TRANSFORMATIVE AI FRAMEWORK FOR WILDFIRE MANAGEMENT

by

SUSHMA DODDALINGEGOWDA, MS, PGDDS, BE

DISSERTATION

Presented to the Swiss School of Business and Management Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

GLOBAL DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

SEPTEMBER OF GRADUATION, 2025

TRANSFORMATIVE AI FRAMEWORK FOR WILDFIRE MANAGEMENT

by

SUSHMA DODDALINGEGOWDA

APPROVED BY

Dissertation chair - Dr. Gualdino Cardoso

RECEIVED/APPROVED BY:

Admissions Director

Acknowledgements

I want to take this opportunity to express my heartfelt appreciation to everyone who played a pivotal role in my dissertation journey.

I would like to take this opportunity to express my gratitude to my mentor and co-chair, Dr. Mario Silic, for his guidance and constructive feedback, which have been pivotal in shaping this process. Additionally, I would like to thank Dr. Gualdino Miguel Cardoso, the chairperson of the dissertation defense committee, and Co-chair Dr. Sasa Petar, whose support and insightful advice have further enriched this journey.

I would also like to acknowledge the dedicated support from the SSBM management staff and the Upgrad staff, who provided the assistance I needed to navigate the challenges.

I want to extend my heartfelt thanks to my family for their unwavering presence and encouragement. Lastly, I thank God for giving me the strength and courage to navigate this journey during challenging times.

ABSTRACT

TRANSFORMATIVE AI FRAMEWORK FOR WILDFIRE MANAGEMENT

SUSHMA DODDALINGEGOWDA
2025

Dissertation Chair: Dr. Gualdino Miguel Cardoso
Co-Chair: Dr. Mario Silic, Dr. Sasa Petar

The purpose of this study is to investigate the methods for AI transformation in the Fire industry, focusing on predicting wildfire characteristics such as wildfire spread and contextual information related to the fire scene. To achieve this, the study utilizes numerical data from various satellite data sources, as this data platform enables scalability for AI-integrated real-world applications on a global scale. The research employs various unsupervised machine learning algorithms on unlabeled data. It proposes new clustering algorithms that predict wildfire characteristics, including contextual information like the threat level to the nearest residence.

This study addresses the challenges associated with integrating AI predictions into real-world applications on time. It proposes a system software architecture designed to effectively schedule GPUs by logically grouping data points from various sources, ensuring that accuracy is maintained. The application is divided into these groupings, enabling the computation of AI predictions from a regional to a national level.

The approach involves continuously acquiring real-time raw fire event data from various sources, including MODIS and VIIRS, and utilizing a deployable cloud platform. This platform is designed to constantly preprocess real-time data and compute predictions using the selected unsupervised algorithms. The study then illustrates the method for exposing the algorithm's predictions and integrating them into a visualization system within a real-world application, ensuring that the information is readily accessible and usable by the firefighting community for effective decision-making and management during wildfire incidents. This research makes a substantial contribution to the business sector by enhancing the management capabilities of firefighting organizations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I:

INTRODUCTION

1.1 Introduction

Artificial intelligence applications are rapidly expanding in various industries, including manufacturing, healthcare, finance, retail, transportation, and education. However, AI applications are less prevalent in the fire industry, especially for wildfire management.

This research focuses on the fire industry as a case study, outlining methods to develop personalized AI solutions that enhance automation and efficiency, tailored to the industry's specific needs. It employs detailed data analysis, utilizing various visualization techniques, to gain a deeper understanding of the data and address challenges in wildfire management. The research involves experimenting with AI models and customizing them to suit the unique requirements of the fire industry. Furthermore, it proposes validation methods to assess the AI models, particularly in terms of improving human efficiency. It outlines a necessary qualification process before deploying AI products in real-world applications. It also suggests a system software architecture designed to meet the accuracy and performance demands of the industry.

This thesis has several essential **business implications** by delivering real-time insights into fire growth and at-risk residences. For insurers, it improves risk assessment, policy pricing, and claims management, while firefighting organizations benefit from operational efficiency through optimized routing and resource deployment. The solution enhances public trust by reducing false alarms and boosting situational awareness, and its scalable, cloud-based design creates **SaaS opportunities** with global expansion potential in wildfire-prone regions. Additional **business implications** include data monetization through partnerships with satellite providers and insurers, as well as CSR and brand reputation gains for enterprises adopting the technology, positioning it as both a high-impact safety tool and a commercially viable innovation.

1.2 Research Problem

Major concerns with existing AI methods stem from their general-purpose focus, which often results in inadequate accuracy and performance when applied to fire industries. These AI solutions tend to be less effective in addressing the unique requirements and complexities inherent to specialized fields, leading to suboptimal outcomes in applications that demand precision and reliability.

These sectors are often highly regulated and operate within systems where failures can lead to catastrophic consequences. Consequently, these industries involve significant human oversight and intervention to ensure safety and reliability.

1.3 Purpose of Research

The purpose of this study is to develop an AI framework that has the potential to revolutionize the fire industry in the area of wildfire management. Traditionally, human experts establish predefined rules that encapsulate the knowledge of the domain related to fire. These systems generate insights by applying specific rules derived from this domain knowledge to the input data received from fire detection systems. When a particular rule condition is satisfied, the system produces output and facilitates decision-making based on those rules. However, a notable limitation of these systems is their inability to learn from new situations or adapt to changing circumstances.

For example, many fire detection systems depend solely on sensors that recognize the presence of a fire but cannot predict the fire’s specific characteristics. Understanding the details of fire scene features is often more important than just detecting the fire. By analysing data from various sensors and contextual sources, humans can gain deeper insights into the fire situation. The AI framework proposed in this study aims to close this gap by improving decision-making in the fire industry. By transitioning from a strict rule-based approach to a more adaptable and responsive method, the framework will facilitate better-informed decisions that account for the complexities of fire emergencies, ultimately enhancing safety and response effectiveness.

1.4 Significance of the Study

The study provides essential guidance on adopting AI in the fire industry in the area of predicting wildfire characteristics and contextual information related to fire scenes:

- Exploratory Data Analysis Approach: This approach outlines a systematic method for conducting exploratory data analysis on unlabeled numeric data produced from the instruments on a satellite. It involves analyzing the history of raw fire events to understand the seasonal and geographic influences on the growth of wildfires.
- Utilizing Data Insights: This study details the effective methods for leveraging insights gained from exploratory data analysis to select appropriate samples. This process is crucial for validating unsupervised machine learning models, ensuring that the data used for testing is relevant and representative.
- Machine Learning Algorithm Selection and Customization: The study focuses on selecting an appropriate unsupervised machine learning algorithm for the specific application. It emphasizes the importance of balancing accuracy with the latency required for computing predictions of growing fire events. This balance is essential for meeting the fire industry's needs to integrate these algorithms into real-world applications. Additionally, the section discusses customizing the machine learning algorithm to enhance predictions with more contextual

information related to the fire scene and the anticipated growth area in near real-time.

- Enhancing the Accuracy of Prediction and Performance: This study emphasizes the importance of improving the accuracy of predicting growing fire events. It highlights the use of multiple data sources and addresses the challenges of reducing latency, particularly when handling large data sets, by integrating various data sources for prediction.
- Real-World Application: This study proposes a framework for continuously integrating and deploying machine learning predictions into real-world products. It focuses on ensuring that the developed algorithms can be effectively applied in practical scenarios, facilitating timely and accurate responses to growing fire events.

1.5 Research Questions: Background and Motivation

Research questions are formulated by considering one of the most challenging areas of the fire industry. These questions drive extensive research in the area of applying AI methods to address the current challenges in wildfire management.

Wildfires pose a significant threat worldwide, particularly affecting the United States, Australia, Canada, Russia, Europe, South America, and Africa. The 2023 Maui wildfires were especially devastating, marking the deadliest wildfires in the U.S. in over a century. In January 2025, the Palisades Fire in California erupted, consuming 23,707 acres and resulting in five deaths while damaging over 12,000 structures and displacing more than 150,000 residents. The Eaton Fire similarly ravaged 14,021 acres, destroying around 5,000 structures and claiming six lives, with both fires together causing estimated losses of $40 billion.

Wildfires often start in remote forest areas rich in fuel, where gusty winds enable rapid spread. Without early detection and swift action, these fires can escalate into uncontrollable infernos requiring vast resources, such as water, skilled firefighters, and air support. Technologies like stationary cameras and sensors exist but are limited by high costs, making satellite monitoring a more practical solution. Unfortunately, current satellites lack the capacity for effective wildfire detection, complicating response efforts. Firefighting agencies often face resource shortages, with containment typically beginning an average of three days after fires first appear in satellite images, underscoring the urgent need for improved detection and response strategies.

1.6 Research Questions

1. Are there any hidden patterns of growing fires in the collected raw data from the history dataset captured from the satellite?

There are various types of equipment used for detecting wildfire events in forests. A significant constraint of conventional equipment, such as fire and smoke detection sensors and cameras, is that their coverage is often less than 50% of the forest area. Additionally, the installation and maintenance costs of these devices can be pretty high, as they require ongoing upkeep in challenging forest environments.

Currently, satellite data from NASA and NOAA is available in the form of geographical coordinates indicating where fires have occurred, along with the date and time, as well as a few other parameters. However, since these satellites are not explicitly designed for fire detection, over 50% of the data consists of noise. This makes it challenging to identify a growing fire on the day it is discovered, as fires are primarily visible in the data only when they expand to a larger area. In the early stages of a fire, distinguishing between noise and actual growing fire events becomes difficult. Therefore, it is crucial to analyze the raw data to uncover hidden patterns. Steps such as preprocessing the data, visualizing it with various plots, performing time series analysis, and applying statistical methods will help to understand the correlation between fire events and other influencing factors.

2. Which machine algorithm is accurate in predicting the fire growth and eliminating the non-growing fires from the raw fire events dataset on the day of discovery?

Fire industries often operate within a more rule-based system that is typically not equipped to handle large datasets effectively. To address this, implementing machine learning techniques is essential for grouping similar fire events and filtering out the noise from the raw fire events from satellite data. Unsupervised learning, a branch of machine learning, focuses on learning from unlabeled data, meaning that it identifies patterns and relationships within the data without any predefined labels or categories. Several unsupervised clustering algorithms can be utilized for this purpose, including: -

Hierarchical Clustering: This method creates a hierarchy of clusters, enabling the exploration of data at varying levels of granularity.

K-Means Clustering: This algorithm partitions the data into a specified number of clusters based on the mean distance between points, making it suitable for identifying similar groups of fire events.

Gaussian Mixture Models (GMMs): GMMs assume that the data points are generated from a mixture of several Gaussian distributions, providing a probabilistic approach to clustering.

Fuzzy C-Means Clustering: Unlike K-means, this approach allows each data point to belong to multiple clusters with varying degrees of membership, which can be beneficial in environments with overlapping characteristics.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN): This algorithm identifies clusters based on the density of data points, making it effective in distinguishing noise from meaningful clusters in the dataset. By applying these techniques, the ability to process and analyze data is ultimately enhanced, leading to more effective wildfire detection and response strategies.

The accuracy of the predictions made by unsupervised clustering algorithms is assessed using various evaluation methods that measure the goodness of the predicted clusters. One such method is inertia, which quantifies the sum of intra-cluster distances. A lower inertia value indicates better accuracy, as it means that the data points within each cluster are closer together. Another vital evaluation metric is the Dunn Index, which assesses the clustering quality based on the compactness and separation of the clusters. A higher Dunn Index signifies better clustering, suggesting that the clusters are well-defined and distinctly separated from one another. By utilizing these evaluation methods, one can effectively gauge the performance of clustering algorithms and optimize their effectiveness in analyzing fire event data.

The Silhouette Score is another valuable evaluation metric used to assess the accuracy of clustering. It measures how similar each data point is to its cluster in comparison to other clusters. The silhouette plot visually represents these scores for each sample, providing insights into the clustering quality. A high silhouette score indicates that clusters are well-separated and defined, which corresponds to better accuracy in the clustering solution. Conversely, a silhouette score close to 0 suggests that the clusters overlap significantly, making it difficult to distinguish between them. A negative silhouette score indicates poor clustering, as it implies that a data point may have been assigned to the wrong cluster. Utilizing the Silhouette Score can help refine clustering algorithms and improve their effectiveness in categorizing fire events.

The Calinski-Harabasz score evaluates clustering by measuring the ratio of the variance between clusters to the variance within clusters. It ranges from 0 to infinity, with higher scores indicating better clustering quality. On the other hand, the Davies-Bouldin index assesses the average similarity between clusters. It also ranges from 0 to infinity, but lower scores signify better clustering. In addition to these quantitative metrics, performing a visual inspection of the predicted clusters is essential for validation. By creating plots, such as 2D or 3D visualizations, one can effectively showcase all fire events within a selected region on a specific date. This visual representation helps in understanding the distribution and separation of clusters, further validating the clustering results and insights derived from the data analysis.

3. Does the accuracy of the machine algorithm in predicting the growing fires vary on real-time fire events data(unseen data)?

The degree to which a machine algorithm performs well on new, unseen data depends on factors such as the quality and diversity of the training data, the complexity of the model, and the effectiveness of model tuning and validation. It's essential to use techniques like cross-validation to get a better understanding of how the algorithm will perform in real-world scenarios.

4. Can machine learning algorithms predict more contextual information about fire scenes in areas expected to experience fire growth in near real-time, such as the threat level to nearby residences from the growing fire?

Machine learning algorithms can be utilized to predict more contextual information about fire scenes, including assessing the threat level to nearby residences from growing fires. By integrating various data sources, such as geographical information, population density, infrastructure proximity, and historical fire data, machine learning models can analyze patterns and relationships that influence the impact of fires on surrounding areas.

5. Is the accuracy of the machine learning algorithm's prediction and the latency of the machine learning algorithm in predicting the fire growth acceptable to integrate into real-world applications in the fire industry?

Machine learning predictions become valuable for integration into real-world applications when they can accurately predict most growing fire events for a given region or country. Therefore, assessing the effectiveness of machine learning models in forecasting growing fire events across various data sources is crucial. It is also essential to identify and implement necessary adjustments when deploying machine learning models for data derived from different satellite sources. One challenge in this process is the latency of machine learning models, which can vary significantly, particularly with large datasets. When predicting diverse contextual information related to the fire scene, this latency can increase further. Generally, while larger datasets tend to enhance model generalization and accuracy by providing a more comprehensive set of examples for the algorithm to learn from, managing latency is essential for making timely predictions.

6. What's the strategy to integrate the AI outputs, fire growth predictions, and additional contextual information into nationwide real-world applications in real time for the fire industry?

Developing a seamless interaction between machine learning models and existing systems is critical for effective implementation. This integration ensures that AI-generated predictions can be readily integrated into current workflows, enhancing overall efficiency and decision-making. Additionally, creating user-friendly interfaces is vital for presenting AI predictions in a manner that is easily understandable for end-users. Such interfaces should prioritize clarity and accessibility, enabling users to interpret complex data without requiring extensive technical expertise.

CHAPTER II:

REVIEW OF LITERATURE

2.1 Theoretical Framework

The literature review focuses on all the technologies under consideration within the chosen target fire domain for wildfire spread and the prediction of contextual information, aiming to understand the application of AI algorithms to complex data structures. It examines the existing literature that guides the approaches and methods for AI-driven real-world applications.

2.1.1 Early Wildfire Detection Technologies in Practice – A Review

According to Ankita et al. (2022), advanced mechanisms that are currently utilized for wildfire detection broadly fall into the following four groups:

Sensor Nodes: Low-power sensors are installed in the forest, which sense the humidity, Temperature, and gases in the near area for fire detection and alert. These sensors are charged by solar energy and are capable of communicating wirelessly. They are also called a wireless sensor network.

Challenges: Initial deployment requires manpower or an Autonomous helicopter. Sensors get damaged during the wildfire, which needs maintenance or Replacements. As wildfires are seasonal, they require maintenance every year or after every wildfire in that location, which is very expensive. It is also not feasible to install the sensors in remote forest areas.

Unmanned Aerial Vehicles (UAVs): UAVs are equipped with cameras; they are remotely operated to fly around the forest to capture Images and Video of the suspected fire scene. The following deep learning algorithms are used for identifying the fire from an image and a Video feed captured by a UAV.

1. A convolutional neural network (CNN) calculates the RGB fire score and the IR image fire score; both scores are combined to establish the presence of fire.
2. YOLOv3 and YOLOv5 deep learning for Fire detection.
3. Recurrent neural networks, long short-term memory neural networks, a generative adversarial network, and a deep belief network also detect the wildfire with reasonable accuracy.

Challenges: UAVs need human involvement throughout their operation. As UAV technology is new, operating costs are high. Flight times are usually a few days or a few hours.

Stationary Camera Networks: Stationary cameras are installed in the forest area of interest; videos are fed to the AI algorithms to detect the fire.

Challenges: Need for a continuous power source. It is Impossible to install this system in a highly remote area.

Satellite Surveillance: NASA and NOAA were two of the first organizations to observe wildfires using an extensive network of polar orbiting (Terra, Aqua) and geostationary (GOES) satellites. Polar satellites scan the entire Earth a few times each day and can monitor the whole planet for fires.

Challenges: Processing this data for wildfire anomalies, especially small fires, poses a challenge due to the lower spatial resolution of satellite images. In addition, smoke can easily appear identical to clouds. High-flying altitudes limit the resolving power of fires to a pixel in the images.

2.1.2 Data-Driven Model for Wildfire Prediction in California.

According to Brennon et al. (2024), Wildfires in California have increased in size, resulting in severe economic and environmental losses. In 2023, they resulted in nearly $1.2 billion in financial losses, while from 2021 to 2022, damages nationwide exceeded $11.2 billion.

The study analyzed a dataset of 128,125 instances with 18 features related to wildfire risk, weather conditions, vegetation, and land use. Also, it included a target parameter indicating whether a fire had occurred. They evaluated several machine learning algorithms, including Naive Bayes, Logistic Regression, SVM, KNN, Decision Trees, and Random Forest. They ultimately selected Random Forest as the final classifier based on the evaluation methods used, namely the ROC curve, confusion matrix, and precision-recall curve. They also examined the feature importance with Random Forest, and the top 5 features contributing to wildfire risk are temperature, wind speed, relative humidity, month of the year, and location. In summary, the Random Forest model demonstrates potential for wildfire prediction; however, its accuracy depends on the quality of the data and the influence of climate and human behavior.

2.1.3 Potential Wildfire Behaviour Characteristics Using Multi-Source Remotely Sensed Data: Towards Wildfire Hazard Assessment

According to Chen et al. (2023), previous research has focused on wildfire modelling, and less attention has been paid to wildfire characteristics such as wildfire speed and intensity. According to them, wildfire spread is affected by weather, fuel, topography, and human intervention. They select two supervised learning models, Random Forest and Extreme Gradient Boosting, to establish the potential wildfire characteristics based on explanatory variables. The Wildfire Dataset analyzes the wildfire characteristics, including probability, speed, and intensity, using a multi-source approach.

Table 1
Variable Importance of Driving Factors

Variables	PWBC models		
	Probability	Speed	Intensity
FMC	0.21	**0.69**	**0.75**
LAI	**0.71**	0.54	0.56
FT	0.03	0.50	0.55
TS	0.05	0.65	0.72
AP	**0.35**	0.15	0.13
RH	0.15	0.62	0.62
AT	0.28	0.50	0.53
WS	**1.00**	0.63	0.71
Elevation	0.23	**1.00**	**1.00**
Slope	0.06	0.65	0.67
Sine of aspect	0.04	0.59	0.60
Cosine of aspect	0.05	0.66	0.66
Distance to roads	0.02	0.68	0.63
Distance to residential areas	0.04	**0.70**	**0.77**
Distance to railways	0.19	0.67	0.72

Source: IEEE Xplore: Modeling Potential Wildfire Behavior Characteristics Using Multi-Source Remotely Sensed Data: Towards Wildfire Hazard Assessment (October 2023).

The study highlights the role of wind speed in predicting probabilities, as well as the relationship between elevation and speed and intensity (Refer to Table 1). They recommend Random Forest for improved management, using overall accuracy and Kappa coefficients for evaluation.

2.1.4 Wildfire Path Predictions Spread Using Machine Learning

According to Mapulane et al. (2022), Wildfires are expected to grow in severity and frequency due to climate change and outdated management practices. Effective models to predict fire spread are essential for wildfire management and disaster preparedness. This paper examines the application of Machine Learning (ML) techniques to predict the spread of wildfires from ignition points to surrounding areas using satellite images from sensors such as MODIS, Sea and Land Surface Temperature Radiometer, Visible Infrared Imaging Radiometer Day Night Band, and SLSTR. The study utilized two machine learning approaches: agent-based (A3C) and supervised learning (LRCN). The A3C model shows significant improvements in predicting fire spread at intermediate time steps, while the LRCN model enhances prediction accuracy. The LRCN model is expected to perform better overall due to its integration of both temporal and spatial properties in modeling wildfire spread.

Both models were evaluated by comparing predicted fire spread regions with a validation dataset to assess their performance and accuracy in identifying burnt and unburnt areas. LRCN shows superior spatial and temporal properties. Thus, they recommend LRCN for future studies focused on effective wildfire management to minimize damage and loss of life.

2.1.5 Satellite Image-Based Wildfire Detection and Alerting System Using Machine Learning

According to Rajalakshmi et al. (2023), Satellite images were collected from Google Images, Open-source initiatives, and Kaggle. After preprocessing the collected images, they utilized supervised learning models, including Support Vector Machine, Random

Forest, Backpropagation Neural Network, and Convolutional Neural Network for wildfire prediction.

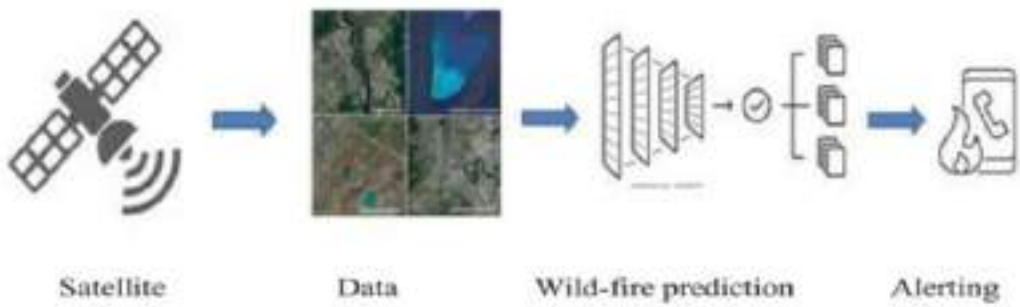


Figure 1
Wildfire prediction system

Source: IEEE Xplore: Satellite Image-Based Wildfire Detection and Alerting System Using Machine Learning (December 2023).

Table 2
Comparison of Five Models Results

Method	P	CE	Recall	OE	Acc	F-measure
Xu	0.483	0.517	0.800	0.200	0.648	0.602
SVM	0.144	0.856	0.336	0.664	0.112	0.202
RF	0.281	0.719	0.776	0.224	0.264	0.413
BP Net	0.000	1.000	0.000	1.000	0.667	—
CNN	0.986	0.267	1.000	0.000	0.974	0.965

Source: IEEE Xplore: Satellite Image-Based Wildfire Detection and Alerting System Using Machine Learning (December 2023).

Table 3
Comparison of Five Models Results Continues

Method	TP	FP	TN	FN
Xu	935	1,001	234	234
SVM	393	2,338	0	776
RF	907	2,318	20	262
BP Net	0	0	2,338	1,169
CNN	1,169	2,338	0	0

Source: IEEE Xplore: Satellite Image-Based Wildfire Detection and Alerting System Using Machine Learning (December 2023).

The model's performance was assessed using the following metrics: Precision (P), Commission Error (CE), Recall, Omission Error (OE), Accuracy, and F-measure. The results of the five models are presented in Tables 2 and 3, with the CNN achieving a 97% accuracy, surpassing the others. This study resulted in a webpage where users can input images for wildfire predictions; notifications are sent to nearby stations upon detection.

2.1.6 Applying Artificial Intelligence (AI) To Improve Fire Response Activities

According to Chang et al. (2022), Firefighting incident commanders are required to make decisions under time constraints and extreme conditions on the front line. This decision-making process necessitates the rapid collection of information regarding current resources and personnel at the fire scene. Firefighting relies heavily on teamwork, and leaders must quickly grasp environmental changes to orchestrate on-site firefighting activities effectively. AI is utilized to continuously calculate the number of firefighters and apparatus on the ground, which aids in maintaining accountability among fireground personnel through ongoing protocols. When firefighters are not present on the scene, AI notifications help commanders determine the need for additional personnel for search and rescue operations. Additionally, AI helps monitor firefighters on-site, identifying signs of fatigue and notifying other team members accordingly.

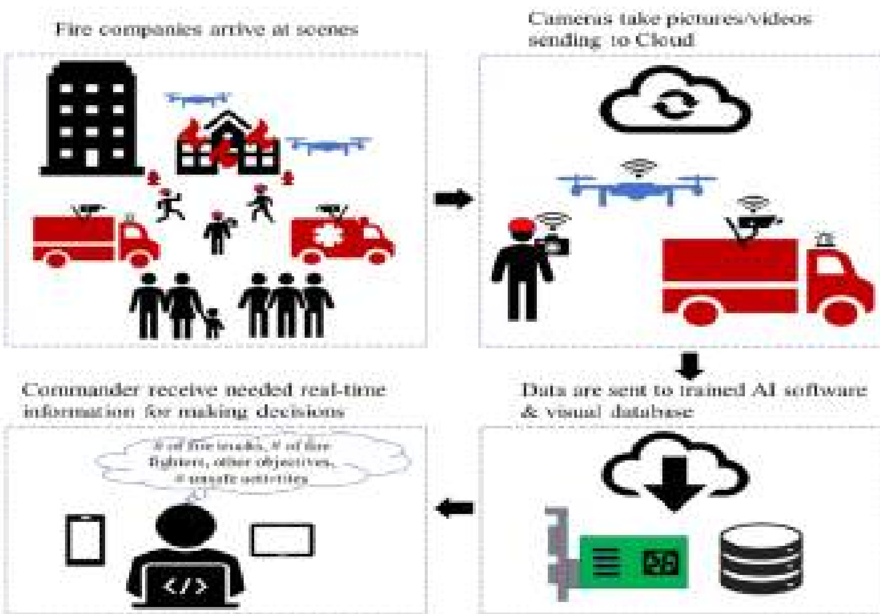


Figure 2
Brainstorm of the procedures of fire safety management practices using AI.

Source: Emergency Management Science and Technology: Applying Artificial Intelligence (AI) to improve fire response activities (January 2022).

Figure 4 illustrates that when the firefighters arrive at the scene,

- The camera begins capturing images/videos, which are installed on drones, vehicles, and firefighters
- Images/videos are then transferred wirelessly or via satellite networks to the Cloud servers at an off-site location.
- AI models use the onsite images and predict the number of fire trucks, the number of firefighters, other objectives, and the number of hazardous activities. The Onsite incident commander receives this information for decision making.

This study recommends training the AI model to recognize additional objects, such as different types of vehicles, fire hydrants on the streets, and specific uniforms worn by firefighters, Emergency Medical Technicians, and police officers. The fire commanders could quickly and precisely grasp the critical information (e.g., number of fire apparatus) on-site. Continuously monitoring firefighting activities onsite and signs of fatigue in firefighters. For instance, firefighters' helmets touching the ground is a clear sign of fatigue, and the AI software will identify the activity and immediately notify other firefighters on the ground.

2.1.7 Improved Real-Time Fire Warning System Based On Advanced Technologies for Visually Impaired People

According to Akmalbek et al. (2022), Smart fire warning systems were developed based on advanced technologies to enhance firefighting safety and protect lives. The proposed AI-based fire-detection method can be applied in various environments, including bright and safe cities, as well as for monitoring fires in urban areas to protect visually impaired individuals. This system application focuses on early fire-detection systems based on cameras and wireless technology for use in housing, as they accommodate patients with impairments and disabilities who live alone, as they are perceived to be more secure from fire-related incidents. Similarly, according to them, the proposed system can be effectively used in the fire safety industry. The system detects and notifies of catastrophic fire outbreaks in real time with high speed and accuracy. Early detection of a fire accelerates the process of eliminating it; thus, the fire poses a lesser threat to the health and lives of people, including the firefighters.

Figure 3
Fire-Detection and Notification Method

Source: MDPI: Improved Real-Time Fire Warning System Based on Advanced Technologies for Visually Impaired People (September 2022).

As illustrated in Figure 3, their proposed system comprises a client-server scheme, with the Smartphone and smart glasses serving as the client and the AI server for image analysis and fire detection.

The client section includes Smart glasses, a Smartphone, and a Home Surveillance Camera. They are used for the following purpose in this literature,

- Smartphone - The smartphone's speaker receives input from the user and sends a command to the smart glass to capture an image.
- Smart glasses - Used for capturing Images when the user gives instructions using a smartphone, this method is used to reduce the power consumption compared to continuous recording from the camera.
- Home Surveillance Camera - Records video constantly.

AI Server - AI server resides on a computer with an 8-core 3.7 GHz processor and 2 1080 GPUs. Receives client-supplied images, processes them using lightweight deep CNN models, and outputs results in Audio format. GPUs in wearable assistive devices are less capable.

According to them, Bluetooth is used for transmitting images captured from the sunglasses to the smartphone, and the transmission time is 0.11 seconds. The Image transmission from a smartphone to an AI server via Wi-Fi or cellular 5G takes 0.32 seconds. The AI server preprocesses the data, predicts the fire, and sends a notification, which takes 0.83 seconds; the total time is 1.26 seconds.

2.1.8 Leveraging The Power Of Internet Of Things And Artificial Intelligence In Forest Fire Prevention, Detection And Restoration: A Comprehensive Survey

According to Sofia et al. (2024), Forest fires pose a significant threat to the planet's ecological balance and human communities. To minimize the damage caused by forest fires and reduce the need for firefighting efforts, it is crucial to predict forest fires by modeling the relationship between fire risk and factors such as weather or fuel availability, and detect them through various monitoring techniques. In response to this growing threat, the field of forest fire prediction and detection has become a topic of ongoing research and development, supporting public policies aimed at controlling forest fires and mitigating the threat they pose.

In conclusion, the use of advanced systems incorporating artificial intelligence (AI) is a promising approach to mitigating the threat posed by forest fires. This study highlights the role of algorithms in forest fire prediction and detection systems, providing a comprehensive overview of the current state-of-the-art in the field. Using these models effectively is crucial in mitigating the adverse effects of forest fires and wildfires, safeguarding human communities, and preserving the resilience of the Earth's ecosystems.

In the context of wildfire spread, this literature suggests that the survey indicates the threat of forest fires continues to grow, resulting in a growing need for more effective and efficient methods for forest fire prevention, detection, and restoration. Fire spread behaviour prediction, including fire spread rates, growth prediction, burned area, and severity, is among the primary areas of concern in these models. The behaviour of fire encompasses a range of physical processes and features, such as combustion rate, flame height, and fuel consumption. Remote sensing data is beneficial in this regard, as it enables a more extensive observation of critical factors that are difficult to assess directly from the field, both in terms of space and time. Landsat land cover data, NOAA weather measurements, and archived MODIS sensor data from several years are employed in these models. Data mining techniques were employed to predict which fires are likely to expand, and satellite monitoring was used to determine if the data collected was sufficient for real-time tracking of Earth phenomena events, such as wildfires. Remote data collection is an effective means of obtaining extensive coverage of essential variables in both space and time, which is difficult to achieve through direct ground measurements. The models employed archived MODIS sensor data from multiple years, combined with Landsat surface cover data and NOAA weather observations,

Accurately understanding and maintaining awareness of a wildfire's dynamic state, including location, type, and features such as the rate of escalation, ignitable material, direction, topography, and weather impacts, is crucial for managing the fire in a systematic and timely manner.

Time limitations, resource management, and exactitude factors affect forest fire spread forecasting in real-time. A framework of cyber for forest fire development forecasting, which merges input data that is collected from various sources like remote meteorological sensors and satellites. To facilitate the instantaneous sharing of outcomes, the gathered data must be structured for simulation tools that utilise parallel programming paradigms and computing platforms. A two-stage prediction framework, comprising the Prediction and Calibration stages, is suggested. The Calibration stage utilizes a Genetic Algorithm (GA) to optimize the most crucial parameters of a forest fire spread model by accurately reproducing recent events through a spatial optimization objective function. The fitness function employed in the Calibration stage strives to minimise the discrepancy between the observed fire spread and the spatial fire development forecasted by FARSITE. However, because the GA is repetitive and the simulations require a significant amount of time, the Calibration stage can be time-consuming. To address this, a Time-Aware Classification (TAC) was integrated into the Calibration stage to allocate the number of cores to each individual in the population, taking into account time limitations. Despite the TAC approach being promising in ensuring that simulations are executed within the distributed time, it may become trapped in local optima in the search space. The RE-TAC approach overcomes the time constraint by using rescaled coarse-resolution data. While the TAC approximation may reject an accurate solution, the ReTAC method produces positive results when dealing with large forest fires. Compared to the TAC version, ReTAC reduces the error and achieves efficiency that is closer to the single-core scheme, where there is no time constraint. The prediction accuracy and time savings of ReTAC improve with increasing computational capacity. ReTAC utilises high-performance computing platforms to leverage parallelism at two levels, with the implementation of a single forest fire propagation forecast that is parallelised using OpenMP. The two-stage prediction plan of ReTAC has been validated and proven to be an effective fire forecasting tool for forest fire function analysts and managers.

2.2 Summary

The literature reviews existing studies to understand how machine learning models are used in predicting wildfire characteristics and what other technologies are utilized for wildfire detection, as well as their limitations, particularly in terms of fire growth and predicting more contextual information from the fire scene. And what are the different types of data used in the existing study for such prediction, and reviewing what methods are used for evaluating the machine algorithm?

While reviewing the existing technology for wildfire detection, it was found that not all the remote areas of the forest are covered by that technology. One of the technologies they use is installing smoke, flame, and gas detecting sensors in the forest, and some forests also install cameras. As they required to be installed in the forest it need lot of human effort installing the sensors, in every wild fire event in the forest they require maintenance post every fire incident in that area, thus increases the maintenance cost using the sensors and cameras in the forest, also these are hard to install on the remote areas of the forest, they also used drone or UAV’s for monitoring the forest, as this technology is new, it needs dedicated remote pilots controlling the UAVs there are still challenges with this technology in the flight time for using them for more than a day. (e.g., Ankita et al. (2022)).

While reviewing the different form of data used for wildfire spread or characteristics prediction, Numerical datasets revolves around environmental factors, such as temperature, humidity, wind speed and direction, soil moisture content, and precipitation levels, presence of vegetation, topography like Elevation, Slope, and land-use patterns, Fuel Parameters like Fuel moisture content, Leaf Area Index, forest type, and tree species, other Weather Parameters like Accumulated Precipitation, Relative humidity, Air Temperature, Infrastructure data like Distance to roads, Distance to residential areas, Distance to railways. Historical wildfire location, dates, and rate of speed from the Global Fire Atlas for the years 2003 to 2016, fire intensity from the MODIS satellite dataset. While preparing these datasets for the model, each instance in the dataset is labelled as either a fire or a non-fire. (e.g., Brennon, 2024; Chen et al., 2023).

Some used image data from the instruments captured by the satellite, including MODIS, Sea and Land Surface Temperature Radiometer, Visible Infrared Imaging Radiometer Day Night Band, and SLSTR, as well as open-source initiatives such as Google Images and Kaggle, which were also collected from the satellite. The dataset is prepared by labeling images as either wildfire or no wildfire (e.g., Mapulane, 2022; Rajalakshmi et al., 2023).

For the numerical dataset captures all the exploratory variable only specific to a particular region or state, machine learning model trained on this dataset which has only specific region data cannot be used for predicting the wild fire for other state or region, this, the approach of collecting region particular data, retraining machine learning model for every new incident event of the same region as the climate changes are not static, might increase the cost of maintaining the accurate dataset and that are region specific. Such solutions are not easily scalable for real-world applications in the global platform (e.g., Brennon et al. (2024)).

Machine learning models, primarily used on numeric datasets, a supervised learning model - Random Forest, show promising results in predicting wildfire characteristics. Evaluation metrics used to evaluate this supervised learning model include the ROC curve, the confusion matrix, the precision-recall curve, and the Kappa coefficient. While using this model, it was found that Wind Speed, Leaf Area Index, Accumulated Precipitation, Elevation, Fuel moisture content, Distance to resident, month of the year, and location are the most influential factors for wildfire spread. (e.g., Chen, 2023; Brennon et al., 2024).

In contrast, the supervised machine learning model used on image-based datasets, specifically Convolutional Neural Networks and LRCN, which combines CNN and RNN, also shows promising results in predicting wildfire characteristics.

Evaluation metrics used are Precision, Commission error, Recall rate, Omission error, Accuracy, and F-measure for CNN, as well as the predicted fire spread regions, which were compared with the validation dataset to determine the overall performance and accuracy of the LRCN model (e.g., Rajalakshmi, 2023; Mapulane et al., 2022).

Many of these studies propose integrating machine learning models to predict wildfire characteristics into real-world applications for future work, but none of these studies have demonstrated methods for incorporating this into real-world applications (e.g., Brennon, 2024; Chen, 2023; Rajalakshmi, 2023; Mapulane et al., 2022).

Additional literature was reviewed to understand the real-world applications of AI prediction in the fire industry. The reviews highlight products designed for real-world applications that utilize AI technology, thereby reducing the threat levels to human health and life, including for firefighters and individuals with visual impairments. (e.g., Sofia, 2024; Akmalbek, 2022; Ray et al., 2022).

Artificial intelligence predictions are applied in real-world products that enhance fire response activities. Firefighter leaders can quickly grasp multiple fire scenes remotely based on AI insights, allowing them to manage resource deployment effectively from one scene to another according to the dynamics of each situation. This approach significantly increases the efficiency of firefighting organizations (e.g., Chang et al., 2022). An AI-based notification system can detect fatigued firefighters at the scene of a fire and alert other firefighters for assistance. This ensures that support is readily available, enabling safer and more effective operations during challenging situations.

Leveraging the Internet of Things (IoT) and Artificial Intelligence for forest fire prevention can significantly reduce the need for firefighting efforts. This approach helps protect human communities and maintain the resilience of Earth's ecosystems by facilitating early detection and proactive measures against potential fires. (e.g., Sofia et al., 2024).

An Artificial Intelligence-based real-time fire warning system detects fires at an early stage, making it suitable for deployment in innovative and safe cities. This system monitors urban areas and supports individuals with impairments and disabilities living alone. By providing real-time notifications of catastrophic fire outbreaks with high speed and accuracy, early detection accelerates the response process, thereby reducing the impact of fire on the health and lives of both residents and firefighters. (e.g., Akmalbek et al., 2022).

According to individuals with impairments and disabilities living alone, an AI-based real-time fire notification system can be effectively utilized in the fire safety industry. This technology offers vital support, ensuring that those who may have difficulty responding to emergencies receive timely alerts and assistance.

CHAPTER III:

METHODOLOGY

3.1 Overview of the Research Problem

Traditionally, software products handle more structured data that is well-organized, primarily used for analysis and reporting. These software applications generally feature dashboards and visualizations. However, humans are responsible for the analysis and decision-making processes. The future of the software business is moving beyond traditional applications; it's about AI-based intelligent applications that augment human capabilities to derive actionable insights.

3.2 Operationalization of Theoretical Constructs

This research focuses on the fire industry as an example, employing a quantitative method to demonstrate the steps and processes involved in transforming a traditional software product into an intelligent software product through the use of AI technology.

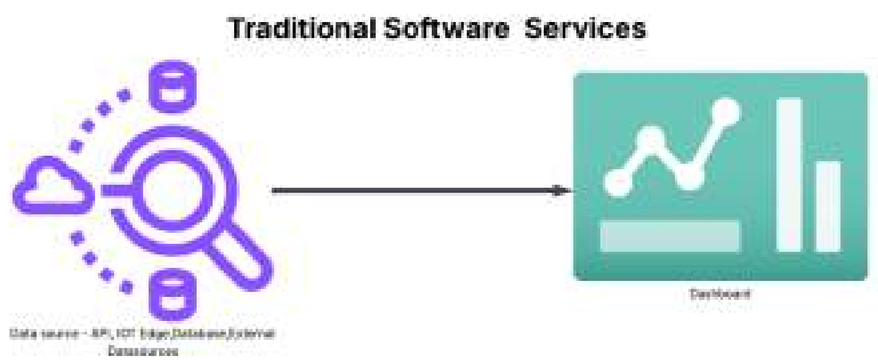


Figure 4
Traditional Software Products (own work)

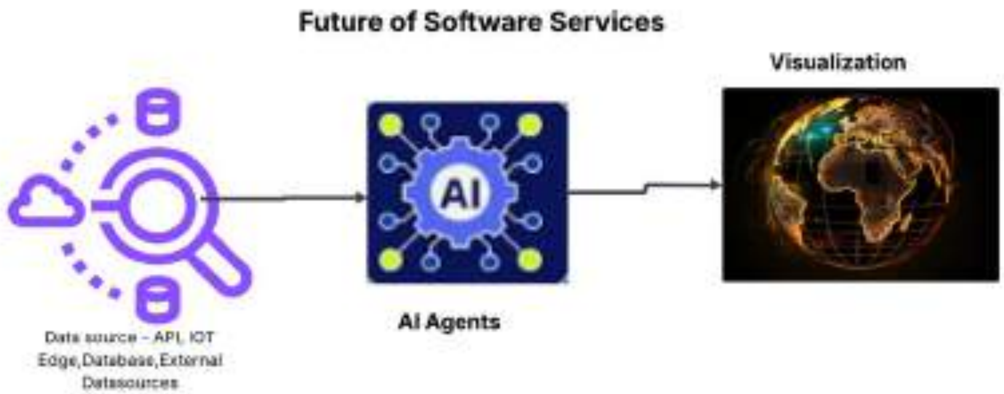


Figure 5
Future of Software Products

Source: Created using Google Images.

This research specifically focuses on developing an AI-based product for the fire industry, which traditionally employs rule-based systems. It aims to analyze data and derive meaningful insights for informed decision-making, predictive modeling, early warning systems, and enhanced situational awareness.

3.3 Research Design

Generally, research can be conducted using both quantitative and qualitative methods. For this research, quantitative methods have been adopted because the study focuses on demonstrating the techniques required for AI transformation within the fire industry. By actually building an AI-based product. These methods will allow for measurable outcomes and a precise evaluation of the product's effectiveness in meeting industry needs.

Quantitative methods will be employed as follows:

- The hypothesis is to predict fire growth in near real-time using the AI model.
- The data collection step involves gathering historical fire data, along with real-time or near-real-time data from various data sources. This comprehensive approach ensures that the model is trained on a robust dataset, enhancing its predictive capabilities and accuracy in identifying potential fire threats.
- AI platform, Framework, and Development.

  AI platform

  This study utilizes the Google Colab cloud-based platform, which is extensively used for data science code development. This choice eliminates the need for a local setup and provides access to GPU and TPU runtimes, which are crucial for accelerating computationally intensive machine learning algorithms. Additionally, Colab integrates seamlessly with other platforms like GitHub, facilitating collaborative work and version control. While other platforms like Microsoft Azure were explored, they are more suitable for general AI applications such as image/video processing and text-based tasks like chatbots. However, Azure does not offer the same level of control for building AI models tailored to domain-specific use cases. Moreover, it can be expensive to use for research-based real-world application projects. This makes Google Colab a more practical and cost-effective choice for the objectives of this study.

  AI Frameworks:

  The Following are the major AI libraries used for experimentation:

    - Pandas is used for processing the data in the CSV format. The CSV format was chosen for processing the data from the data sources.

- matplotlib - This is used for exploratory data analysis.
- Scikit-Learn libraries - This is used for experimenting with traditional clustering machine learning models for handling unstructured forms of data.
- PyGithub – This library enables a mechanism to access the stored dataset in GitHub and store the predicted dataset for the chosen Google Colab platform.

Development and Code Management using Python: Python is the most widely used programming language for AI, supporting all major AI libraries, and it is also suitable for real-world applications.

- The data analysis step includes performing exploratory analysis on the history dataset to understand the population and choose the sample.
- From Exploratory data analysis, identify if there are any hidden patterns in the data.
- Analyse and derive the dataset to include the parameters that potentially contribute to detecting the hidden patterns
- Exercise the various machine learning algorithms that can predict the fire anomalies in the various sample datasets.
- Evaluate machine learning algorithms using the right metric to meet the accuracy and performance.
- Choose the best-performing machine learning model.

Statistical Detailed Design

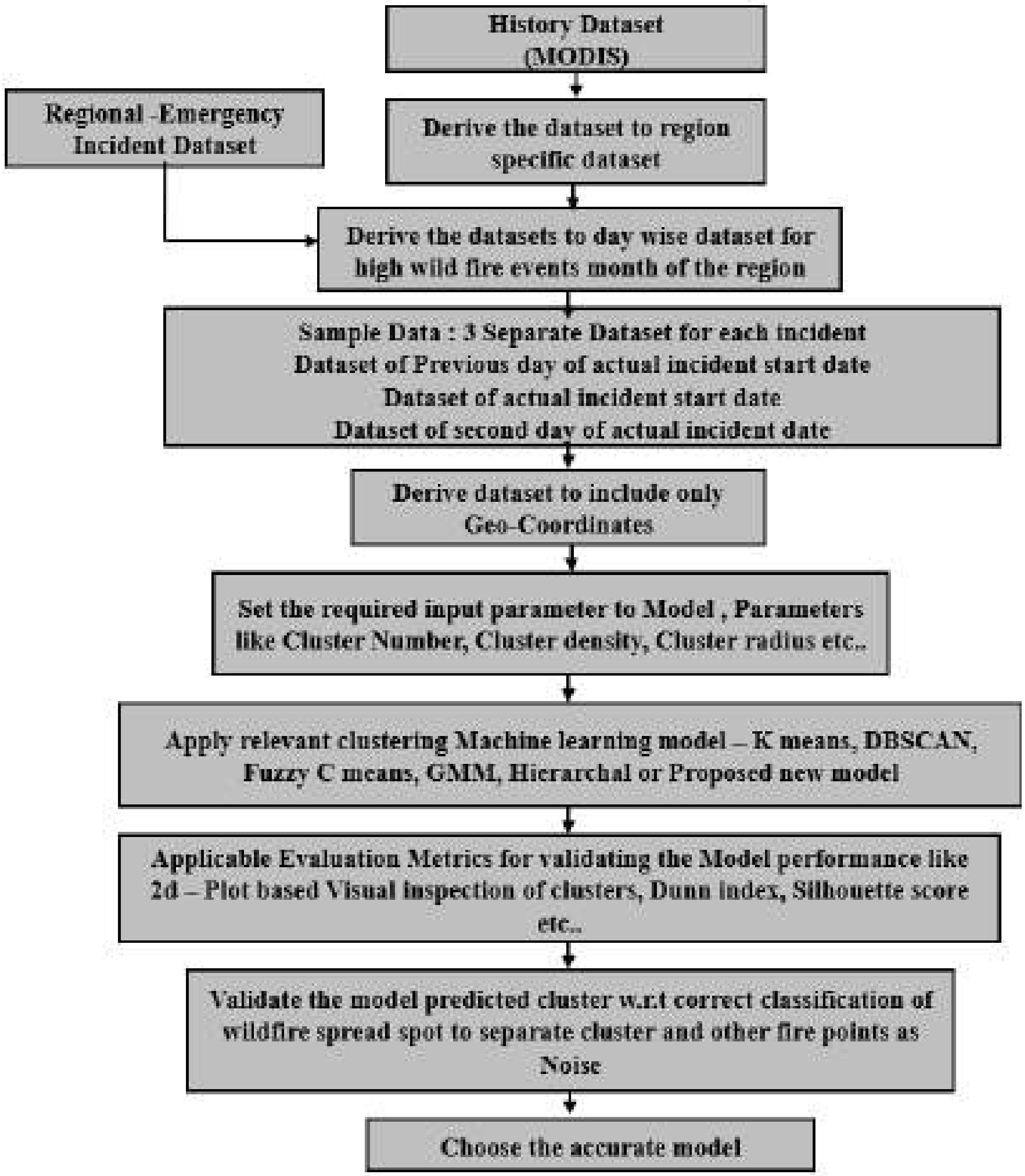


Figure 6
Statistical Detailed Design – For AI Model Selection & Evaluation.

Figure 6 illustrates the statistical design used to select the most accurate model for predicting the spread of wildfires.

For the initial analysis, concentrate on the historical dataset from MODIS. There's no need to include datasets from multiple instruments in this phase. This focused approach will help ensure clarity and precision in analyzing wildfire events.

Extract the dataset specifically for the region identified through exploratory data analysis (EDA) as having the highest number of wildfire events over the past years. This targeted approach will enable a more precise investigation of wildfire patterns and trends in that particular area. Refine the dataset to create a day-wise dataset specifically for the months that experienced the most significant wildfire events in the selected region. This will involve isolating data from those months and breaking it down to the daily level, allowing for a more detailed analysis of wildfire occurrences. Construct a sample dataset that includes a 3-day time frame centered around the actual dates of wildfire incidents. This dataset should consist of: - One day before the day of the wildfire occurrence, - The day of the wildfire occurrence itself, - One day following the day of the occurrence.

Clean the dataset by removing all parameters except for the geo-coordinates. This will result in a streamlined dataset that focuses solely on the spatial data relevant to the analysis of wildfire incidents. By retaining only the geo-coordinates, you can better assess the spatial distribution of wildfire occurrences. Define the key input parameters for the clustering models as follows:

- Number of Clusters: Expected number of clusters based on prior knowledge or methods like the elbow method.
- Minimum Cluster Density: Minimum number of points required to form a dense region.
- Cluster Radius: Maximum distance between points to be in the same cluster.

- Distance Metric: The method for measuring similarity (e.g., Euclidean, Manhattan).
- Max Iterations: Maximum iterations for algorithms that refine clusters.
- Random Seed: For reproducibility in random processes.
- Outlier Detection Threshold: To identify and exclude outliers.

These parameters will guide the clustering process, enabling effective analysis of wildfire patterns. Apply all relevant machine learning clustering models to the selected dataset. This may include: -

- K-means Clustering: For partitioning data into distinct clusters based on distance.
- DBSCAN: To identify clusters of varying shapes and sizes while filtering out noise.
- Hierarchical Clustering: To create a tree of clusters that can be cut at different levels.
- Gaussian Mixture Models (GMM): For probabilistic clustering based on the assumption that data points are generated from a mixture of several Gaussian distributions.
- By leveraging these models, conduct a thorough analysis of the wildfire data to uncover meaningful patterns and insights related to wildfire spread and behavior.
- Evaluate the accuracy of the clustering models using a combination of methods, including:
- 2D Visual Inspection: Create scatter plots to visually assess the clustering results.

- Dunn Index: Measures the ratio of the smallest distance between points in different clusters to the greatest intra-cluster distance.
- Silhouette Score: Evaluates how similar an object is to its cluster compared to other clusters.
- Calinski-Harabasz Index: Assesses cluster validity based on the ratio of the sum of between-cluster dispersion to within-cluster dispersion.
- Davies-Bouldin Index: Evaluates the average similarity ratio of each cluster with its most similar cluster.

Using these metrics will provide a comprehensive assessment of the clustering performance of wildfire data analysis. Select the model that effectively clusters the wildfire spread points while categorizing all other data as noise. This model should demonstrate strong performance based on the evaluation metrics used, ensuring that it accurately identifies significant clusters of wildfire incidents and segregates irrelevant data. Choosing the right model will enhance the understanding of wildfire patterns and inform better management strategies.

System Design – AI-Enabled Real-World Application

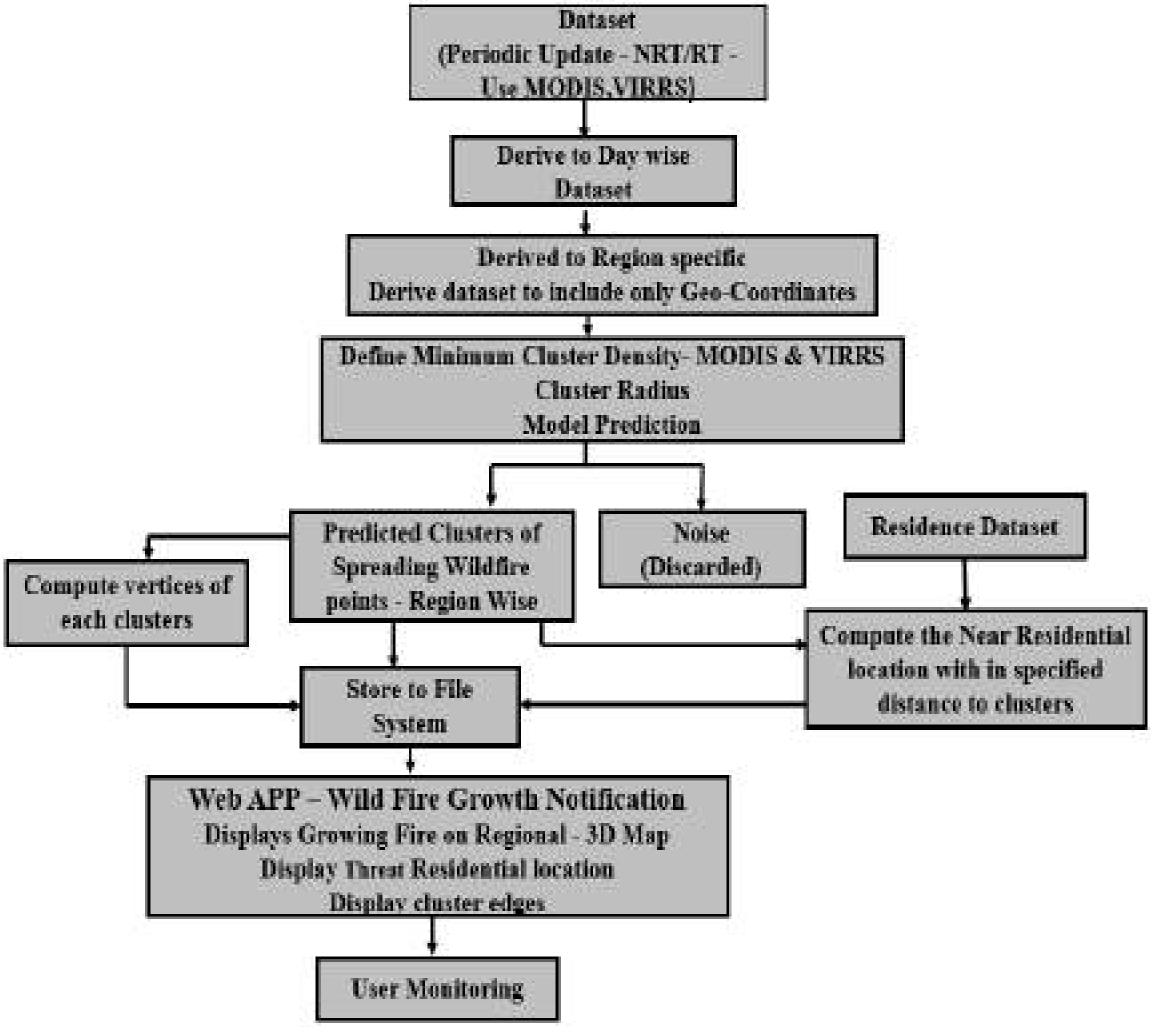


Figure 7
System Design - AI-enabled Real-World Application

Figure 7 depicts the system design for building an AI-enabled real-world application. The architecture illustrates how different components interact to process and analyze data, facilitating intelligent decision-making. Key elements include data ingestion, processing pipelines, machine learning models, and user interface layers, all working together to deliver insights and functionality to end-users.

The system process for handling real-time or near-real-time datasets from MODIS and VIIRS instruments involves the following steps:

- Data Acquisition: Collect the last 24 hours of datasets from MODIS and VIIRS, updating them daily.
- Region-Specific Dataset: Derive datasets specific to regions, maintaining separate collections for MODIS and VIIRS.
- Parameter Optimization: Eliminate all parameters in the dataset, retaining only the geo-coordinates to focus on location-based analysis.
- Clustering Parameters: Define minimum cluster density and cluster radius, noting that cluster density may vary based on the dataset source (MODIS or VIIRS).
- Clustering Model Application: Apply a suitable machine learning clustering model that accurately predicts clusters of wildfire spread points and noise.
- Cluster Analysis: Compute vertices for each identified cluster and determine the nearest residential locations using a dataset of residential geo-coordinates.
- Integration: Integrate all region-specific predicted clusters into a comprehensive national overview.
- Data Storage: Store the predicted clusters, their vertices, and the nearby residence locations in a shared file system for easy access and analysis.
- Web Application Development: Develop a web application that displays a 3D map visualizing wildfire spread clusters, nearby residences at risk, and cluster border points.

- Deployment Platform: Utilize Google Colab Pro and other platforms for deployment.
- Development Tools: Build the web app using HTML, CSS, JavaScript, and Google Map services to ensure interactive and user-friendly visualization.

3.4 Research Design Limitations

The following are some of the key limitations:

- Model performance cannot be assured when working with large datasets.
- The specific cloud environment constrains the effectiveness of the selected algorithm; different cloud platforms or edge devices may produce slightly varying results.
- Tuning parameters for the models differ when utilizing data points from various sources, due to the varying resolutions of the data collection instruments.

3.5 Data Collection Procedures

For performing exploratory data analysis and predicting wildfire growth, the following instruments' raw fire events data are utilized.

- Moderate Resolution Imaging Spectroradiometer (MODIS) is installed on the Aqua and Terra satellites.
- Visible Infrared Imaging Radiometer Suite (VIIRS) is installed on the satellites Suomi NPP, NOAA-20, and NOAA-21.

Background of data capturing instruments:

MODIS instrument views the entire Earth's surface in 1 to 2 days, acquires data in 36 spectral bands with Wavelengths from 0.4 µm to 14.4 µm, and data is available in the following spatial resolutions,

- 250 m (band 1 to 2) and 500m (band 3 to 7) -The primary use of the data in this band is for viewing Land/Cloud/Aerosols Boundaries.
- 1000m (band 8 to 36) - The primary use of data from this band is for viewing the Ocean Color / Phytoplankton/Biogeochemistry. Atmospheric, Surface/Cloud Temperature, Water Vapor, Atmospheric Temperature, Cirrus Clouds, Water Vapor, Cloud Properties, Ozone, Surface/Cloud Temperature, Cloud Top Altitude.

Visible Infrared Imaging Radiometer Suite (VIIRS), installed on the satellite Suomi NPP, NOAA-20. This instrument detects fire events at different densities compared to MODIS. This instrument is primarily used for global Earth observations, imagery, and Radiometric measurements of land, atmosphere, cryosphere, and oceans. Data from both MODIS and VIIRS instruments are utilized, which ensures a greater coverage of the areas for predicting fire anomalies.

3.5.1 Data obtaining steps - Satellite data - Raw Fire events from MODIS & VIRRS.

Data is obtained from NASA's Fire Information for Resource Management System (FIRMS). Historic and Real/near-real-time data are obtained from the following webpage.

Historic Data Main Web Page Title: NASA FIRMS Fire Information for Resource Management System, select DOWNLOAD ARCHIVED DATA, follow the sequence to get to the final webpage for getting the historical data.

- Webpage Subtitle - Archive Download, Authentication through Email.
- Webpage Subtitle - Download Requests email ID - Select Create request.
- Webpage Subtitle - Download Request.

Enter the following field, from the webpage with the subtitle 'Download Request'

- Select 'Country' from the drop-down

- Select Fire Source ‘MODIS’ from the drop-down
- Select Date from Jan 2020 to the current date of 2025
- Select Comma-Separated - Text (.csv)

The dataset from the requested instrument, country, selected duration, and requested format will be sent to the email address provided.

Real-time data Main Web Page Title: NASA FIRMS Fire Information for Resource Management System, select ‘Web services’, follow the sequence to access the final webpage, where you can obtain the URL for real-time data.

Webpage Subtitle - Web Services,

- Select the link ‘API - Application Programming Interface’.
- Webpage Subtitle – API, Choose the link ‘country’
- Webpage Subtitle - API / country

Enter the following details.

- Select Country – United States (USA) from the drop-down
- Select the following Source one at a time from the drop-down
  a) MODIS (URT +NRT)
  b) VIIRS NOAA-20 (URT +NRT)
  c) VIIRS NOAA-21 (URT +NRT)
  d) VIIRS S-NPP (URT +NRT)
- Enter - MAP KEY
- Select Day Range – 1 from the drop-down
- Select the button ‘Display Results’
- Copy the link and use it in the code.
- Change to all different sources as specified above, and copy the individual links and put them in your code.

3.5.2 Data obtaining steps - Emergency Incidents: Download the actual wildfire emergency incidents from the **Current Emergency Incidents Web Page**. This data is captured from the state government of California and consists of actual wildfire emergencies in the state for the past 13 years.

- Go to the subtitle – ‘Incident Data’ on this webpage
- Select ‘ALL DATA AS CSV’ to download the actual emergency incident.

3.5.3 Data obtaining steps - Residential data of the USA: For predicting the contextual information of fire scenes, residential data from the US Census Bureau and the neighbourhood database are used. This data is captured as follows:

- Download the residential data from the United States Census Bureau. From the webpage with the title ‘Gazetteer Files’, select 2024
    a) Go to the Webpage subtitle ‘Census Tracts’
    b) Select the state from ‘Download a Single State Census Tracts Gazetteer File ‘from the dropdown.
    c) Select the link “Download the National Census Tracts Gazetteer Files[2.3MB]’.
- Download residential data from the United States Neighbourhoods database available from the website with the webpage title ‘simple maps, Interactive maps and data’.

3.5.4 Parameter and Description - Satellite Dataset (MODIS/VIRRS)

Table 4 provides a comprehensive overview of all the parameters included in the MODIS/VIIRS datasets used for exploratory data analysis of the raw fire event and wildfire growth predictions after preprocessing.

Table 4
Satellite Dataset Parameters

Parameter	Description	Type and Range
latitude	Geo Location of fire spots	Latitude ranges from 18°N to 72°N for the USA
longitude	Geo Location of fire spots	Longitude ranges from 67°W to 179°W.
brightness	Brightness temperature data of the fire spots	300-510
scan	Represents the along-scan pixel size, which is the spatial resolution in the East-West direction of the scan.	1 - 5
track	Represents the along-track pixel size, which is the spatial resolution in the North-South direction of the scan.	1-2
acq_date	The date on which this fire was active	01-01-2020 to 03-31-2025
acq_time	Time at which the fire was active	00 to 2359
satellite	Terra - Terra satellite Aqua – Aqua satellite N20 - NOAA-20 satellite N21 - NOAA-21 satellite	NA

	N - Suomi NPP satellite	
instrument	Refers to the instrument MODIS or VIIRS, which captured the fire spot.	NA
confidence	This access quality of fire pixels, assigning confidence levels to gauge the reliability of detected hotspots	0-100
frp	Fire Radiative Power refers to the rate of outgoing thermal radiative energy emitted from a burning spot	0-16146.4
bright_t31	It refers to the brightness temperature of the fire pixel, measured in Kelvin, specifically from channel 31 of the MODIS instrument; this refers to the intensity of the fire.	Ranges from 264 to 401 Kelvin, Data type is float32,
daynight	Indicates the Day or night fire spot detected.	D, N
Type	It refers to different land cover classification schemes, including IGBP (International Geosphere-Biosphere	1 to 3,

	Programme), LAI/fPAR, and NPP (Net Primary Production), which are used to categorize and map various land surface features.	

3.5.5 Parameter And Description - Emergency Incident Dataset

This dataset has been meticulously collected by the regional incident management team in California and has received approval from local government authorities. It serves a critical purpose: validating the predictions of machine learning models regarding the emergence of fire locations by directly comparing predicted fire growth with actual wildfire occurrences in this dataset.

Table 5
Emergency Incident Dataset

Parameter	**Description**	**Type and Range**
incident_date_created	This is the date and time when the wildfire incident was reported	Date 2013 to Date 2025
incident_acres_burned	This indicates the acres burned in the wildfire incident	Max 1032648.0 Min 0
incident_latitude	Geo Location of fire spots	
incident_longitude	Geo Location of fire spots	
incident_type	This indicates the Type of emergency	Wildfire, Fire, flood, earthquake.

3.5.6 Parameter And Descriptions - Census and Neighbourhood Dataset

The United States Census Bureau - U.S. Gazetteer Files dataset, urban areas, and The U.S. Neighbourhoods website dataset comprised most of the residential addresses and Geographical data coordinates of the USA.

3.6 Data Analysis

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is performed on the independent parameters to gather insights into fire events within the MODIS dataset, focusing on their relationship with the Date (Season) and geographic location. The analysis examines explicitly:

- Latitude
- Longitude
- Date

General Analysis

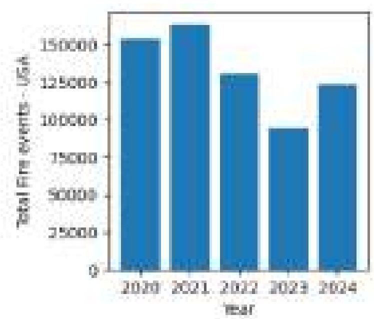


Figure 8
Fire events from 2020 to 2024, USA - MODIS Raw Fire events

Plot Interpretation: In Figure 8, the y-axis indicates the total number of fire events, while the x-axis represents the Year.

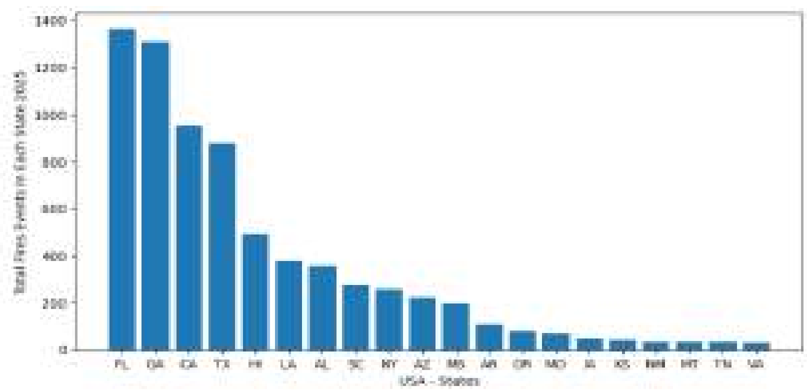
Plot Inference:

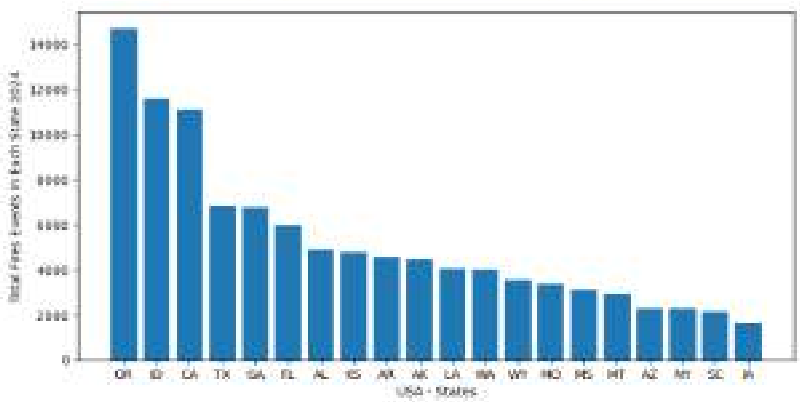
General Insights on fire events across the USA in the past 5 years:

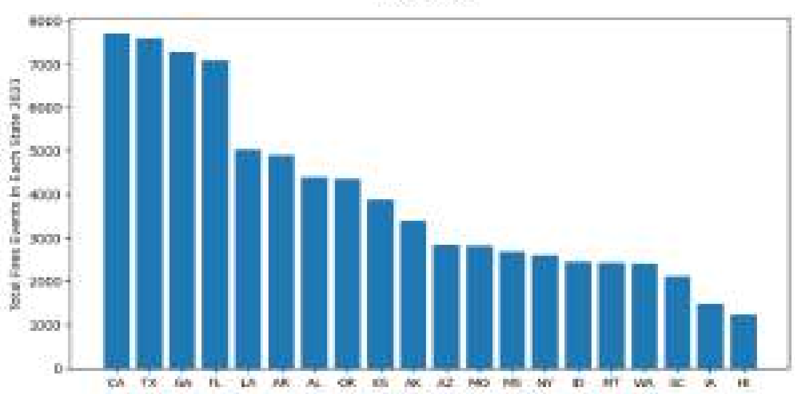
- In the past 5 years, 2021 recorded the highest number of fire events in the USA.

- Conversely, 2023 saw the lowest number of fire events.
- An overall trend from 2021 to 2023 indicates a decline in fire events.
- However, projections for 2024 suggest a resurgence, with an expected increase in fire events compared to 2023.

Total fire events across various geographic locations in the USA:







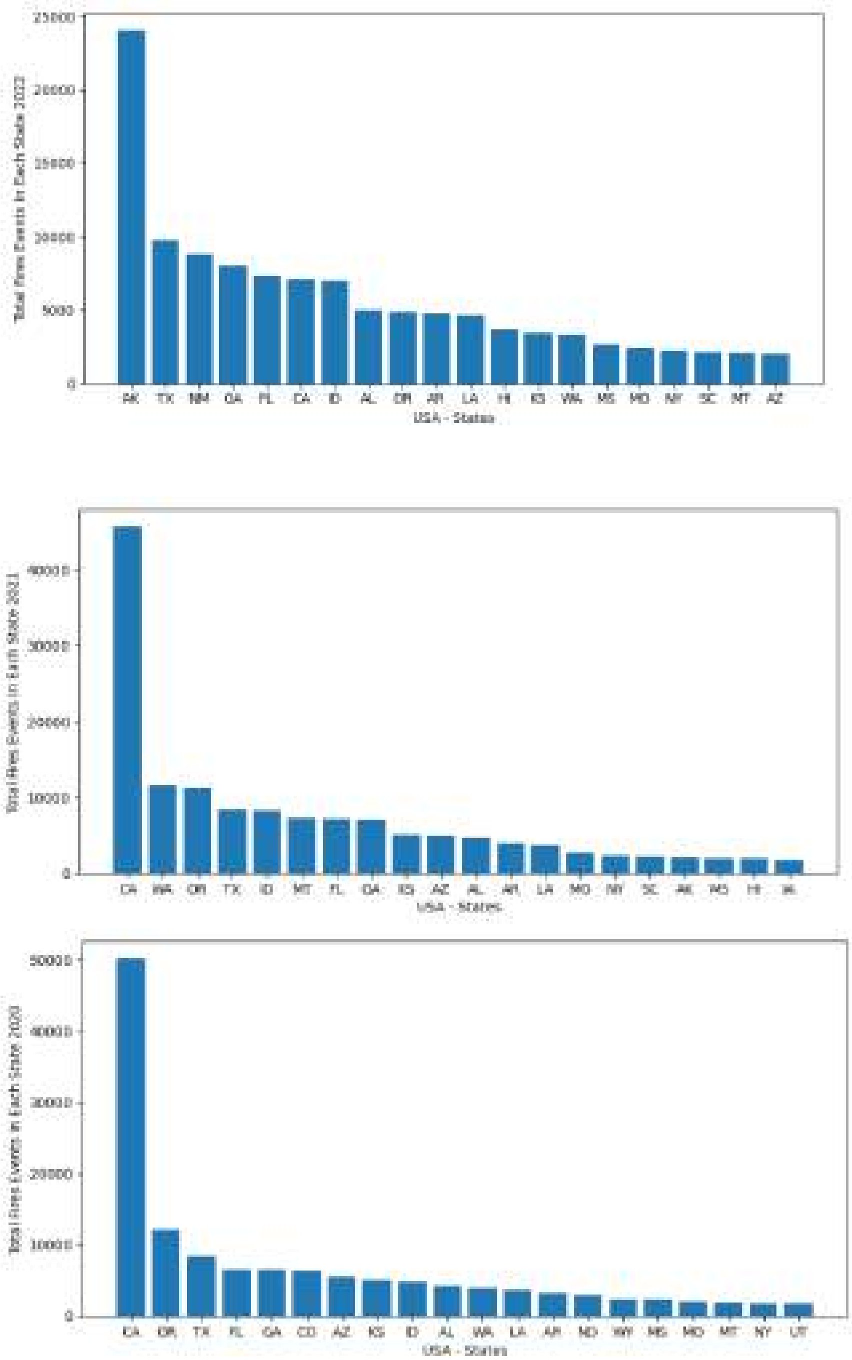


Figure 9
USA State-wise Raw Fire Events MODIS - 2020 to 2025

Plot Interpretation: In Figure 10, the Y-axis represents the total number of fire events across different states in the USA, while the X-axis lists the states themselves.

Plot Inference: Insights into fire events in various geographic locations of the USA over the past five years include: -

- 2020: California, Oregon, Texas, Florida, Georgia.
- 2021: California, Washington, Oregon, Texas, Idaho.
- 2022: Alaska, Texas, Georgia, Florida, New Mexico.
- 2023: California, Texas, Georgia, Florida, Louisiana.
- 2024: Oregon, Idaho, California, Texas, Georgia.
- 2025: Florida, Georgia, California, Texas, Hawaii.

From 2020 to 2025, the states listed above reported the highest number of fire events in the USA. Further exploratory analysis will be conducted on these regions to examine the behavior of fire events in each geographic location and date (season).

3.6.1 Fire Event In California

3.6.1.1 Seasonal Influence on Fire Events

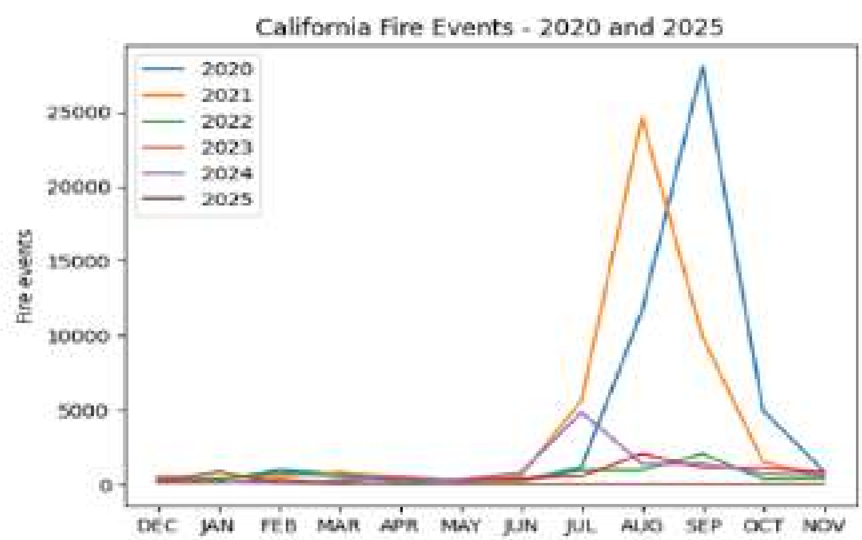


Figure 10
California Fire events - 2020 to 2025

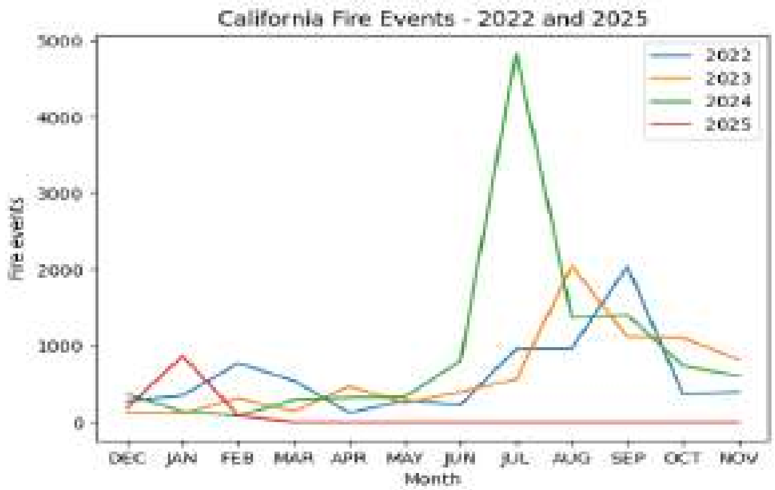


Figure 11
California Fire events - 2022 to 2025

Plot Interpretation:

Figure 11: This figure presents multiple subplots, illustrating the total number of fire events on the Y-axis and the months on the X-axis for the years 2020 to 2025.

Figure 12: This figure also comprises multiple subplots, displaying the total number of fire events on the Y-axis and the months on the X-axis for the years 2022 to 2025.

Plot Inference:

Insights of Fire events in California:

- In the last 5 years (2020-2025), fire events peak from June to October.
- Over the last 3 years (2022-2024), fire events have been lower in comparison to the years 2021 and 2022.
- Notably, both 2022 and 2025 exhibit a slight spike in fire events during the first quarter of the year.

3.6.1.2 Geographic location influence on Fire events

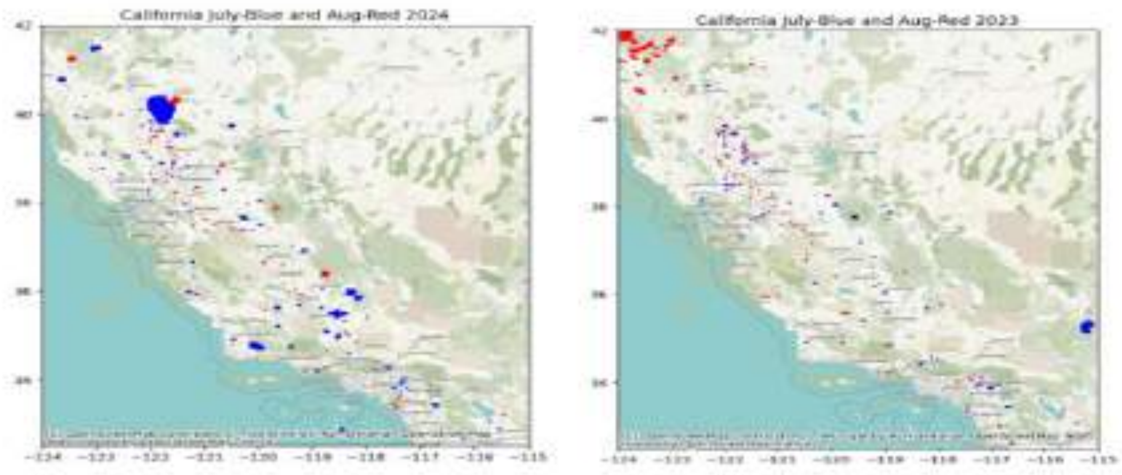


Figure 12
California Map - Fire events Jul Aug 2023 and Jul Aug 2024 Map

Plot Interpretation:

Figure 13 is represented as follows,

- Fire events for July 2023 and July 2024 are depicted on the California map in blue.
- Fire events for August 2023 and August 2024 are shown in red on the California map.
- July and August are highlighted explicitly due to being the peak wildfire months in California.

The years 2023 and 2024 are chosen to represent wildfire events, as they are the most recent data available. The subsequent plots in this EDA section will follow a similar pattern.

Plot Inference:

- Fire events in this location demonstrate the following behaviors:
- Spotting: Fire events tend to occur in isolated locations.
- Spreading: Fire events show a tendency to expand across the area.

3.5.2 Fire Event in Washington

3.5.2.1 Seasonal Influence on Fire Events

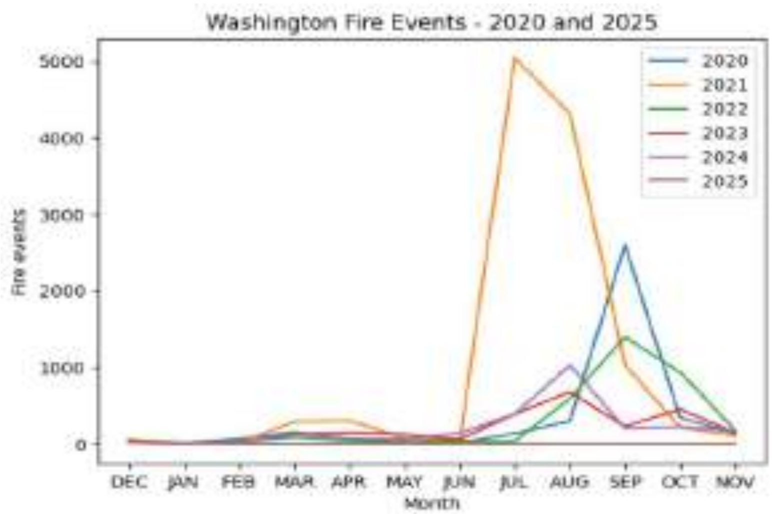


Figure 13
Washington fire events Monthly - 2020 to 2025

Plot Interpretation:

Figure 14: This figure represents multiple subplots. It indicates the total fire events in the Y axis and Months in the X-axis for the years 2020 to 2025.

Plot Inference:

Insights into fire events in Washington:

- In the last 5 years (2020-2025), fire events peak between June and October.
- During this period, the year 2021 recorded the highest number of fire events.

3.5.2.2 Geographic location influence on Fire events

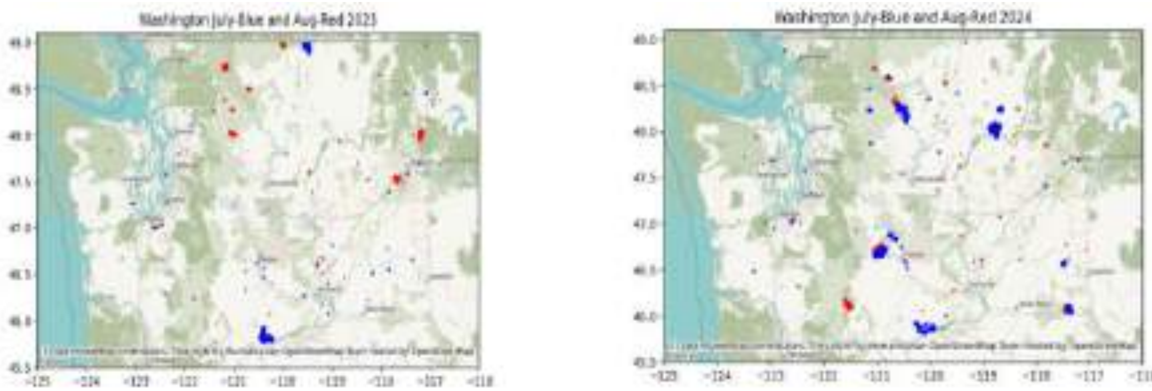


Figure 14
Washington Map - Fire events Jul Aug 2023 and Jul Aug 2024

Plot Interpretation:

Figure 15 is represented as follows,

- Fire events for July 2023 and July 2024 are depicted on the Washington map in blue.
- Fire events for August 2023 and August 2024 are shown in red on the Washington map.
- July and August are selected for this representation, as they are the peak wildfire months in Washington.

Plot Inference:

Fire events in this location demonstrate the following behaviors:

- Fire events tend to occur in isolated spots.
- Fire events tend to spread.
- In 2024, approximately 12 fire events in this location expanded into larger areas.
- In 2023, roughly seven fire events in this location spread to wider areas.

3.5.3 Fire Events in Idaho

3.5.3.1 Seasonal Influence on Fire Events

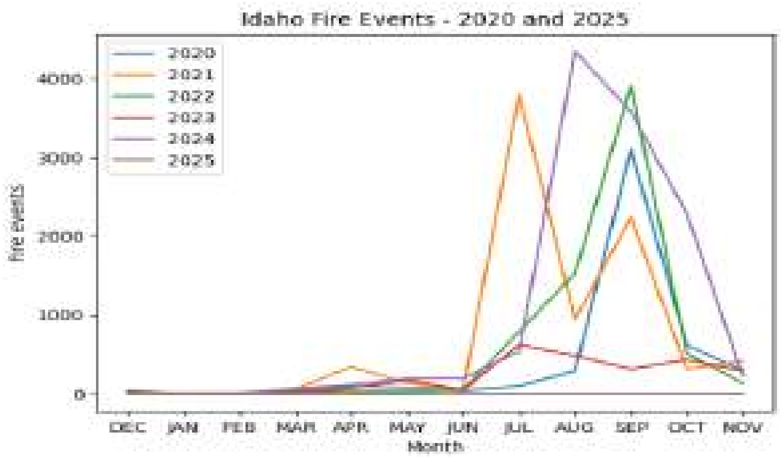


Figure 15
Idaho fire events Monthly - 2020 to 2025

Plot Interpretation:

Figure 16: This figure represents multiple subplots. It indicates the total fire events in the Y-axis and Months in the X-axis for the years 2020 to 2025.

Plot Inference:

Insights into fire events in California:

- In the past 5 years (2020-2025), fire events peaked between June and October.
- Within this timeframe, the year 2024 recorded the highest number of fire events.

3.5.3.2 Geographic location Influence on Fire events

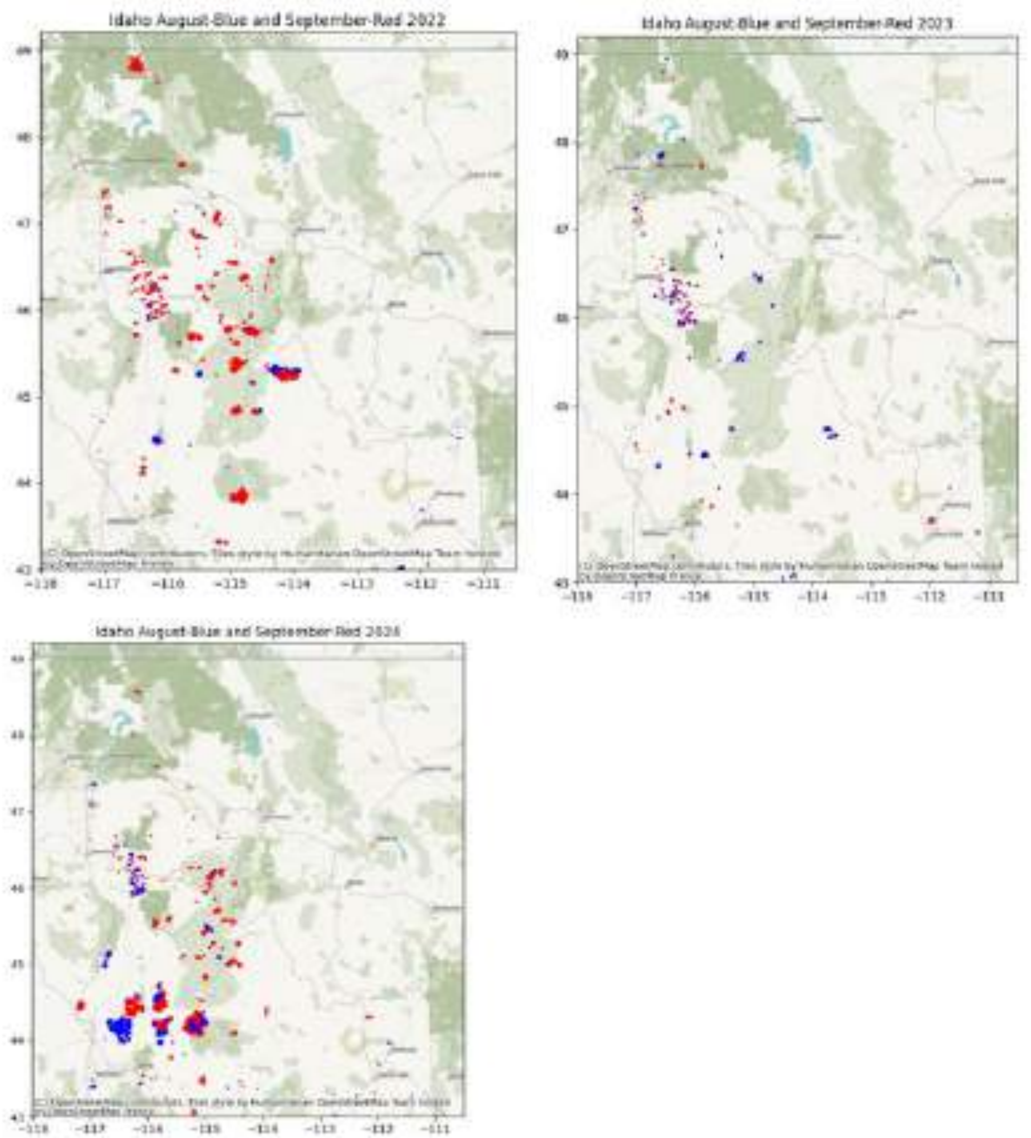


Figure 16
Idaho Map - Fire events of Aug Sep 2022, Aug Sep 2023, and Aug Sep 2024

Plot Interpretation:

Figure 17 is represented as follows,

- All fire events from August 2022, August 2023, and August 2024 are depicted on the California map in red.
- All fire events from September 2022, September 2023, and September 2024 are represented on the California map in blue.
- August and September are noted as the peak months for fire events in this location.

Plot Inference:

Fire events in this location demonstrate the following behaviours:

- Fire events tend to occur in isolated spots.
- Fire events tend to spread.
- In 2024, approximately 10 fire events in this location expanded into larger areas.
- In 2022, roughly nine fire events in this location spread to wider areas.

3.5.4 Fire Events in Oregon

3.5.4.1 Seasonal Influence on Wildfire Events

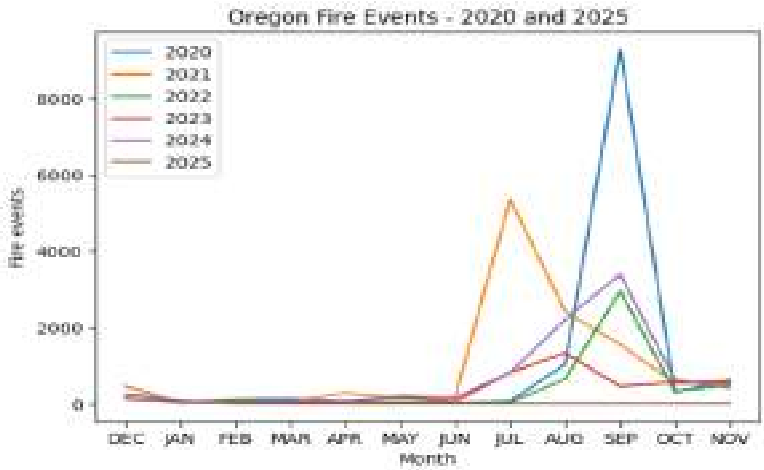


Figure 17
Oregon fire events Monthly - 2020 to 2025

Plot Interpretation

Figure 18: This figure presents multiple subplots, illustrating the total number of fire events on the Y-axis and the months on the X-axis for the years 2020 to 2025.

Plot Inference:

Insights into fire events in Oregon:

- In the last 5 years (2020-2024), fire events peak between June and October.
- During this period, the year 2021 recorded the highest number of fire events.
- In the past 3 years (2022-2024), the year 2024 reports the highest number of fire events.

3.5.4.2 Geographic location influence on Wildfire events

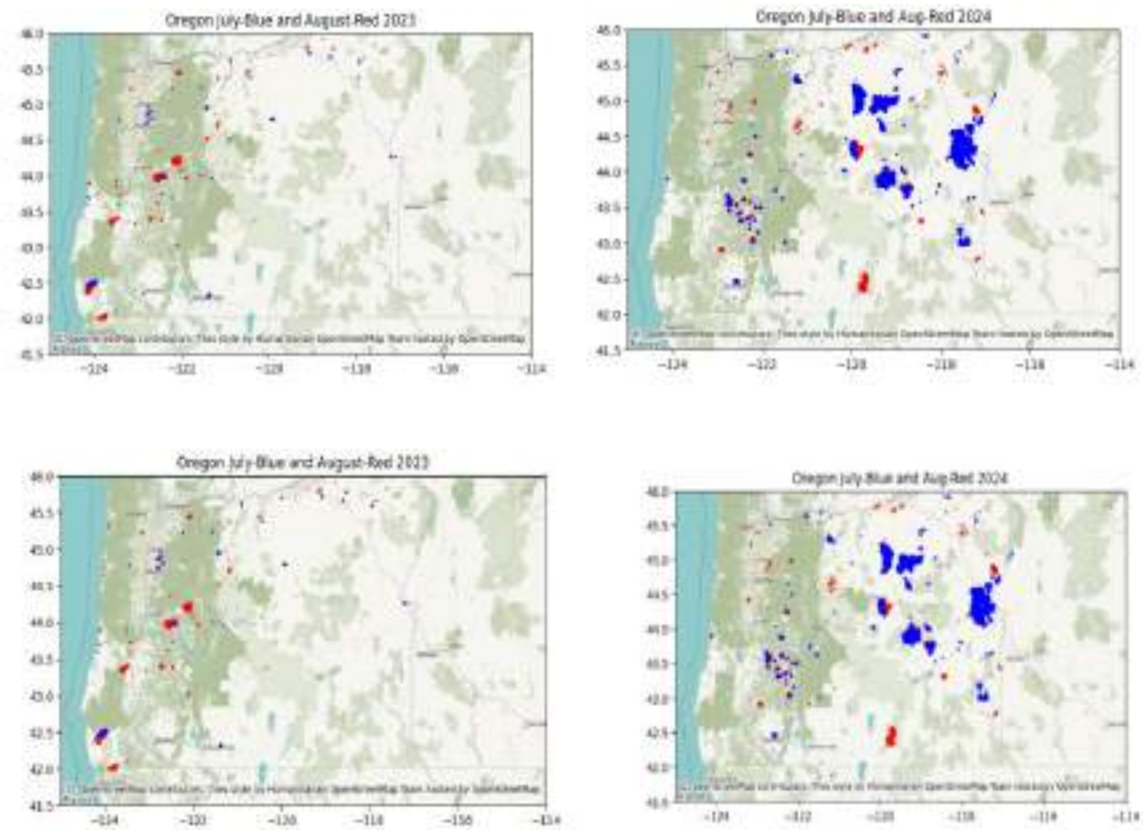


Figure 18
Oregon Map - Fire events of July-Aug 2023 and July-Aug 2024

Plot Interpretation:

Figure 19 is represented as follows,

- All fire events from July 2023 and July 2024 are depicted on the Oregon map in blue.
- All fire events from August 2023 and August 2024 are represented on the Oregon map in red.
- July and August are selected for this representation, as they are the peak fire months for this location.

Plot Inference:

Fire events in this location demonstrate the following behaviours:

- Fire events tend to occur in isolated spots.
- Fire events tend to spread.
- In 2024, approximately nine fire events in this location expanded into larger areas.
- In 2023, roughly five fire events in this location spread to wider areas.

3.5.5 Fire Events in Texas

3.5.5.1 Seasonal Influence on the Fire Events

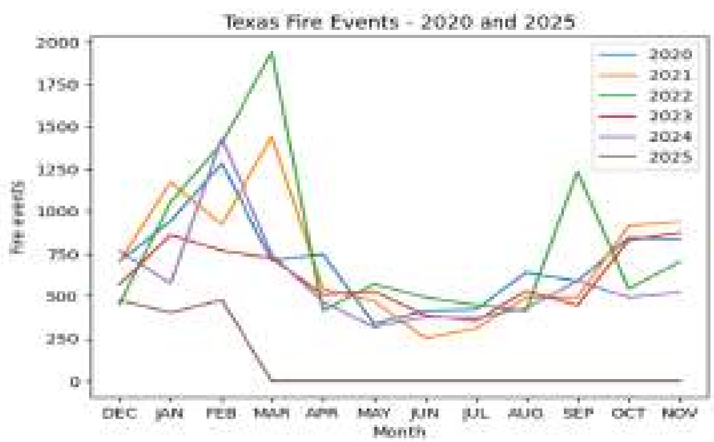


Figure 19
Texas fire events Monthly - 2020 to 2025

Plot Interpretation:

Figure 20: This figure presents multiple subplots, illustrating the total number of fire events on the Y-axis and the months on the X-axis for the years 2020 to 2025.

Plot Inference:

Insights into Wildfire Events in Texas

- In the last 6 years (2020-2025), fire events peak between January and April.
- In the last 5 years (2020-2024), fire events have remained moderate from April to September.

- In the last 5 years (2020-2024), there has been a slight increase in fire events from September to November.
- In the past 5 years (2020-2024), the year 2022 reported the highest number of fire events.
- In the past 3 years (2022-2024), the year 2024 reports the highest number of fire events.

3.5.5.2 Geographic location influence on Fire events

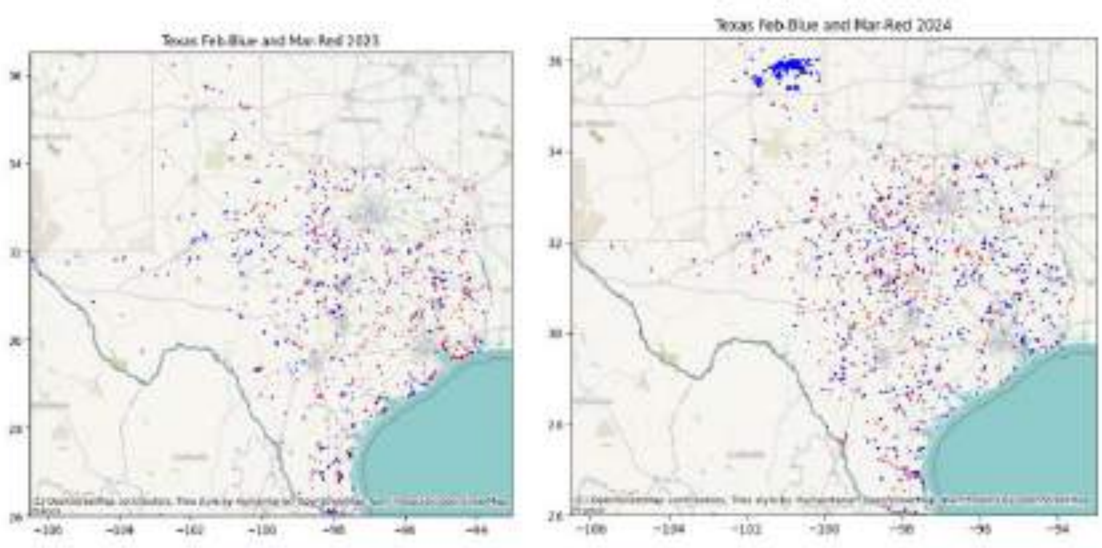


Figure 20
Texas Map - Fire events of Feb Mar 2023 and Feb Mar 2024

Plot Interpretation

Figure 21 is represented as follows,

- All wildfire events from July 2023 and July 2024 are depicted on the Texas map in blue.
- All wildfire events from August 2023 and August 2024 are represented on the Texas map in red.
- February and March are selected for this depiction, as they are the peak wildfire months for this location.

Plot Inference:

Fire events in this location demonstrate the following behaviour:

- Fire events tend to occur in isolated spots.
- In 2023 and 2024, there were no visible observations of fire spread in this location.

3.5.6 Fire Events in Georgia

3.5.6.1 Seasonal Influence on Fire Events

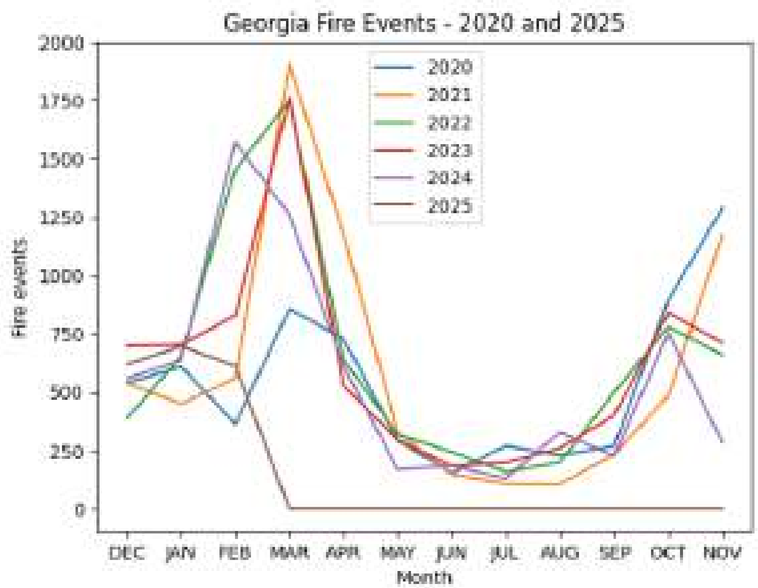


Figure 21
Georgia Fire events Monthly - 2020 to 2025

Plot Interpretation:

Figure 22: This figure presents multiple subplots, illustrating the total number of fire events on the Y-axis and the months on the X-axis for the years 2020 to 2025.

Plot Inference:

Plot Inference:

- In the last 5 years (2020-2024), fire events have shown a peak from January to May.
- Additionally, there has been an increase in fire events from September to November, followed by a decrease in December.

3.5.6.2 Geographic location influence on Fire events

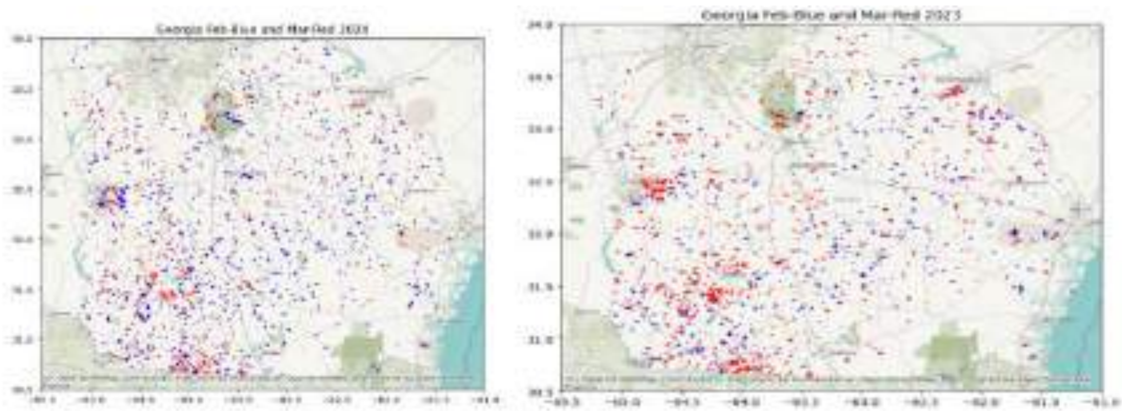


Figure 22
Georgia Map - Fire events of Feb Mar 2023 and Feb Mar 2024

Plot Interpretation:

Figure 23 is represented as follows,

- All the Fire events of July 2023 and July 2024 are represented in the Georgia Map in blue.
- All the Fire events of August 2023 and August 2024 are represented in the Georgia Map in red.
- February and March are considered for the depiction as they are the peak fire months in Georgia.

Plot Inference:

Fire event in this location exhibits the following behavior,

- Fire event exhibits a spot
- In 2023 and 2024, no visible observation of the fire spread in this location.

3.5.7 Fire Events In Florida

3.5.7.1 Seasonal Influence on Fire Events

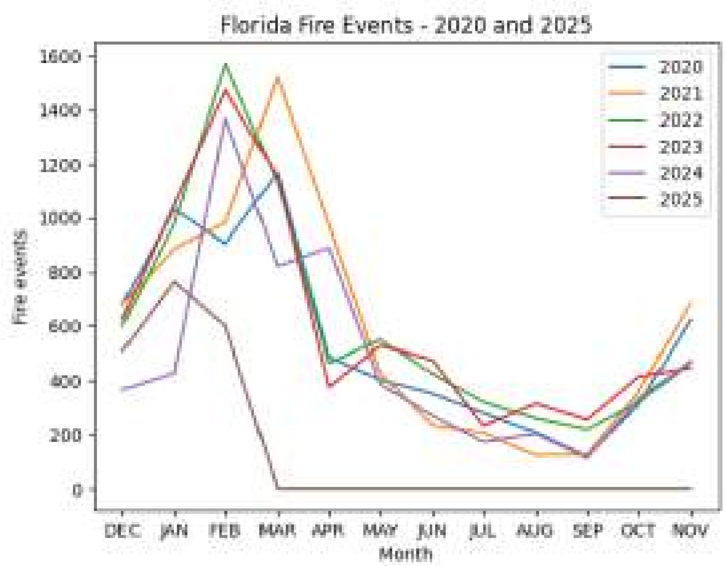


Figure 23
Florida Fire events Monthly - 2020 to 2025

Plot Interpretation:

Figure 24 demonstrates multiple subplots, showcasing the total number of fire events represented on the Y-axis against the months on the X-axis for the years 2020 to 2025.

Plot Inference:

Insights of Fire events in Florida:

- In the last 5 years (2020-2024), fire events have shown peaks between January and May.
- Additionally, there has been a gradual increase in fire events from September to January, followed by another peak.

Geographic location influence on Fire events

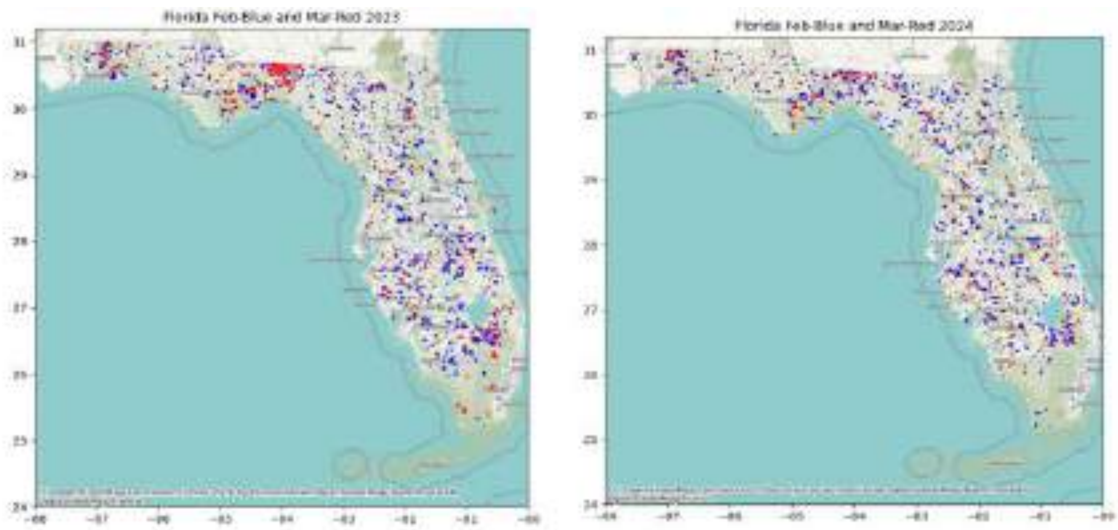


Figure 24
Florida Map - Fire events of Feb Mar 2023 and Feb Mar 2024

Plot Interpretation

Figure 25 is represented as follows,

- All the wildfire events from July 2023 and July 2024 are represented in blue on the Florida map.
- Similarly, all wildfire events from August 2023 and August 2024 are shown in red.
- Additionally, February and March are highlighted, as these months are considered peak wildfire periods in Florida.

Plot Inference:

The fire in this location exhibits the following behavior:

It shows spot fires. However, in 2023 and 2024, there have been no visible observations of fire spread in this area.

3.5.8 Fire Events In Pennsylvania

3.5.8.1 Seasonal Influence on Fire Events

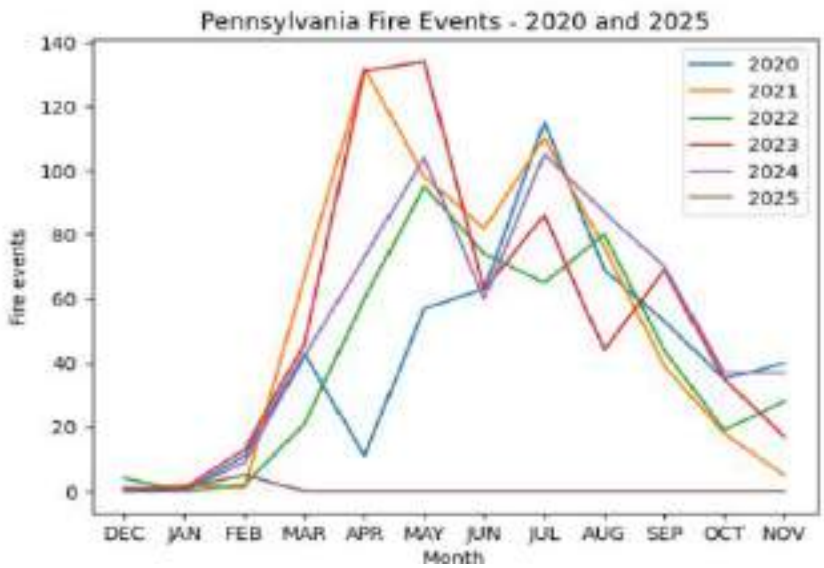


Figure 25
Pennsylvania fire events Monthly - 2020 to 2025

Plot Interpretation: Figure 26 displays multiple subplots that indicate the total number of fire events on the Y-axis, with months represented on the X-axis for the years 2020 to 2025.

Plot Inference: In the last 5 years (2020-2025), Pennsylvania has reported one of the lowest numbers of fire events. This state is selected for analysis to understand better the seasonal patterns of fire events in the northeastern part of the USA.

3.5.8.2 Geographic location influence on Fire events

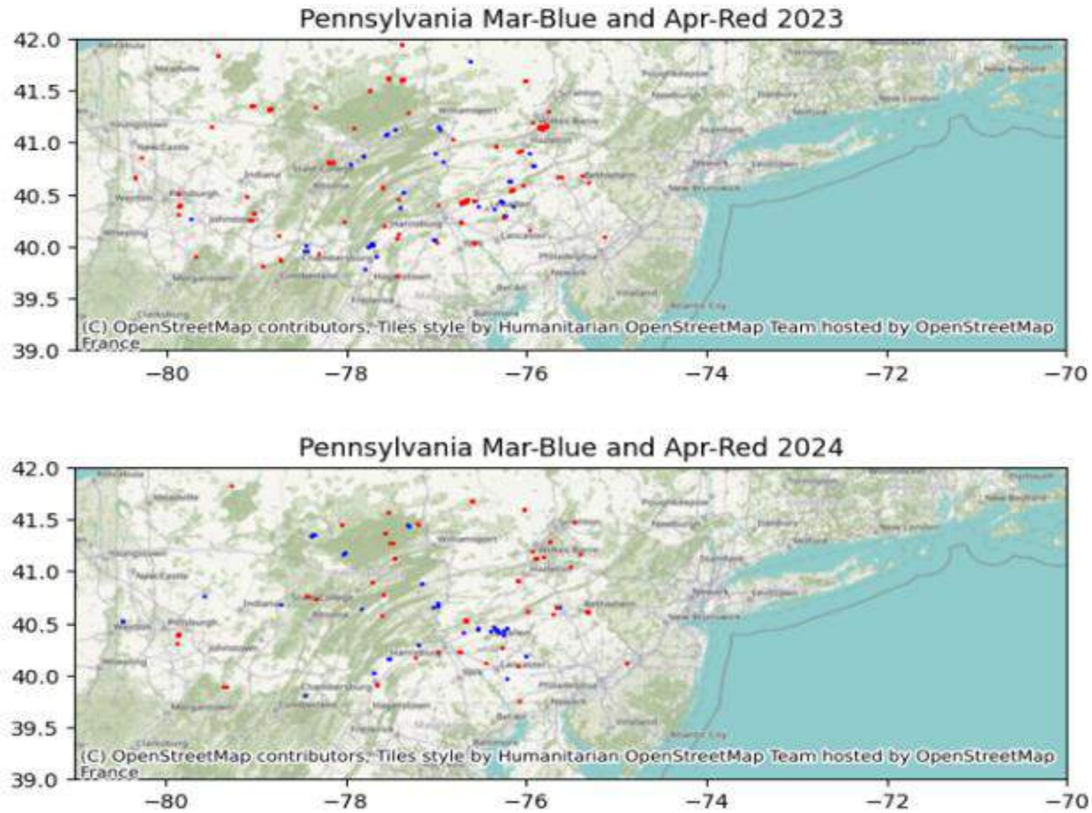


Figure 26
Pennsylvania Map - Fire events of Mar Apr 2023 and Mar Apr 2024

Plot Interpretation:

Figure 27 is represented as follows,

- Fire events from March 2023 and March 2024 are represented in blue on the Pennsylvania map.
- In contrast, fire events from April 2023 and April 2024 are shown in red on the same map.

Plot Inference:

Fire in this location exhibits the following behavior:

Spot fires characterize them. Additionally, there have been no visible observations of fire spread in this area during 2023 and 2024.

3.6 Data Analysis Summary

Exploratory Data analysis reveals that wildfires are spreading at specific longitudes towards the western part of the country, and have a low tendency to spread at specific longitudes towards the eastern part of the country. Wildfire spread behavior did not vary across different latitudes of the country. Southwest to northwest exhibited the exact wildfire-spreading nature across all latitudes, while southeast to northeast exhibited a low tendency of wildfire spread across all latitudes of the country. Possible causes of spreading could be highly flammable vegetation and heavy winds specific to those regions, or a lack of efficiency in the human operation to control the wildfire in some areas.

States in the USA show a seasonal pattern, western states experience growing wildfires from June to November, and the Eastern part of the USA experiences mostly isolated fire events from January to May.

The historical raw fire event datasets from California and Idaho, both located in the western region of the USA, have been selected for AI model experimentation. These states are among the top five in the USA that have experienced wildfires over the past five years. The fire events included in the dataset encompass all combinations of small and large fire growths, noises, nearby fire growth areas, and the gradual increase in fire growth from the initial discovery day to the following days.

3.7 Data Preparation

3.7.1 Elimination of Variables

This research aims to predict “Fire growth”. The dataset comprised the parameters’ latitude, longitude, acq_date, acq_time, satellite, instrument, brightness, scan, track, confidence, frp, bright_t31, and day-night for the entire globe.

Parameters brightness, scan, track, confidence, frp, bright_t31, and day-night are eliminated from the dataset.

3.7.2 Derived Metrics

To filter the raw fire events for the selected region, derive a dataset that captures explicitly only the fire events occurring within the defined boundaries of that region. This process involves applying a geographical filter to the raw data, ensuring that only relevant fire events are included. Once the filtering is complete, create a more focused dataset that can be used for further analysis and insight into fire occurrences in the specified area.

California:

- Latitude - 32.5 to 42.
- Longitude -124 to -115.

Idaho:

- Latitude - 43 to 49.
- Longitude -124 to -110.5.

To further analyze the dataset, separate it into individual datasets corresponding to each day. This step is essential to illustrate the variation in fire points from the day the fire spread was first identified, as well as the subsequent days reflecting the fire's growth.

By doing this, we can visually validate the clustering of fire growth observed on the initial day and compare it with the clusters from the following days.

California Datasets:

- Filter by date - January 7, 2025
- Filter by date - January 8, 2025
- Filter by date - January 9, 2025

Idaho Datasets:

- Filter by date - August 5, 2025
- Filter by date - August 6, 2025
- Filter by date - August 7, 2025

Remove the parameters acq_date, acq_time, satellite, and instrument from the datasets. Consequently, each dataset will consist of just two parameters: Latitude and Longitude for the selected dates and the specified region.

3.8 Implementation and evaluation

3.8.1 Machine Learning Model and Evaluation

The following unsupervised learning models are experimented with and evaluated.

- K-Means Clustering.
- Fuzzy C-means Clustering.
- Gaussian Mixture Models Clustering.
- Agglomerative Hierarchical Clustering.
- DBSCAN clustering machine learning algorithm.
- New Proposed Model: Multi-level multi-criteria clustering algorithm.

3.8.2 K-means Clustering

The following Python Packages are used for the development:

- General Libraries: These are necessary for reading the dataset in CSV format and for deriving the dataset based on the timestamp of the fire event occurrence.
- K-means Clustering Libraries: These libraries are required for computing 'k' clusters from the input dataset.
- Clustering Evaluation Libraries: These libraries are necessary for evaluating the clusters generated by the K-means algorithm.

General Import Libraries

- import pandas as pd
- from datetime import datetime
- import time

K-means Clustering Libraries

- from sklearn.cluster import KMeans
- import math
- import numpy as np

K-means Clustering Evaluation Libraries

- from sklearn.metrics import silhouette_score
- from sklearn.metrics import pairwise_distances
- import matplotlib.pyplot as plt

Dataset Preparation

Firstly, derive the dataset based on the specified region by filtering using the geographic coordinates. For Region 1, which is California, USA, select the relevant geo-coordinates to isolate the dataset pertaining to this area.

- 'latitude’ < 42 'latitude' > 32.5
- 'longitude' > -124 'longitude' < -115.5

Region 2: Idaho, USA

- 'latitude' < 49   'latitude' > 43
- 'longitude' > -118 'longitude' < -110.5

Secondly, derive the dataset for the required dates, focusing specifically on the incident days. Sample Dataset 1, Region 1: California, USA: The dataset includes fire events from MODIS for the following dates:

- January 6, 2025
- January 7, 2025
- January 8, 2025

Sample Dataset 2, Region 1: California, USA: The dataset includes fire events from MODIS for the following dates:

- January 21, 2025
- January 22, 2025
- January 23, 2025

Sample Dataset 3, Region 2: Idaho, USA: The dataset includes fire events from MODIS for the following dates:

- August 5, 2025
- August 6, 2025
- August 7, 2025

Modeling the Data Using K-Means

To initialize the K-means algorithm, the following input parameters need to be set:

Randomly choose the value of 'k,' which represents the desired number of clusters. You can set 'k' to either 2, 3, or 4, depending on the analytical requirements for the dataset.

K-means Clustering - Evaluation

The following evaluation metrics are used to evaluate the K-means clustering algorithm.

- Silhouette score

- Inertia
- Dunn Index

3.8.3 Fuzzy C-Means Clustering

The following Python packages are used for the development:

General Libraries: These libraries are essential for reading the dataset in CSV format and for deriving the dataset based on the timestamp of the fire event occurrence.

Fuzzy C-Means Libraries: These libraries are required for computing Fuzzy C-Means clusters from the input dataset.

Fuzzy C-Means Clustering Evaluation Libraries: These libraries are necessary for evaluating the clusters derived from the Fuzzy C-Means algorithm.

General Import Libraries

- import pandas as pd
- from datetime import datetime
- import time

K-means Clustering and Evaluation Libraries

- import skfuzzy as fuzz
- from skfuzzy import control as ctrl
- from sklearn.metrics import pairwise_distances
- import matplotlib.pyplot as plt

Dataset Preparation

Firstly, derive the dataset based on the specified region by filtering with the geographic coordinates.

For Region 1: California, USA, select the relevant geo-coordinates to isolate the dataset about this area. This step involves ensuring that only the data points that fall within the defined latitude and longitude ranges for California are included in the dataset.

- 'latitude’ < 42 'latitude' > 32.5
- 'longitude' > -124 'longitude' < -115.5

Region 2: Idaho, USA

- 'latitude’ < 49   'latitude' > 43
- 'longitude' > -118 ‘longitude' < -110.5

Secondly, derive the dataset for the required dates, specifically on the incident days.

Sample Dataset 1, Region 1: California, USA: The dataset consists of fire events from MODIS for the following dates:

- January 6, 2025
- January 7, 2025
- January 8, 2025

Sample Dataset 2, Region 1: California, USA: The dataset consists of fire events from MODIS for the following dates:

- January 21, 2025
- January 22, 2025
- January 23, 2025

Sample Dataset 3, Region 2: Idaho, USA: The dataset consists of fire events from MODIS for the following dates:

- August 5, 2025
- August 6, 2025
- August 7, 2025

Modeling the Data Using Fuzzy C-Means Clustering

The following input parameters need to be initialized for the Fuzzy C-means clustering algorithm:

‘c’: This parameter represents the number of clusters and can be set to 2, 3, or 4, depending on the density of the dataset.

‘m’: This is the fuzziness parameter, which controls the degree of overlap or fuzziness between clusters. It typically ranges from 1.5 to 2.5 and can be set to values such as 1.5, 2, or 2.5.

Fuzzy C-Means Clustering - Evaluation: After clustering, various evaluation metrics can be applied to assess the clustering results, ensuring the effectiveness of the clustering process.

The following evaluation metrics are used for evaluating Fuzzy C-means clustering.

- Fuzzy Partition Coefficient (FPC)
- Partition Entropy Coefficient (PEC)
- Dunn Index

3.8.4 Gaussian Mixture Models Clustering

The following Python packages are used for the development:

General Libraries: These libraries are essential for reading the dataset in CSV format and for deriving the dataset based on the timestamp of the fire event occurrence.

Gaussian Mixture Models Libraries: These libraries are required for computing the probabilities and clusters of Gaussian Mixture Models from the input dataset.

Gaussian Mixture Models Evaluation Libraries: These libraries are necessary for evaluating the clusters derived from the Gaussian Mixture Models.

General Import Libraries

- import pandas as pd
- from datetime import datetime
- import time

Gaussian Mixture Models, Clustering and Evaluation Libraries

- from sklearn.mixture import GaussianMixture
- from sklearn.metrics import silhouette_score
- import matplotlib.pyplot as plt

Dataset Preparation

Firstly, derive the Dataset based on the region, Filter by regional Geo coordinates

Region 1: California, USA

- 'latitude’ < 42 'latitude' > 32.5
- 'longitude' > -124 'longitude' < -115.5

Region 2: Idaho, USA

- 'latitude’ < 49 'latitude' > 43
- 'longitude' > -118 ‘longitude' < -110.5

Secondly, derive the dataset for the required dates, specifically on the incident days.

Sample Dataset 1, Region 1: California, USA: The dataset consists of fire events from MODIS for the following dates:

- January 6, 2025
- January 7, 2025
- January 8, 2025

Sample Dataset 2, Region 1: California, USA: The dataset consists of fire events from MODIS for the following dates:

- January 21, 2025
- January 22, 2025
- January 23, 2025

Sample Dataset 3, Region 2: Idaho, USA: The dataset consists of fire events from MODIS for the following dates:

- August 5, 2025

- August 6, 2025
- August 7, 2025

Modeling the Data Using Gaussian Mixture Models Clustering

The following input parameters need to be initialized for the Gaussian Mixture Models clustering algorithm:

Component: This parameter represents the number of clusters (or elements) in the model. It can be set to 3. The model will compute the probability of each data point belonging to each cluster.

Gaussian Mixture Models Clustering - Evaluation

The following evaluation metrics are used for evaluating the Gaussian Mixture Model cluster.

- Visual Inspection using the Plots
- Silhouette score

3.8.5 Agglomerative Hierarchical Clustering

The following Python packages are used for the development:

General Libraries: These libraries are essential for reading the dataset in CSV format and for deriving the dataset based on the timestamp of the fire event occurrence.

Agglomerative Hierarchical Clustering Libraries: These libraries are required for performing Agglomerative Hierarchical Clustering on the input dataset.

General Import Libraries

- import pandas as pd
- from datetime import datetime
- import time

Agglomerative Hierarchical Clustering and Evaluation Libraries

- from sklearn.cluster import AgglomerativeClustering

- from scipy.cluster.hierarchy import dendrogram, linkage
- from sklearn. metrics import silhouette_score
- import matplotlib.pyplot as plt.

Dataset Preparation

Firstly, derive the Dataset based on the region, Filter by regional Geo coordinates

Region 1: California, USA

- 'latitude' < 42 'latitude' > 32.5
- 'longitude' > -124 'longitude' < -115.5

Region 2: Idaho, USA

- 'latitude' < 49   'latitude' > 43
- 'longitude' > -118 'longitude' < -110.5

Secondly, derive the dataset for the required dates, specifically on the incident days and near real-time.

Sample Dataset 1, Region 1: California, USA: The dataset consists of fire events from MODIS for the following dates:

- January 6, 2025
- January 7, 2025
- January 8, 2025

Sample Dataset 2, Region 1: California, USA: The dataset consists of fire events from MODIS for the following dates:

- January 21, 2025
- January 22, 2025
- January 23, 2025

Sample Dataset 3, Region 2: Idaho, USA: The dataset consists of fire events from MODIS for the following dates:

- August 5, 2025
- August 6, 2025
- August 7, 2025

Modeling the Data Using Agglomerative Hierarchical Clustering

The following input parameters need to be initialized for the Agglomerative Hierarchical Clustering algorithm:

- Set the number of clusters to 3.
- Linkage method set to ‘complete’ (it can also be set to other values like ‘ward method’, ‘single’, or ‘average’)

Agglomerative Hierarchical Clustering - Evaluation

The following evaluation metrics are used to evaluate the Agglomerative Hierarchical Clustering.

- Visual Inspection using the Plots
- Silhouette score

3.8.6 DBSCAN Clustering

The following Python packages are used for the development:

General libraries: These are required for reading the dataset in CSV format and for deriving the dataset based on the timestamp of the fire events. DBSCAN Clustering libraries: These are necessary for computing DBSCAN on the input dataset.

DBSCAN Clustering Evaluation Libraries: These are required for evaluating the results of the DBSCAN clustering.

General Import Libraries

- import pandas as pd
- from datetime import datetime
- import time

DBSCAN Clustering Libraries

- from sklearn.cluster import DBSCAN
- import matplotlib.pyplot as plt
- import math
- import numpy as np

DBSCAN Clustering Evaluation Libraries

- from sklearn.metrics import silhouette_score
- from sklearn.metrics import calinski_harabasz_score
- from sklearn.metrics import davies_bouldin_score

Dataset Preparation

Firstly, derive the Dataset based on the region, Filter by regional Geo coordinates

Region 1: California, USA

- 'latitude’ < 42 'latitude' > 32.5
- 'longitude' > -124 'longitude' < -115.5

Region 2: Idaho, USA

- 'latitude’ < 49   'latitude' > 43
- 'longitude' > -118 ‘longitude' < -110.5

Secondly, derive a separate dataset for the required days and date, specifically focused on the incident days and the near real-time.

Incident 1, Region 1: California, USA: The dataset consists of fire events from MODIS for the following dates:

- January 6, 2025
- January 7, 2025
- January 8, 2025

Incident 2, Region 1: California, USA: The dataset consists of fire events from MODIS for the following dates:

- January 21, 2025
- January 22, 2025
- January 23, 2025

Incident 3, Region 1: California, USA, Near Real Time: The dataset consists of near-real-time fire events from:

- MODIS
- VIIRS NOAA-21
- VIIRS NOAA-20
- Date: April 21, 2025

Incident 4, Region 2: Idaho, USA

The dataset consists of fire events from MODIS for the following dates:

- August 5, 2025
- August 6, 2025
- August 7, 2025

Modeling The Data Using DBSCAN

The following input parameters are set for the DBSCAN models:

- Set the radius of the neighbourhood around a data point to 0.1 meters.
- Set the minimum points required to form a dense region/cluster.
- If the captured instrument is MODIS, set the minimum points to 5.
- If the captured instrument is VIIRS, set the minimum points to 10.

This input parameter configuration must be tailored to the specific input device in use. Adjustments should be made to ensure optimal clustering based on the characteristics of the data captured by different instruments.

DBSCAN Clustering - Evaluation

The following evaluation metrics are used for evaluating the DBSCAN cluster.

- Silhouette score
- Calinski-Harabasz Index
- Davies-Bouldin Index

3.8.7 Newly Proposed Multi-Level Multi-Criteria Clustering Algorithm

Proposes a new clustering algorithm that first marks all the data points that have more than a specified nearest neighbors, determines the nearest neighbors based on the specified distance between the points, and excludes the data points that do not have nearest neighbors or the nearest neighbors count is less than the specified nearest neighbors. It then regroups, points to a new cluster by identifying one core point from previously marked points, and determines the nearest points to this core point based on the specified distance.

From the core point and rest all the marked points, form a cluster with all the neighboring points nearer to the core points, and exclude the remaining marked points that are not nearer. Mark these points as ‘unknown cluster’, then identify another core point in the remaining points, then determine the nearest neighbors to this core point, and form a new cluster. This process continues until all the points form a new cluster or are eliminated from the group. Then, it further determines the approximate area of each cluster. If the distance between any two clusters is less than the specified inter-cluster distance, then it forms a bigger cluster by combining the two nearest clusters.

It determines the area using a bounding box (Rectangular Approximation), a convex hull(polygon/polytope). The design of the proposed new clustering algorithm is as follows.

- Cluster the data points that share similar characteristics.

- Cluster the data points that are close to each other, and measure the Euclidean distance between these data points.
- Exclude the data points from the Cluster that are far from the other fire spots in The Cluster.
- Form a Cluster when at least 5 fire points are at a shorter distance to each other
- Compute the area of the cluster.
- The higher the fire data points in a smaller area, the higher the probability that the cluster is the wildfire spreading spot.

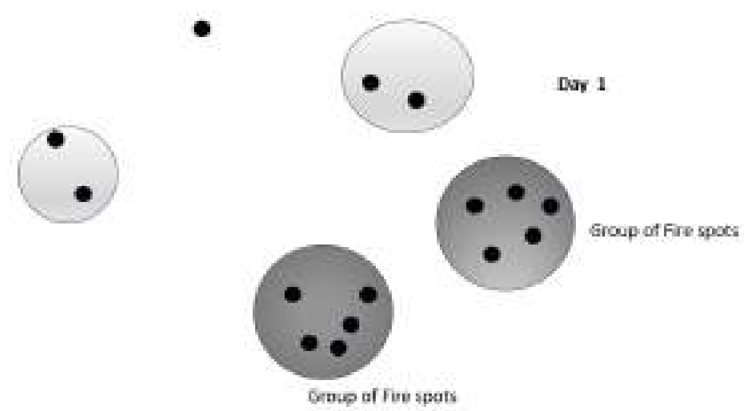


Figure 27
Fire spot - Potential for Clustering

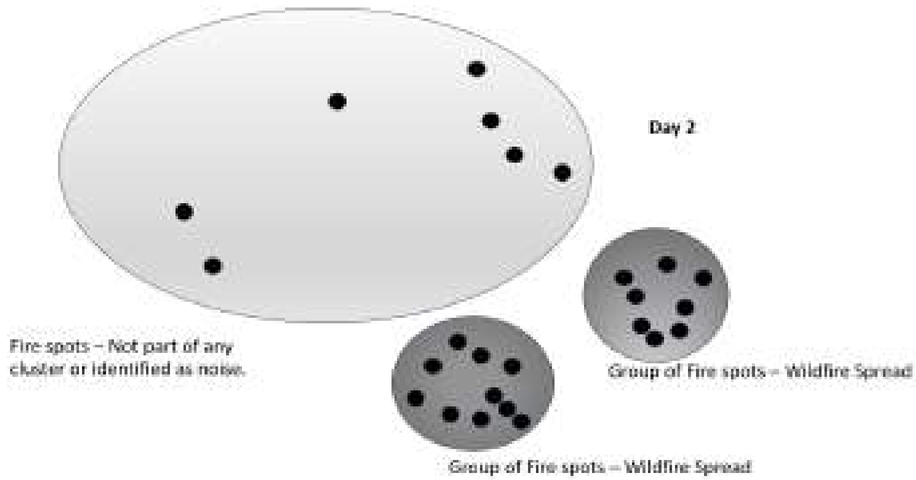


Figure 28
Higher probability of Wildfire spreading to a spot

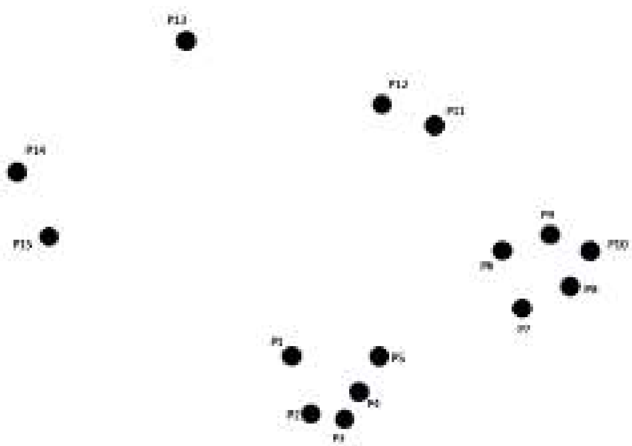


Figure 29
Day 1 - Fire spot - on 2-D space

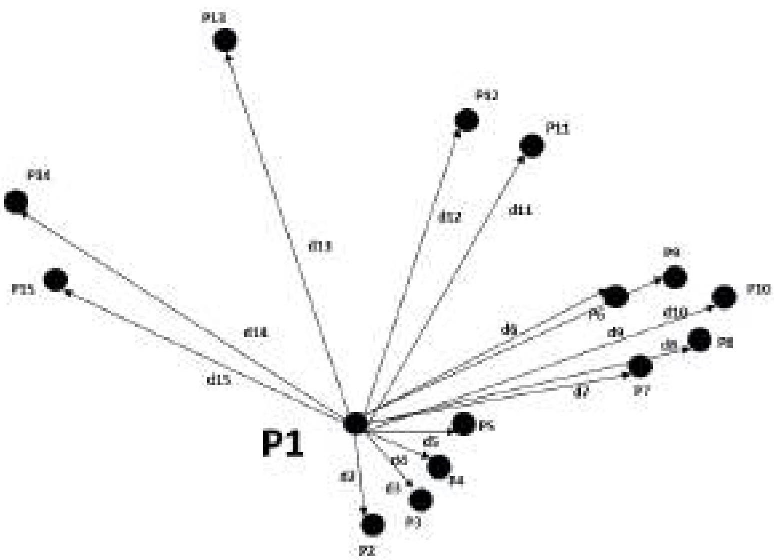


Figure 30
Day 1 - P1 to Pmax - Euclidean distance

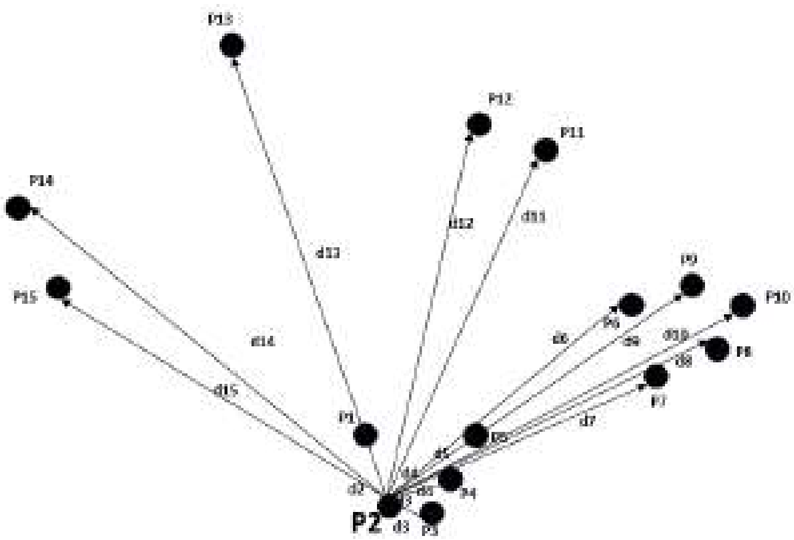


Figure 31
Day 1 - P2 to all other points - Euclidean distance

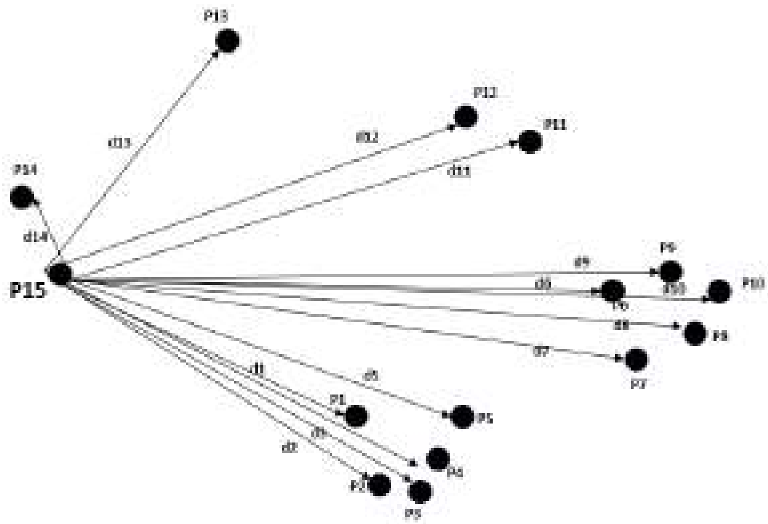


Figure 32
Day 1 - P15 to all other points - Euclidean distance

- Consider Points in the space/dataset P1(x1, y1), P2(x2, y2), P3(x3, y3), P4(x4, y4), ..., P15(x15, y15).
- Compute the distance from P1 to rest all the points in the space/dataset
  d1,n = SQRT [(xn x1)2 +(yn - y1)2], where 'n' ranges from 1 to max points in the space/dataset.
- Compute the distance from P2 to all the points in the space/dataset
  d2,n = SQRT [(xn x2)2 +(yn – y2)2], where 'n' ranges from 1 to a maximum point in the space/dataset.
- Compute the distance from P3 to all the points in the space/dataset
  d3,n = SQRT [(xn x3)2 +(yn – y3)2], where 'n' is the 1 to max points in the space/dataset.
- Compute the distance from P4 to all the points in the space/dataset
  d4,n = SQRT [(xn x4)2 +(yn – y3)2], where 'n' is the 1 to max points in the space/dataset.

- Compute the distance from P5 to rest all the points in the space/dataset

  d5, n = SQRT [(xn x5)2 +(yn – y5)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.

  Similarly, compute the distance from all the points

- Compute the distance from P6 to all the points in the space/dataset

  d6, n = SQRT [(xn x6)2 +(yn – y6)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.

- Compute the distance from P7 to all the points in the space/dataset

  d7, n = SQRT [(xn x7)2 +(yn – y7)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.

- Compute the distance from P8 to all the points in the space/dataset

  d8, n = SQRT [(xn x8)2 +(yn – y8)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.

- Compute the distance from P9 to all the points in the space/dataset

  d9, n = SQRT [(xn x9)2 +(yn – y9)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.

- Compute the distance from P10 to all the points in the space/dataset

  d10, n = SQRT [(xn x10)2 +(yn – y10)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.

- Compute the distance from P11 to rest all the points in the space/dataset

  d11, n = SQRT [(xn x11)2 +(yn – y11)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.

- Compute the distance from P12 to all the points in the space/dataset

  d12, n = SQRT [(xn x12)2 +(yn – y12)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.

- Compute the distance from P13 to all the points in the space/dataset

  d13, n = SQRT [(xn x13)2 +(yn – y13)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.
- Compute the distance from P14 to all the points in the space/dataset

  d14, n = SQRT [(xn x14)2 +(yn – y14)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.
- Compute the distance from P14 to rest all the points in the space/dataset, dn, n-1 = SQRT [(xn-1 xn)2 +(yn-1 – yn)2], where ‘n’ varies from 1 to the maximum number of data points in the space/dataset.

  dmin(1… n).

Table 6

Computation of distances - all points of the dataset

Points	Distance with other Points in Space or Dataset													
P1	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$	$d_{1,7}$	$d_{1,8}$	$d_{1,9}$	$d_{1,10}$	$d_{1,11}$	$d_{1,12}$	$d_{1,13}$	$d_{1,14}$	$d_{1,15}$
P2	$d_{2,1}$	$d_{2,3}$	$d_{2,4}$	$d_{2,5}$	$d_{2,6}$	$d_{2,7}$	$d_{2,8}$	$d_{2,9}$	$d_{2,10}$	$d_{2,11}$	$d_{2,12}$	$d_{2,13}$	$d_{2,14}$	$d_{2,15}$
P3	$d_{3,1}$	$d_{3,2}$	$d_{3,4}$	$d_{3,5}$	$d_{3,6}$	$d_{3,7}$	$d_{3,8}$	$d_{3,9}$	$d_{3,10}$	$d_{3,11}$	$d_{3,12}$	$d_{3,13}$	$d_{3,14}$	$d_{3,15}$
P4	$d_{4,1}$	$d_{4,2}$	$d_{4,3}$	$d_{4,5}$	$d_{4,6}$	$d_{4,7}$	$d_{4,8}$	$d_{4,9}$	$d_{4,10}$	$d_{4,11}$	$d_{4,12}$	$d_{4,13}$	$d_{4,14}$	$d_{4,15}$
P5	$d_{5,1}$	$d_{5,2}$	$d_{5,3}$	$d_{5,4}$	$d_{5,6}$	$d_{5,7}$	$d_{5,8}$	$d_{5,9}$	$d_{5,10}$	$d_{5,11}$	$d_{5,12}$	$d_{5,13}$	$d_{5,14}$	$d_{5,15}$
P6	$d_{6,1}$	$d_{6,2}$	$d_{6,3}$	$d_{6,4}$	$d_{6,5}$	$d_{6,7}$	$d_{6,8}$	$d_{6,9}$	$d_{6,10}$	$d_{6,11}$	$d_{6,12}$	$d_{6,13}$	$d_{6,14}$	$d_{6,15}$
P7	$d_{7,1}$	$d_{7,2}$	$d_{7,3}$	$d_{7,4}$	$d_{7,5}$	$d_{7,6}$	$d_{7,8}$	$d_{7,9}$	$d_{7,10}$	$d_{7,11}$	$d_{7,12}$	$d_{7,13}$	$d_{7,14}$	$d_{7,15}$
P8	$d_{8,1}$	$d_{8,2}$	$d_{8,3}$	$d_{8,4}$	$d_{8,5}$	$d_{8,6}$	$d_{8,7}$	$d_{8,9}$	$d_{8,10}$	$d_{8,11}$	$d_{8,12}$	$d_{8,13}$	$d_{8,14}$	$d_{8,15}$
P9	$d_{9,1}$	$d_{9,2}$	$d_{9,3}$	$d_{9,4}$	$d_{9,5}$	$d_{9,6}$	$d_{9,7}$	$d_{9,8}$	$d_{9,10}$	$d_{9,11}$	$d_{9,12}$	$d_{9,13}$	$d_{9,14}$	$d_{9,15}$
P10	$d_{10,1}$	$d_{10,2}$	$d_{10,3}$	$d_{10,4}$	$d_{10,5}$	$d_{10,6}$	$d_{10,7}$	$d_{10,8}$	$d_{10,10}$	$d_{10,11}$	$d_{10,12}$	$d_{10,13}$	$d_{10,13}$	$d_{10,15}$
P11	$d_{11,1}$	$d_{11,2}$	$d_{11,3}$	$d_{11,4}$	$d_{11,5}$	$d_{11,6}$	$d_{11,7}$	$d_{11,8}$	$d_{11,10}$	$d_{11,11}$	$d_{11,12}$	$d_{11,13}$	$d_{11,13}$	$d_{11,15}$
P12	$d_{12,1}$	$d_{12,2}$	$d_{12,3}$	$d_{12,4}$	$d_{12,5}$	$d_{12,6}$	$d_{12,7}$	$d_{12,8}$	$d_{12,10}$	$d_{12,11}$	$d_{12,12}$	$d_{12,13}$	$d_{12,13}$	$d_{12,15}$
P13	$d_{13,1}$	$d_{13,2}$	$d_{13,3}$	$d_{13,4}$	$d_{13,5}$	$d_{13,6}$	$d_{13,7}$	$d_{13,8}$	$d_{13,9}$	$d_{13,10}$	$d_{13,11}$	$d_{13,12}$	$d_{13,14}$	$d_{13,15}$
P14	$d_{14,1}$	$d_{14,2}$	$d_{14,3}$	$d_{14,4}$	$d_{14,5}$	$d_{14,6}$	$d_{14,7}$	$d_{14,8}$	$d_{14,9}$	$d_{14,10}$	$d_{14,11}$	$d_{14,12}$	$d_{14,13}$	$d_{14,15}$
P15	$d_{15,1}$	$d_{15,2}$	$d_{15,3}$	$d_{15,4}$	$d_{15,5}$	$d_{15,6}$	$d_{15,7}$	$d_{15,8}$	$d_{15,9}$	$d_{15,10}$	$d_{15,11}$	$d_{15,12}$	$d_{15,13}$	

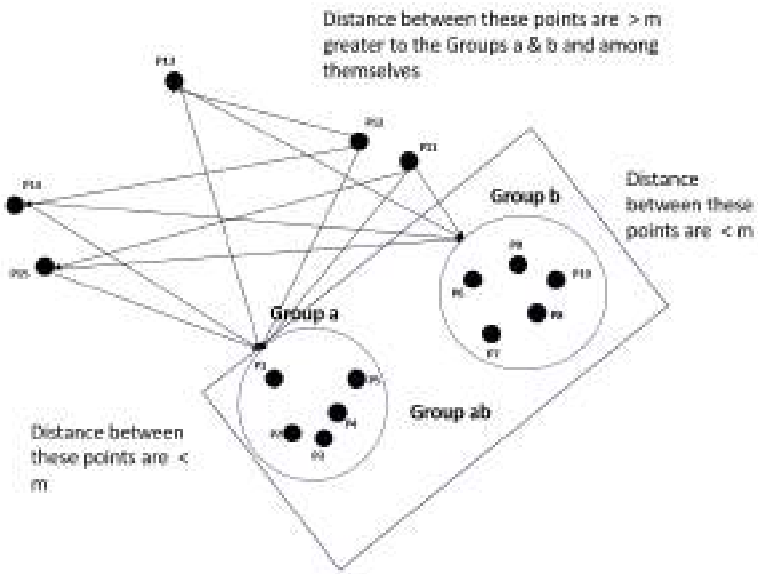


Figure 33
Grouping the points to cluster based on the farthest point distance

- Choose the distance that is based on the business need. The basis for choosing this distance is the historical event and the area of spread.
- Choose 3 miles, 5 Miles, or 10 miles, **Cluster Farthest Points Distance** = m.
- The basis for choosing this distance is the historical event and the area of spread.
- Then filter out all the points above this distance; these are most likely the fire spots that are not spreading and are farthest apart, so eliminate those points.
- Compare the distance computed between the points < **Cluster Farthest Points Distance**.
- Choose the **Nearest neighbor count** based on historic events. Basis: In the history data set, fire events that are nearer on day 1 and then sped up from the second day, 4 or 10, based on the history.
- Eliminate all the points from the dataset that do not have any nearest neighbors.
- From the cluster (Cluster ab) of points where each point has the **Nearest neighbor points >= the Nearest neighbor count.**

Table 7
Distance between points that are less than the clusters' farthest points

Point	Nearest Neighbor p= Nearest Neighbor Count	Distance	Nearest Neighbor	Distance	Nearest Neighbor	Distance	Nearest Neighbor	Distance	Nearest Neighbor	Distance	Nearest Neighbor	Distance	Nearest Neighbor	Distance	Nearest Neighbor	Distance	Nearest Neighbor		Nearest Neighbor					
P1	Yes	$d_{1,2}$	Yes	$d_{1,3}$	Yes	$d_{1,4}$	Yes	$d_{1,5}$	Yes	$d_{1,6}$	No	$d_{1,7}$	No	$d_{1,8}$	No	$d_{1,9}$	No	$d_{1,10}$	No	$d_{1,11}$	$d_{1,12}$	$d_{1,13}$	$d_{1,14}$	$d_{1,15}$
P2	Yes	$d_{2,1}$	Yes	$d_{2,3}$	Yes	$d_{2,4}$	Yes	$d_{2,5}$	Yes	$d_{2,6}$	No	$d_{2,7}$	No	$d_{2,8}$	No	$d_{2,9}$	No	$d_{2,10}$	No	$d_{2,11}$	$d_{2,12}$	$d_{2,13}$	$d_{2,14}$	$d_{2,15}$
P3	Yes	$d_{3,1}$	Yes	$d_{3,2}$	Yes	$d_{3,4}$	Yes	$d_{3,5}$	Yes	$d_{3,6}$	No	$d_{3,7}$	No	$d_{3,8}$	No	$d_{3,9}$	No	$d_{3,10}$	No	$d_{3,11}$	$d_{3,12}$	$d_{3,13}$	$d_{3,14}$	$d_{3,15}$
P4	Yes	$d_{4,1}$	Yes	$d_{4,2}$	Yes	$d_{4,3}$	Yes	$d_{4,5}$	Yes	$d_{4,6}$	No	$d_{4,7}$	No	$d_{4,8}$	No	$d_{4,9}$	No	$d_{4,10}$	No	$d_{4,11}$	$d_{4,12}$	$d_{4,13}$	$d_{4,14}$	$d_{4,15}$
P5	Yes	$d_{5,1}$	Yes	$d_{5,2}$	Yes	$d_{5,3}$	Yes	$d_{5,4}$	Yes	$d_{5,6}$	No	$d_{5,7}$	No	$d_{5,8}$	No	$d_{5,9}$	No	$d_{5,10}$	No	$d_{5,11}$	$d_{5,12}$	$d_{5,13}$	$d_{5,14}$	$d_{5,15}$
P6	Yes	$d_{6,1}$	No	$d_{6,2}$	No	$d_{6,3}$	No	$d_{6,4}$	No	$d_{6,5}$	Yes	$d_{6,7}$	Yes	$d_{6,8}$	Yes	$d_{6,9}$	Yes	$d_{6,10}$	Yes	$d_{6,11}$	$d_{6,12}$	$d_{6,13}$	$d_{6,14}$	$d_{6,15}$
P7	Yes	$d_{7,1}$	No	$d_{7,2}$	No	$d_{7,3}$	No	$d_{7,4}$	No	$d_{7,5}$	Yes	$d_{7,6}$	Yes	$d_{7,8}$	Yes	$d_{7,9}$	Yes	$d_{7,10}$	Yes	$d_{7,11}$	$d_{7,12}$	$d_{7,13}$	$d_{7,14}$	$d_{7,15}$
P8	Yes	$d_{8,1}$	No	$d_{8,2}$	No	$d_{8,3}$	No	$d_{8,4}$	No	$d_{8,5}$	Yes	$d_{8,6}$	Yes	$d_{8,7}$	Yes	$d_{8,9}$	Yes	$d_{8,10}$	Yes	$d_{8,11}$	$d_{8,12}$	$d_{8,13}$	$d_{8,14}$	$d_{8,15}$
P9	Yes	$d_{9,1}$	No	$d_{9,2}$	No	$d_{9,3}$	No	$d_{9,4}$	No	$d_{9,5}$	Yes	$d_{9,6}$	Yes	$d_{9,7}$	Yes	$d_{9,8}$	Yes	$d_{9,10}$	Yes	$d_{9,11}$	$d_{9,12}$	$d_{9,13}$	$d_{9,14}$	$d_{9,15}$
P10	Yes	$d_{10,1}$	No	$d_{10,2}$	No	$d_{10,3}$	No	$d_{10,4}$	No	$d_{10,5}$	Yes	$d_{10,6}$	Yes	$d_{10,7}$	Yes	$d_{10,8}$	Yes	$d_{10,9}$	Yes	$d_{10,11}$	$d_{10,12}$	$d_{10,13}$	$d_{10,14}$	$d_{10,15}$
P11	No	$d_{11,1}$	No	$d_{11,2}$	No	$d_{11,3}$	No	$d_{11,4}$	No	$d_{11,5}$	No	$d_{11,6}$	No	$d_{11,7}$	No	$d_{11,8}$	No	$d_{11,9}$	No	$d_{11,10}$	$d_{11,12}$	$d_{11,13}$	$d_{11,14}$	$d_{11,15}$
P12	No	$d_{12,1}$	No	$d_{12,2}$	No	$d_{12,3}$	No	$d_{12,4}$	No	$d_{12,5}$	No	$d_{12,6}$	No	$d_{12,7}$	No	$d_{12,8}$	No	$d_{12,9}$	No	$d_{12,10}$	$d_{12,11}$	$d_{12,13}$	$d_{12,14}$	$d_{12,15}$
P13	No	$d_{13,1}$	No	$d_{13,2}$	No	$d_{13,3}$	No	$d_{13,4}$	No	$d_{13,5}$	No	$d_{13,6}$	No	$d_{13,7}$	No	$d_{13,8}$	No	$d_{13,9}$	No	$d_{13,10}$	$d_{13,11}$	$d_{13,12}$	$d_{13,14}$	$d_{13,15}$
P14	No	$d_{14,1}$	No	$d_{14,2}$	No	$d_{14,3}$	No	$d_{14,4}$	No	$d_{14,5}$	No	$d_{14,6}$	No	$d_{14,7}$	No	$d_{14,8}$	No	$d_{14,9}$	No	$d_{14,10}$	$d_{14,11}$	$d_{14,12}$	$d_{14,13}$	$d_{14,15}$
P15	No	$d_{15,1}$	No	$d_{15,2}$	No	$d_{15,3}$	No	$d_{15,4}$	No	$d_{15,5}$	No	$d_{15,6}$	No	$d_{15,7}$	No	$d_{15,8}$	No	$d_{15,9}$	No	$d_{15,10}$	$d_{15,11}$	$d_{15,12}$	$d_{15,13}$	$d_{15,14}$	

$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$	$d_{1,7}$	$d_{1,8}$	$d_{1,9}$	$d_{1,10}$	$d_{1,11}$	$d_{1,12}$	$d_{1,13}$	$d_{1,14}$	$d_{1,15}$
$d_{2,1}$	$d_{2,3}$	$d_{2,4}$	$d_{2,5}$	$d_{2,6}$	$d_{2,7}$	$d_{2,8}$	$d_{2,9}$	$d_{2,10}$	$d_{2,11}$	$d_{2,12}$	$d_{2,13}$	$d_{2,14}$	$d_{2,15}$
$d_{3,1}$	$d_{3,2}$	$d_{3,4}$	$d_{3,5}$	$d_{3,6}$	$d_{3,7}$	$d_{3,8}$	$d_{3,9}$	$d_{3,10}$	$d_{3,11}$	$d_{3,12}$	$d_{3,13}$	$d_{3,14}$	$d_{3,15}$
$d_{4,1}$	$d_{4,2}$	$d_{4,3}$	$d_{4,5}$	$d_{4,6}$	$d_{4,7}$	$d_{4,8}$	$d_{4,9}$	$d_{4,10}$	$d_{4,11}$	$d_{4,12}$	$d_{4,13}$	$d_{4,14}$	$d_{4,15}$
$d_{5,1}$	$d_{5,2}$	$d_{5,3}$	$d_{5,4}$	$d_{5,6}$	$d_{5,7}$	$d_{5,8}$	$d_{5,9}$	$d_{5,10}$	$d_{5,11}$	$d_{5,12}$	$d_{5,13}$	$d_{5,14}$	$d_{5,15}$
$d_{6,1}$	$d_{6,2}$	$d_{6,3}$	$d_{6,4}$	$d_{6,5}$	$d_{6,7}$	$d_{6,8}$	$d_{6,9}$	$d_{6,10}$	$d_{6,11}$	$d_{6,12}$	$d_{6,13}$	$d_{6,14}$	$d_{6,15}$
$d_{7,1}$	$d_{7,2}$	$d_{7,3}$	$d_{7,4}$	$d_{7,5}$	$d_{7,6}$	$d_{7,8}$	$d_{7,9}$	$d_{7,10}$	$d_{7,11}$	$d_{7,12}$	$d_{7,13}$	$d_{7,14}$	$d_{7,15}$
$d_{8,1}$	$d_{8,2}$	$d_{8,3}$	$d_{8,4}$	$d_{8,5}$	$d_{8,6}$	$d_{8,7}$	$d_{8,9}$	$d_{8,10}$	$d_{8,11}$	$d_{8,12}$	$d_{8,13}$	$d_{8,14}$	$d_{8,15}$
$d_{9,1}$	$d_{9,2}$	$d_{9,3}$	$d_{9,4}$	$d_{9,5}$	$d_{9,6}$	$d_{9,7}$	$d_{9,8}$	$d_{9,10}$	$d_{9,11}$	$d_{9,12}$	$d_{9,13}$	$d_{9,14}$	$d_{9,15}$
$d_{10,1}$	$d_{10,2}$	$d_{10,3}$	$d_{10,4}$	$d_{10,5}$	$d_{10,6}$	$d_{10,7}$	$d_{10,8}$	$d_{10,9}$	$d_{10,11}$	$d_{10,12}$	$d_{10,13}$	$d_{10,14}$	$d_{10,15}$
$d_{11,1}$	$d_{11,2}$	$d_{11,3}$	$d_{11,4}$	$d_{11,5}$	$d_{11,6}$	$d_{11,7}$	$d_{11,8}$	$d_{11,9}$	$d_{11,10}$	$d_{11,12}$	$d_{11,13}$	$d_{11,14}$	$d_{11,15}$
$d_{12,1}$	$d_{12,2}$	$d_{12,3}$	$d_{12,4}$	$d_{12,5}$	$d_{12,6}$	$d_{12,7}$	$d_{12,8}$	$d_{12,9}$	$d_{12,10}$	$d_{12,11}$	$d_{12,13}$	$d_{12,14}$	$d_{12,15}$
$d_{13,1}$	$d_{13,2}$	$d_{13,3}$	$d_{13,4}$	$d_{13,5}$	$d_{13,6}$	$d_{13,7}$	$d_{13,8}$	$d_{13,9}$	$d_{13,10}$	$d_{13,11}$	$d_{13,12}$	$d_{13,14}$	$d_{13,15}$
$d_{14,1}$	$d_{14,2}$	$d_{14,3}$	$d_{14,4}$	$d_{14,5}$	$d_{14,6}$	$d_{14,7}$	$d_{14,8}$	$d_{14,9}$	$d_{14,10}$	$d_{14,11}$	$d_{14,12}$	$d_{14,13}$	$d_{14,15}$
$d_{15,1}$	$d_{15,2}$	$d_{15,3}$	$d_{15,4}$	$d_{15,5}$	$d_{15,6}$	$d_{15,7}$	$d_{15,8}$	$d_{15,9}$	$d_{15,10}$	$d_{15,11}$	$d_{15,12}$	$d_{15,13}$	$d_{15,14}$	

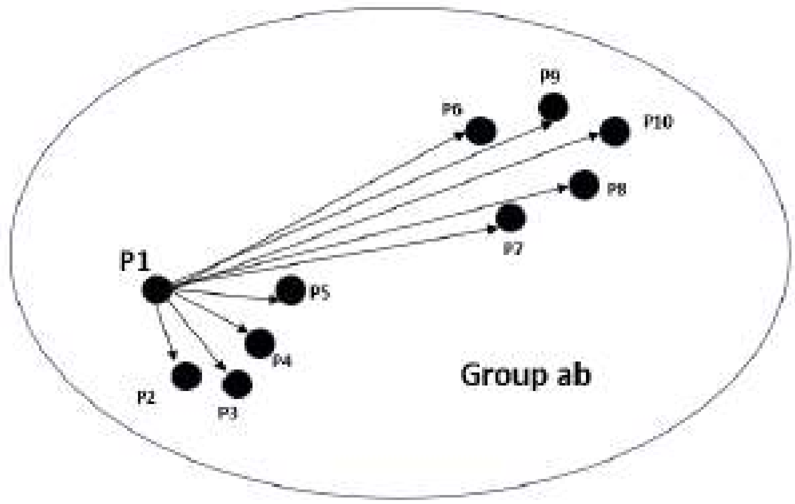


Figure 34 Group AB regrouping - P1

Repeat these steps for the Cluster ab until each point in the cluster is nearer to each other by less than the Cluster Farthest Points Distance and has nearest neighbors equivalent to or greater than the **Nearest neighbor count.**

- Choose randomly any one point from the large cluster ab and determine how many points are close to this point.
- Randomly chosen point is P1, compute the distance from P1 to all the points in the Cluster ab.
- d1,n = SQRT [(xn x1)2 +(yn - y1)2]
- where 'n' is the remaining points of the Cluster ab.
- If the distance from P1 to other points in cluster ab is less than the Cluster Farthest Points Distance, then set that point as the Nearest Neighbor of P1.
- Form a new Cluster A that is set as the Nearest neighbors of P1.

Points	Distance to P2	Nearest Neighbour of P1	Distance to P3	Nearest Neighbour of P1	Distance to P4	Nearest Neighbour of P1	Distance to P5	Nearest Neighbour of P1	Distance to P6	Nearest Neighbour of P1	Distance to P7	Nearest Neighbour of P1	Distance to P8	Nearest Neighbour of P1	Distance to P9	Nearest Neighbour of P1	Distance to P10	Nearest Neighbour of P1
P1	$d_{1,2}$	Yes	$d_{1,3}$	Yes	$d_{1,4}$	Yes	$d_{1,5}$	Yes	$d_{1,6}$	No	$d_{1,7}$	No	$d_{1,8}$	No	$d_{1,9}$	No	$d_{1,10}$	No

Table 8
P1 - Closest & Final Grouping - Cluster 1

- Determine whether the total points in the new Cluster A are greater than or equal to the **Nearest neighbor count.**

Repeat the same steps for the remaining points of Cluster AB.

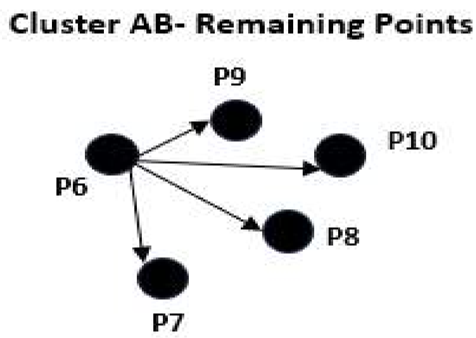


Figure 35
P6 Cluster

Figure 36 depicts the remaining points of the cluster AB.

- Randomly choose the new point from the cluster AB
- Randomly choose point P6, and determine the distance from this point to the rest of the points in the cluster AB.
- Determine the total points that are close to P6 are greater than or equal to the **Nearest neighbor count.** Then form this as a new cluster B.

Table 9
Closest & Final Grouping - Cluster 2

Points	Distance to P7	Nearest Neighbour of P6	Distance to P8	Nearest Neighbour of P6	Distance to P9	Nearest Neighbour of P6	Distance to P10	Nearest Neighbour of P6
P6	$\mathbf{d_{6,7}}$	Yes	$\mathbf{d_{6,8}}$	Yes	$\mathbf{d_{6,9}}$	Yes	$\mathbf{d_{6,10}}$	Yes

- Repeat these steps 10 times to identify the existence of 10 different clusters.
- It's assumed that for the chosen country based on EDA, there cannot be more than 10 active spreading wildfire points at different locations in the selected state.
- If there are more than 5 clusters formed for the selected state, then drop the clusters from the results if any cluster has data points less than 3% of the data points in the dataset.

Determine the area and vertices of each cluster

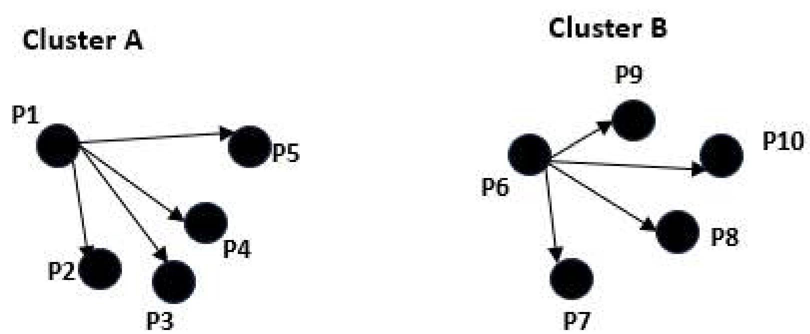


Figure 36
New model Predicted - Final Clusters.

- Clusters are assumed to be convex hulls.

- Vertices of the convex hull are determined using the algorithm Graham Scan or Jarvis March (Gift Wrapping).
- The area of the convex hull is computed from Gauss's Shoelace Formula.

  Area = 0.5 * abs(sum (xi * (yi+1 - yi-1) for i in range(n))). Here, where (xi, yi) are the coordinates of the vertices in counterclockwise order.

Specify the distance between the vertices of the clusters

Forms and other bigger clusters consist of small clusters where any of the vertices of Cluster A is less than the specified inter-distance from Cluster B.

Determine the distance from the cluster to the nearest Residence area.

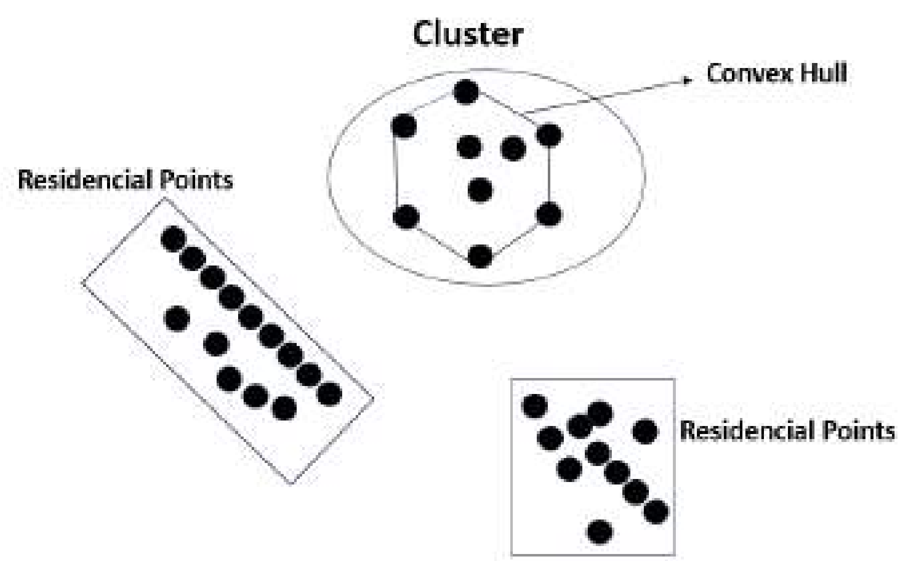


Figure 37
Cluster and Residential Proximity Computation

- Compute the Euclidean distance between the six vertices of the identified clusters. Convex hulls with the data points in the Census and Neighbourhood dataset of the Selected region.

  dv,n = SQRT [(xn xv)2 +(yn – yv)2]

  xv, yv – ‘v’ refers to 6 vertices, varies from 1 to 6 of the convex hulls.

Where ‘n’ refers to all the datapoints that correspond to the selected region in the Census and Neighborhood dataset.

Inter Distance, Intra distance, fixed density Clustering - Evaluation

The following evaluation metrics are used for evaluating this Clustering with real wildfire incidents.

- Visual Inspection using the Plot

CHAPTER IV:

RESULTS

4.1 Research Question One

Are there any hidden patterns of growing fires in the collected raw data from the history dataset captured from the satellite?

In the Data Analysis section, a comprehensive method for analyzing historical data is outlined, followed by its application to real-time scenarios. Understanding historical data is crucial, as it enables the identification of constraints and patterns that correlate with actual incidents. This analysis provides a foundation for assessing how these patterns may behave in real-time situations. The process begins with testing AI methods on historical datasets where the outcomes of past incidents are already known. If the AI demonstrates accurate predictions based on this historical data, the learned configurations can then be applied to real-time data with similar input settings. However, it is essential to note that the insights gained from historical data in a specific region may not apply to all areas. In such cases, AI methods must be configured with region-specific learnings derived from the historical data relevant to that locale. Additionally, the use of different instruments for data capture can lead to variations in data resolution. Consequently, the learnings from one instrument's historical data may differ from those of another. To address this, the configuration of AI methods must take into account the specific instrument used for data collection. The same configurations that yielded accurate results with historical data should then be replicated for real-time predictions.

Firstly, understand the history of fire event data from the MODIS instrument over the past 5 years for the USA.

Table 10
Raw Fire events 2020 to 2024 - USA - MODIS

YEAR	Total Raw Fire Events
2020	153916
2021	163735
2022	130445
2023	94000
2024	123652

Table 11
Total Raw Fire events - Top 5 states, USA, 2020 to 2024

Year 202	Oregon	Idaho	Californi	Texas	Georgia
Raw Fire	14702	11566	11096	6788	6775
Year 202	California	Texas	Georgia	Florida	Louisiana
Raw Fire	7685	7581	7269	7083	5026
Year 202	Alaska	Texas	w Hampsl	Georgia	Florida
Raw Fire	24045	9782	8815	8057	7332
Year 202	California	Washingto	Oregon	Texas	Idaho
Raw Fire	45589	11472	11332	8398	8325
Year 202	California	Oregon	Texas	Florida	Georgia
Raw Fire	50155	12156	8466	6566	6495

Table 12
Total Raw Fire events - Monthly Top 5 State USA - 2020 to 2024

Year	State	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2023	California	117	311	141	467	249	393	560	2046	1114	1101	820	366
2024	California	137	83	291	318	327	803	4836	1369	1404	738	603	187
2023	Washington	6	17	128	151	125	59	391	678	227	454	145	20
2024	Washington	2	8	133	67	66	139	1936	1029	204	221	171	21
2023	Idaho	0	1	14	69	178	47	624	482	311	432	285	9
2024	Idaho	2	2	64	118	193	194	536	4344	3572	2284	229	28
2023	Oregon	42	42	15	64	179	61	822	1343	460	570	592	167
2024	Oregon	21	22	81	62	219	163	7214	2210	3384	643	421	262
2023	Texas	862	768	724	509	529	387	354	530	446	830	875	767
2024	Texas	575	1431	747	466	317	376	371	428	589	492	521	475
2023	Georgia	704	828	1762	528	295	183	198	260	400	840	712	55
2024	Georgia	636	1572	1260	612	169	184	130	328	226	753	285	620
2023	Florida	1051	1472	1152	375	528	475	232	315	255	414	447	367
2024	Florida	428	1368	823	891	387	269	172	201	111	332	472	508
2023	Pennsylvania	0	9	42	73	104	60	105	87	70	37	37	0
2024	Pennsylvania	1	13	46	131	134	63	86	44	69	35	17	1

Table 13
Total Raw Fire events - Monthly Top 5 states, USA - 2025

Year	State	JAN	FEB
2025	California	867	86
2025	Washington	10	0
2025	Idaho	1	0
2025	Oregon	81	1
2025	Texas	403	476
2025	Georgia	696	613
2025	Florida	765	600
2025	Pennsylvania	1	5

Table 10 lists the total raw fire events in the MODIS dataset in the past 5 years in the USA. Table 11 lists the top 5 states in the USA that had the highest fire events in the past 5 years based on the MODIS dataset. Tables 12 and 13 illustrate the seasonal spread of raw fire events across different geographical locations in the USA for the years 2025, 2024, and 2023. It provides insights into how fire incidents vary by season in various regions, highlighting trends and patterns throughout the year. Here's a breakdown of the geographical locations of raw fire events based on seasons:

- South West: California (Coastal)
- North West:
  a) Oregon (Coastal)
  b) Washington (Coastal)
  c) Idaho (Non-Coastal)
- South East:
  a) Georgia (Coastal)
  b) Florida (Coastal)
- Northeast: Pennsylvania (Non-Coastal)
- Northeast

This classification helps in understanding the distribution of fire events across various regions and their seasonal patterns.

Based on the historical data analysis of MODIS raw fire events, it was found that

- States in the USA show a seasonal pattern
- Western states experience growing wildfires from June to November
- The eastern part of the USA experiences wildfires from January to May

Also, by referring to raw fire events of the peak months in 2D-plots in the following figures, that was discussed in detail in the data analysis section,

- Figure 25 Florida Map – Fire events of Feb Mar 2023 and Feb Mar 2024
- Figure 23 Georgia Map - Fire events of Feb Mar 2023 and Feb Mar 2024
- Figure 21 Texas Map – Fire events of Feb Mar 2023 and Feb Mar 2024

States in the eastern part of the USA did not exhibit fire-spreading characteristics in their monthly depictions of peak months. This also indicates that wildfire spreading characteristics are more prevalent in the higher longitudinal region of the USA. In contrast, they are less observed or rarely observed in the lower longitudinal region of the USA, with no impact across the different latitudes of the USA.

California and Idaho have been selected for further AI experimentation, as these states are among the top five states that have experienced significant fire events over the past five years. They are also located in both coastal and non-coastal areas in the western United States.

Additionally, in California, data is available from the California Current Emergency Incidents organization, which lists all actual wildfire events in the state. This data is available for the past 10 years, which is used to validate the actual wildfire incidents against the machine learning model's predicted wildfire growth events.

Table 14
The 2025 deadliest wildfire events in California

incident_name	incident_date_created	incident_dateonly_extinguished	incident_acres_burned	incident_longitude	incident_latitude	incident_id
Palisades Fire	2025-01-07T10:30:00Z	31-01-2025	23448	-118.545	34.0702	abfed7a3-794a-4294-9852-43176dcbe18a
Eaton Fire	2025-01-07T18:18:00Z	31-01-2025	14021	-118.069	34.2035	dbfa574e-1b10-4467-b5e5-7d1e06cbcfde
Hughes Fire	2025-01-22T10:53:02Z		10425	-118.567	34.553	bf194e6a-78cb-46be-840b-567bceaabbb6
Border 2 Fire	2025-01-23T13:58:00Z		6625	-116.844	32.5947	a9719e0a-db5d-48d4-895c-5b096ae1802d
Silver Fire	2025-03-30T14:11:55Z		1611	-118.353	37.3997	014151c4-c77d-4fe5-9ffc-d90909ad8f4b

Table 14 represents the deadliest wildfire events in California in early 2025. The Palisades Fire, Eaton Fire, and Hughes Fire incident dates are chosen to understand any hidden patterns of growing fires in the MODIS raw dataset on the day of discovery and their progression on the second day.

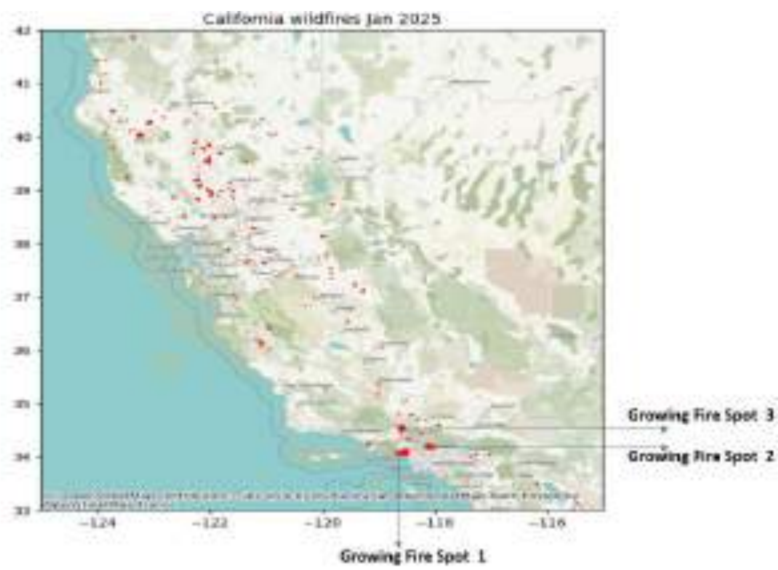


Figure 38
Wildfires pattern - JAN CA 2025

Figure 39 illustrates all the raw fire events captured by the MODIS instrument for January 2025 in state California, highlighting three significant fire incidents that spread to larger areas.

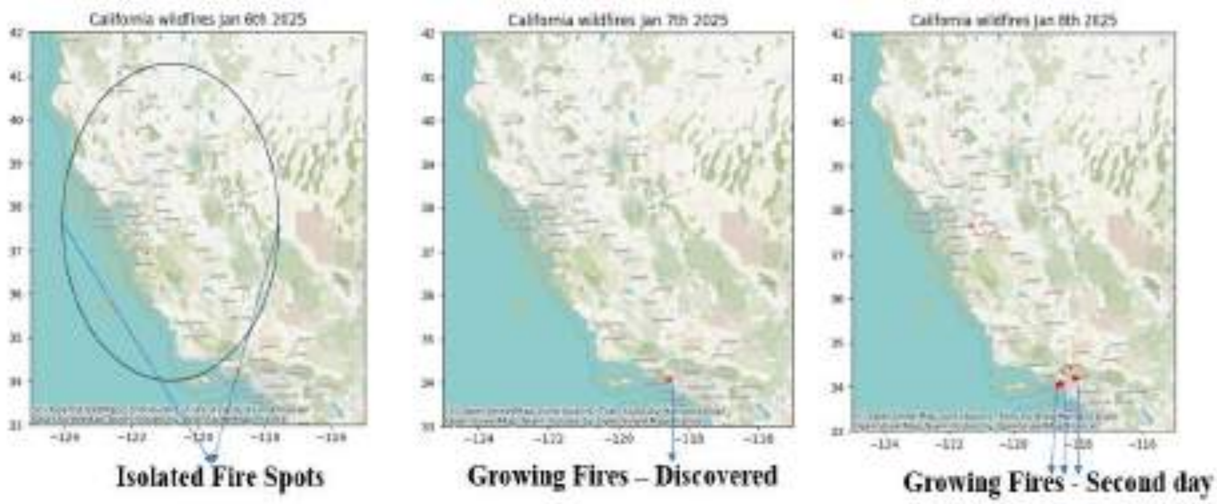


Figure 39
Growing fire and Isolated fires - CA JAN 2025 Incident One

Figure 40 illustrates raw fire event data as observed by the MODIS instrument of the Palisades Fire and Eaten Fire, referred to in Table 14, which shows a small dense pattern on the day it was discovered, and expanded to a larger area on the second day. On the previous day, only isolated fire incidents were recorded, characterized by small fire spots. The last day's raw data was also plotted to ensure that the actual wildfire discovery dates in the manually collected dataset are accurate.

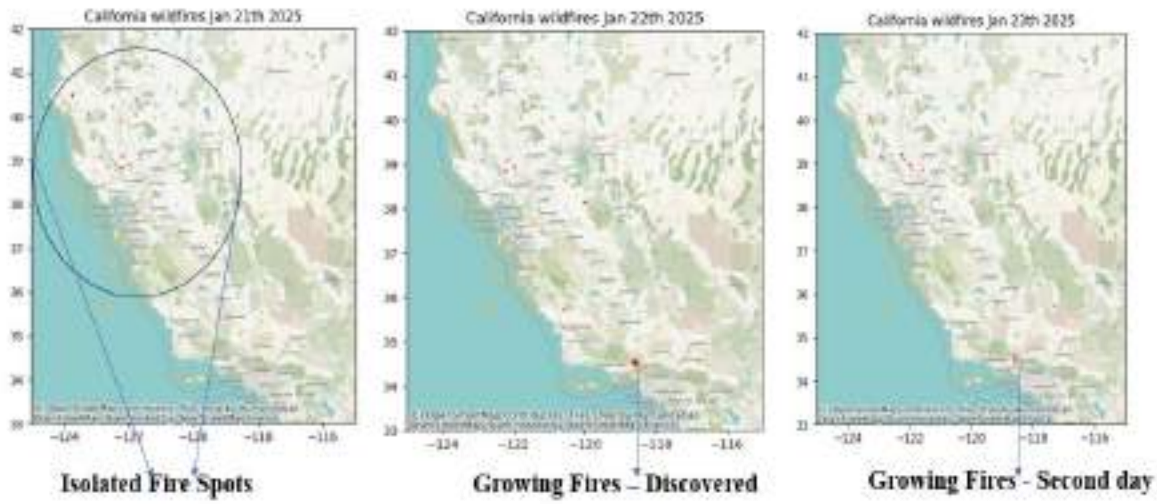


Figure 40
Growing fire and Isolated fires - CA JAN 2025 Incident Two

Figure 41 illustrates raw fire event data as observed by the MODIS instrument of the Hughes Fire in the state of California, referred to in Table 14, which shows a small dense pattern on the day it was discovered. However, this fire did not show growth on the following day, which may be attributed to effective control measures implemented by fire management or may have occurred due to natural suppression influenced by weather phenomena.

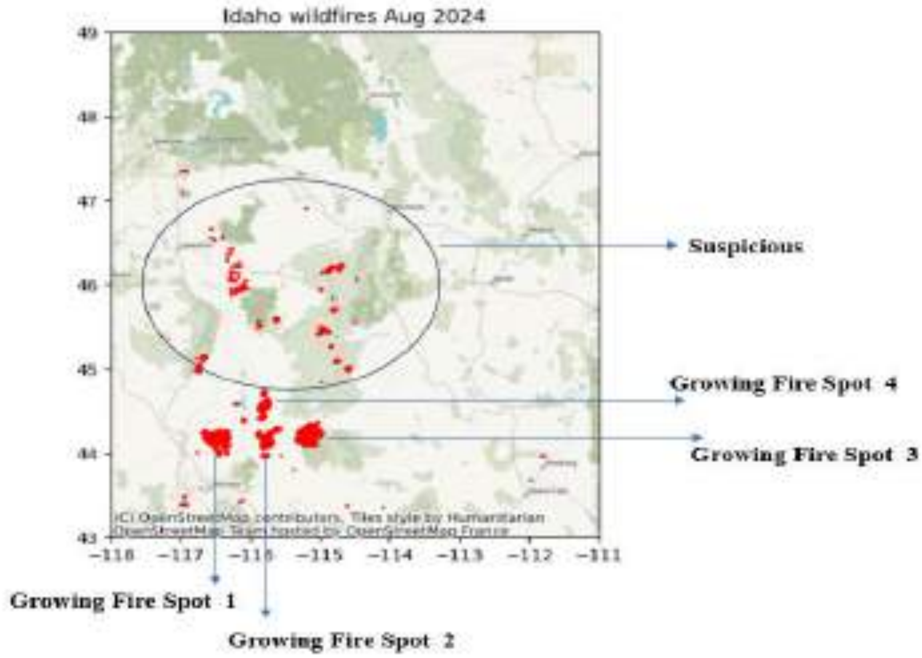


Figure 41
Wildfires pattern - AUG Idaho 2025

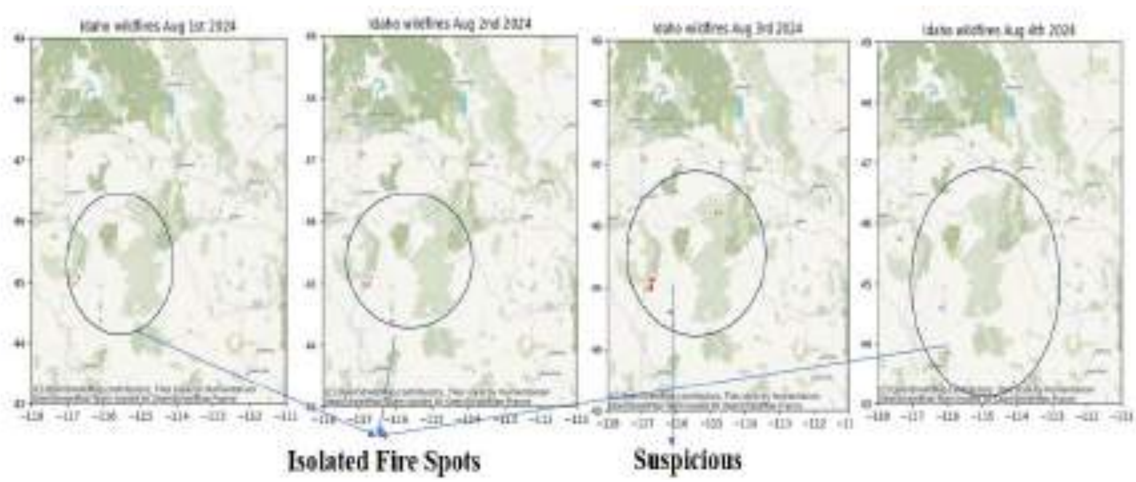


Figure 42
Time Series Raw Fire events MODIS AUG 1 to 4 – 2024

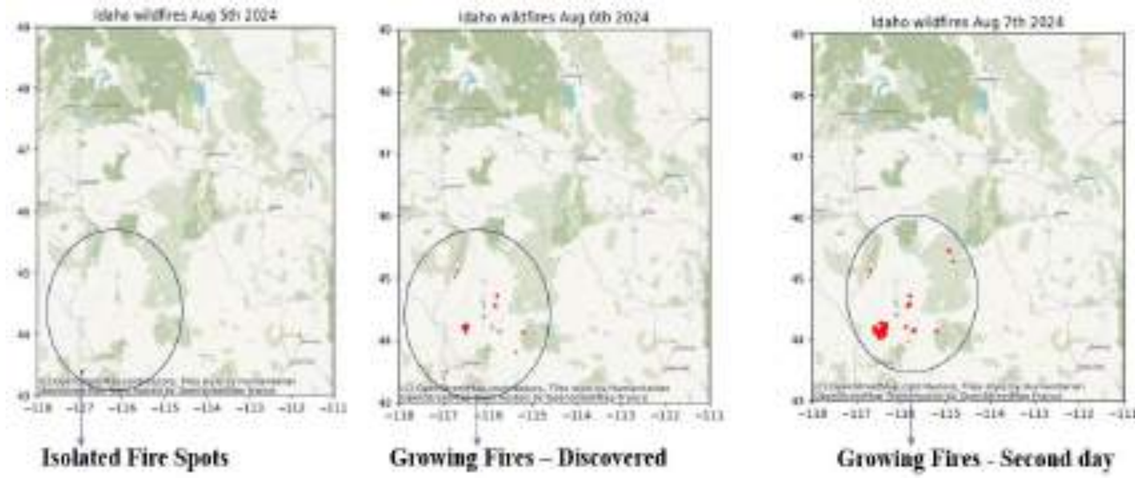


Figure 43
Time Series Raw Fire events MODIS AUG 5 to 7, 2024

Figure 42 illustrates growing wildfire patterns in the state of Idaho for the peak month of August. Visually, this pattern aligns with the real wildfire incident pattern in California, as depicted in the Figure. 39.

For the state of Idaho, there is no manually collected data available on the real wildfire incidents. Hence, MODIS raw fire events for each day from the beginning of August are depicted in a separate 2D map in Figures 43 and 44.

On the 3rd day of August, although it visually appeared to be a small, dense fire, it did not grow the next day, so it was discarded for AI model experimentation.
On August 6, growing fires were discovered, exhibiting a similar pattern to that observed in the 2D plot of the California wildfire on the day it was found, Figures 40 and 41.

4.2 Research Question Two

Which machine algorithm is accurate in predicting the fire growth and eliminating the non-growing fires from the raw fire events dataset on the day of discovery?

Unsupervised clustering machine learning algorithms are primarily utilized for unlabeled data, and several algorithms have been experimented with in this context:

K-Means Clustering: A widely used method that partitions data into a specified number of clusters based on the mean distance between points.

Fuzzy C-Means Clustering: This approach allows data points to belong to multiple clusters with varying degrees of membership, accommodating uncertainty in cluster assignments.

Gaussian Mixture Models (GMM): A probabilistic model that assumes data points are generated from a mixture of several Gaussian distributions, providing a flexible approach to clustering.

Agglomerative Hierarchical Clustering: A method that builds clusters iteratively by merging smaller clusters into larger ones based on distance measures.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise): This algorithm identifies clusters based on the density of data points, effectively distinguishing noise from meaningful clusters.

In addition to these established methods, a new machine learning algorithm, the Multi-Level Multi-Criteria Clustering Algorithm, has been developed. This innovative model showcases the substantial applications of AI in analyzing complex data. It retains all the capabilities of DBSCAN while introducing additional features that enhance its ability to predict contextual information related to growing fire scenes. The development of this new model represents a significant advancement in clustering techniques, aiming to improve the accuracy and relevance of predictions in fire management and other applications.

4.2.1 K-Means Clustering

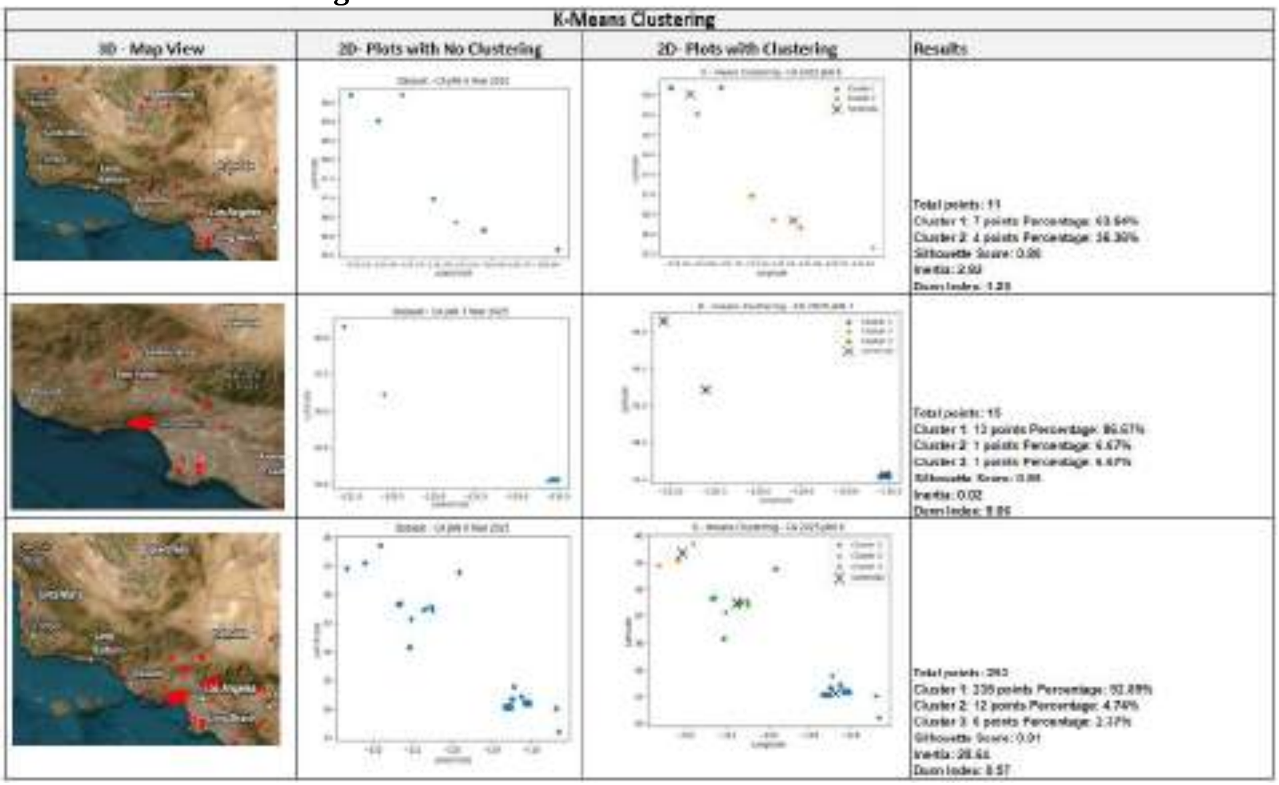


Figure 44
K-Means Clustering Results -Sample Data 1

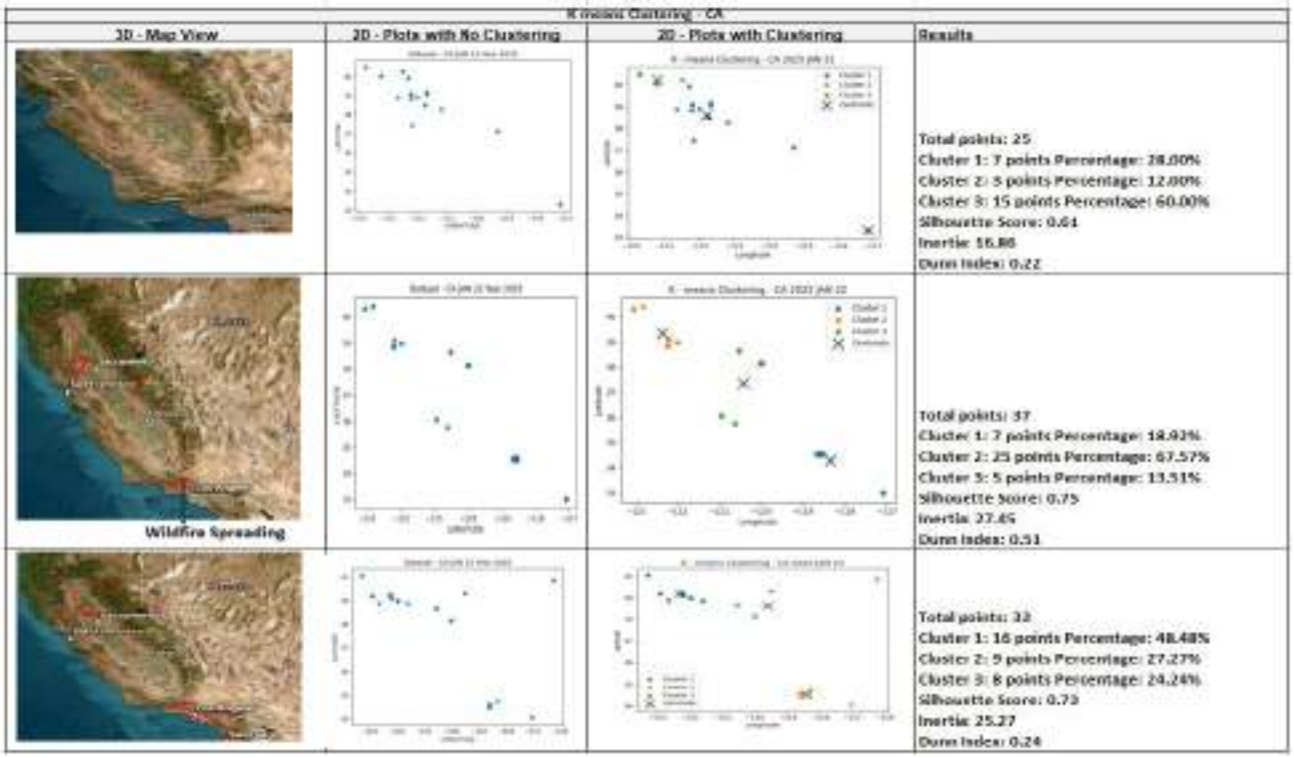


Figure 45
K-Means Clustering Results -Sample Data 2

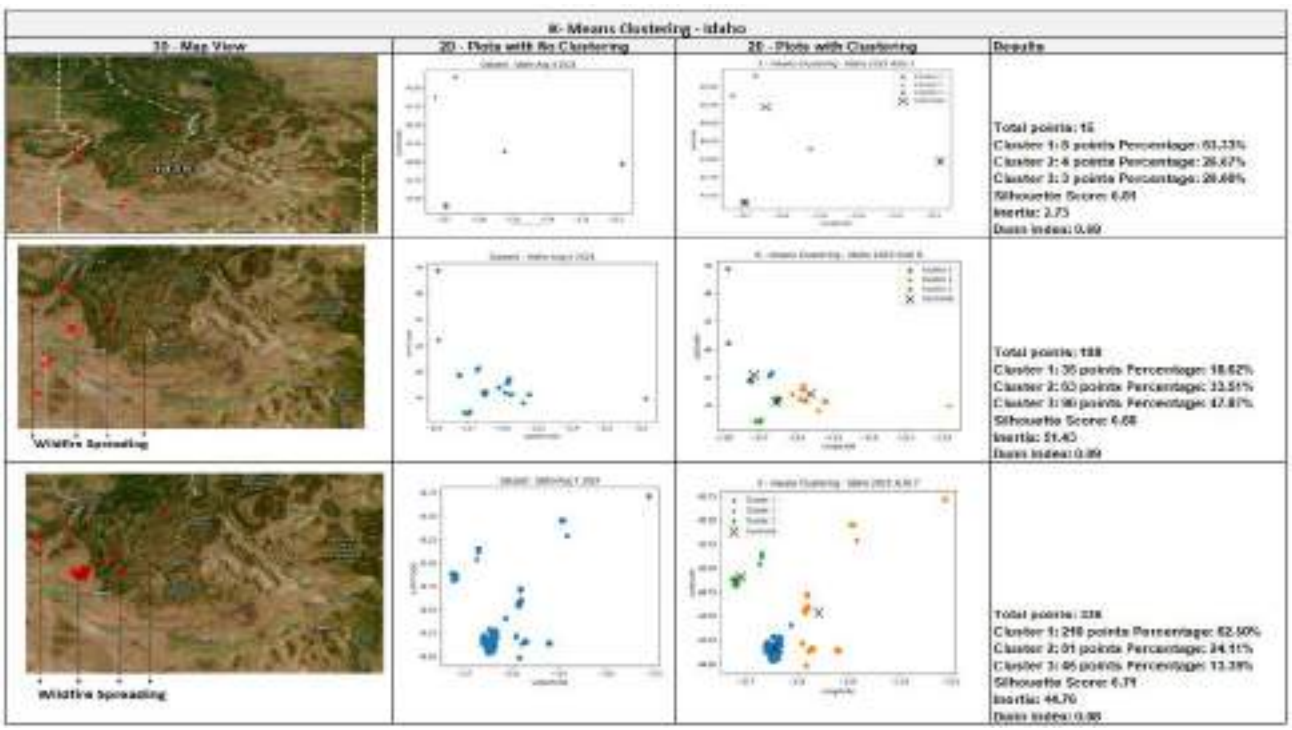


Figure 46
K-Means Clustering Results - Sample Data 3

<u>K-means Results Interpretation:</u>

Visual Inspection:

3D Map view:

Depicts all the fire events that occurred for the selected date in the selected region.

Figure 45: Represents the January 6, 7, and 8, 2025 fire events of California.

Figure 46: Represents the January 21, 22, 23, 2025 fire events of California.

Figure 47: Represents the August 5, 6, and 7, 2024 fire events of Idaho.

2D plot with no clustering:

Follows the exact depiction as specified in the 3D Map View,

The X axis represents the Longitude, and the Y axis represents the Latitude of the fire events.

This Plot is utilized for the Visual Inception by comparing it with the resultant clusters from the K-means prediction.

2D plot with clustering and Results:

Figures 45, 46, 47: K-Means results are depicted as follows

January 6, 2025 dataset (Fig. 45):

- K-Means incorrectly clusters 63.64% of non-spreading fire events in cluster 1.
- K-Means incorrectly clusters 36.36% of non-spreading fire events in cluster 2.

January 7, 2025 dataset (Fig. 45) Incident Day:

- K-Means accurately clusters 86.67% of spreading fire events in cluster 1.
- K-Means incorrectly clusters 6.67% of non-spreading fire events to cluster 2.
- K-Means incorrectly clusters 6.67% of non-spreading fire events to cluster 3.

January 8, 2025 dataset (Fig. 45):

- K-Means almost accurately clusters 89% of spreading fire events in cluster 1, although a few points in this cluster are far away, appearing much isolated from the actual fire growth area.
- K-Means incorrectly clusters 4.74% of non-spreading fire events to cluster 2.
- K-Means incorrectly clusters 2.37% of non-spreading fire events to cluster 3.

January 21, 2025 dataset (Fig. 46):

- K-Means incorrectly clusters 28% of non-spreading fire events in cluster 1.
- K-Means incorrectly clusters 12% of non-spreading fire events in cluster 2.
- K-Means incorrectly clusters 60% of non-spreading fire events in cluster 3.

January 22, 2025 dataset (Fig. 46) Incident Day:

- K-Means almost accurately clusters 18.92% of spreading fire events to cluster 1. Few points in this cluster are far away and do not appear to be spreading spots.
- K-Means incorrectly clusters 67.57% of non-spreading fire events in cluster 2.
- K-Means incorrectly clusters 13.51% of non-spreading fire events in cluster 3.

January 23, 2025 dataset (Fig. 46):

- K-Means almost accurately clusters 48.48% of spreading fire events to cluster 1. Few points in this cluster are far away and do not appear to be spreading spots.
- K-Means incorrectly clusters 27.27% of non-spreading fire events in cluster 2.
- K- K-Means incorrectly clusters 24.24% of non-spreading fire events to cluster 3.

August 5, 2024 dataset (Fig. 47):

- K-Means incorrectly clusters 53.33% of non-spreading fire events to cluster 1.
- K-Means incorrectly clusters 26.67% of non-spreading fire events to cluster 2.
- K-Means incorrectly clusters 20% of non-spreading fire events in cluster 3.

August 6, 2024 dataset (Fig. 47) Incident Day:

- K-Means almost accurately clusters 18.62% of spreading fire events to cluster 1. Few points in this cluster are far away and do not appear to be spreading spots.
- K-Means almost accurately clusters 33.51% of spreading fire events to cluster 2. Few points in this cluster are far away and do not appear to be spreading spots.
- K-Means almost accurately clusters 47.87% of spreading fire events to cluster 3.

August 7, 2024 dataset (Fig. 47):

- K-Means almost accurately clusters 62.50% of spreading fire events to cluster 1.
- K-Means almost accurately clusters 24.11% of spreading fire events to cluster 2. Few points in this cluster are far away and do not appear to be spreading spots.
- K-Means almost accurately clusters 13.39% of spreading fire events to cluster 3. Few points in this cluster are far away and do not appear to be spreading spots.

Silhouette score:

The silhouette score is notably high for the K-Means clustering results for the sample dataset on the day of the incident, as well as the second day following the incident. Interestingly, this high score is also observed for the dataset from the day before the incident, although this particular dataset is largely noisy and lacks dense fire events.

This observation raises a significant concern: the silhouette score may not be an effective measure when clusters do not contain a substantial number of dense events. In datasets characterized by noise, such as those with sparse fire events, the silhouette score might misleadingly suggest a good clustering outcome, even when the clusters are not truly meaningful. This highlights a critical need for more robust evaluation metrics or adjusted clustering approaches that can better handle noisy datasets and provide reliable insights, particularly in scenarios where dense fire events are essential for practical analysis and prediction.

Inertia:

In the sample dataset for the Incident Day, the inertia value is notably low, indicating that the clusters formed through K-Means clustering are compact and well-defined. This suggests that the clustering is effective for this particular dataset. However, for the other sample datasets, the inertia values are high, which is primarily attributed to the presence of outliers. High inertia in these cases indicates that the data points are more dispersed within the clusters, suggesting that the K-Means clustering results may not be accurate or meaningful. The presence of outliers can significantly impact the clustering process, as K-Means is sensitive to such anomalies. Therefore, when interpreting the clustering results, it is essential to consider the effect of outliers and explore potential preprocessing steps, such as outlier removal or alternative clustering methods, to improve the accuracy of clustering for datasets that exhibit such characteristics.

Dunn Index:

A higher Dunn Index is desirable as it indicates better clustering quality, with well-separated and compact clusters. In this case, the Dunn Index is high specifically for the Sample Data 1 (Incident Day) dataset clusters, suggesting that the K-Means clustering performed well in distinguishing these clusters. Conversely, for all other sample datasets (Sample Data 2 and Sample Data 3), the Dunn Index is low. This low score indicates that the K-Means clusters for these datasets lack separation and compactness, suggesting that the clustering results are not accurate.

Limitations of K-Means:

The K-Means clustering algorithm requires further manual intervention to accurately differentiate between clusters representing spreading fire events and those representing non-spreading fire events. Unfortunately, it tends to incorrectly merge normal days and non-spreading fire events into a single cluster. On the actual incident day, the algorithm also mistakenly includes some distant fire points as part of the same spreading fire cluster. This misclassification of outliers can lead to inaccuracies in computing the vertices or border area of the cluster, potentially resulting in incorrect interpretations of the fire scene. Additionally, the performance metrics for this type of dataset show a lower Dunn Index and higher inertia, further indicating that K-Means clustering is not well-suited for such scenarios. The low Dunn Index suggests poor separation between clusters, while the high inertia indicates that data points are widely dispersed within the clusters. These factors underscore the challenges encountered when applying K-Means clustering to datasets with complex patterns and outliers, highlighting the need for alternative clustering approaches or more sophisticated preprocessing techniques to achieve accurate and meaningful results.

4.2.2 Fuzzy C Means Clustering

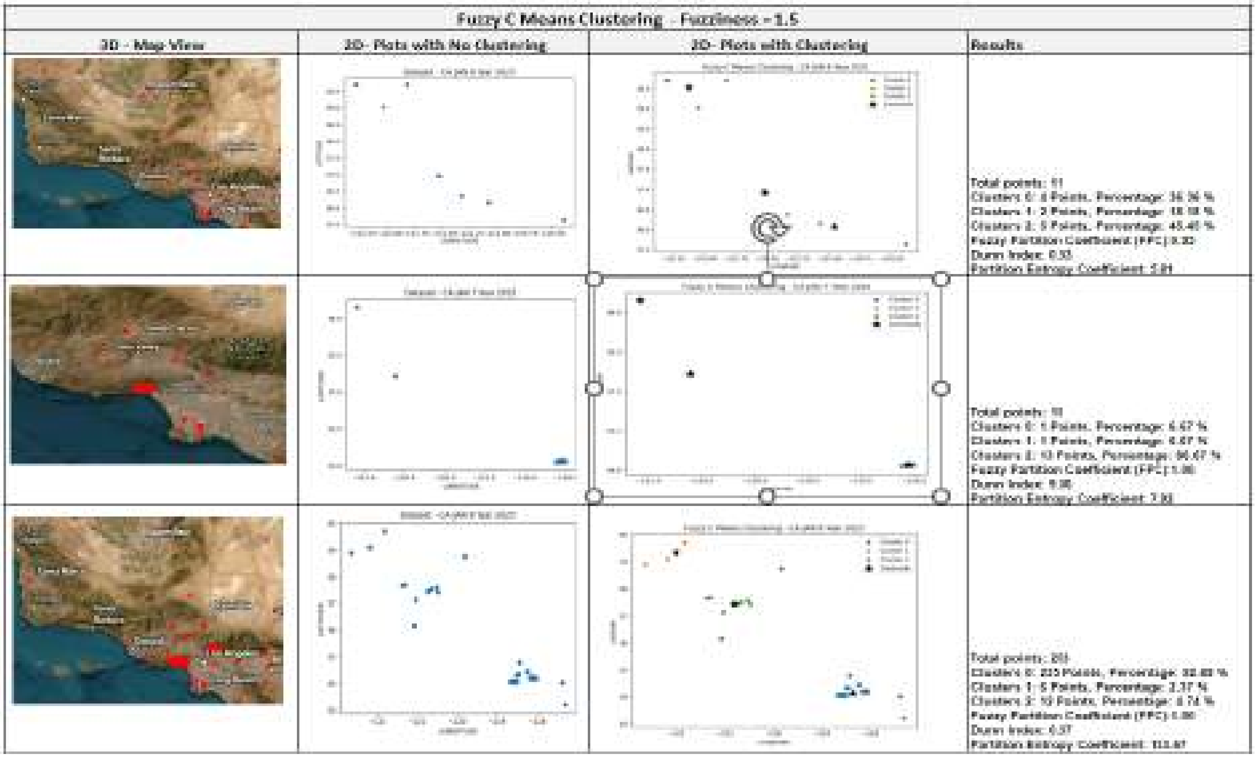


Figure 47

Fuzzy C-means Clustering Results Fuzziness 1.5 - Sample Data 1

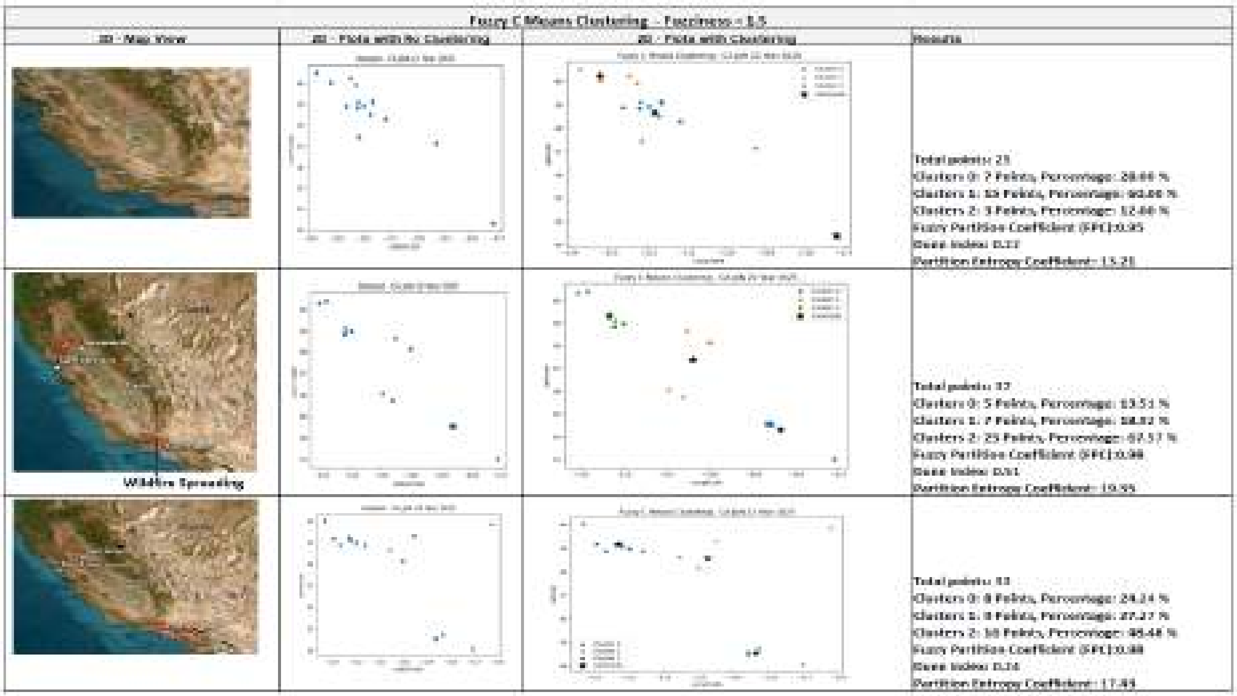


Figure 48
Fuzzy C-means Clustering Results Fuzziness 1.5 - Sample Data 2

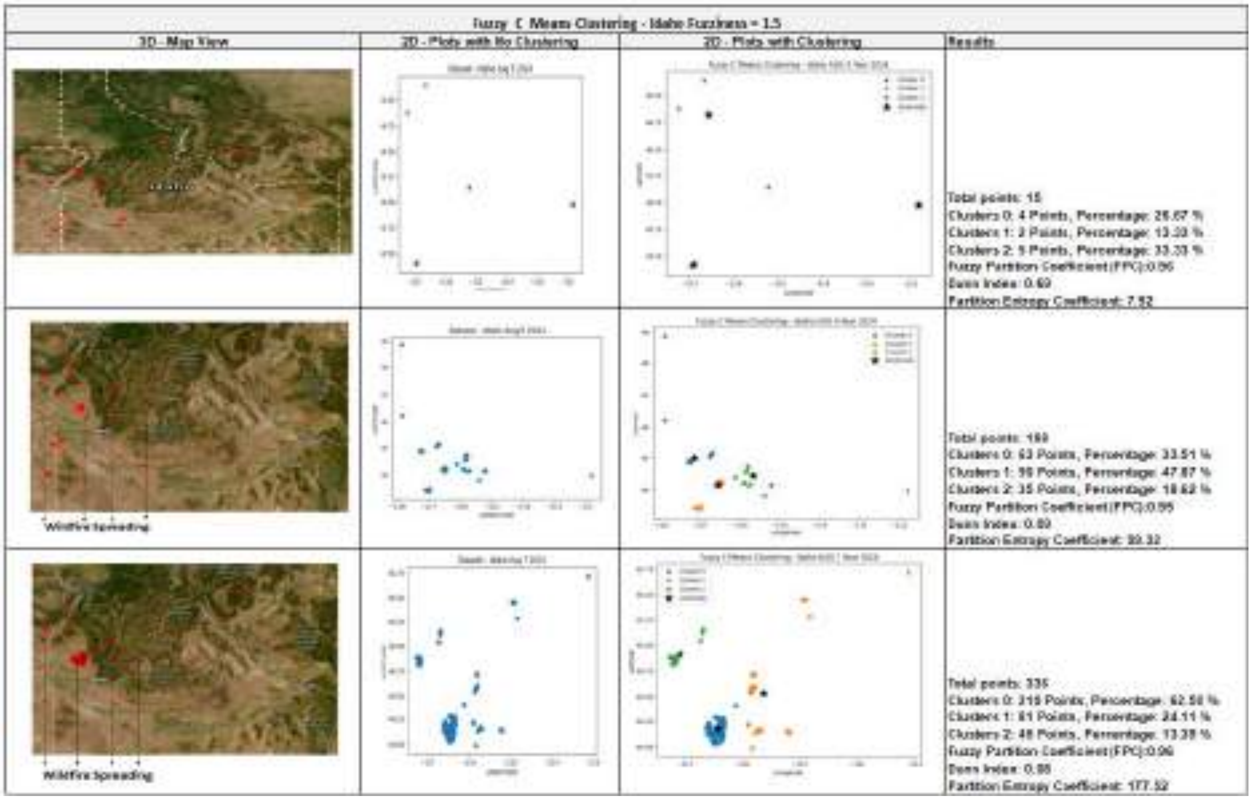


Figure 49
Fuzzy C-means Clustering Results Fuzziness 1.5 - Sample Data 3

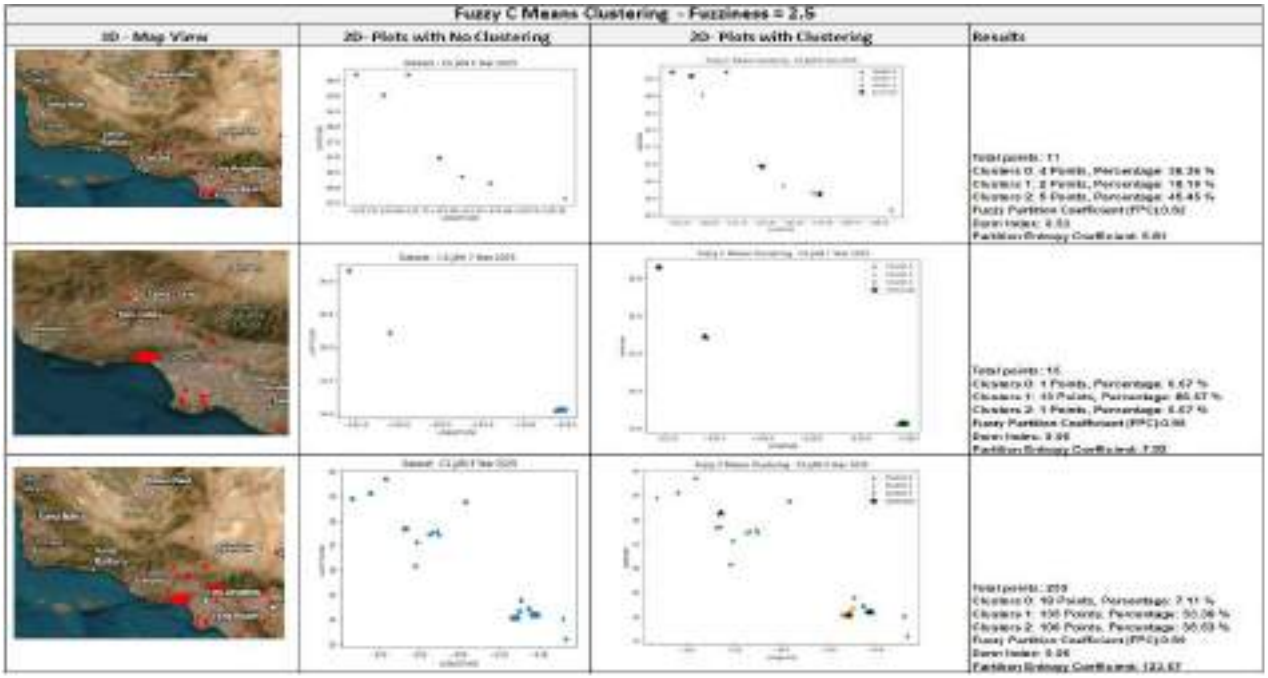


Figure 50
Fuzzy C-means Clustering Results Fuzziness 2.5 - Sample Data 1

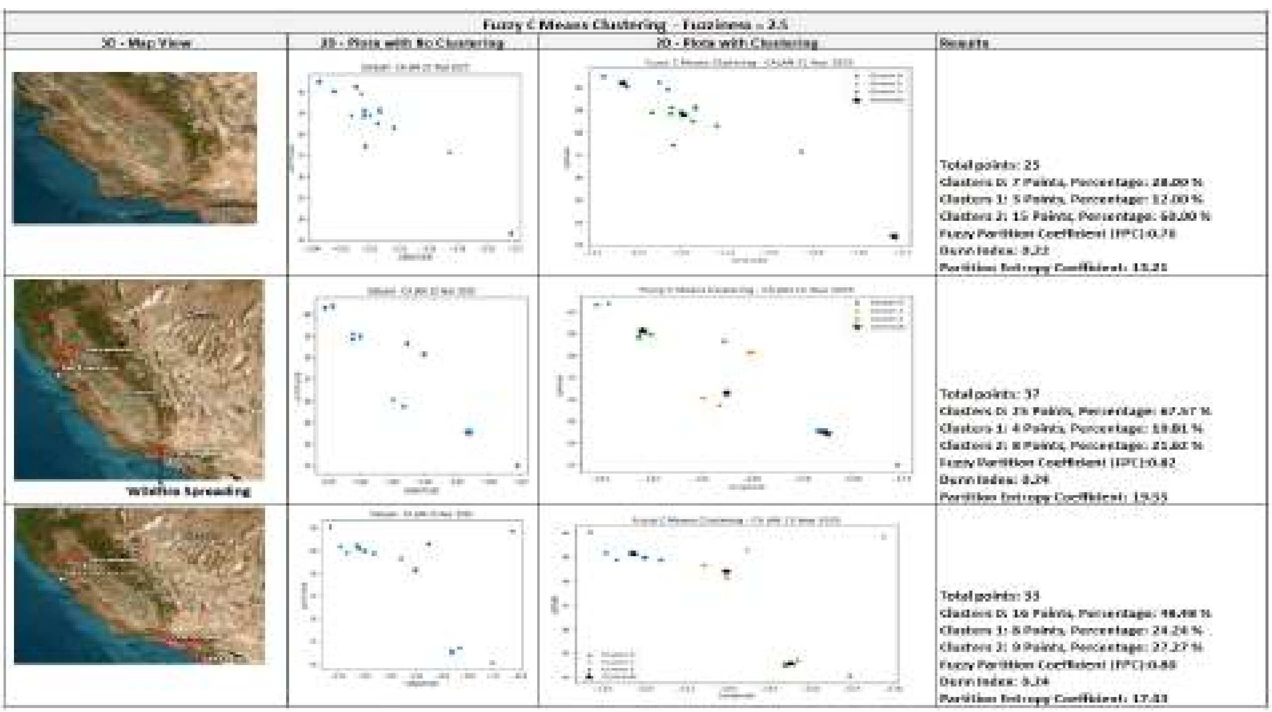


Figure 51
Fuzzy C-means Clustering Results Fuzziness 2.5 - Sample Data 2

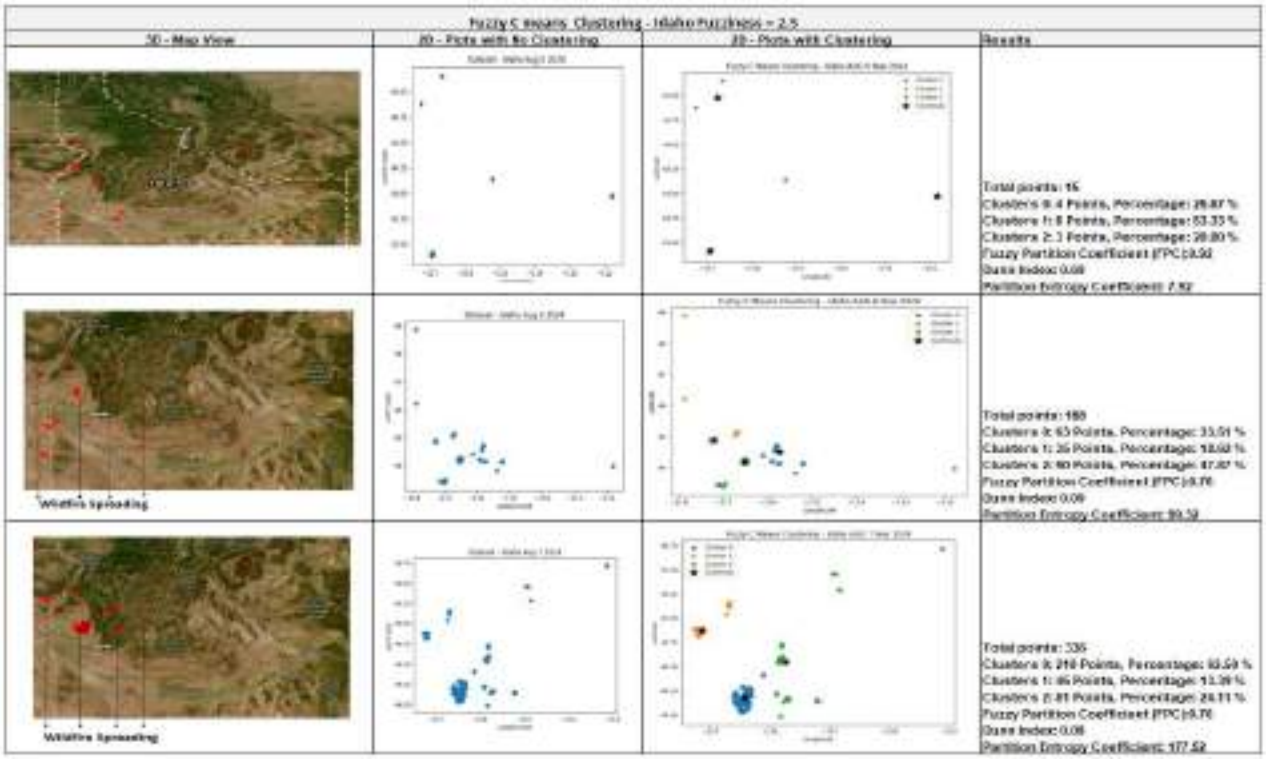


Figure 52
Fuzzy C-means Clustering Results Fuzziness 2.5 - Sample Data 3

Fuzzy C Means Results Interpretation:

Visual Inspection

3D Map view:

Depicts all the fire events that occurred for the selected date in the selected region.

Figure 48, 51: Represents the January 6, 7, 8, 2025 fire events of California.

Figure 49, 52: Represents the January 21, 22, 23, 2025 fire events of California.

Figure 50, 53: Represents the August 5, 6, 7, 2024 fire events of Idaho.

2D plot with no clustering:

Follows the exact depiction as specified in the 3D Map View,

The X axis represents the Longitude, and the Y axis represents the Latitude of the fire events.

This Plot is utilized for the Visual Inception by comparing it with the resultant clusters from the K-means prediction.

2D plot with clustering and Results:

Figures 48, 49, 50, 51, 52, 53: Fuzzy C Means results are depicted as follows

January 6, 2025 dataset (Fig. 48, 51):

- Fuzzy C Means incorrectly clusters 36.36% of non-spreading fire events to cluster 0 for Fuzziness = 1.5 and 2.5.
- Fuzzy C Means incorrectly clusters 18.18% of non-spreading fire events into the cluster for Fuzziness = 1.5 and 2.5.
- Fuzzy C Means incorrectly clusters 45.45% of non-spreading fire events to cluster 2 for Fuzziness = 1.5 and 2.5.
- Fuzzy Partition coefficient 0.95 for Fuzziness = 1.5
- Fuzzy Partition coefficient 0.82 for Fuzziness = 2.5

No change to cluster distribution based on the fuzziness in this dataset.

January 7, 2025 dataset (Fig. 48, 51) Incident Day:

- Fuzzy C Means incorrectly clusters 6.67% of non-spreading fire events to cluster 0 for Fuzziness = 1.5.
- Fuzzy C Means incorrectly clusters 6.67% of non-spreading fire events to cluster 1 for Fuzziness = 1.5.
- Fuzzy C Means clusters accurately cluster 86.67% of spreading fire events to cluster 2 for Fuzziness = 1.5.
- Fuzzy C Means incorrectly clusters 6.67% of non-spreading fire events to cluster 0 for Fuzziness = 2.5.
- Fuzzy C Means accurately clusters 86.67% of spreading fire events to cluster 1 for Fuzziness = 2.5.

- Fuzzy C Means incorrectly clusters 6.67% of non-spreading fire events to cluster 2 for Fuzziness = 2.5.
- Fuzzy Partition coefficient one at Fuzziness = 1.5
- Fuzzy Partition coefficient 0.95 at Fuzziness = 2.5
- No change to cluster distribution based on the fuzziness in this dataset.

January 8, 2025 dataset (Fig. 48, 51):

- Fuzzy C Means almost accurately clusters 92.89% of spreading fire events to cluster 0 for Fuzziness = 1.5.
- Fuzzy C Means incorrectly clusters 2.37% of non-spreading fire events to cluster 1 for Fuzziness = 1.5
- Fuzzy C Means incorrectly clusters 4.74% of non-spreading fire events to cluster 2 for Fuzziness = 1.5
- Fuzzy C Means almost accurately clusters 53.36% of spreading fire events to cluster 1 at Fuzziness = 2.5. A Few points in this cluster are far away and are not close to other dense, spreading fire points.
- Fuzzy C Means almost accurately clusters 39.53% of spreading fire events to cluster 2 at Fuzziness = 2.5. A few points in this cluster are far away and do not appear to be spreading spots
- Fuzzy C Means incorrectly clusters 7.11% of non-spreading fire events to cluster 0 at Fuzziness = 2.5
- Fuzzy Partition coefficient 1 at Fuzziness = 1.5
- Fuzzy Partition coefficient 0.89 at Fuzziness = 2.5
- January 21, 2025 dataset (Fig. 49, 52):
- Fuzzy C Means incorrectly clusters 28% of non-spreading fire events to cluster 0.

- Fuzzy C Means incorrectly clusters 60% of non-spreading fire events to cluster 1.
- Fuzzy C Means incorrectly clusters 12% of non-spreading fire events to cluster 2.
- Fuzzy Partition coefficient 0.95 at Fuzziness = 1.5.
- Fuzzy Partition coefficient 0.76 at Fuzziness = 2.5.

January 22, 2025 dataset (Fig.49, 52) Incident Day:

- Fuzzy C Means almost accurately clusters 67.57% % of spreading fire events to cluster 0 at Fuzziness = 2.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy C Means incorrectly clusters 10.81% of non-spreading fire events to cluster 1 at Fuzziness = 2.5.
- Fuzzy C Means incorrectly clusters 21.62% of non-spreading fire events to cluster 2 at Fuzziness = 2.5.
- Fuzzy C Means incorrectly clusters 13.15% of non-spreading fire events to cluster 0 at Fuzziness = 1.5.
- Fuzzy C Means incorrectly clusters 18.92% of non-spreading fire events to cluster 1 at Fuzziness = 1.5.
- Fuzzy C Means incorrectly clusters 67.57% of spreading fire events to cluster 2 at Fuzziness = 1.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy Partition coefficient 0.98 at Fuzziness = 1.5.
- Fuzzy Partition coefficient 0.82 at Fuzziness = 2.5.

January 23, 2025 dataset (Fig. 49, 52):

- Fuzzy C Means incorrectly clusters 48.48% of non-spreading fire events to cluster 0 at Fuzziness = 2.5
- Fuzzy C Means incorrectly clusters 24.24% of non-spreading fire events to cluster 1 at Fuzziness = 2.5.
- Fuzzy C Means almost accurately clusters 27.27% of spreading fire events to cluster 2 at Fuzziness = 2.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy C Means incorrectly clusters 24.24% of non-spreading fire events to cluster 0 at Fuzziness = 1.5.
- Fuzzy C Means incorrectly clusters 27.27% of non-spreading fire events to cluster 1 at Fuzziness = 1.5.
- Fuzzy C Means incorrectly clusters 48.48% of spreading fire events to cluster 2 at Fuzziness = 1.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy Partition coefficient 0.98 at Fuzziness = 1.5.
- Fuzzy Partition coefficient 0.80 at Fuzziness = 2.5.

August 5, 2024 dataset (Fig. 50, 53):

- Fuzzy C Means incorrectly clusters 26.67% of non-spreading fire events to cluster 0 at Fuzziness = 1.5.
- Fuzzy C Means incorrectly clusters 13.33% of non-spreading fire events to cluster 1 at Fuzziness = 1.5.
- Fuzzy C Means incorrectly clusters 33.33% of non-spreading fire events to cluster 2 at Fuzziness = 1.5.

- Fuzzy C Means incorrectly clusters 26.67% of non-spreading fire events to cluster 0 at Fuzziness = 2.5.
- Fuzzy C Means incorrectly clusters 53.33% of non-spreading fire events to cluster 1 at Fuzziness = 2.5.
- Fuzzy C Means incorrectly clusters 20% of non-spreading fire events to cluster 2, at Fuzziness = 2.5.
- Fuzzy Partition coefficient 0.96 at Fuzziness = 1.5.
- Fuzzy Partition coefficient 0.92 at Fuzziness = 2.5.

August 6, 2024 dataset (Fig. 50, 53) Incident Day:

- Fuzzy C Means almost accurately clusters 33.51% of spreading fire events to cluster 0 at Fuzziness = 1.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy C Means almost accurately clusters 47.89% of Spreading fire events to cluster 1 at Fuzziness = 1.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy C Means almost accurately clusters 18.62% of spreading fire events to cluster 2 at Fuzziness = 1.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy C Means almost accurately clusters 33.51% of spreading fire events to cluster 0 at Fuzziness = 2.5. Few points in this cluster are far away and do not appear to be spreading spots
- Fuzzy C Means accurately clusters 18.62 % of spreading fire events to cluster 1 at Fuzziness = 2.5. Few points in this cluster are far away and do not appear to be spreading spots.

- Fuzzy C Means accurately clusters 47.87% of spreading fire events to cluster 2 at Fuzziness = 2.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy Partition coefficient 0.95 at Fuzziness = 1.5.
- Fuzzy Partition coefficient 0.76 at Fuzziness = 2.5.

August 7, 2024 dataset (Fig. 50, 53):

- Fuzzy C Means clusters 62.50% of spreading fire events to cluster 0 at Fuzziness = 1.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy C Means clusters 24.11% of spreading fire events to cluster 1 at Fuzziness = 1.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy C Means clusters 13.39% of spreading fire events to cluster 2 at Fuzziness = 1.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy C Means accurately clusters 62.50% of spreading fire events to cluster 0 at Fuzziness = 2.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy C Means accurately clusters 13.39 % of spreading fire events to cluster 1 at Fuzziness = 2.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy C Means accurately clusters 24.11% of spreading fire events to cluster 2 at Fuzziness = 2.5. Few points in this cluster are far away and do not appear to be spreading spots.
- Fuzzy Partition coefficient 0.96 at Fuzziness = 1.5.

- Fuzzy Partition coefficient 0.76 at Fuzziness = 2.5.

Other evaluation methods:

Fuzzy Partition Coefficient

The Fuzzy Partition coefficient was better for the lowest fuzziness at 1.5.

Dunn Index:

The Dunn Index is suitable for this dataset. The Dunn index is high only for the sample data 1 Incident Day dataset clusters; for all other sample data clusters, the Dunn index is low, indicating that Fuzzy C-means clustering is not accurate.

Limitations of Fuzzy C-Means:

The Fuzzy C-means clustering algorithm also requires additional manual intervention to accurately distinguish between clusters representing spreading fire events and those indicating non-spreading fires. One significant issue is that it tends to incorrectly cluster non-spreading fire events from normal days into the same cluster. Furthermore, adjusting the fuzziness parameter from 1.5 to 2.5 complicates the process of identifying distinct clusters, leading to increased confusion in the clustering results. On the actual incident day dataset, the algorithm includes some farthest fire events within the same cluster. This misclassification of outliers can negatively impact the computation of the cluster's vertices or border area, consequently resulting in incorrect interpretations of the fire scene. The Fuzzy C-means clustering also shows a lower Dunn Index for this type of dataset, indicating poor separation between clusters. This suggests that while the algorithm is designed to handle ambiguity in data, its current configuration is ineffective for accurately capturing the complex patterns present in fire event data. As such, this highlights the need for further refinement of the algorithm or exploration of alternative clustering methods to improve accuracy in such scenarios.

4.2.3 Gaussian Mixture Models Clustering

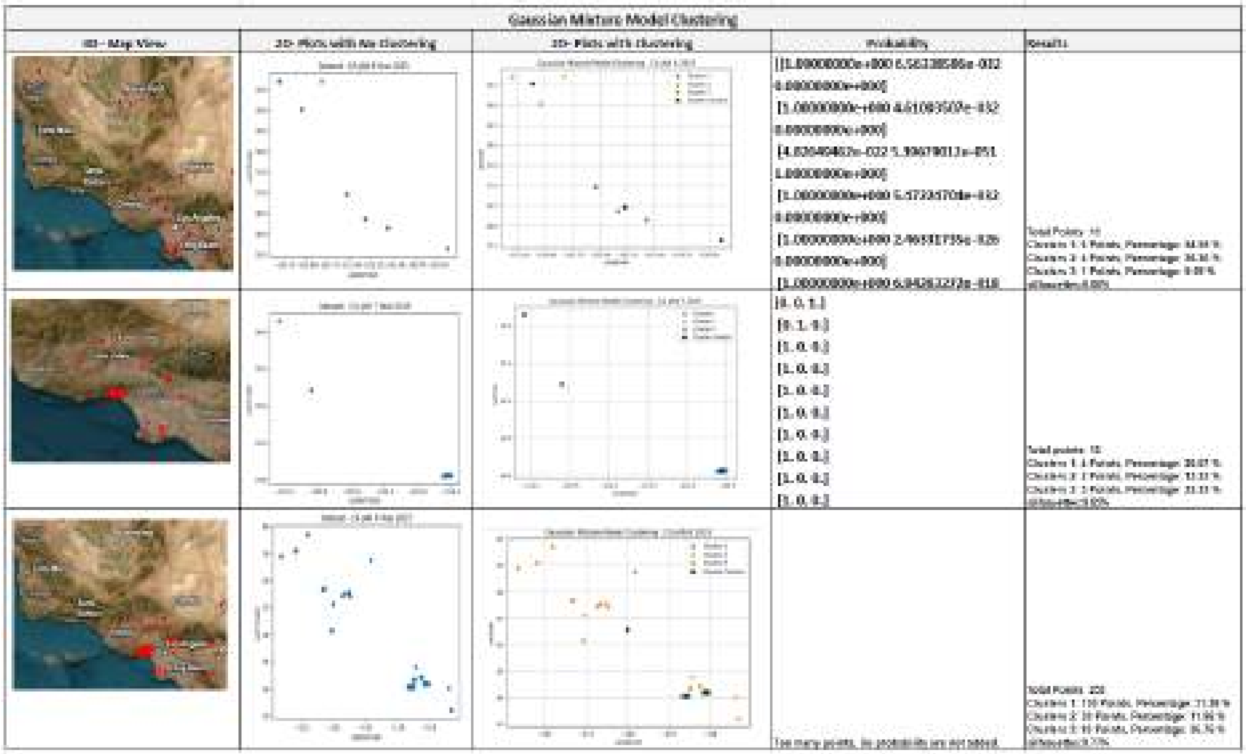


Figure 53

Gaussian Mixture Models Clustering Results - Sample Data 1

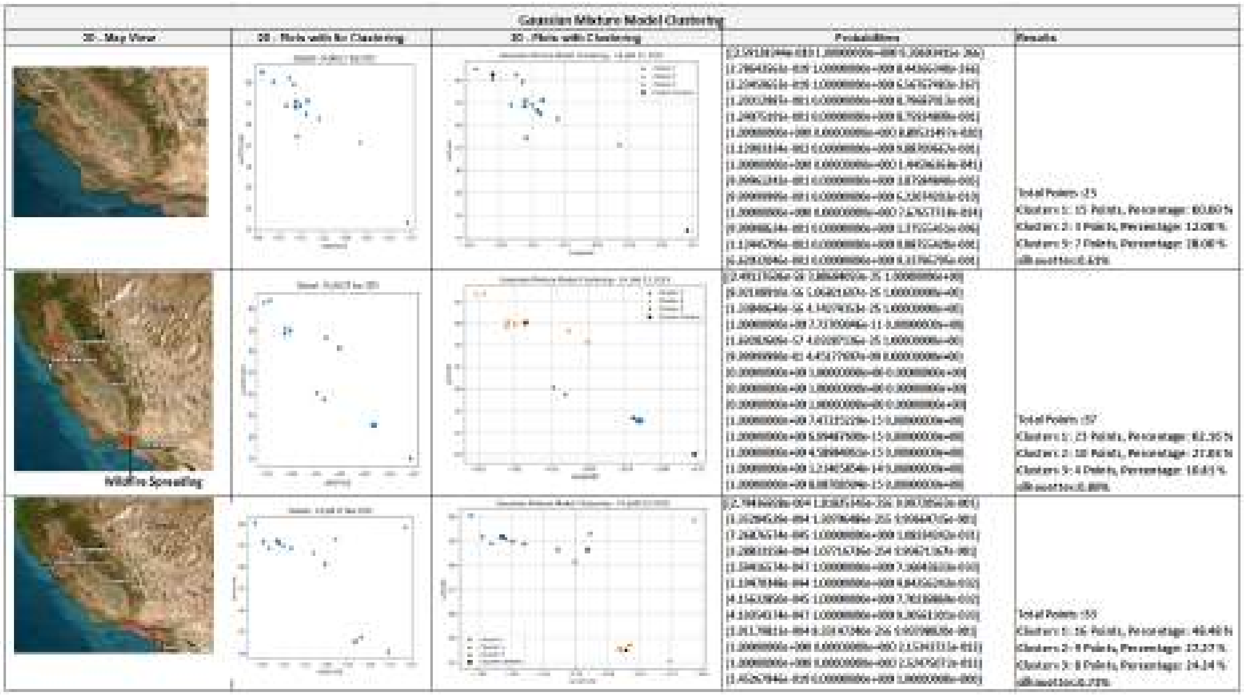


Figure 54

Gaussian Mixture Models Clustering Results - Sample Data 2

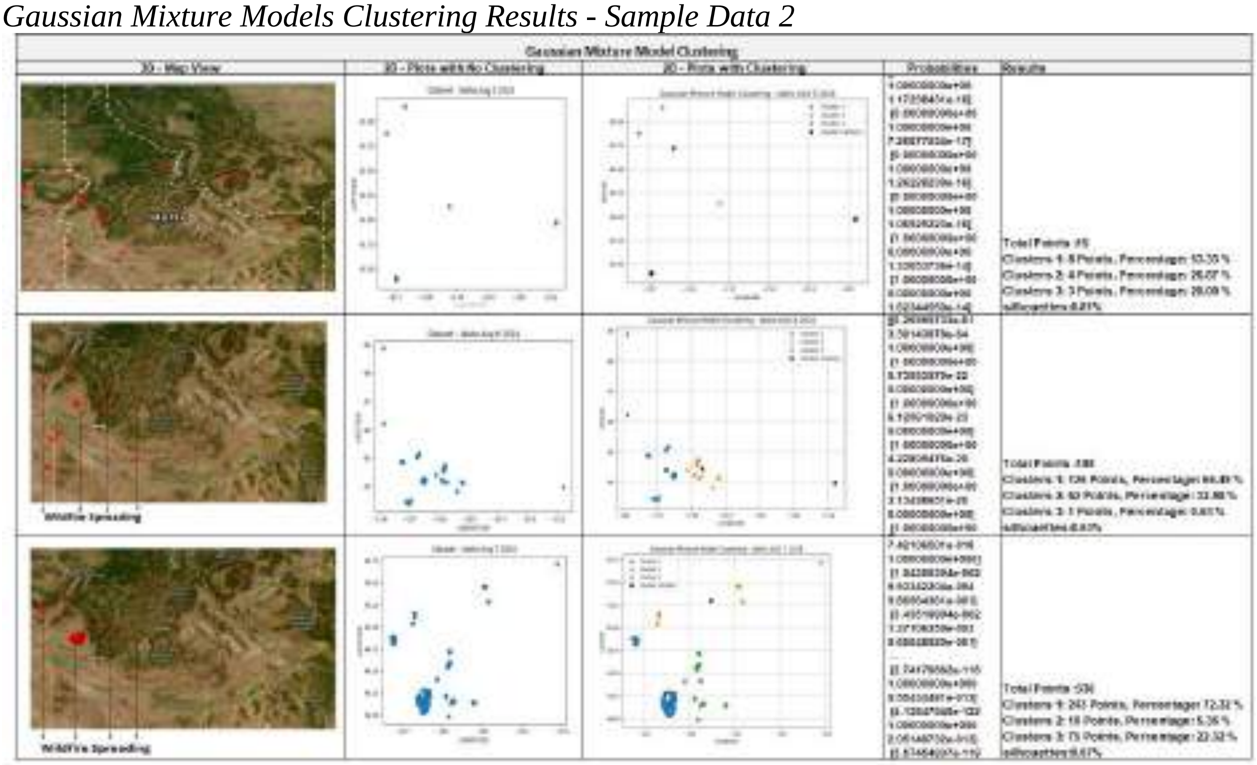


Figure 55

Gaussian Mixture Models Clustering Results - Sample Data 3

Gaussian Mixture Models Results Interpretation:

Figures 54, 55, and 56 represent the results of the Gaussian Mixture Models, with the silhouette score indicating high values on the incident date in sample data 1 and 2, particularly when the dataset contains fewer datapoints. However, this metric has a lower score in sample 3, which has more datapoints. Hence, this metric is not reliable for the evaluation where there are more datapoints in the dataset. The models' prediction results are almost the same as those of K-means; therefore, the results explanation is not repeated here.

4.2.4 Agglomerative Hierarchical Clustering

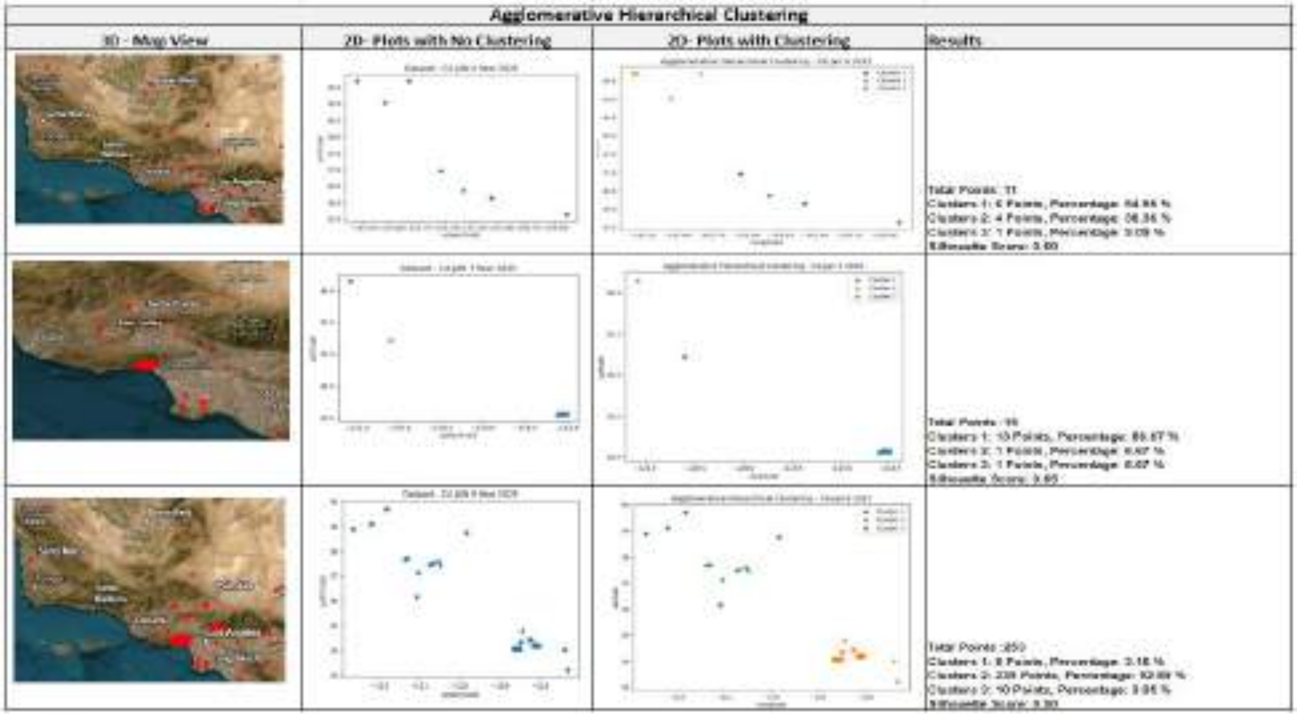


Figure 56
Agglomerative Hierarchical Clustering Results - Sample Data 1

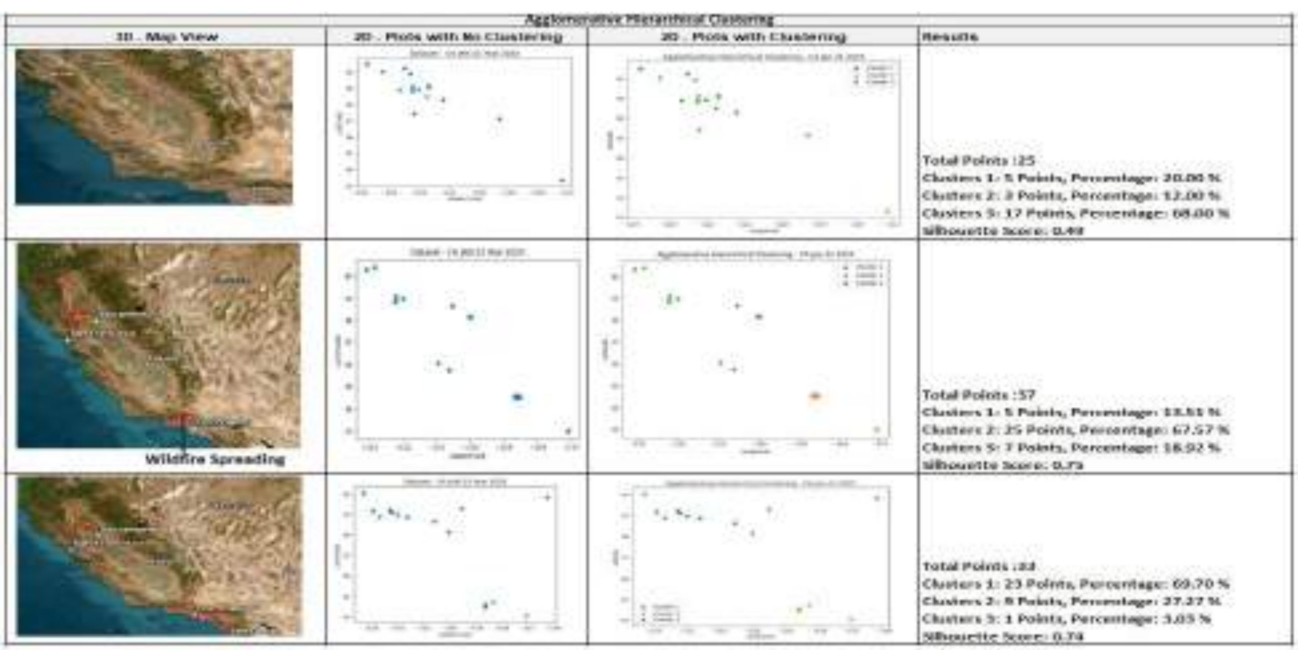


Figure 57
Agglomerative Hierarchical Clustering Results - Sample Data 2

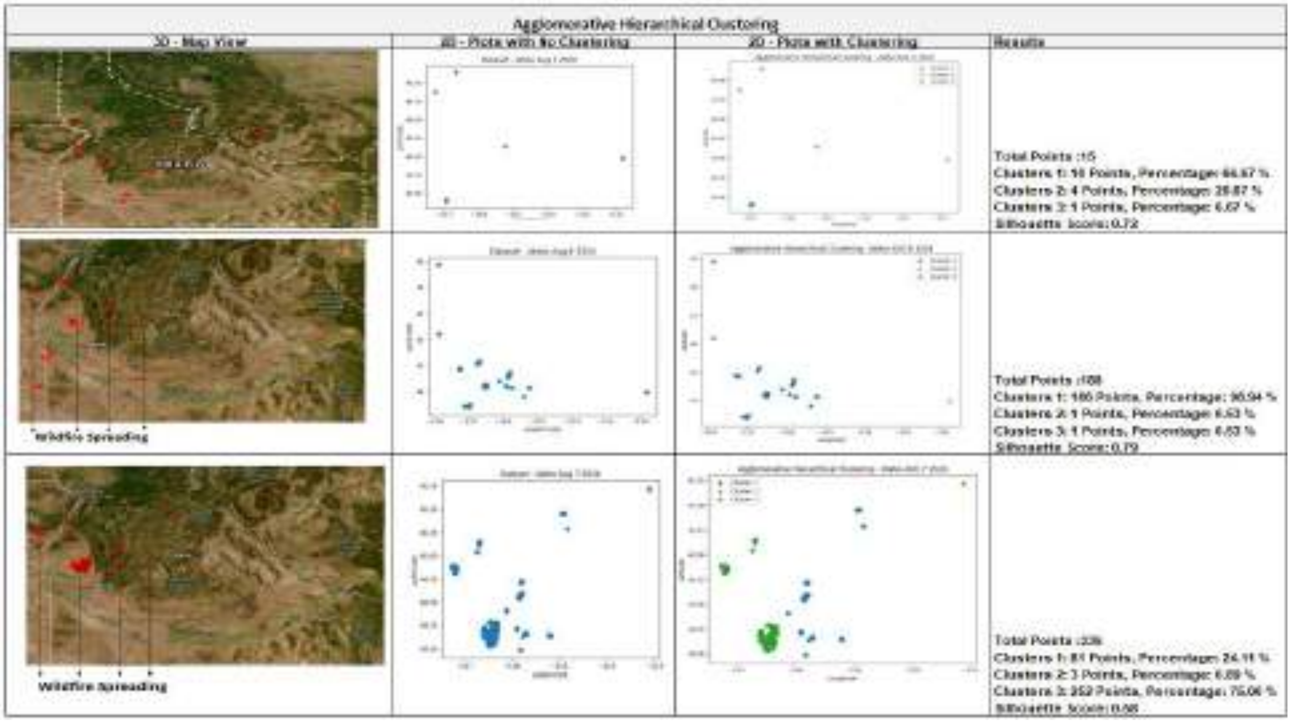


Figure 58
Agglomerative Hierarchical Clustering Results - Sample Data 3

Agglomerative Hierarchical Clustering Results Interpretation:

Figures 57, 58, and 59 represent the results of the Agglomerative Hierarchical Clustering, with Silhouette scores indicating high values on the incident date in the sample data (1, 2), particularly when there are fewer datapoints. However, this metric has a lower score in the sample 3 dataset on the second day of the incident, when there are more datapoints. Hence, this metric is not reliable for the evaluation where there are more datapoints in the dataset.

Models' prediction results are almost the same as K-means. Hence, the results explanation is not repeated here.

4.2.5 DBSCAN Clustering

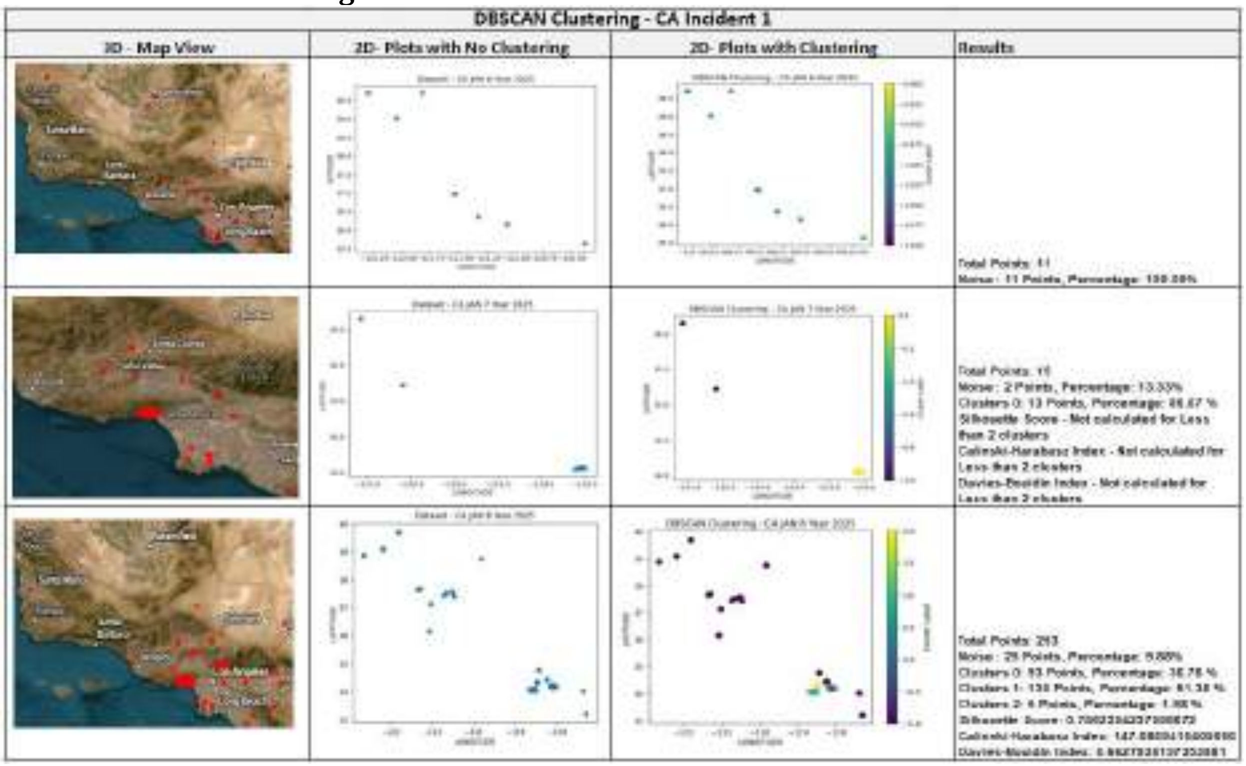


Figure 59

DBSCAN Clustering Results -Sample Data 1

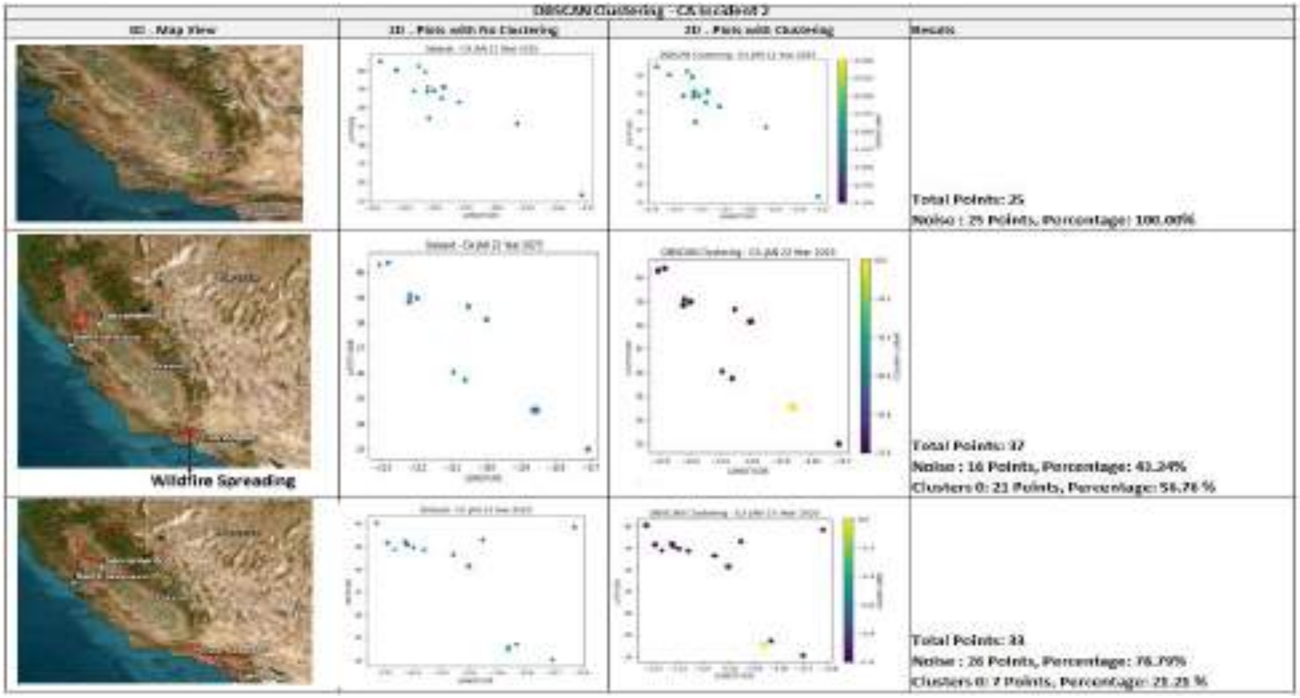


Figure 60
DBSCAN Clustering Results - Sample Data 2

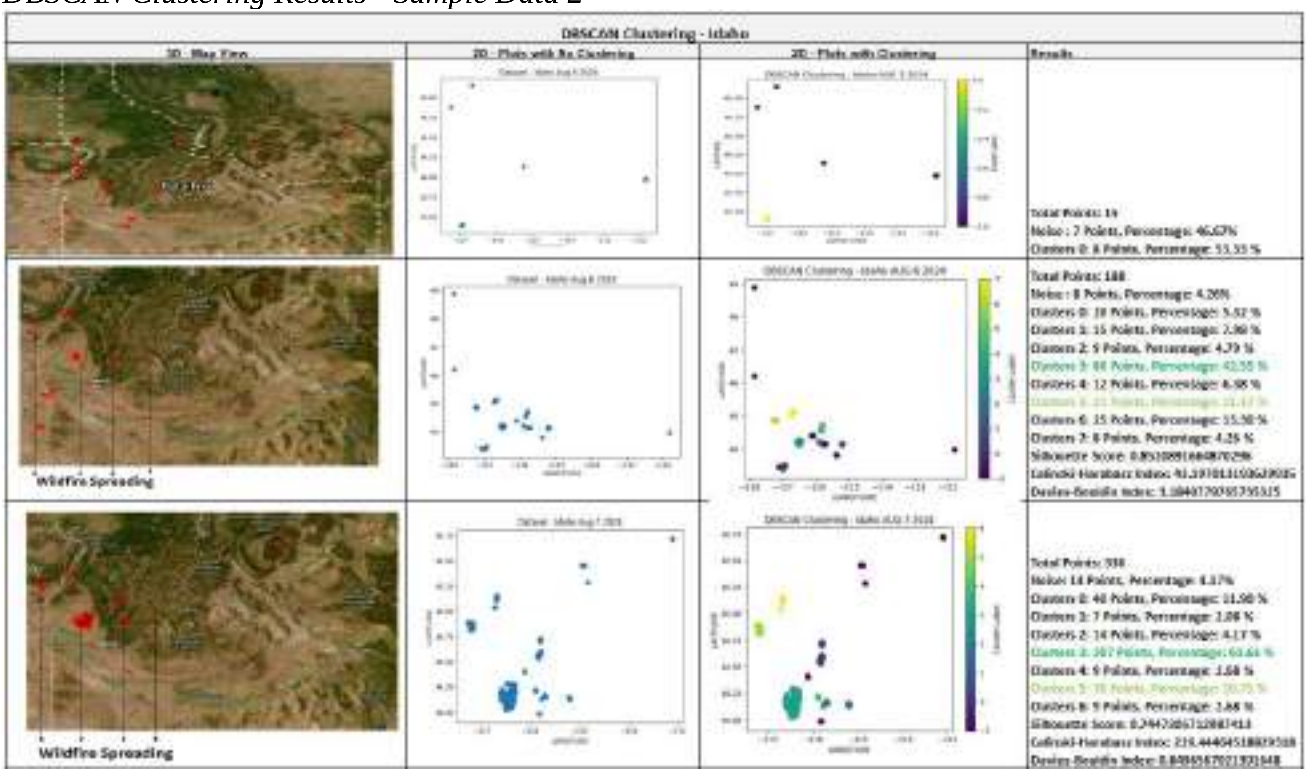


Figure 61
DBSCAN Clustering Results - Sample Data 3

DBSCAN Results Interpretation:

Visual Inspection

3D Map view:

Depicts all the fire events that occurred for the selected date in the selected region.

Figure 60: Represents the January 6, 7, and 8, 2025 fire events of California.

Figure 61: Represents the January 21, 22, 23, 2025 fire events of California.

Figure 62: Represents the August 5, 6, 7, 2024 fire events of Idaho.

2D plot with no clustering:

Follows the exact depiction as specified in the 3D Map View,

- The X axis represents the Longitude, and the Y axis represents the Latitude of the fire events.
- This Plot is utilized for the Visual Inception by comparing it with the resultant clusters from the DBSCAN prediction.

2D plot with clustering and Results:

Figures 60, 61, 62: DBCAN results are depicted as follows

January 6, 2025 dataset (Fig. 60):

- Indicates all the non-spreading fire events are represented in '– ve', indicating 100% noise.

January 7, 2025 dataset (Fig. 60) Incident Day:

- It accurately clusters 86.67% of spreading fire events to cluster 0.
- Other non-spreading fire events are represented in -ve, 13.33% indicates noise.

January 8, 2025 dataset (Fig. 60):

- It accurately clusters 36.76% of spreading fire events to cluster 0.
- It accurately clusters 51.38% of spreading fire events to cluster 1.
- It accurately reports 1.98% of spreading fire events to cluster 2.

- Other non-spreading fire events are represented in -ve, 9.88% indicates noise.

January 21, 2025 dataset (Fig. 61):

All the fire events are represented in '– ve', indicating 100% noise.

January 22, 2025 dataset (Fig. 61) Incident Day:

- The DBSCAN Algorithm accurately clusters 56.76% of spreading fire events into cluster 0.
- Other non-spreading fire events are represented in -ve, 43.22% indicates noise.

January 23, 2025 dataset (Fig. 61):

- The algorithm accurately clusters 21.21% of spreading fire events to cluster 0.
- Other non-spreading fire events are represented in -ve, 78.79 % indicates noise.

August 5, 2024 dataset (Fig. 62):

- DBCAN accurately clusters 53.33% of suspicious spreading fire events in cluster 0.
- Other fire events are represented in -ve, 46.67 % indicates noise.

August 6, 2024 dataset (Fig. 62) Incident Day:

- The DBSCAN Algorithm accurately clusters 5.32% of spreading fire events into cluster 0.
- The DBSCAN Algorithm accurately clusters 7.98% of spreading fire events into cluster 1.
- The DBSCAN Algorithm accurately clusters 4.79% of spreading fire events into cluster 2.
- The DBSCAN Algorithm accurately clusters 42.55 % of spreading fire events into cluster 3.
- The DBSCAN Algorithm accurately clusters 6.8 % of spreading fire events into cluster 4.

- The DBSCAN Algorithm accurately clusters 11.17 % of spreading fire events into cluster 5.
- The DBSCAN Algorithm accurately clusters 13.30 % of spreading fire events into cluster 6.
- The DBSCAN Algorithm accurately clusters 4.26 % of spreading fire events into cluster 7.
- Other non-spreading fire events are represented in -ve, 4.26 % indicates noise.

August 7, 2024 dataset (Fig. 62):

- The DBSCAN Algorithm accurately clusters 11.90% of spreading fire events into cluster 0.
- The DBSCAN Algorithm accurately clusters 2.01% of spreading fire events into cluster 1.
- The DBSCAN Algorithm accurately clusters 4.17% of spreading fire events into cluster 2.
- The DBSCAN Algorithm accurately clusters 61.61 % of spreading fire events into cluster 3.
- The DBSCAN Algorithm accurately clusters 2.68 % of spreading fire events into cluster 4.
- The DBSCAN Algorithm accurately clusters 10.71 % of spreading fire events into cluster 5.

The algorithm accurately clusters 2.68 % of spreading fire events into cluster 6.

Other non-spreading fire events are represented in -ve, 4.17 % indicates noise.

Silhouette score:

This score is not computed for sample datasets 1 and 2 in the DBSCAN results; this evaluation metric requires at least two clusters to score the clustering. The silhouette score is not suitable for the DBSACN evaluation, as this type of dataset is most likely to have fewer clusters on the day of the incident.

Calinski-Harabasz Index:

The Calinski-Harabasz Index score is high on the Sample Data 4 when there are more fire events spreading clusters, which indicates a good score for the clusters. The Calinski-Harabasz Index is not computed on the remaining sample data; it requires at least two clusters for scoring.

For this application, the dataset is likely to have less fire event data on the day of the incident, which may result in single cluster predictions from DBSCAN. Therefore, the Calinski-Harabasz Index evaluation is not suitable for this application.

Davies-Bouldin index

Davies-Bouldin index is low on the Sample Data 4 second day of incident compared to the first day of incident, but the evaluation metric does not compute the score when there are fewer than 2 cluster, as this type of dataset most likely have fewer clusters on the day of incident, hence it's not suitable for the DBSCAN results evaluation.

Limitations of DBSCAN:

When there are nearby spreading clusters, the algorithm tends to separate them into two distinct small clusters if they meet the specified criteria for cluster density and radius. However, it fails to account for the potential to merge these clusters when the intra-cluster distance is small, which can lead to an inaccurate representation of the overall fire event dynamics.

Additionally, the algorithm demonstrates inconsistency when faced with multiple clusters; it often shifts some data points across the nearest clusters. This inconsistency can lead to instability in clustering outcomes, making it challenging to rely on the algorithm for the precise delineation of fire event clusters. These issues underscore the need for enhancements to the algorithm's logic to better account for proximity and cluster integrity, thereby ensuring a more accurate and meaningful analysis of spreading fire events.

4.2.6: Multi-Level Multi-Criteria Clustering Algorithm

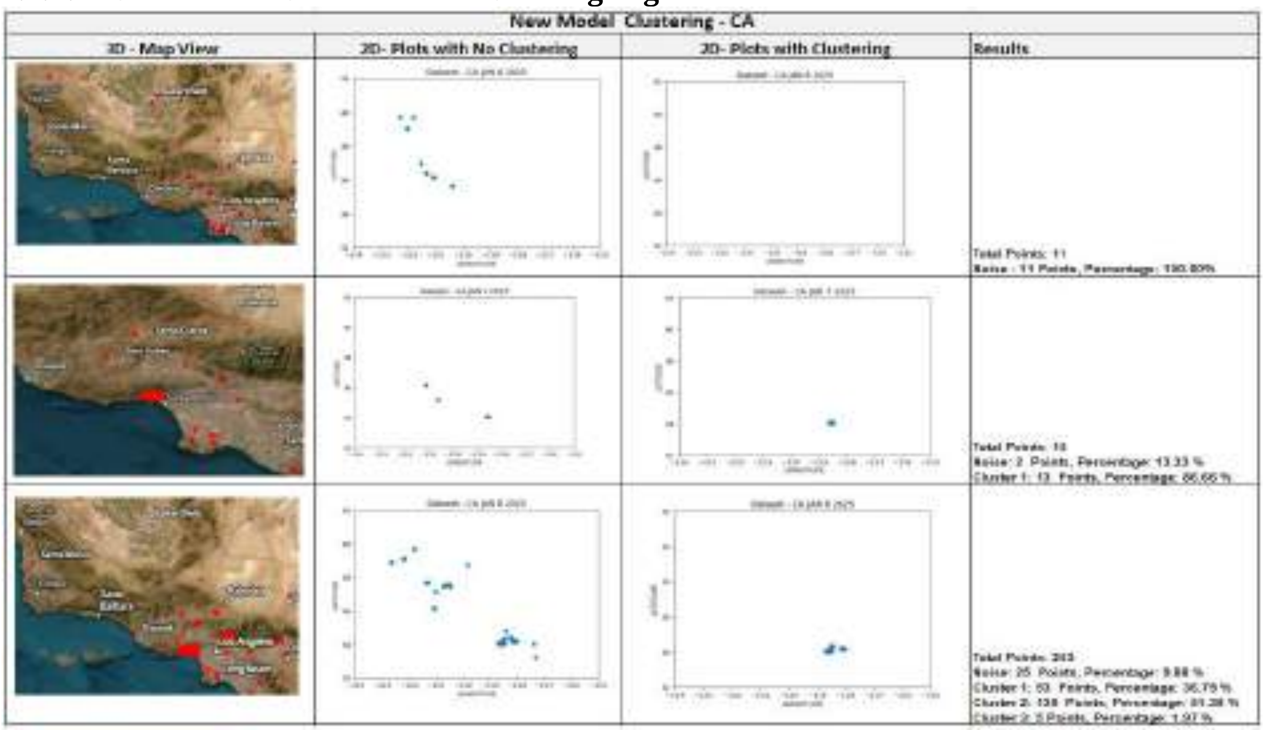


Figure 62
Proposed New Model Clustering Results - Sample Data 1

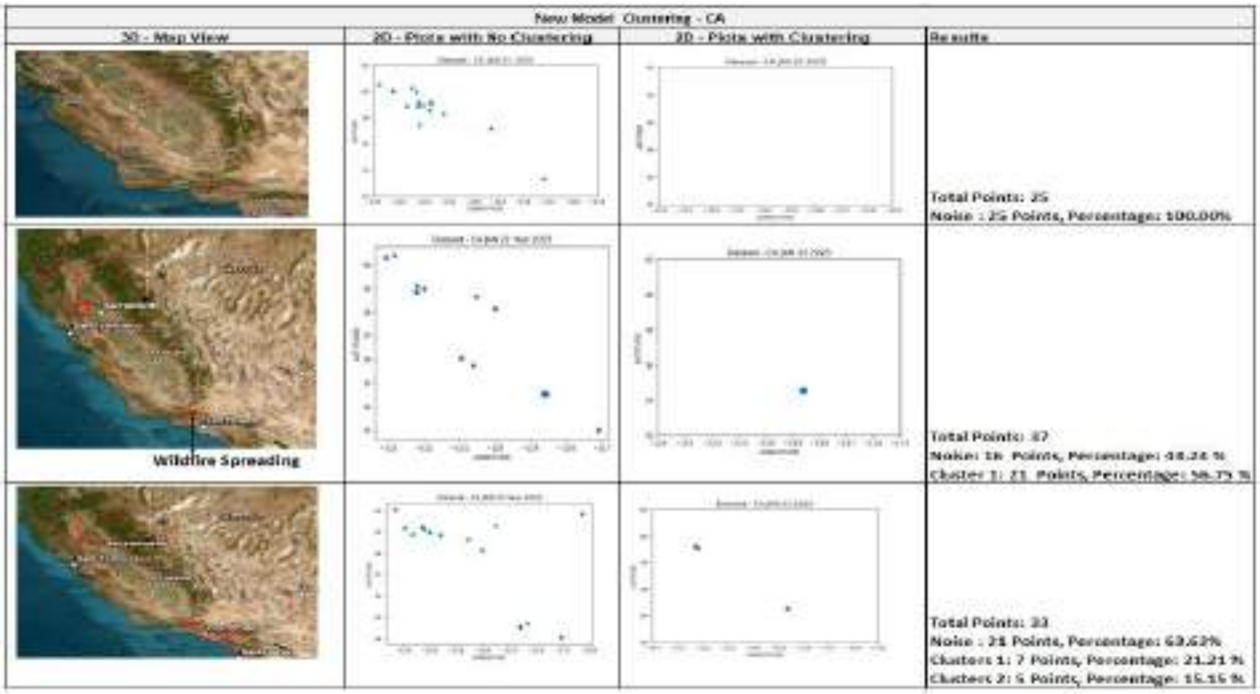


Figure 63

Proposed New Model Clustering Results - Sample Data 2

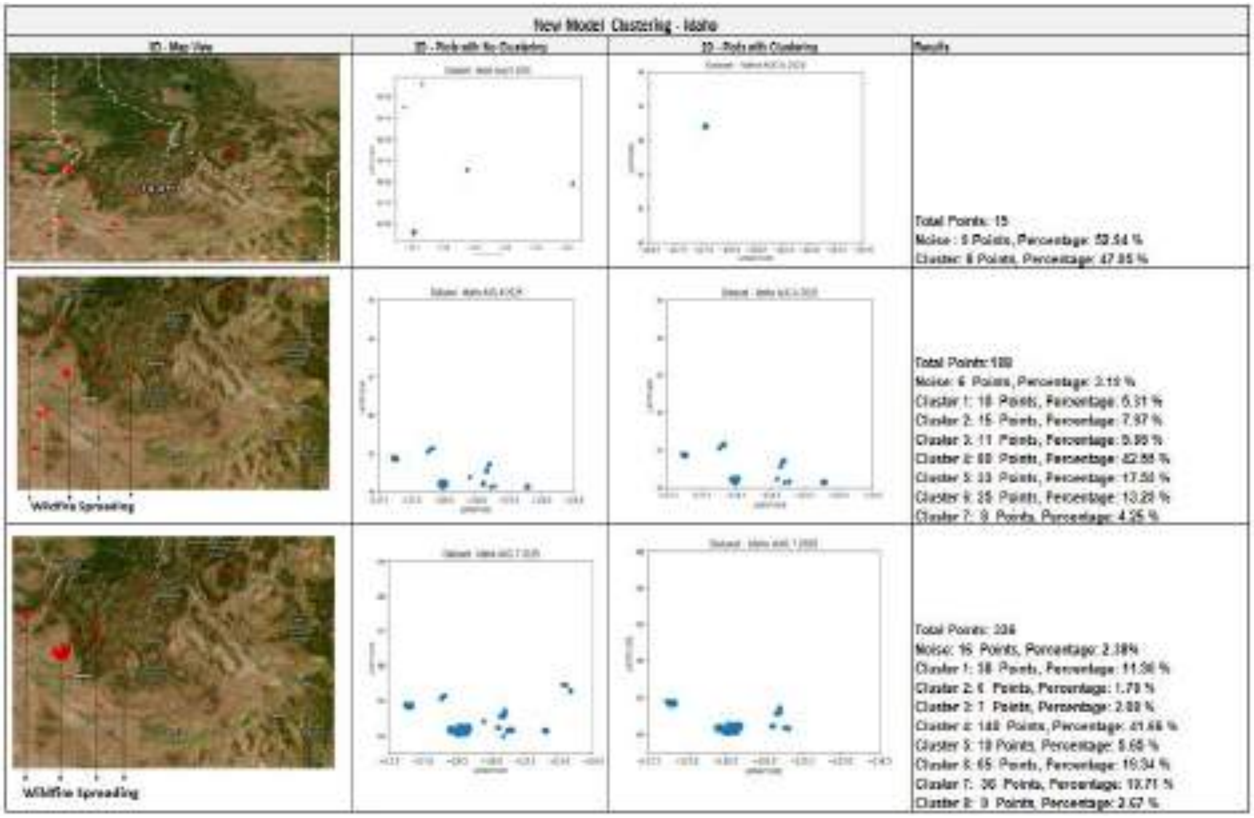


Figure 64
Proposed New Model Clustering Results - Sample Data 3

Proposed New Model Results Interpretation:

Figures 63, 64, and 65 represent the result of the Proposed New Model Clustering.

2D plot with clustering and Results:

Figures 63, 64, 65: clustering results are depicted as follows

January 6, 2025 dataset (Fig. 63):

- No clusters are formed, 100% of the datapoints are noise.

January 7, 2025 dataset (Fig. 63) Incident Day:

- Cluster 1 has 86.66 fire events, and the rest of the datapoints are noise.
- Other 13.33% of fire events are identified as noise.

January 8, 2025 dataset (Fig. 63):

- Cluster 1 has 36.76% of fire events
- Cluster 2 has 51.38% of fire events
- Cluster 3 has 1.97% of fire events.
- 9.88% of fire events are noise.

January 21, 2025 dataset (Fig. 64):

- No clusters are formed, 100% of the datapoints are noise.

January 22, 2025 dataset (Fig. 64) Incident Day:

- Cluster 1 has 56.75% of fire events
- 43.24% of fire events are noise

January 23, 2025 dataset (Fig. 64):

- Cluster 1 has 56.75% of fire events
- 43.24% of fire events are noise.

August 5, 2024 dataset (Fig. 65) :

- Cluster 1 has 47.05% of fire events
- 52.94% of fire events are noise.

August 6, 2024 dataset (Fig. 65) Incident Day:

- Cluster 1 has 5.31% of fire events.
- Cluster 2 has 7.97% of fire events
- Cluster 3 has 5.85% of fire events
- Cluster 4 has 42.55% of fire events
- Cluster 5 has 17.55% of fire events
- Cluster 6 has 13.29% of fire events
- Cluster 7 has 4.25% of fire events
- 3.19% of fire events are noise.

August 7, 2024 dataset (Fig. 65):

- Cluster 1 has 11.30% of fire events.
- Cluster 2 has 1.78% of fire events
- Cluster 3 has 2.08% of fire events
- Cluster 4 has 41.66% of fire events
- Cluster 5 has 5.65% of fire events
- Cluster 6 has 19.34% of fire events
- Cluster 7 has 10.71% of fire events
- Cluster 8 has 2.67% of fire events
- 2.38 % of datapoints are noise.

The newly proposed model effectively eliminates all noise, isolating only the fire events that are spreading within the cluster. The results show high accuracy on the incident day. In the following days, as the fire expands, over 95% of the data points predominantly represent the spreading fire events, as can be observed in sample dataset 3, where the noise percentage is less than 5%. In situations where the fire becomes uncontrollable, indicated by the occurrence of more than five clusters nearby, the distinction between smaller, separate clusters and a few larger clusters with smaller adjacent clusters becomes significant.

This model excels in several key areas:

- It accurately identifies 100% of the data points as noise when there are no spreading fire events present in the dataset.
- It successfully recognizes all spreading fire events and accurately assigns them to clusters on the incident day, as well as at other times, despite the presence of noise.
- The results produced by the model are consistent and reliable.

4.3 Research Question Three

Does the accuracy of the machine algorithm in predicting the growing fires vary on real-time fire events data (unseen data)?

From Research Question 1, DBSCAN was tested on the real-time dataset from MODIS and VIIRS (on the satellites NOAA-20 and NOAA-21).

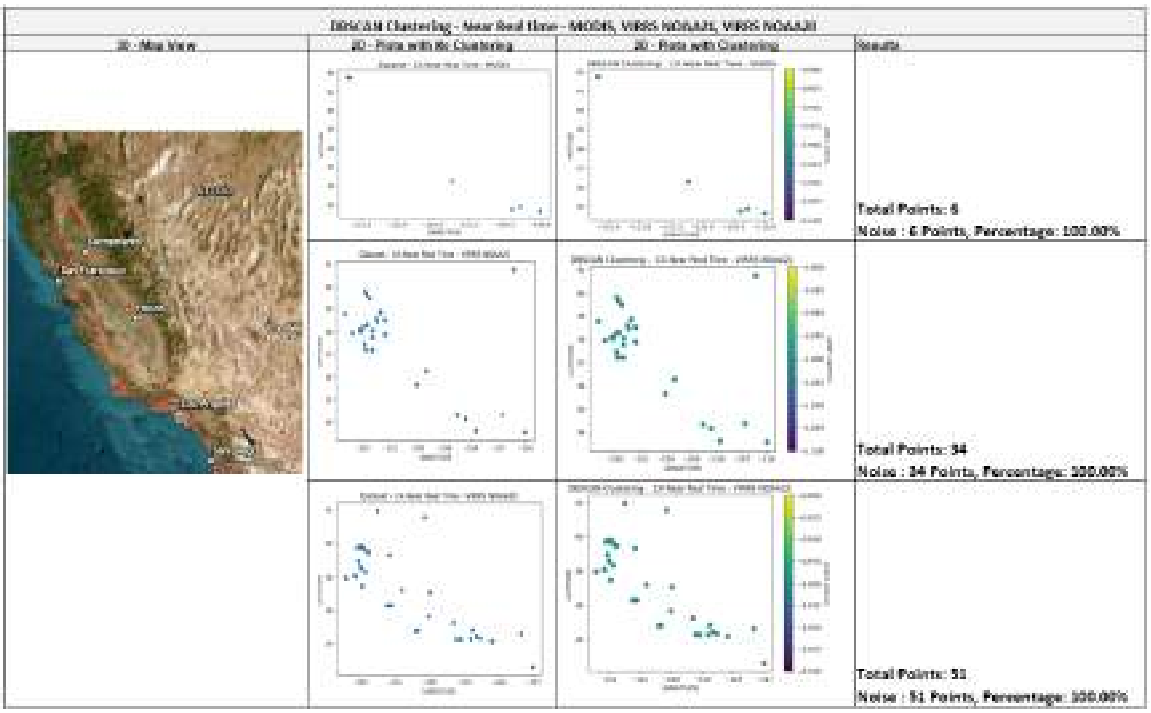


Figure 65
DBSCAN Clustering Results -Near real-time dataset

Figure 66: Represents near-real-time/real-time fire events of California captured on April 21 /2025.

Visual Inspection

Real Time /Near real time dataset (Fig. 66):

Modis Dataset:

- All the non-spreading fire events are represented in '- ve', which indicates 100% noise.

VIRRS NOAA 20:

- All the non-spreading fire events are represented in '- ve', which indicates 100% noise.

VIRRS NOAA 21:

- All the non-spreading fire events are represented in '- ve', which indicates 100% noise.

Based on the visual inspection of results from the DBSCAN algorithm, performance and accuracy remain the same on the unseen real-time dataset tested from the source MODIS.

But accuracy decreased when DBSCAN was experimented on the unseen dataset from different sources VIIRS on satellite NOAA21, hence the experiment on VIRRS dataset was repeated by tuning the input parameter 'minimum points in density' to higher values as the resolution of instrument VIRRS is higher than the MODIS, accuracy of the DBSCAN increased on the VIIRS dataset at the higher 'minimum points in density', experiment was again repeated for the same day dataset from the VIIRS dataset equipped on the different satellite NOAA20, DBSCAN algorithm accuracy remained same as VIRRS dataset of NOAA21, there is no change required for the input parameter 'minimum points in density' between VIRRS on NOAA20 and VIRRS NOAA21, hence it can be concluded that as far the source remain same accuracy of the DBSCAN algorithm remained same between the sample dataset and the unseen dataset, when there is change of source of dataset input Parmeter requires to be tuned to improve the accuracy.

The Proposed New Multilevel multicriteria clustering algorithm was also tested on the unseen near-real-time dataset (refer to Web App Figure 74, 75 in Research Question 6), similar to DBSCAN, the same machine learning input configuration applied for the MODIS dataset was required to be changed for the VIRRS, with the ‘minimum points in density’ increased to a higher value, 2 times the MODIS configured value.

With these changes, this model illustrates the higher accuracy for both the unseen data set from MODIS and VIRRS. There was no change required between the VIRRS on different satellites, SNPP, or on NOAA 20 or NOAA21. The input configuration applied for the VIRRS SNPP dataset remains the same for NOAA 20 and NOAA 21.

No other customization or experimentation is required for the different data sources, including VIRRS on SNPP, NOAA-20, and NOAA-21, as this instrument produces these data and is similar to MODIS, except for the resolution, which necessitated changes only in the minimum points in density.

4.4 Research Question Four

Can machine learning algorithms predict more contextual information about fire scenes in areas expected to experience fire growth in near real-time, such as the threat level to nearby residences from the growing fire?

Refer to Figure 74 of the Web App, which represents residences under threat in different colors, specifically pink. This prediction is a result of the newly proposed model, a multi-level, multi-criteria clustering algorithm. This model first predicts fire growth clusters and then utilizes the residential dataset to predict contextual information for the expected fire growth scene, identifying residences located near the predicted growing fire.

4.5 Research Question Five

Is the accuracy of the machine learning algorithms in predicting the fire growth and the latency in predicting the fire growth and its additional contextual information acceptable to real-world applications in the fire industry?

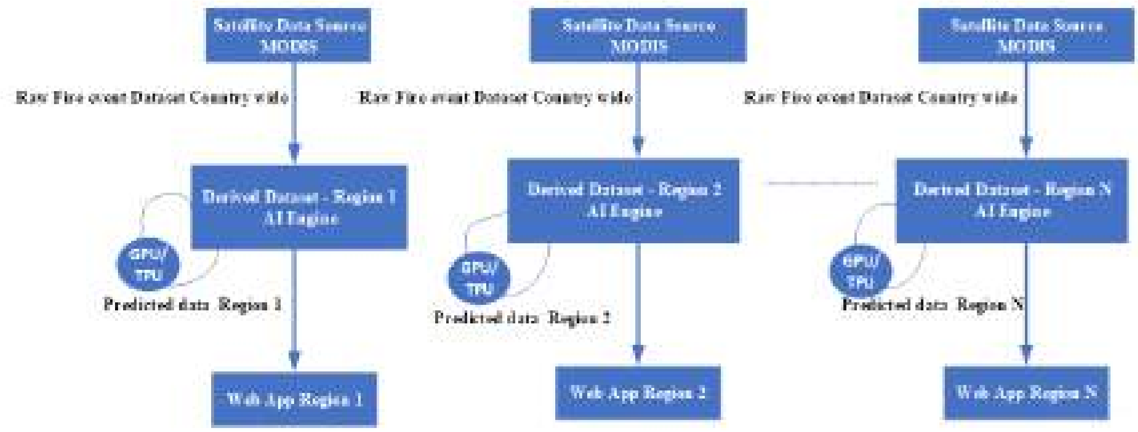


Figure 66
Region-Based Prediction- Multiple GPUs - Parallel computing

The initial hypothesis was to compute the prediction at the regional dataset as depicted in Figure 67, which uses an individual GPU for each region and was hosted on a separate web URL. However, the algorithm's prediction accuracy was higher, and the latency of computing the growing fire predictions and the additional contextual information was within a minute; this was an acceptable performance for the fire industry. However, maintaining individual GPUs for each regional app and hosting on the particular web URL for each region is an expensive solution; calling to the globe is more costly in terms of GPU utilization from each app.

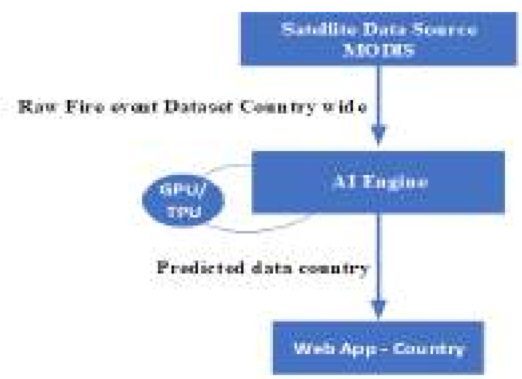


Figure 67
Country-based - Single GPU Prediction

When the algorithm was experimented on the country-wide dataset using a single GPU, as represented in Figure 68, the latency of the algorithms increased to > 15 min for predicting the growing fire in the country-wide dataset, and for predicting the contextual information for each predicted growing fire in the country-wide dataset, the latency increased to more than 30 minutes.

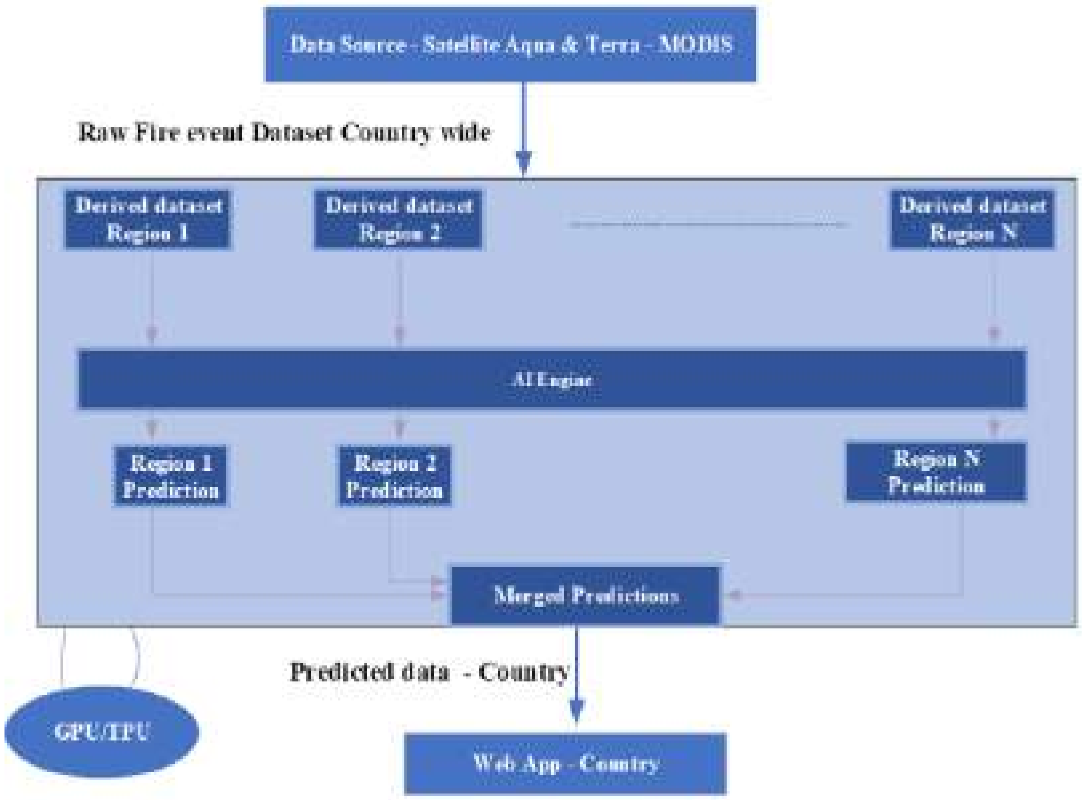


Figure 68
Derived metric - Region-based prediction and Integrated prediction - Sequential GPU scheduling

To reduce latency, the dataset is first divided into regional datasets (Figure 69). The machine learning model is then sequentially scheduled to predict the growing fire in each regional dataset, utilizing a single GPU, once all the regional predictions are completed. The machine learning model is scheduled sequentially to predict the contextual information of each predicted growing fire scene, utilizing the regional residence dataset. Then, all the predicted regional growing fire data is merged to obtain country-wide predictions.

Contextual information of every growing fire scene is merged to get country-wide predictions. With this approach, latency has improved from 30 minutes to less than 2 minutes, which is still acceptable for the fire industry application.

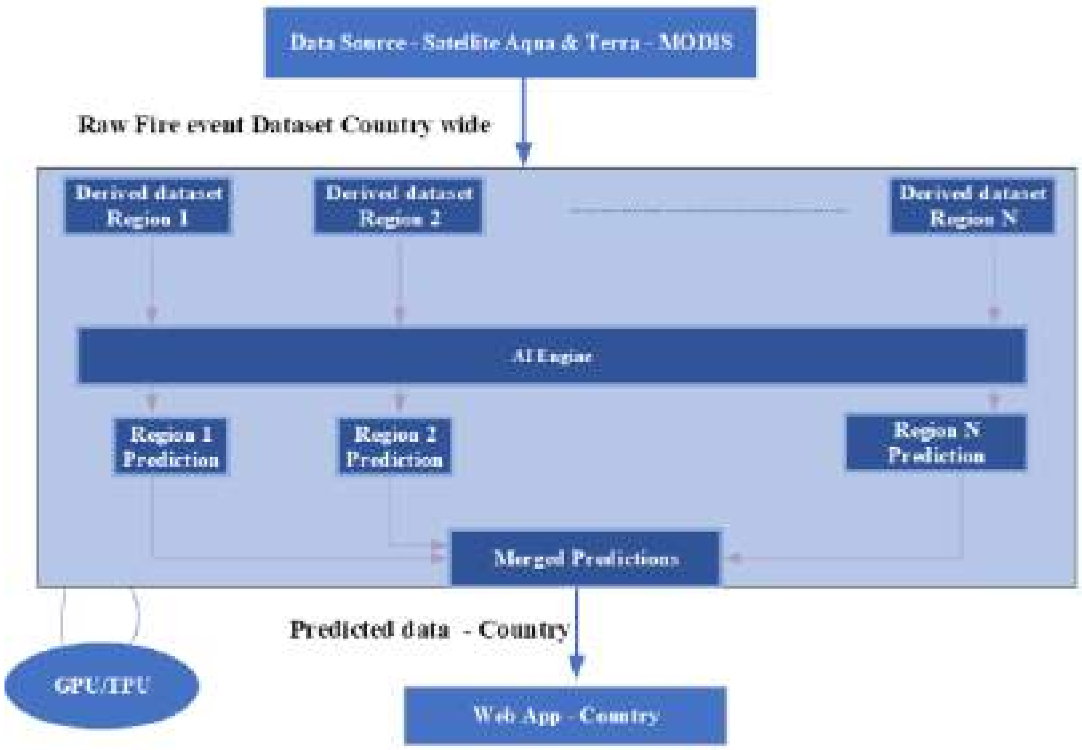


Figure 69
System Software Architecture - Performance and Accuracy - Application A Data Source MODIS

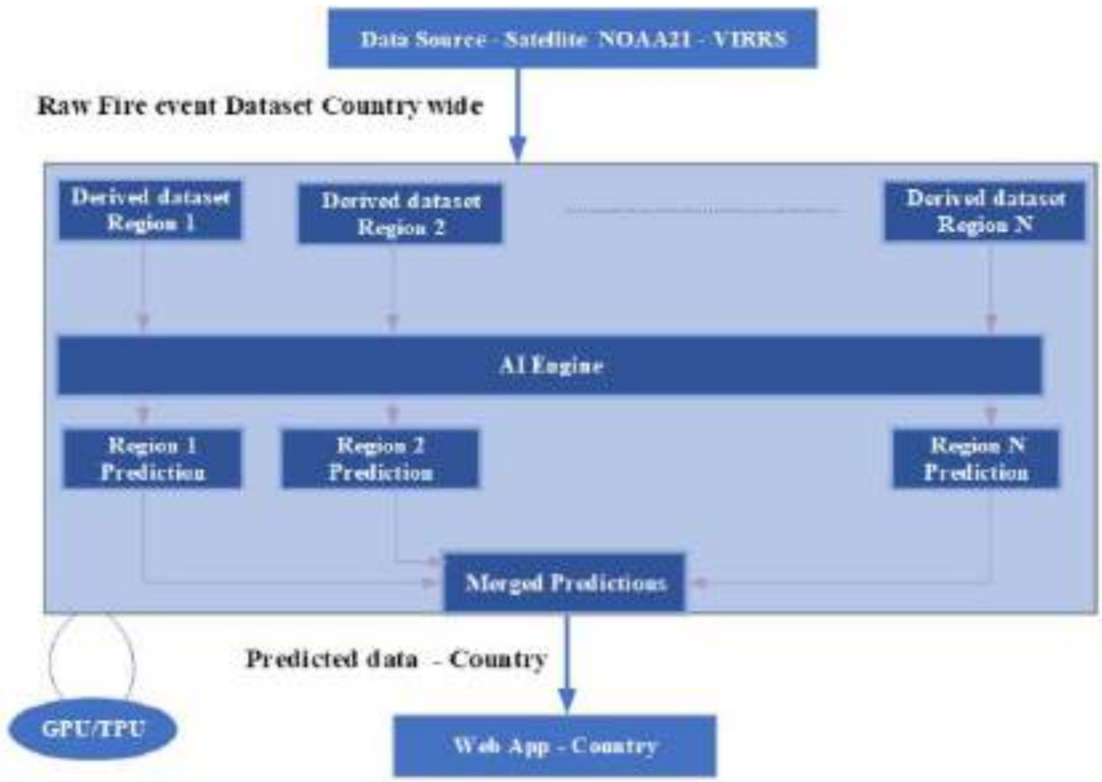


Figure 70
System Software Architecture - Performance and Accuracy Aspects - Application B Data Source VIRRS NOAA21

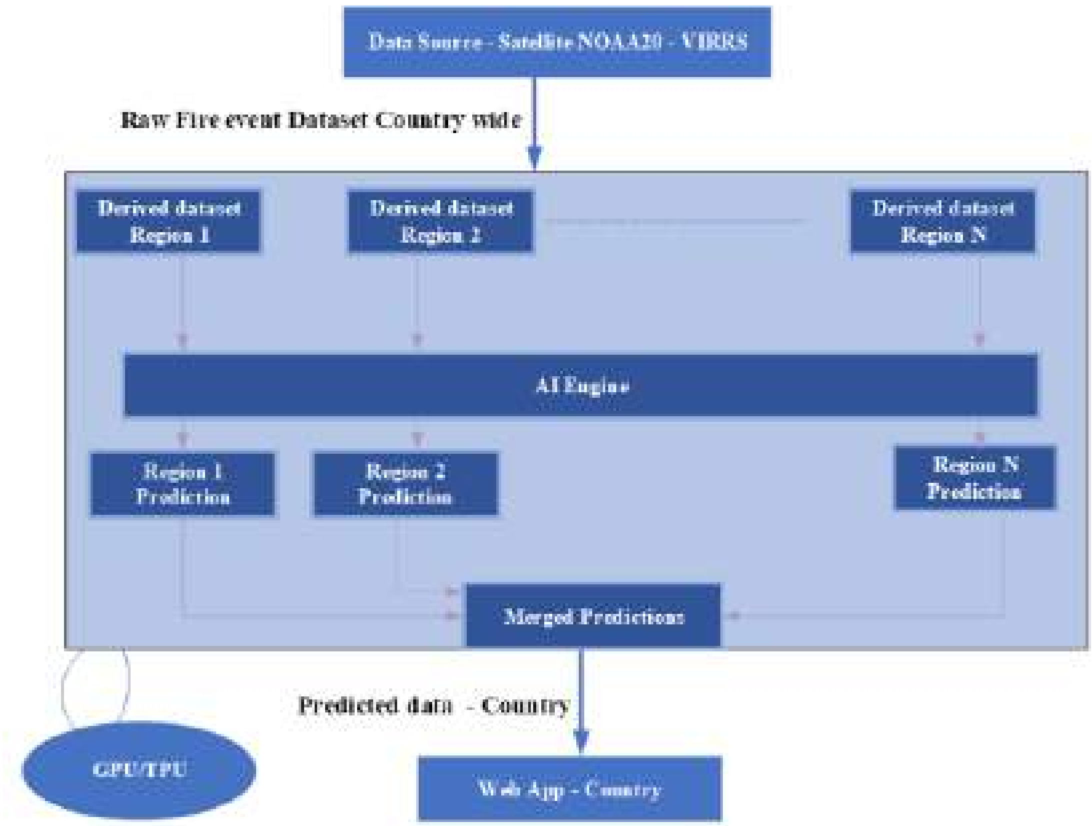


Figure 71
System Software Architecture - Performance and Accuracy Aspects - Application C Data Source VIRRS NOAA20

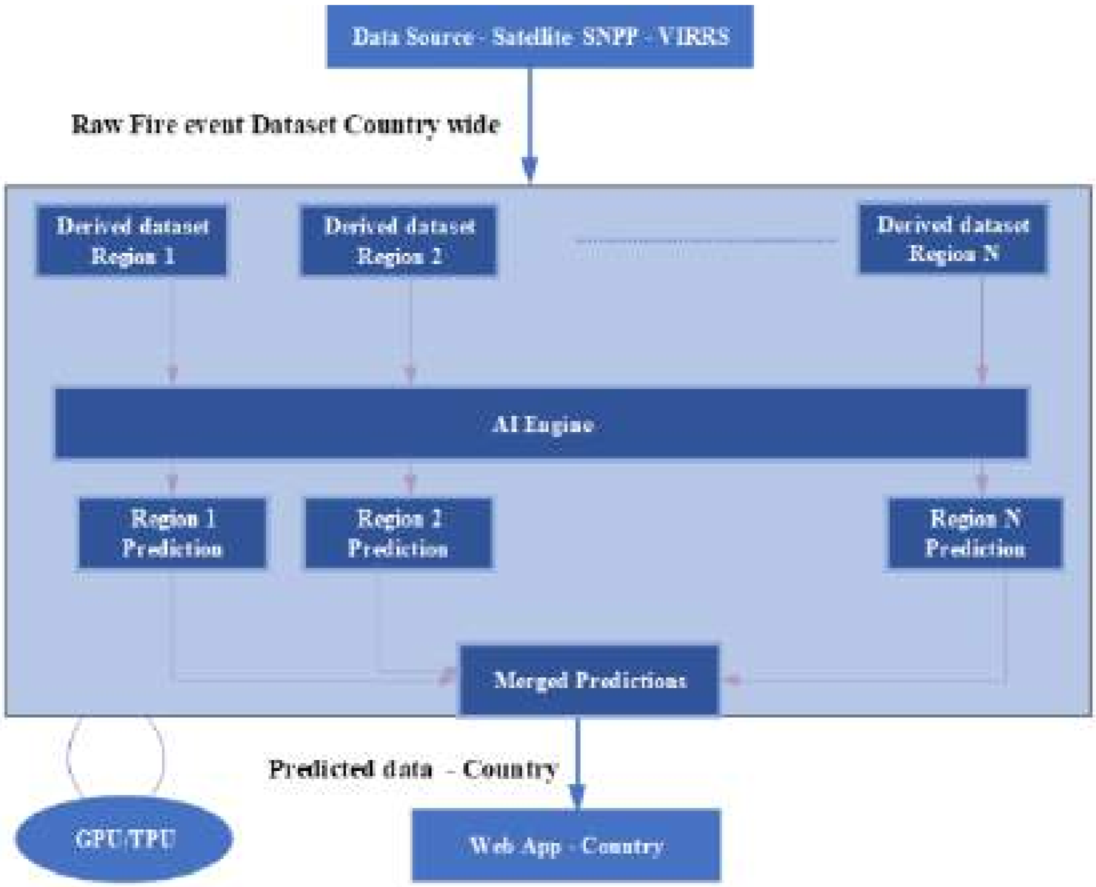


Figure 72
System Software Architecture - Performance and Accuracy Aspects -Application D Data Source VIRRS SNPP

Accuracy is another important factor for the fire industry. Though using the MODIS dataset, accuracy was achieved to a greater extent w.r.t to eliminating noises from the raw fire events dataset and by identifying growing fires, but due to the satellite blind spots and resolution, not all the fire growth can be detected by one instrument, there are possibilities that one among the other satellites SNPP or NOAA 20 or NOAA 21 that cover the USA could have observed those growing fires in initial stage, hence to achieve higher level of accuracy machine learning models are implemented in real/near real time dataset from other satellites. With the increased number of data points, the latency of the machine learning algorithms deteriorates to more than 60 minutes for predicting growing fires from four different data sources, including the raw fire event dataset, and for predicting contextual information for each detected growing fire scene.

To further improve latency, four separate applications are created, each of which is executed using a dedicated GPU, and each application independently processes datasets from its respective data source.

Figure 70 illustrates that Application A processes the dataset from the MODIS instrument installed on the satellites Aqua and Terra. Application A is associated with a single GPU; the machine learning model predicts all growing fires and their contextual information within 2 minutes. Similarly, Figure 71 illustrates that Application B processes the dataset from the VIRRS instrument installed on the NOAA-21 satellite. Application B is associated with one GPU; the machine learning model predicts all the growing fires and their contextual information within 2 minutes. Figure 72 illustrates Application C processing the dataset from the VIRRS instrument installed on the satellite NOAA-20. Application C is associated with a single GPU; the machine learning model predicts all growing fires and their contextual information within 2 minutes. Figure 73 illustrates Application D, which utilizes the dataset from the VIRRS instrument, installed on the SNPP satellite. Application D is associated with a single GPU, the machine learning model predicts all the growing fires and their contextual information within 2 minutes, when Applications A, B, C and D are scheduled parallelly to compute the prediction from their dedicated data sources, all the predicted growing fires for all the 4 data sources are calculated within 2 min, this latency is acceptable to fire industry application.

4.6 Research Question Six

What's the strategy to integrate the AI outputs, fire growth predictions, and additional contextual information into nationwide real-world applications in real time for the fire industry?

Figure 73
Web App Fire Growth Points

Figure 74
Web App - Fire Growth Point & Residence threat

Figures 74 and 75 represent the Web App that was built as part of this research. AI prediction is integrated into this web app. This web app primarily renders a 3D map of the USA, developed using Google Maps services and programming languages such as HTML, CSS, and JavaScript. This map displays only the growing fires across the country and marks the areas near residences under threat with pink circles. The traditional display comprises all the raw fire events from the satellite.

Figure 75
AI Integration to Web APP

Improvements in this software product are achieved by leveraging AI prediction. This map displays only growing fires in near real-time and provides additional contextual information by highlighting residences near the growing fires that require attention from firefighters. Satellite real-time raw fire events data is continuously read from the AI engine using the FIRMS web API services. The AI engine is scheduled to run periodically in Google Colab Pro, although another cloud platform can also be used. The AI engine continuously computes predictions and additional contextual information, writing them to the shared file system.

When the user accesses the web app through the URL, the Web App fetches the prediction from the file system and displays the growing fire prediction along with contextual information. Real/Near-real-time predicted fire events include the numeric values of latitude and longitude of the growing fire events, labeled as growing fires in the file system. Additionally, the predicted contextual information consists of the numeric values of latitude and longitude of residential information near the growing fires, labeled as "threat residence." The web application utilizes a distinct color palette and pattern to represent these elements based on their labels visually.

4.6 Summary of Findings

Unsupervised machine learning algorithms play a crucial role in detecting fire growth within our dataset, which contains two primary types of data:

- Type 1 Data (Noise): This category includes latitude and longitude coordinates of isolated fire events that lack nearby occurrences.
- Type 2 Data: This involves latitude and longitude coordinates of fire events that are closely and densely located.

Several important objectives guide the analysis:

- Accurate Identification of Type 2 Data: The algorithm should reliably identify Type 2 data on the day fire growth is observed.
- Clustering Separation: When multiple clusters of Type 2 data are present, the algorithm should adeptly distinguish between them.
- Exclusion of Type 1 Data: All Type 1 data must be excluded from active clusters exhibiting significant growth.
- Autonomous Cluster Determination: Given the unknown number of clusters, the clustering algorithm should automatically determine the number of existing clusters.
- Threat Level Assessment: The algorithm should evaluate the threat level for the nearest residences in areas prone to fire.

To meet these objectives, a variety of unsupervised machine learning algorithms were considered, including:

- Density-Based Spatial Clustering (DBSCAN)
- K-Means Clustering
- Fuzzy C-means Clustering
- Gaussian Mixture Models Clustering
- Agglomerative Hierarchical Clustering
- New Model: Multilevel Multicriteria Clustering Algorithm

While K-means clustering, Fuzzy C-means clustering, Gaussian Mixture Models, and Agglomerative Hierarchical Clustering are established techniques, they present some challenges in this context. These algorithms necessitate a predefined number of clusters, which is not known in our case. Additionally, they may struggle to classify Type 1 data as noise, leading to its unwarranted inclusion in clusters.

The accuracy of these algorithms can also vary, especially when small clusters are positioned near each other. Their outcomes are influenced by the specified 'number of clusters'; a lower setting may result in the formation of broader clusters, while a higher setting tends to create smaller, distinct ones.

Ultimately, this means K-means and similar methods may not effectively address objectives 3, 4, and 5. Moreover, objectives 1 and 2 may also fall short, as these algorithms often fail to successfully differentiate between noise and genuine fire spread event clusters, rendering their outputs less applicable in real-world scenarios.

In contrast, Density-Based Spatial Clustering has shown promise in successfully addressing the key objectives (1, 2, 3, and 4) established for this study. This algorithm does not require a predetermined number of clusters; instead, it leverages two essential parameters: minimum cluster density and maximum distance between endpoints. Both parameters are derived through exploratory data analysis, allowing for a more adaptive and practical approach to clustering in this context. However, the results produced by this algorithm may exhibit inconsistencies, particularly in cases where multiple nearer fire growing points exist within the dataset. Furthermore, it does not satisfy Objective 5. In comparison, the newly proposed model achieves a high prediction accuracy of 95% and effectively meets all specified objectives.

When these algorithms were tested on the nationwide dataset, their performance was found to be suboptimal. Consequently, the dataset was segmented into smaller subsets based on regional divisions. The machine learning algorithms were employed to predict fire growth within each regional dataset, and these predictions were subsequently integrated to form a comprehensive nationwide forecast. This approach resulted in improved model performance.

Similarly, the performance of the machine learning model declined when tasked with predicting contextual information about the fire-growing scene, specifically assessing the threat level to nearby residences from the growing fire. To enhance performance, the algorithms first identified residences at risk at the regional level and then consolidated these regional predictions to generate a national-level assessment. This integration of predictions into real-world applications further increased the effectiveness of the model.

The proposed software architecture in the research enhances accuracy. It reduces latency by logically dividing a large application into multiple smaller applications and scheduling the GPU for these independent applications. This approach minimizes latency, as each application is dedicated to processing data from a separate source. The results of these independent applications are then integrated, thereby improving the accuracy of the predictions. This research presents a method for continuously integrating and deploying AI-predicted fire growth in real-time, raw fire event data from MODIS and VIRRS, along with predicted contextual information of the fire scene, leveraging a cloud platform. A web app was also developed as part of this research, alongside the study. AI prediction is integrated into this web app in real time. This web app primarily renders a 3D map of the USA This map displays AI-predicted growing fires across the country and marks the predicted near residences under threat with pink circles.

CHAPTER V:

DISCUSSION

5.1 Discussion of Research Question One

Are there any hidden patterns of growing fires in the collected raw data from the history dataset captured from the satellite?

Based on the findings from the exploratory data analysis of fire events obtained from the MODIS instruments aboard the Aqua and Terra satellites, the following insights emerged from the time series analysis of fire event geographic coordinates:

1. On the day the fire originated, there was a high density of fire events concentrated in a specific area.
2. In the days that followed, these fire events expanded to cover a larger area, leading to an increase in the number of geographic coordinates reflecting the fire events.
3. Visualizing the geographic coordinates of these fire events on a 2D plot for the day of the incident revealed that the coordinates were densely clustered, forming a darker, thicker area on the regional map of the USA, specifically around the known site of the fire. Refer to figures 39, 40, and 41.
4. The MODIS dataset indicated that numerous geographical points, greater than five and within a 13-mile radius, demonstrated fire growth in the following days. This observation was particularly evident in the California region during two real incidents, the Palisades Fire and the Eaton Fire, detailed in Table 14. A similar pattern of increasing fire density was observed for the Idaho region on the incident day, as illustrated in Figure 44.
5. When examining the 2D plots of fire event coordinates from the day before the incident, the events appeared isolated, with distances exceeding 13 miles between

them. As shown in Figures 40, 41, 43, and 44, these isolated fire events did not present as dense spots in the 2D map, as there were no more than five events close together.

The significance of utilizing satellite datasets lies primarily in their scalability on a global level, which streamlines data preprocessing and enhances the efficiency of AI models. The standardized format of these datasets enables consistent application of AI methods across different regions and contexts. Although satellites can occasionally miss early fire detections due to limitations such as sensor resolution or blind spots, they tend to provide more reliable indicators when fire events exhibit a growing pattern. This correlation increases confidence that these events represent developing wildfires, allowing for better monitoring and response strategies.

Performing exploratory data analysis on the MODIS dataset is significant for several reasons. A key aspect is that it helps determine the most suitable machine learning algorithms for the dataset. Additionally, it facilitates the selection of a relevant subset of historical data, specifically those records of raw fire events that span over a decade. This targeted approach is essential because running algorithms on the entire dataset can be pretty time-consuming. By focusing on a carefully chosen subset that encapsulates the patterns and behaviors of past raw fire events, we can ensure more reliable results when applying the model to unseen data. Ultimately, this enhances the model's effectiveness in real-time applications, aiding in the timely prediction of wildfire occurrences.

Existing literature utilizes satellite images of wildfires gathered from various resources, including Google Images, open-source initiatives, and Kaggle, as well as data from MODIS, Sea and Land Surface Temperature Radiometer, Visible Infrared Imaging Radiometer Day Night Band, and SLSTR. These datasets are manually labeled as either wildfire or non-wildfire. A limitation in such literature is that it does not determine the characteristics of growing wildfires from satellite image datasets. (e.g., Mapulane, 2022; Rajalakshmi et al., 2023).

Some of the existing literature utilizes data, with the majority focusing on weather and environmental conditions, such as temperature, humidity, wind speed and direction, soil moisture content, and precipitation levels- all factors that significantly determine the likelihood of a fire. Additionally, the presence of vegetation, topography, and land-use patterns influences the risk of wildfires in a given area. It had a target parameter of fire or no fire, based on real wildfire incidents in that region. (e.g., Brennon et al., 2024).

The limitation of this dataset is that it focuses only on a particular region of the USA. The dataset is not scalable, which increases the effort required to generate the dataset for each state and nation. Revising the models for every new incident also requires significant effort to maintain the accuracy of these models.

Another existing literature also focuses on fuel parameters, Weather Parameters, Infrastructure, Topography, the Global Fire Atlas, and Fire Intensity from MODIS for the specific region—the southwestern border of China—for predicting wildfire characteristics. A limitation of this approach is that collecting data across all areas of the country and the globe is a challenging task. (e.g., Chen et al., 2023).

5.2 Discussion of Research Question Two

Which machine algorithm is accurate in predicting the fire growth and eliminating the non-growing fires from the raw fire events dataset on the day of discovery?

While answering Research Question 1, it was found that the data available from MODIS is not labelled; hence, unsupervised algorithms were implemented to predict growing wildfire events from the raw fire events dataset.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and proposed a new unsupervised clustering algorithm, the Multi-Criteria Clustering Algorithm, was able to predict the growing fire accurately and eliminate isolated fire events, i.e., noise from the predicted growing fire cluster. Other unsupervised algorithms, including K-Means, Fuzzy C-Means, Gaussian Mixture Models, and Agglomerative Hierarchical Clustering, were also experimented with; however, they were unable to eliminate the isolated fire events from the clusters accurately.

These algorithms are validated using three sample datasets from the recent MODIS history dataset, two sample datasets for the California region, and a third sample dataset for the Idaho area in the United States.

For the sample datasets 1, 2, and 3, DBSCAN results are indicated in the Figures 60,61,62. DBSCAN was able to cluster all the fire-growing events on the day the fire was discovered and also on the subsequent day when they exhibited a growing pattern. This algorithm was able to effectively mark 100% isolated fire events as noise on the previous day of the fire growth. Similar behavior was observed in all three sample datasets, as indicated in Figures 63, 64, and 65, by the newly proposed algorithm, the Multi-Criteria Clustering Algorithm. One difference between the DBSCAN and the Multi-Criteria Clustering Algorithm is that, when there are multiple clusters of growing fires on the same day, the Multi-Criteria Clustering Algorithm can consistently cluster the growing fire events into the same cluster. Every execution of this algorithm on the same dataset produces the same output, but DBSCAN results were not consistently repeatable. The border fire event of one cluster is often transferred between two neighboring clusters. However, this algorithm accurately classifies the growing fire events. Still, it fails to provide the same repeatable output, as it frequently moves the border of growing fire points between the two nearest clusters. This inconsistency is unacceptable for use in a real-world application, particularly when deriving more contextual information for the detected fire scenes. The Multi-Criteria Clustering Algorithm is identified as an accurate model for determining fire growth in the MODIS dataset.

Existing literature utilizes labeled numeric data; therefore, they employ supervised machine learning models to predict wildfire characteristics. Random Forest performed better compared to Decision Trees, KNN, Support vector machine, logistic regression, and Naive Bayes on the labeled numeric dataset that comprised of the weather and environmental conditions, such as temperature, humidity, wind speed and direction, soil moisture content, and precipitation levels, the presence of vegetation, topography, and land-use patterns influences. The support vector machine was eliminated because it required excessive time for prediction (e.g., Chen et al., 2023).

Another existing literature also recommends Random Forest for predicting wildfire characteristics over other machine learning models, such as extreme gradient boosting. They also had a similar type of region-specific labeled numeric dataset with a few additional parameters, such as the Global Fire Atlas, which contains historical locations, dates, Rates of speed between 2003 and 2016, and fire intensity data from MODIS, along with weather and environmental conditions parameters (e.g., Brennon et al., 2024).

A limitation of the existing literature is that although the recommended supervised algorithm, random forest, demonstrated higher accuracy on the test dataset, which is historically collected, it has not been tested on real-time data. The authors suggested that enhancing the dataset with local vegetation data, such as chaparral, grassland, or coniferous forests, as well as incorporating regional population data, could further improve the accuracy of the machine learning algorithm. However, a significant limitation remains with the dataset, as the random forest was only tested for one specific region, using an input configuration parameter of max depth 20, defined solely for this regional California dataset in the USA, to achieve the highly accurate results. In contrast, another study described a maximum depth of 10 for Yunnan Province, China, while the maximum depth for other regions in the country remains undefined. This ultimately hampers the scalability of the model at both national and global levels.

5.3 Discussion of Research Question Three

Does the accuracy of the machine algorithm in predicting the growing fires vary on real-time fire events data (unseen data)?

DBSCAN and the Multilevel Multicriteria Clustering Algorithm were tested on real-time data over several days, fetching raw MODIS fire event data directly from the FIRMS web API. The accuracy in predicting wildfire growth remained consistent on both real-time data and unseen data, as illustrated in Figures 66 and 74.

The input parameter for minimum cluster density points was set to 5 during experiments on the historical dataset, which did not necessitate any changes for the real-time data, as shown in Figure 74. The analysis indicated a growing fire predicted across all regions of the United States based on real-time data from various states.

Additionally, both algorithms were tested using raw fire events from alternative sources, including VIRRS data from the NOAA-20, NOAA-21, and SNPP satellites. Real-time data was directly retrieved from the FIRMS web API for all areas in the U.S. For the raw fire event dataset from VIRRS, the minimum cluster density points parameter was adjusted to 10 from the initial 5 to enhance the accuracy of wildfire growth predictions.

Overall, the findings indicate that the accuracy of the algorithms remains high for both real-time data and unseen datasets, provided the source remains the same. This shows a robust model for predicting wildfire growth across various regions of the country and potentially for datasets from other countries as well.

The only adjustment necessary pertains to the use of different sensors based on their resolution. The study highlighted that when utilizing data from different sensors, the minimum cluster density should be modified accordingly. Higher-resolution sensors require a higher minimum cluster density to ensure accurate predictions, whereas applying the same lower resolution settings could lead to inaccuracies. Thus, when incorporating data points from new satellite-based sensors, it is essential to make minimal adjustments to maintain model accuracy. Once calibrated for one region, the same input values can be applied at a global level.

By answering this question, this research has demonstrated its significance by adapting the proposed unsupervised AI model framework to function effectively with any satellite-based sensor raw fire numeric dataset, thereby predicting wildfire spreading characteristics. The approach has been proven to achieve higher accuracy, making it applicable to various regions worldwide.

Existing literature emphasizes that the random forest algorithm achieves higher accuracy on test datasets compared to other supervised learning models when working with labeled numeric data. However, there is a notable gap in the experimentation of these models using real-time data for the same region or other regions of the country. While some studies have employed Convolutional Neural Networks (CNNs) and customized CNN models for predicting wildfires using image datasets, they still struggle to demonstrate accuracy in predicting wildfire characteristics in real-time or with unseen data (e.g., Brennon, 2024; Chen, 2023; Mapulane, 2022; Rajalakshmi et al., 2023).

5.4 Discussion of Research Question Four

Can machine learning algorithms predict more contextual information about fire scenes in areas expected to experience fire growth in near real-time, such as the threat level to nearby residences from the growing fire?

The chosen Multilevel Multicriteria Clustering Algorithm predicts more contextual information about the fire scene. This prediction is shown in Figure 75, which represents the predicted latitude and longitude of the threat residence in pink.

Predicting contextual information about fire scenes plays a vital role in assessing the threat levels of wildfires to human safety. As wildfires continue to escalate, it becomes increasingly critical for firefighters to prioritize containment efforts to mitigate risks to life.

Additionally, the proposed Multilevel Multicriteria Clustering Algorithm framework (Refer Section 3.8.7) offers an easily customizable solution for gathering more contextual details about the fire scene, which is essential for effective wildfire management.

Existing literature has not addressed the prediction of contextual information regarding growing fire scenes, as most studies have primarily focused on wildfire characteristics or detection. (e.g., Brennon, 2024; Chen, 2023; Mapulane, 2022; Rajalakshmi et al., 2023). Another existing literature that utilizes AI for other applications in the fire industry recommends training the AI model to recognize additional context in fire scenes for future work, such as different types of vehicles, fire hydrants on streets, and specific uniforms worn by firefighters, Emergency Medical Technicians, and police officers. The fire commanders could quickly and precisely grasp the AI-predicted critical information (e.g., number of fire apparatus) on-site. Continuously monitoring firefighting activities onsite and signs of fatigue in firefighters.

For instance, when a firefighter's helmet touches the ground, it is a clear sign of fatigue, and the AI software immediately notifies other firefighters on the ground. (e.g., Chang, 2022).

5.5 Discussion of Research Question Five

Is the accuracy of the machine learning algorithms in predicting the fire growth and the latency in predicting the fire growth acceptable to real-world applications in the fire industry?

The proposed system software architecture, depicted in Figures 70, 71, 72, and 73, outlines the framework for implementing unsupervised machine learning models on national-level datasets and from multiple satellite sensors to enhance accuracy suitable for real-world applications in the fire industry on a cloud platform. It also provides guidance on cost-effectively improving prediction latency, ensuring that the implementation of machine learning models meets the necessary efficiency for practical use in this sector. By utilizing increased data points, the architecture aims to achieve the desired accuracy levels essential for these applications.

The wildfire growth initially starts small, and by employing this proposed architecture, it is possible to achieve wildfire growth prediction across the entire country using raw fire event data obtained from multiple satellite sensors and contextual information of the fire scene within 2min, as it does not immediately result in catastrophe or jeopardize life safety, predicting with in 2min should be acceptable for the fire industry.

The latency requirement for the fire alarm, as specified by the National Fire Protection Association (NFPA), was reviewed. NFPA is a global, self-funded, non-profit organization dedicated to eliminating death, injury, property damage, and economic loss due to fire, electrical, and related hazards, according to multiple sources. NFPA develops and publishes consensus-based codes and standards that are widely used to prevent and mitigate these hazards, over 300 codes and standards that address various aspects of fire and electrical safety, building design, hazardous materials, and more. These standards are developed through a consensus process involving technical committees and subject matter experts. Code NFPA 72 pertains to fire detection, signalling, and emergency communication systems. This code outlines the requirements for fire alarm systems, encompassing design, installation, inspection, testing, and maintenance, for all types of buildings. The primary goal is to protect life and property from fire and related hazards. According to NFPA 72, fire alarm signals must be received and confirmed at a central monitoring station/remote station within 90 seconds.

As wildfires generally start small and do not immediately impact life safety, predicting their growth and providing contextual information within 2 minutes should be acceptable for the fire industry, even from a regulatory standpoint, as we closely achieve NFPA code specifications for building fires. Building fire detection time is more aggressive in the NFPA code, as it immediately affects life and property.

Additionally, while reviewing the existing literature on wildfire characteristic prediction, researchers often fail to integrate their findings into practical, real-world applications; instead, most of them suggest these integrations as future work (e.g., Brennon, 2024; Chen, 2023; Mapulane, 2022; Rajalakshmi et al., 2023).

Some existing literature that focuses on the application of AI in other areas of the fire industry has demonstrated effective integrations of AI predictions into real-world applications (e.g., Akmalbek 2022).

For instance, they ran an AI model on computers equipped with 2 GPUs, serving as a server that utilizes images sent from IoT devices to predict fires using an image-based AI model. They found that using 2 GPUs enables fire predictions from images and can send notifications within 0.83 seconds. However, a limitation of this literature is the lack of a clear illustration of the location of the AI server computer, its load capacity, or the number of buildings or coverage areas it can manage.

If the AI model takes 0.83 seconds to predict fire from a single image, predicting fire from the entire city's building camera images would become computationally expensive. Utilizing 2 GPU-based servers for the town as a whole could lead to several hours required to complete the predictions. Thus, the methods proposed in the existing literature currently do not operate effectively on a large scale and can be very expensive to operate with a larger number of GPUs.

Additionally, this fixed GPU-based system, which is used for achieving real-time response, may not be suitable, as there will often be a need to derive more contextual information about the fire scene. A fixed GPU-based system typically requires hardware upgrades, which can be expensive, especially when incorporating additional AI-based features into the existing product.

5.6 Discussion of Research Question Six

What's the strategy to integrate the AI outputs, fire growth predictions, and additional contextual information into nationwide real-world applications in real time for the fire industry?

The AI engine resides on a cloud computing platform. It is scheduled to run AI models continuously to read raw fire events in real-time, as detected by sensors on a satellite through the FIRM's web API services, preprocess the data, and make predictions. This predicted output of fire growth and contextual information is integrated in real-time into the shared file system and API.

For a real-world application, a web app has been built. This web app displays real-time AI-predicted data from the shared file system/API on a 3D map of the United

States. This web app can be accessed from any smartphone, iPad, tablet, or Personal computer using the URL. Firefighting organizations can utilize this web app for improved wildfire management.

While reviewing the existing literature on wildfire characteristic prediction, researchers often fail to integrate their findings into practical, real-world applications; instead, most of them suggest these integrations as future work (e.g., Brennon, 2024; Chen, 2023; Mapulane, 2022; Rajalakshmi et al., 2023).

Some existing literature that focuses on the application of AI in other areas of the fire industry has demonstrated the effective integration of AI predictions into real-world applications in real-time (e.g., Chang 2022; Akmalbek et al., 2022).

In this literature, images of the fire scene are captured in real time from the cameras installed on the drone, vehicles, and firefighters, utilizing the Wi-Fi or satellite network these images are transferred to cloud server from camera, AI models on the cloud platform use the onsite images preprocess the data and predict the number of fire fighters in the fire scene and send the information to onsite incident commander which is used for decision making.

In another existing literature, they use a client-server scheme for integrating the AI prediction to real world application in real time, where the clients are used for collecting the data, smart glass that consist of camera and Home surveillance Camera are the clients in the client-server scheme, they are used for capturing images in the building, Smart glass sends image using Bluetooth to the smartphone(client), Smart phone and camera sends image to the AI server using the cellular or 5G, AI model in the AI server preprocess the image and predict the fire presence in the building and sends the notification in real time. (e.g., Akmalbek 2022).

Following Business Implications of AI-Driven Wildfire Growth Prediction

1. Enhanced Risk Management for Insurers

Insurance companies can leverage the predictive model to assess wildfire risks more accurately at both regional and national scales. Real-time identification of threatened residential zones enables dynamic adjustment of risk portfolios, more precise pricing of homeowner policies, and faster claims management after fire incidents.

2. Operational Efficiency for Firefighting Organizations

The integration of real-time predictions into a web application reduces uncertainty in field operations. Logistics managers can optimize deployment routes for vehicles, thereby lowering response times, fuel costs, and resource waste. This efficiency translates directly into cost savings while improving response capacity during critical incidents.

3. Technology-Enabled Public Safety Services

Emergency response agencies gain access to a practical decision-support tool that provides real-time situational awareness. This strengthens public trust and positions agencies to justify investments in advanced technology platforms. The reduced rate of false alarms further improves credibility and operational reliability.

4. Market Opportunities in SaaS and Cloud Solutions

The modular, GPU-optimized architecture offers a scalable business model for cloud-based SaaS solutions. Technology providers can commercialize this platform by offering subscription services to governments, NGOs, and private companies in forestry and land management. Expansion opportunities exist globally, particularly in regions prone to wildfires.

5. Data Monetization and Partnerships

The integration of MODIS and VIRRS real-time data streams opens avenues for data monetization. Partnerships with satellite data providers, insurance companies, and environmental consultancies could generate recurring revenue streams through API access, analytics dashboards, and custom risk reports.

6. Corporate Social Responsibility (CSR) and Brand Reputation

Enterprises adopting this technology, especially in insurance, utilities, and telecommunications, can enhance their CSR profiles by actively contributing to disaster prevention and community safety. Demonstrating proactive adoption of AI-driven wildfire prediction enhances stakeholder confidence and brand value.

7. Global Expansion Potential

While the study is applied to U.S. wildfire data, the underlying methodology is transferable to other regions with similar environmental risks (e.g., Australia, Mediterranean Europe, South America). This creates significant potential for global partnerships, licensing agreements, and technology exports.

CHAPTER VI:

IMPLICATIONS AND RECOMMENDATIONS

6.1 Implications

Performing exploratory data analysis on the MODIS dataset is significant for several reasons. A key aspect is that it helps determine the most suitable machine learning algorithms for the dataset. Additionally, it facilitates the selection of a relevant subset of historical data, specifically those records of raw fire events that span over a decade. This targeted approach is essential because running algorithms on the entire dataset can be pretty time-consuming. By focusing on a carefully chosen subset that encapsulates the patterns and behaviors of past raw fire events, we can ensure more reliable results when applying the model to unseen data/real-time data. Ultimately, this enhances the model's effectiveness in real-time applications, aiding in the timely prediction of wildfire occurrences. The significance of utilizing satellite datasets lies primarily in their scalability on a global level, which streamlines data preprocessing and enhances the efficiency of AI models. The standardized format of these datasets enables consistent application of AI methods across different regions and contexts. Although satellites can occasionally miss early fire detections due to limitations such as sensor resolution or blind spots, they tend to provide more reliable indicators when fire events exhibit a growing pattern. This correlation increases confidence that these events represent developing wildfires, allowing for better monitoring and response strategies.

This research has demonstrated its significance by adapting the proposed unsupervised AI model framework to function effectively with any satellite-based sensor raw fire numeric dataset, thereby predicting wildfire spreading characteristics. The approach has been proven to achieve higher accuracy, making it applicable to various regions worldwide. Additionally, the proposed Multilevel Multicriteria Clustering

Algorithm framework offers an easily customizable solution for gathering more contextual details about the fire scene, which is essential for effective wildfire management. This Novel clustering model makes a significant contribution to both theory and practical application within the fire industry.

An AI framework that leverages an efficient data derivation strategy and further employs cost-effective parallel and sequential computation methods to schedule GPUs and fulfill the real-time computation demand for practical applications, aligning closely with the regulatory timing requirements of the fire industry.

The practical application of the research is evident in a real-world product developed as part of this study: a web application designed to monitor wildfire characteristics and threat levels to residences in real-time. This monitoring system is particularly beneficial for wildfire management organizations, enabling them to make informed decisions, respond swiftly to emerging threats, and implement effective strategies for fire prevention and control. By leveraging advanced analytics and real-time data, these organizations can enhance their operational efficiency and improve safety outcomes for communities at risk.

6.2 Recommendations for Future Research

Future research can explore deriving more contextual insights of the fire scene that might be required for more effective wildfire management using AI methods.

Exploring the data captured by sensors on other satellites, such as Landsat, Orbital Tech, and the GOES-R Series, can further enhance the accuracy of predicting wildfire characteristics. Explore this data for additional use cases using AI in the fire industry. Additionally, examining the feasibility of deploying AI as a redundant system alongside a primary rule-based system will be crucial for managing high-risk areas within these industries.

REFERENCES

Abdusalomov, A.B., Mukhiddinov, M., Kutlimuratov, A. and Whangbo, T.K., 2022. Improved real-time fire warning system based on advanced technologies for visually impaired people. *Sensors*, *22*(19), p.7305.

Abid, F., 2021. A survey of machine learning algorithms based forest fires prediction and detection systems. Fire technology, 57(2), pp.559-590.

Ali, A.W. and Kurnaz, S., 2025. Optimizing Deep Learning Models for Fire Detection, Classification, and Segmentation Using Satellite Images. *Fire*, *8*(2), p.36.

Alkhatib, R., Sahwan, W., Alkhatieb, A. and Schütt, B., 2023. A brief review of machine learning algorithms in forest fires science. *Applied Sciences*, *13*(14), p.8275.

Ananthi, J., Sengottaiyan, N., Anbukaruppusamy, S., Upreti, K. and Dubey, A.K., 2022. Forest fire prediction using IoT and deep learning. *International Journal of Advanced Technology and Engineering Exploration*, *9*(87), pp.246-256.

Bergado, J.R., Persello, C., Reinke, K. and Stein, A., 2021. Predicting wildfire burns from big geodata using deep learning. *Safety science*, *140*, p.105276.

Bjånes, A., De La Fuente, R. and Mena, P., 2021. A deep learning ensemble model for wildfire susceptibility mapping. *Ecological Informatics*, *65*, p.101397.

Bot, K. and Borges, J.G., 2022. A systematic review of applications of machine learning techniques for wildfire management decision support. *Inventions*, *7*(1), p.15.

Chang, R.H., Peng, Y.T., Choi, S. and Cai, C., 2022. Applying Artificial Intelligence (AI) to improve fire response activities. *Emergency Management Science and Technology*, *2*(1), pp.1-6.

Chaturvedi, S., Thakur, P.S., Khanna, P., Ojha, A., Song, Y. and Awange, J.L., 2025. Satellite Image-Based Surveillance and Early Wildfire Smoke Detection Using a Multiattention Interlaced Network. *IEEE Transactions on Industrial Informatics*.

Chen, R., Li, Y., Fan, C., Yin, J., Zhang, Y., He, B. and Zhang, Q., 2023, July. Modeling Potential Wildfire Behavior Characteristics Using Multi-Source Remotely Sensed Data: Towards Wildfire Hazard Assessment. In *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium* (pp. 2366-2369). IEEE.

Cheng, S., Jin, Y., Harrison, S.P., Quilodrán-Casas, C., Prentice, I.C., Guo, Y.K. and Arcucci, R., 2022. Parameter flexible wildfire prediction using machine learning techniques: Forward and inverse modelling. *Remote Sensing*, *14*(13), p.3228.

Gain, M., Raha, A.D., Biswas, B., Bairagi, A.K., Adhikary, A. and Debnath, R., 2024, May. LEO Satellite Oriented Wildfire Detection Model Using Deep Neural Networks: A Transfer Learning Based Approach. In *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)* (pp. 214-219). IEEE.

Ghali, R., Akhloufi, M.A. and Mseddi, W.S., 2022. Deep learning and transformer approaches for UAV-based wildfire detection and segmentation. *Sensors*, *22*(5), p.1977.

Gholamnia, K., Gudiyangada Nachappa, T., Ghorbanzadeh, O. and Blaschke, T., 2020. Comparisons of diverse machine learning approaches for wildfire susceptibility mapping. Symmetry, 12(4), p.604.

Ghorbanzadeh, O., Valizadeh Kamran, K., Blaschke, T., Aryal, J., Naboureh, A., Einali, J. and Bian, J., 2019. Spatial prediction of wildfire susceptibility using field survey GPS data and machine learning approaches. *Fire*, *2*(3), p.43.

Giannakidou, S., Radoglou-Grammatikis, P., Lagkas, T., Argyriou, V., Goudos, S., Markakis, E.K. and Sarigiannidis, P., 2024. Leveraging the power of the Internet of Things and artificial intelligence in forest fire prevention, detection, and restoration: A comprehensive survey. *Internet of Things*, *26*, p.101171.

Gonçalves, L.A.O., Ghali, R. and Akhloufi, M.A., 2024. YOLO-Based models for smoke and Wildfire Detection in Ground and aerial images. *Fire*, *7*(4), p.140.

Grari, M., Idrissi, I., Boukabous, M., Moussaoui, O., Azizi, M. and Moussaoui, M., 2022. Early wildfire detection using machine learning model deployed in the fog/edge layers of IoT. *Indones. J. Electr. Eng. Comput. Sci*, *27*(2), pp.1062-1073.

Hahs, B., Sood, K. and Gomez, D., 2024, May. A Data-Driven Model for Wildfire Prediction in California. In *2024 International Conference on Smart Applications, Communications and Networking (SmartNets)* (pp. 1-6). IEEE.

Huot, F., Hu, R.L., Goyal, N., Sankar, T., Ihme, M. and Chen, Y.F., 2022. Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, pp.1-13.

Iban, M.C. and Sekertekin, A., 2022. Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces, Turkey. *Ecological Informatics*, *69*, p.101647.

Jain, P., Coogan, S.C., Subramanian, S.G., Crowley, M., Taylor, S. and Flannigan, M.D., 2020. A review of machine learning applications in wildfire science and management. *Environmental Reviews*, *28*(4), pp.478-505.

Kondylatos, S., Prapas, I., Ronco, M., Papoutsis, I., Camps-Valls, G., Piles, M., Fernández-Torres, M.Á. and Carvalhais, N., 2022. Wildfire danger prediction and understanding with deep learning. *Geophysical Research Letters*, *49*(17), p.e2022GL099368.

Makhaba, M. and Winberg, S., 2022, July. Wildfire path prediction spread using machine learning. In *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)* (pp. 1-5). IEEE.

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.

Mazzeo, G., Falconieri, A., Filizzola, C., Genzano, N., Pergola, N. and Marchese, F., 2025. Wildfire detection and mapping by satellite with an enhanced configuration of the Normalized Hotspot Indices: results from Sentinel-2 and Landsat 8/9 data integration. *IEEE Transactions on Geoscience and Remote Sensing*.

Mohapatra, A. and Trinh, T., 2022. Early wildfire detection technologies in practice—a review. *Sustainability*, *14*(19), p.12270.

Pang, Y., Li, Y., Feng, Z., Feng, Z., Zhao, Z., Chen, S. and Zhang, H., 2022. Forest fire occurrence prediction in China based on machine learning methods. *Remote Sensing*, *14*(21), p.5546.

Pham, B.T., Jaafari, A., Avand, M., Al-Ansari, N., Dinh Du, T., Yen, H.P.H., Phong, T.V., Nguyen, D.H., Le, H.V., Mafi-Gholami, D. and Prakash, I., 2020. Performance evaluation of machine learning methods for forest fire modeling and prediction. *Symmetry*, *12*(6), p.1022.

Prapas, I., Kondylatos, S., Papoutsis, I., Camps-Valls, G., Ronco, M., Fernández-Torres, M.Á., Guillem, M.P. and Carvalhais, N., 2021. Deep learning methods for daily wildfire danger forecasting. *arXiv preprint arXiv:2111.02736*.

Preeti, T., Kanakaraddi, S., Beelagi, A., Malagi, S. and Sudi, A., 2021, June. Forest fire prediction using machine learning techniques. In *2021 International Conference on Intelligent Technologies (CONIT)* (pp. 1-6). IEEE.

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.

Radke, D., Hessler, A. and Ellsworth, D., 2019, August. FireCast: Leveraging Deep Learning to Predict Wildfire Spread. In *IJCAI* (pp. 4575-4581).

Rajalakshmi, R., Sivakumar, P., Krishnakumari, L. and Subhashini, G., 2023, December. Satellite Image-Based Wildfire Detection and Alerting System Using Machine Learning. In *2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)* (pp. 1-5). IEEE.

Ramos, L., Casas, E., Bendek, E., Romero, C. and Rivas-Echeverría, F., 2024. Hyperparameter optimization of YOLOv8 for smoke and wildfire detection: Implications for agricultural and environmental safety. *Artificial Intelligence in Agriculture*, *12*, pp.109-126.

Reid, C.E., Jerrett, M., Petersen, M.L., Pfister, G.G., Morefield, P.E., Tager, I.B., Raffuse, S.M. and Balmes, J.R., 2015. Spatiotemporal prediction of fine particulate matter during the 2008 northern California wildfires using machine learning. *Environmental science & technology*, *49*(6), pp.3887-3896.

Rodrigues, M. and De la Riva, J., 2014. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software*, *57*, pp.192-201.

Sayad, Y.O., Mousannif, H. and Al Moatassime, H., 2019. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire safety journal*, *104*, pp.130-146.

Thangavel, K., Spiller, D., Sabatini, R., Amici, S., Sasidharan, S.T., Fayek, H. and Marzocca, P., 2023. Autonomous satellite wildfire detection using hyperspectral imagery and neural networks: A case study on Australian wildfire. *Remote Sensing*, *15*(3), p.720.

Xu, N., Lovreglio, R., Kuligowski, E.D., Cova, T.J., Nilsson, D. and Zhao, X., 2023. Predicting and assessing wildfire evacuation decision-making using machine learning: Findings from the 2019 kincade fire. Fire Technology, 59(2), pp.793-825.

Zaidi, A., 2023. Predicting wildfires in Algerian forests using machine learning models. *Heliyon*, *9*(7).

Zhang, L., Zhang, Q., Yang, Q., Yue, L., He, J., Jin, X., and Yuan, Q., 2025. Near-real-time wildfire detection approach with Himawari-8/9 geostationary satellite data integrating multi-scale spatial–temporal features. *International Journal of Applied Earth Observation and Geoinformation*, *137*, p.104416.