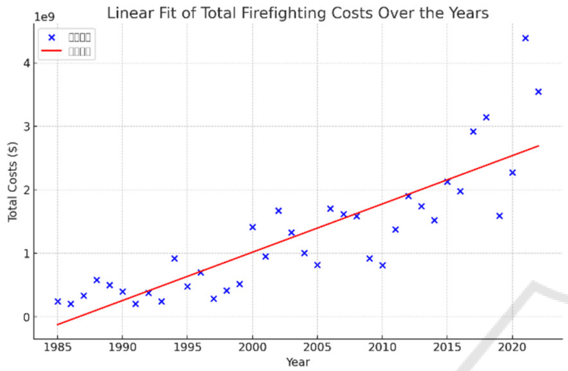

Figure 1: Linear fit of costs (Picture credit: Original)

4 RESULTS OF THE EXPERIMENT

4.1 Analysis of the data

With historical statistics to back up the analysis, it is easy to see that the level of human activity is different from a century ago, excessive behavioural activity has already had a non-negligible impact on the results of the model, and the anthropogenic variables, as reported in (Mann et al.,2016), are shown to at least make the significance of climatic factors misjudged i.e. climatic conditions are misjudged to be the trigger of the events such as those that lead to the mountain fires probability would increase by 24% close to a quarter. And that would lead the study in a direction skewed away from the facts. So this emphasizes the model and tests the ability of the model to combine multiple factor variables and also puts a higher demand on the accuracy of the correlation of the variables.

Figure 1 shows a positive correlation, meaning that as the years progress, the total firefighting costs also exhibit an upward trend. The trend is represented by the red linear fit line. Contrasting the data with years, the number of wildfires in 2023 is below the five-year and 10-year averages, and down from previous years. The Correlation (R-value) shown in the graph represents a linear fit, and visually, it shows a linear relationship, indicating that the data points generally follow an increasing trend. There are also outliers that could not be ignored. During the preprocessing part of the data before fitting the linear relationship, it was calculated that the outliers occur more frequently in the early 1980s and 2000s. They are completely off the predicted straight line and outside the IQR (Interquartile Range). The straight line fitted by the data calculation, analysis, and most of the cases are between the real data high and low sides, the deviation is not big, which shows that the linear model is still good for the prediction of the expenditure model fitting effect.

4.2 Prediction and Conclusion

In the U.S. statistics show that most of the triggers of life, in fact, are man-made, and common causes of fires include high temperatures, low humidity in forested vegetation, and thunderstorms in which sky lightning strikes flammable materials or man-made, but surprisingly the area burned and covered by mountain fires caused by man-made causes is more than twice the area burned by natural causes (Balch et al.,2017).

The predictive results of the entire model appear to be quite good and should continue to be successful in predicting fire protection expenditures in the coming years, but if more complex models or variables can be considered, the predictions and models will be more reliable.

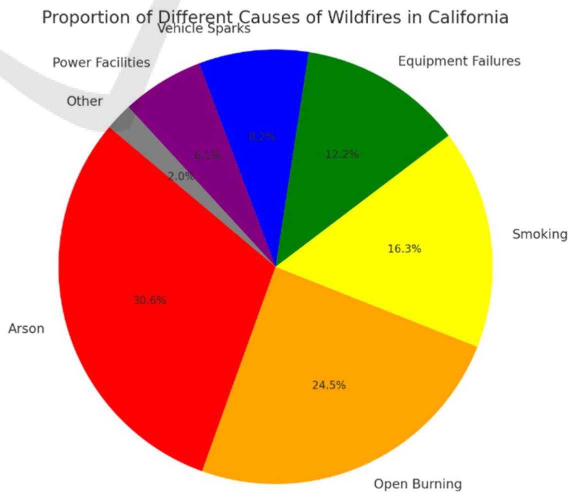

Figure 2: Proportion of the causes (Picture credit: Original)

Figure 2 gives a very clear picture of the different types of causes of the California fires, and the percentages. Of these, 71.9% are attributed to Arson, Smoking, and Open Burning, which together make up the most significant contributors, with the remainder being mostly caused by aging machinery and equipment, or damaged energy equipment.

The above statement does not mean that there is no link between climate change and the frequency of hill fires. On the contrary, in Turco et al (2023), it is argued that for nearly a quarter of a century, between 1996 and 2020, the number of hill fires in the California region of the United States was five times higher than the number of fires that had occurred between 1971 and 1995, and that these changes were most directly attributable to the climate change that was evident in the investigations and modelling analyses.

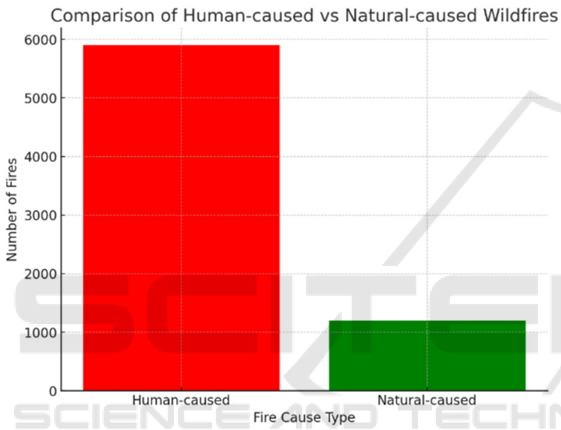

Figure 3: Comparison of two kinds of wildwires (Picture credit: Original)

Figure 3 shows the number of California fires caused by both human-caused and natural factors over the years, with the former portion being significantly higher than the latter portion. Human-caused factors (especially Arson, Open Burning, and Smoking) are the primary drivers Human-caused factors (especially Arson, Open Burning, and Smoking) are the primary drivers behind wildfires in California, accounting for over 75% of total fires.

In short, based on the data provided by CAL FIRE each year, it is easy to see that fire protection expenditures have been on the rise, and accordingly, some policies or adjustments may be needed to help better manage the situation, such as severe penalties for arsonists, harsher fines, or more maintenance and improvement of forests. For those outliers, although they may not be used by the model, if they occur consecutively, they indicate that the model is inappropriate or that something very bad is going to happen, which can serve as a good warning.

However, because the mountain fires occurring in early 2025 are different from previous cases, they will be ignored as outliers during data processing so as not to affect the results of the analyses, and the paper will concentrate on analyzing the complete data before 2023. As the figure 4 and figure 5 below shows.

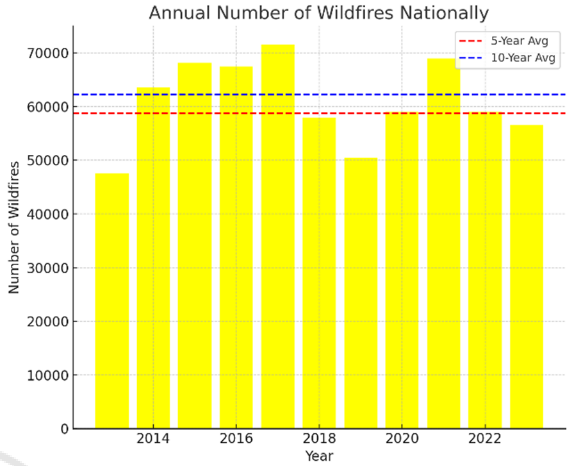

Figure 4: Total number of the fires (Picture credit: Original)

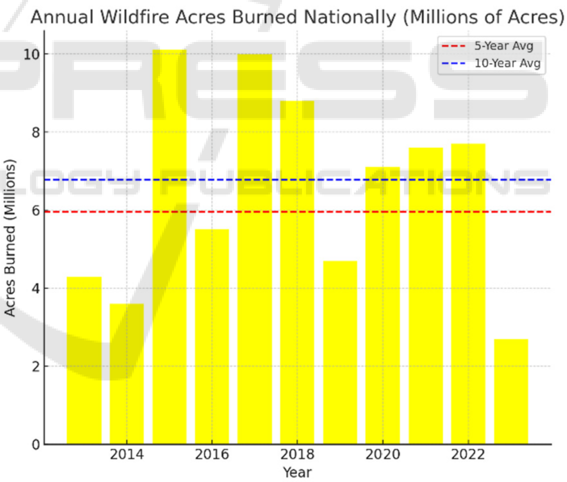

Figure 5: Area covered by hill fires (Millions of Acres) (Picture credit: Original)

The horizontal dashed lines represent averages over the past five and ten years, providing a clear visual reference for 2023 comparison with historical data.

5 SUGGESTIONS

5.1 Enhanced Fire Prevention and Awareness Campaigns

Enhanced Fire Prevention and Awareness Campaigns

Strengthening the publicity and education on fire prevention awareness from childhood, and reasonable community-focused education are very targeted preventive activities, which can effectively reduce the number of hill fires caused by the human factor, because the human factor is very often unconsciously caused by human beings seemingly small behavior. This is an approach that requires long-term persistent supervision, but the effect will not be immediate.

5.2 Stronger Regulations and Enforcement

Human-caused fires must be dealt with seriously, and regulations must be strengthened, while utilities must reinforce infrastructure such as fences, power grids and walls to minimize the possibility of human influence. Secondly, clearer regulations and penalties should be added for the use of pyrotechnics and automotive equipment.

5.3 Improve Firefighting and Emergency Response

While maintaining current firefighting expenditures, the use of various drone technologies, similar to the class of drones, replaces humans in a variety of high-risk rescue activities, data-monitoring activities, emergency response, and other tasks. All of these need more funding as support. Because robots can detect fires earlier and more timely, at an earlier time to contain the spread of fire.

Secondly, for accident prone areas, the use of cameras should be more widespread and the implementation of remote sensing technology is imperative. Similarly, in the face of a lack of materials, the optimal allocation of resources is also necessary, as long as this can be avoided, the dilemma of emergency rescue supplies. Emergency evacuation drills, while enhancing people's self-protection, can also make the face of disaster, human beings are harmed can be reduced to a minimum.

6 CONCLUSIONS

Based on the statistical results, graphs, etc., it is clear that the frequency of hill fires in California has been escalating in recent decades. In particular, the area covered by each fire has been expanding, and the summer burned area in northern and central California has increased about five-fold from 1996-2021 compared to 1971-1995. Secondly, for a seasonal problem such as hill fires, the extreme seasonal anomalies that are generated illustrate how the severity of hill fires has repeatedly challenged the limits of human control over nature. These anomalies need to be explained, and the dramatic changes in climatic conditions leading to higher temperatures and drying out of the ground as a direct result of the fires are lengthening the duration of the high fire season.

And while trying to find ways to control the fire, humans should also face their problems, according to the results of the data mediation shows that about 85% of the mountain fires are for ignition caused in, human causes including accidents as well as negligence but are not an excuse for humans to ignore the problem.

For the existing mountain fire prediction models, continuous optimisation and technological updates are also important. First of all, AI algorithms should be integrated into the process of predictive algorithms to improve the ability to sense the state of affairs, so that he can be the first time through the network to obtain a variety of data at the same time, using their arithmetic power to analyse satellite imagery, mapping, providing rescue options route, and so on. Similarly, it is important to continue to develop hybrid modelling, as the models I have used in this thesis are the most basic data analysis models, which can vary greatly in accuracy when faced with multivariate, dynamic data problems. In conclusion, improving the entire algorithmic model is also a multifaceted endeavour that needs to be taken in tandem to better protect humans from themselves.

REFERENCES

Abatzoglou, J. T., & Kolden, C. A. 2013. Climate change in western US deserts: Potential for increased wildfire. *International Journal of Wildland Fire*, *22*(5), 635-645.

Abatzoglou, J. T., & Williams, A. P. 2016. Impact of anthropogenic climate change on wildfire across western US forests. Proceedings of the National Academy of Sciences, 113(42), 11770-11775.

Balch, J. K., Bradley, B. A., Abatzoglou, J. T., Nagy, R. C., Fusco, E. J., & Mahood, A. L. 2017. Human-started wildfires expand the fire niche across the United States. *Proceedings of the National Academy of Sciences, 114*(11), 2946–2951.

Faramarzi, H., Hosseini, S. M., Pourghasemi, H. R., & Farnaghi, M. 2021. Forest fire spatial modelling using ordered weighted averaging multi-criteria evaluation. *Journal of Forest Science*, *67*(2), 87-100.

Green, M. E., Kaiser, K., & Shenton, N. 2020. *Modeling wildfire perimeter evolution using deep neural networks*. *arXiv*. https://arxiv.org/abs/2009.03977

Hino, M., & Field, C. B. 2023. Fire frequency and vulnerability in California. *PLOS Climate*, 2(2), e0000087.
Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. 2020. A review of machine learning applications in wildfire science and management. *Environmental Reviews, 28*(4), 478–505.
Mann, M. L., Batllori, E., Moritz, M. A., Waller, E. K., Berck, P., Flint, A. L., Flint, L. E., & Dolfi, E. 2016. Incorporating anthropogenic influences into fire probability models: Effects of human activity and climate change on fire activity in California. *PLOS ONE, 11*(4), e0153589.
Southern, K., Harding, D., Hurley, B., & Blanco, A. 2025. California fire forces evacuation of Pacific Palisades in Los Angeles. *The Times*.
Turco, M., Abatzoglou, J. T., Herrera, S., Zhuang, Y., Jerez, S., Lucas, D. D., AghaKouchak, A., & Cvijanović, I. 2023. Anthropogenic climate change impacts exacerbate summer forest fires in California. *Proceedings of the National Academy of Sciences, 120*(25), e2213815120.