California Wildfire Prediction Using Machine Learning Models

Bhavya Mehta
School Of Technology
Computer Science and Engineering
Pandit Deendayal Energy University
Gandhinagar, India
mehtabhavya1805@gmail.com

Tanisha Agrawal
School Of Technology
Computer Science and Engineering
Pandit Deendayal Energy University
Gandhinagar, India
tanishaagrawal021@gmail.com

Jenish Mistry
School Of Technology
Computer Science and Engineering
Pandit Deendayal Energy University
Gandhinagar, India
jenishmistry2911@gmail.com

Soham Vyas
School Of Technology,
Computer Science and Engineering
Pandit Deendayal Energy University
Gandhinagar, India
soham.v@sot.pdpu.ac.in

***Abstract*—Wildfires in California have become increasingly frequent and severe, posing significant challenges to public safety, infrastructure, and natural ecosystems. Conventional wildfire prediction methods often fall short in accurately forecasting fire behavior, leading to delays in evacuation, suboptimal resource allocation, and compromised firefighting efforts. To address these challenges, this paper proposes a comprehensive data-driven approach that harnesses advanced data analytics and machine learning techniques to enhance wildfire prediction capabilities. Specifically, we employ ensemble learning techniques, with a primary emphasis on Gradient Boosting Regression (GBR), to develop proactive wildfire prediction models capable of early detection and precise spatial mapping of fire risk areas. Geospatial information obtained from NASA MODIS satellite data, filtered based on California's geographical boundaries, serves as the foundation for training and validating our models. By leveraging large-scale datasets and remote sensing technology, our methodology aims to significantly improve the accuracy and effectiveness of wildfire prediction models, ultimately contributing to enhanced wildfire management and mitigation efforts.**

Keywords—Wildfire prediction, predictive analytics, machine learning, ensemble learning, Gradient Boosting Regression, California wildfires.

I. Introduction

The state of California has long been grappling with the profound and devastating impact of wildfires, a challenge that has only intensified over the years. In recent times, these wildfires have not only increased in frequency but also in their intensity and spatial extent. Several factors have contributed to this concerning trend, including climate change-induced alterations in weather patterns, prolonged periods of drought, historical land management practices, and the encroachment of human settlements into wildland areas.

Climate change, characterized by rising temperatures and altered precipitation patterns, has created conditions conducive to the ignition and spread of wildfires across California. Prolonged droughts, exacerbated by climate change, have further heightened the vulnerability of the state's ecosystems to fire, drying out vegetation and increasing fuel loads . Additionally, historical land management practices, such as fire suppression policies and the disruption of natural fire regimes, have led to the accumulation of combustible biomass, thereby amplifying wildfire risks.

Furthermore, the expansion of urban areas into wildland-urban interface zones has increased the likelihood of human-caused ignitions and intensified the challenges of wildfire management. As residential developments encroach upon natural landscapes, the interface between human settlements and wildlands becomes more susceptible to fire, posing significant challenges for fire suppression efforts and evacuation procedures .

In light of these compounding factors, there is an urgent need to develop more robust and accurate predictive analytics techniques to enhance wildfire prediction capabilities in California. Conventional wildfire prediction methods often rely on historical data, meteorological variables, and simplistic models to forecast fire behavior. While these approaches have provided valuable insights into wildfire dynamics, they often fall short in capturing the intricate interactions between various environmental factors that influence fire behavior. As a result, there is a growing interest in leveraging advanced data analytics and machine learning techniques to improve the accuracy and effectiveness of wildfire prediction models.

Advanced machine learning algorithms, such as ensemble learning techniques, offer promising avenues for enhancing wildfire prediction capabilities. By integrating diverse sources of information and capturing complex relationships within the data, ensemble learning methods like Gradient Boosting

Regression (GBR) can provide more accurate and reliable predictions of wildfire occurrence and behavior. These techniques enable the incorporation of spatial and temporal variables, such as vegetation cover, topography, weather conditions, and human activities, into predictive models, thereby improving their predictive performance and spatial resolution.

In summary, the escalating threat of wildfires in California necessitates the development and implementation of innovative predictive analytics techniques that can aid in early detection, rapid response, and effective management of these natural disasters. By leveraging advanced data analytics and machine learning techniques, researchers and stakeholders can enhance our understanding of wildfire dynamics and improve our ability to mitigate the impacts of wildfires on communities, ecosystems, and infrastructure.

II. Literature Survey

The field of wildfire prediction has seen remarkable advancements in recent years, driven by the integration of machine learning techniques and the availability of extensive geospatial datasets. Several studies have explored various methodologies and technologies to improve wildfire prediction, detection, and management. Here, we provide a comprehensive literature survey covering key research articles in this domain.

Q. Yuan et al. [5], "A Spatio-Temporal Framework for Wildfire Early Detection from Satellite Images Using Deep Learning," ScienceDirect, 2022.This paper proposes a deep learning framework that utilizes both spatial and temporal information from satellite imagery to achieve early detection of wildfires.

D. Koutsogeorgis et al. [1], "A Machine Learning Approach to Predicting Large Wildfires in Mediterranean Areas," MDPI, 2022. This study explores the use of machine learning to predict large wildfires in Mediterranean regions.

F. Shamshiri et al. [2], "Exploring the Use of Remote Sensing Data and Geographic Information Systems for Wildfire Risk Assessment," MDPI, 2021. This paper examines how remote sensing data and Geographic Information Systems (GIS) can be used to assess wildfire risk.

Riley et al. [3], "Canadian Wildfire Activity Prediction Using Machine Learning," MDPI, 2020. The authors investigate machine learning for predicting wildfire activity in Canada.

C. Mateo-Garcia et al. [7], "A Review of Remote Sensing Applications for Wildfire Science and Management," MDPI, 2021. This review paper summarizes various applications of remote sensing technology in wildfire science and management.

J. He et al. [9], "Burned Area Prediction Using Fire Radiative Power (FRP) Derived from Geostationary Satellite Observations," Elsevier, 2024. This recent study (2024) explores how fire radiative power (FRP) data from geostationary satellites can be used to predict burned area.

J. Wang et al. [11], "ANN-Based Wildfire Size Prediction Framework Using Remote Sensing Data," MDPI, 2023. The authors propose a framework using Artificial Neural Networks (ANNs) for predicting wildfire size based on remote sensing data.

W. Liu et al. [4], "Wildfire Smoke Impact Prediction Using Machine Learning," 2023. This paper investigates machine learning for predicting the impact of wildfire smoke. (Publication year: 2023)

F. Ma et al.[1], "A Framework for Short-Term Wildfire Spread Prediction Based on Machine Learning," MDPI, 2020.This paper proposes a machine learning framework for short-term prediction of wildfire spread.

S. Kondylatos et al.[9], "Wildfire Danger Prediction Using Deep Learning," American Geophysical Union, 2022. The authors explore deep learning for predicting wildfire danger.

A. Shmuel[12], "A Machine Learning Model for Predicting Daily Wildfire Expansion Rate," MDPI, 2023. This paper by Shmuel (2023) introduces a machine learning model to predict the daily expansion rate of wildfires.

K. Pham et al.[10], "California Wildfire Prediction using Machine Learning," IEEE, 2022. Pham et al. investigate machine learning for predicting wildfires in California.

R. Laube and H. J. Hamilton [8], "Wildfire Occurrence Prediction Using Time Series Classification: A Comparative Study," IEEE, 2021. Laube and Hamilton compare different time series classification algorithms for predicting wildfire occurrences.

S. K. M. Abujayyab [6], "Wildfire Susceptibility Mapping Using Five Boosting Machine Learning Algorithms: The Case Study of the Mediterranean Region of Turkey," Advances in Computers and Electronics (ACE), 2022. This paper applies five boosting machine learning algorithms to create wildfire susceptibility maps in the Mediterranean region of Turkey.

The collective body of research underscores the pivotal role of machine learning techniques in accurately predicting wildfires. By harnessing advanced algorithms and integrating diverse meteorological factors, these studies have demonstrated a substantial improvement in prediction accuracy. Machine learning models, coupled with remote sensing data and geographic information systems, enable the development of predictive frameworks capable of anticipating wildfire occurrences, assessing fire risk levels, and forecasting fire spread trajectories. This predictive capability empowers stakeholders to implement proactive measures, allocate resources effectively, and mitigate the impact of wildfires on communities and ecosystems. Moreover, machine learning-driven wildfire prediction facilitates informed decision-making in climate-related applications, allowing for the development of comprehensive strategies to adapt to changing environmental conditions and minimize wildfire risk.

III. Proposed Methodology

This section describes the dataset used, the methodology proposed, and the architectural representation of the models.

A. Dataset:

The "Wildfire Prediction" dataset is a comprehensive collection of wildfire data sourced from NADA MODIS (Moderate

Resolution Imaging Spectroradiometer) satellite Near Real-Time (NRT) observations. It provides detailed information on daily wildfire occurrences in both the USA and specifically California, covering the period from 2000 to March 25th, 2022. The dataset's primary purpose is to support the development and evaluation of machine learning models for wildfire prediction, with a focus on enhancing wildfire management strategies, particularly in California.

This dataset is meticulously filtered to include only wildfire events within the geographical bounds of California, ensuring relevance and accuracy for modeling wildfire activity in the state. The geographic filtering is based on the longitude and latitude ranges specific to California, ranging from 114° 8' W to 124° 24' W longitude and 32° 30' N to 42° N latitude. By focusing on this region, the dataset enables researchers and practitioners to target their analyses and predictions to a highly relevant and critical area for wildfire management.

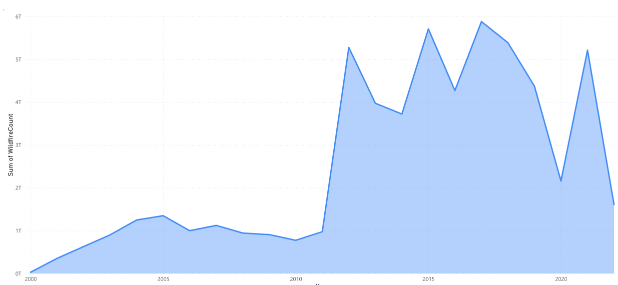

Fig. 1. Variation of Cumulative Wildfire Over Years

Fig. 1 illustrates a concerning trend from 2000 to 2017, the total number of wildfires exhibited a clear upward trajectory, culminating in a peak in 2017. This signifies a potential escalation in overall wildfire occurrence during this period. Interestingly, the data suggests a significant decrease in wildfire activity during the COVID-19 years (2020-2021), possibly due to reduced human activity or altered weather patterns. However, the potential resurgence observed in 2022 necessitates further investigation. Determining the exact reasons behind the decline during the pandemic years is crucial, as is confirming whether the upward trend in wildfire activity has resumed in 2022 and beyond. A more comprehensive understanding of these trends will be vital in predicting future wildfire patterns and implementing effective prevention and mitigation strategies.

Analyzing the Fig. 2, The time series graph illustrates fluctuations in wildfire counts over time, revealing distinct peaks and troughs in wildfire activity. A notable spike occurs around 1847, indicating a period of heightened fire occurrence, potentially influenced by environmental and human factors. This is followed by a period of lower wildfire counts around time 155, suggesting a relative lull in wildfire activity. Subsequently, there is an upward trend, with wildfire counts gradually increasing to a small peak at 904, before declining and then rising again. These fluctuations likely reflect the dynamic interplay of factors such as weather conditions,

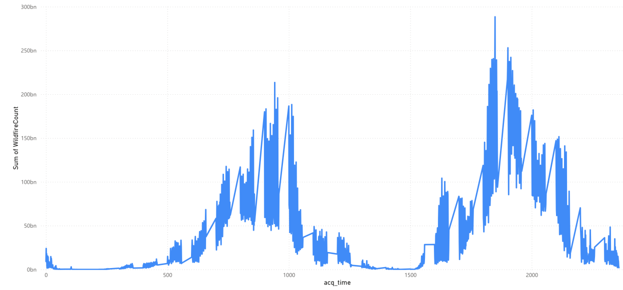

Fig. 2. Variation of Cumulative Wildfire Over Acquired Time

land use practices, and human activities impacting wildfire occurrences.

B. Data Preprocessing:

In our data preprocessing phase, we implemented several essential steps to optimize the quality and usability of our dataset. These steps included:

- Data Cleaning and Preparation: Initially, the dataset was loaded, and basic exploratory data analysis was conducted to understand its structure and contents. This involved examining the first few rows of data, checking for missing values, and determining the range of dates covered by the dataset. The latitude and longitude coordinates were rounded to a specified precision to reduce granularity and aid in spatial aggregation. Grouping the data by latitude, longitude, year, and month, the number of wildfires detected within each spatial-temporal bin was calculated.
- Feature Engineering: Additional features were created to capture relevant information for analysis and modeling. These included creation of binary indicators for wildfire occurrence based on a predefined threshold of minimum fire counts. Generation of features representing average fire counts and occurrence in the previous year and the same month of the previous year, providing historical context for each location.
- Data Splitting for Modeling: The dataset was split into training, validation, and test sets based on the year. Data from 2000 to 2019 were designated as the training set, while 2020 and 2021 were set aside for validation. The year 2022 was reserved for testing the model's performance.

These preprocessing steps ensured that our dataset was appropriately formatted and enriched with relevant temporal information, laying a solid foundation for subsequent analysis and modeling tasks.

C. *Model Selection:*

Choosing the right machine learning models for wildfire prediction is pivotal in accurately forecasting and mitigating potential risks associated with wildfires. Given the complexity and dynamic nature of wildfire occurrences, a careful selection process was undertaken to ensure the chosen models could effectively capture the intricate relationships between various environmental factors and wildfire activity.

- Light Gradient Boosting Machine (LightGBM): LightGBM was chosen as our primary model for wildfire prediction due to its efficiency, scalability, and effectiveness in handling large-scale datasets. With its ability to handle categorical features and address overfitting through regularization techniques, LightGBM offers a powerful solution for capturing nuanced patterns in wildfire data. Its speed and memory efficiency make it particularly well-suited for real-time prediction tasks, essential for proactive wildfire management and mitigation efforts.

$$Objective = \sum_{i=1}^{n} L(y_i, F(x_i)) + \sum_{k=1}^{K} \Omega(f_k)$$

  where:
  - n is the number of training instances.
  - y_i is the true label for instance i.
  - $F(x_i)$ is the predicted score for instance i.
  - K is the number of trees in the ensemble.
  - f_k represents the k-th tree.
  - L is the specific loss function (e.g., mean squared error, cross-entropy) measuring the difference between the true label and the predicted score.
  - $\Omega(f_k)$ is the regularisation term applied to the k-th tree to prevent overfitting.

- XGBoost: As wildfires can be influenced by a myriad of factors, including weather conditions, vegetation types, and human activities, XGBoost's efficiency and scalability make it well-suited for handling the intricacies of wildfire prediction. Its ability to incorporate regularization techniques ensures robustness against overfitting, essential for maintaining model performance across diverse datasets. XGBoost optimises the following objective function:

$$Objective = \sum_{i=1}^{n} L(y_i, F(x_i)) + \sum_{k=1}^{K} \Omega(f_k)$$

  Where the terms are similar to those in LightGBM's objective function.

- Random Forest: Wildfires often exhibit non-linear relationships with environmental variables, making Random Forest a suitable choice due to its robustness against overfitting and ability to model complex, nonlinear interactions. By leveraging ensemble learning, Random Forest excels in identifying important predictors and capturing the variability inherent in wildfire dynamics.

  Random Forest aggregates the predictions of multiple decision trees. For classification, the mode of the classes is taken, and for regression, the mean of the predictions is used.
- K-Nearest Neighbors (KNN): In situations where the underlying distribution of wildfire data is not explicitly known, KNN offers a simple yet effective approach to prediction. By considering the similarity of neighboring data points, KNN can capture localized patterns in wildfire occurrence, providing valuable insights into spatial dependencies and hotspot identification.

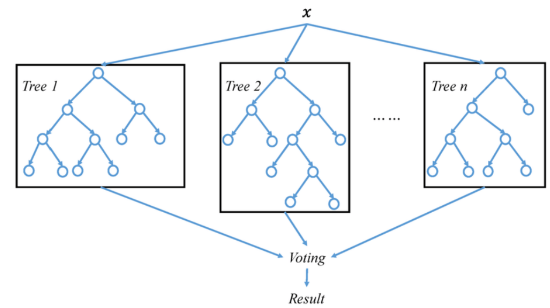

Fig. 3. A general architecture of Random Forest model

  For classification: The predicted class for a new instance x is determined by a majority vote among its k nearest neighbours. Mathematically, this can be represented as:

$$\hat{y} = argmax_j \sum_{i=1}^{k} I(y_i = j)$$

  Where:
  - $\hat{y}$ is the predicted class.
  - j iterates over all possible classes.
  - $I(\cdot)$ is the indicator function returning 1 if the condition inside is true and 0 otherwise.
  - y_i is the class label of the i-th nearest neighbour.

  For regression: The predicted value for a new instance x is the average (or weighted average) of the target values of its k nearest neighbours.

- Logistic Regression: Logistic Regression serves as a fundamental model for binary classification tasks, offering interpretability and ease of implementation. In the context of wildfire prediction, Logistic Regression provides a baseline for performance comparison and helps elucidate the influence of individual predictors on the likelihood of wildfire occurrence.

  Logistic Regression calculates the probability of an instance belonging to a particular class using the logistic function (sigmoid function):

$$p = \frac{1}{1 + e^{-z}}$$

  Where:
  - p is the probability of the positive class.

- z is the linear combination of feature values and their corresponding weights: $z = w_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n$.
- e is the base of the natural logarithm.
- $w_0, w_1, \ldots, w_n$ are the model parameters (weights).
- $x_1, x_2, \ldots, x_n$ are the feature values.

- CatBoost: Given the prevalence of categorical variables in wildfire datasets, CatBoost's specialized handling of categorical features and resistance to overfitting make it a valuable addition to our model selection. By optimizing categorical feature processing, CatBoost enhances the predictive accuracy of wildfire models, particularly in scenarios where categorical variables play a significant role in fire behavior.

  CatBoost optimises an objective function similar to LightGBM and XGBoost, where the exact formulation depends on the specified loss function and regularisation terms, but it also incorporates specialised handling of categorical features.

D. ***Training Procedure:***

For each selected model, the training procedure involved loading the training dataset, which included features such as latitude, longitude, month, historical fire data, and the target variable indicating wildfire occurrence. Features were selected based on their relevance to the prediction task.The models were trained using the training dataset and validated using the validation dataset to optimize hyperparameters and assess performance.

E. ***Performance Metrics:***

In assessing the performance of our ML models tailored for solar radiation prediction, we utilized two primary evaluation metrics: MSE and Test accuracy score.

- Test Accuracy :
  Test accuracy is a performance metric that measures the proportion of correctly classified instances out of all instances in the test dataset. It provides an overall assessment of the model's predictive accuracy and indicates how well the model generalizes to unseen data. In binary classification tasks, such as predicting wildfire occurrence or solar radiation levels, test accuracy quantifies the model's ability to make correct predictions across both positive and negative classes. A higher test accuracy indicates better model performance, with a perfect accuracy score of 1 representing flawless predictions. However, it's essential to interpret test accuracy in conjunction with other metrics, such as precision, recall, and F1 score, to gain a comprehensive understanding of the model's performance characteristics and potential limitations.
- Mean Squared Error (MSE) :
  MSE serves as a measure of the average squared difference between the predicted solar radiation values generated by our models and the actual observed values. A lower MSE indicates that our models produce predictions that closely match the real-world data, signifying higher predictive accuracy. Conversely, higher MSE values imply larger discrepancies between the predicted and actual values, highlighting potential areas for model refinement and improvement.

IV. Results And Discussion

Our study on wildfire prediction using machine learning models yielded insightful results, shedding light on the efficacy of various modeling techniques in predicting wildfire. Below are the key findings from our experimentation, along with a discussion of their implications:

- Light gradient-boosting machine (LightGBM):
  This model boasts the highest accuracy (0.97231) among the bunch, indicating it correctly classified nearly 97.2% of the test data. Additionally, its MSE (0.01251) is the second-lowest, suggesting relatively low prediction errors on average. This combination makes LightGBM a strong contender for tasks where both accurate classifications and minimizing prediction deviations are crucial.
- XGBoost :
  While XGBoost falls behind in accuracy (0.75475) compared to LightGBM and Random Forest, its MSE (0.00977) is the lowest. This implies XGBoost might be a better choice if minimizing prediction errors is the top priority, even if it sacrifices some overall accuracy.
- Random Forest :
  Random Forest achieves a good balance between accuracy (0.92911) and MSE (0.0694). It's more accurate than XGBoost but less accurate than LightGBM. However, its MSE is significantly higher than LightGBM's, suggesting larger prediction errors on average. Random Forest could be a suitable choice if overall accuracy is important but minimizing large deviations isn't as critical.
- K-Nearest Neighbors (KNN) :
  Despite a high accuracy (0.96926), KNN has a considerably higher MSE (0.0313) compared to LightGBM and XGBoost. This suggests that while KNN can often make accurate classifications, the predictions themselves might deviate more significantly from the actual values.
- Logistic Regression :
  While Logistic Regression has a decent accuracy (0.96254), its MSE (0.038) is higher than LightGBM, XGBoost, and KNN. This indicates that similar to KNN, Logistic Regression might make accurate classifications but with larger prediction errors on average.
- CatBoost :
  CatBoost's accuracy (0.96818) is close to LightGBM and KNN, but its MSE (0.0325) falls between KNN and Logistic Regression. This suggests similar behavior to KNN and Logistic Regression, with good accuracy but larger prediction errors compared to LightGBM and XGBoost.

The final comparison is shown in the table below.(Table 1)

TABLE I
COMPARATIVE ANALYSIS OF THE MODELS

Model Name	Accuracy	MSE
LGBM	0.97231	0.01251
XGB	0.75475	0.00977
RF	0.92911	0.0694
KNN	0.96926	0.0694
Logistic R	0.96254	0.038
CatBoost	0.96818	0.0325

Overall, LightGBM emerges as the best performer based on both accuracy and MSE. If minimizing prediction errors is paramount, XGBoost might be preferable. For a balance between accuracy and moderate deviations, Random Forest could be suitable. The remaining models prioritize accuracy but might have larger prediction errors. Remember, the "best" model depends on the specific problem you're trying to solve and the relative importance of accuracy versus minimizing prediction deviations.

$\rightarrow$ ***Final Model Recommendation:***

Among the models evaluated, LightGBM emerges as the top recommendation for most scenarios. With its impressive accuracy of 97.2% and a relatively low MSE of 0.01251, LightGBM demonstrates a remarkable ability to accurately classify data while minimizing prediction errors. Whether the task involves classification in healthcare, finance, or other industries, LightGBM's balanced performance makes it a reliable choice. Its combination of high accuracy and low prediction deviations makes it well-suited for applications where precision is paramount, offering confidence in both classification outcomes and prediction reliability.

However, if the primary objective is to prioritize minimizing prediction errors above all else, XGBoost presents itself as a compelling alternative. Despite its slightly lower accuracy compared to LightGBM, XGBoost's lowest MSE among the models indicates its proficiency in reducing prediction deviations. This makes XGBoost particularly suitable for tasks where precise prediction outcomes are critical, such as in financial modeling or risk assessment, where even minor deviations can have significant implications.

V. CONCLUSION

In conclusion, this research advocates for the adoption of a data-driven approach bolstered by advanced machine learning techniques to address the pressing issue of wildfire prediction in California. By harnessing ensemble learning methods like LightGBM, Random Forest, XGBoost, KNN, Logistic Regression, and CatBoost, alongside geospatial data from NASA MODIS satellite, we can significantly enhance the accuracy of wildfire prediction and spatial mapping of fire risk areas. This approach is crucial given the compounding factors contributing to increased wildfire severity, offering stakeholders the tools they need to make informed decisions and take proactive measures to mitigate the impacts of wildfires on communities and ecosystems.

Among the models evaluated, LightGBM stands out for its exceptional accuracy, while XGBoost provides a compelling alternative with its focus on error minimization. This research not only contributes to advancing wildfire management strategies but also underscores the transformative potential of data-driven approaches in addressing real-world challenges. Continued research and collaboration are paramount for further refining predictive models and ensuring resilience in the face of the escalating threat of wildfires in California.

REFERENCES

[1] Nikita Agrawal, Peder V. Nelson, and Russanne D. Low. A machine learning approach to predicting large wildfires in mediterranean areas. *Remote Sensing*, 15(23), 2023.

[2] M. Razu Ahmed, Khan Rubayet Rahaman, and Quazi K. Hassan. Remote sensing of wildland fire-induced risk assessment at the community level. *Sensors*, 18(5), 2018.

[3] Muhammad Bilal, Muhammad Arshad, Muhammad Asad Khan, Shadi R. Alharbi, Matthias Dehmer, Marco Beccuti, and Muhammad Faisal. Covid-19 classification via deep learning using ct images: a comprehensive review. *Nature Communications*, 13(1):757, 2022.

[4] Muhammad Bilal, Muhammad Arshad, Muhammad Asad Khan, Shadi R. Alharbi, Matthias Dehmer, Marco Beccuti, and Muhammad Faisal. Covid-19 classification via deep learning using ct images: a comprehensive review. *Nature Communications*, 13(1):757, 2022.

[5] Abdelmalek Bouguettaya, Hafed Zarzour, Amine Mohammed Taberkit, and Ahmed Kechida. A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. *Signal Processing*, 190:108309, 2022.

[6] Shanshan Fu, Han Yu, Zhenyu Zhang, Lei Deng, and Hui Wang. Wildfire susceptibility mapping using five boosting machine learning algorithms:the case study of the mediterranean region of turkey. *Advances in Civil Engineering*, 2022:1–14, 2022.

[7] Stella Girtsou, Alexis Apostolakis, Giorgos Giannopoulos, and Charalampos Kontoes. A machine learning methodology for next day wildfire prediction. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 8487–8490, July 2021.

[8] Ryan Laube and Howard J. Hamilton. Wildfire occurrence prediction using time series classification: A comparative study. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 4178–4182, Dec 2021.

[9] Carmine Maffei and Massimo Menenti. Predicting forest fires burned area and rate of spread from pre-fire multispectral satellite measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:263–278, 2019.

[10] Kaylee Pham, David Ward, Saulo Rubio, David Shin, Lior Zlotikman, Sergio Ramirez, Tyler Poplawski, and Xunfei Jiang. California wildfire prediction using machine learning. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 525–530, Dec 2022.

[11] Dario Spiller, Andrea Carbone, Stefania Amici, Kathiravan Thangavel, Roberto Sabatini, and Giovanni Laneve. Wildfire detection using convolutional neural networks and prisma hyperspectral imagery: A spatial-spectral analysis. *Remote Sensing*, 15(19), 2023.

[12] Ningzhe Xu, Ruggiero Lovreglio, Erica Kuligowski, Thomas Cova, Daniel Nilsson, and Xilei Zhao. Predicting daily wildfire expansion rate with machine learning. *Fire Technology*, 01 2023.