Forest Fire Predictions using Fuzzy Systems

Muhammad Haris Saad
LUMS, Lahore, Pakistan
24100055@lums.edu.pk

Taha Sarwat
LUMS, Lahore, Pakistan
25100287@lums.edu.pk

Abstract

Forest fires represent a significant threat to ecosystems and human lives, necessitating effective detection and management systems. In this study, we propose a comprehensive approach to forest fire detection utilizing a fuzzy rule-based system with Sugeno inference, an Adaptive Neuro-fuzzy Inference System (ANFIS), and Genetic Algorithms. Initially, a fuzzy rule-based system is developed, incorporating domain knowledge to establish rules for fire detection. Subsequently, the ANFIS model is employed to enhance detection accuracy by learning from a forest fire dataset, assessing key variables such as Fire Weather Index (FWI), Rain, Temperature etc. Performance metrics including accuracy, precision, and recall are computed to evaluate the efficacy of the ANFIS model. Furthermore, we introduce a novel approach integrating Genetic Algorithms with Fuzzy Logic Controllers to automatically generate fuzzy rules, enhancing the adaptability and robustness of the detection system. Through experimental validation and comparative analysis, our approach demonstrates promising results in forest fire detection, offering a potential solution for early warning systems and proactive fire management strategies.

1 Introduction

Forest fires represent a pervasive and multifaceted challenge with profound ecological, environmental, and socio-economic ramifications globally. As climate change accelerates, the frequency, intensity, and spatial extent of forest fires have escalated, posing unprecedented threats to biodiversity, ecosystems, and human communities. According to the Global Forest Watch, an estimated 4.46 million hectares of forest were lost to fires in 2020 alone, underscoring the urgency for effective fire detection and management strategies.

Traditional methods of forest fire detection have relied on human observation, satellite imagery, and remote sensing technologies. While these approaches have provided valuable insights, they often suffer from limitations such as latency, insufficient coverage, and susceptibility to environmental factors. To address these challenges, researchers and practitioners have increasingly turned to advanced computational techniques and artificial intelligence (AI) algorithms for more accurate and timely fire detection.

In this context, our research focuses on the development and evaluation of an integrated framework for forest fire detection, leveraging fuzzy logic, adaptive neuro-fuzzy inference systems (ANFIS), and genetic algorithms (GA). By harnessing the power of these computational tools, we aim to enhance the efficiency, accuracy, and reliability of forest fire detection systems, thereby mitigating the impact of wildfires on ecosystems and human lives.

The key objectives of this study include the following steps:

Designing a fuzzy rule-based system for forest fire detection, incorporating domain knowledge and expert heuristics in which we first make the rules by our own knowledge referencing with the dataset. Implementing ANFIS models to learn from historical forest fire data and improving the detection accuracy. Integrating genetic algorithms to automatically generate and refine fuzzy rules, enhancing the adaptability and robustness of the detection system. In the last we will be evaluating the performance of the proposed framework in terms of accuracy, precision and recall through comparative analysis for better understanding.

Through a comprehensive investigation and empirical evaluation of our proposed framework, this research aims to contribute to the advancement of forest fire prediction technologies and foster a deeper understanding of fire dynamics. By providing insights into the effectiveness and efficiency of computational methods in forest fire management, we aspire to empower decision-makers with actionable intelligence for proactive fire prevention and emergency response strategies.

2 Literature Review

As this topic is of great importance, we must first also explore a reasonable amount of work that already exists in this field. By doing this, we get a better idea of how much work has already been done, and if there are any other possible contributions to be made, with the addition of getting a better feel for just how many possible solutions exist to a problem of such relevance.

Machine Learning and Artificial Intelligence have been extensively studied in the context of forest fire detection. Various research papers have explored the use of Machine Learning algorithms and Deep Learning methods for predicting and detecting forest fires. These studies highlight the potential of AI-based systems in enhancing forest fire detection, which is crucial for effective forest fire management and mitigation efforts. As such, we have used fuzzy systems to aid in forest fire detection.

The following are a number of research papers that align with our study, and highlight the usage of the aforementioned AI techniques specifically in the context of the problem of forest fire detection.

2.1 Machine Learning

An insightful review[6] on the Machine Learning techniques used in forest fires detection showed that several studies have explored the use of Convolutional Neural Networks (CNNs), Radial Basis Function Networks (RBFNs), and Multilayer Perceptron (MLP), for early fire detection using data from wireless sensor networks, UAV imagery, and video-based systems, with reported detection accuracies ranging from 81% to 98%. Machine Learning models, including Logistic Regression, Random Forest, and Artificial Neural Networks, have been used to predict the likelihood of fire occurrence by analyzing factors such as weather, vegetation, topography, and human activities, with the reported prediction accuracies ranging from 74% to 94%. Lastly, Machine Learning algorithms, including Random Forest, Support Vector Machines, and Logistic Regression, have been employed to map fire severity, burned areas, and fire susceptibility using remote sensing data, such as satellite imagery and UAV-based data, and the reported mapping accuracies range from 79% to 98%.

One similar review[4] categorized Machine Learning into three main types: supervised learning (decision trees, random forests, boosted ensembles, support vector machines, and artificial neural networks/deep learning), unsupervised learning (clustering algorithms like k-means and Gaussian mixture models), and agent-based learning (reinforcement learning). The authors conducted a scoping review of 300 relevant publications on ML applications in wildfire science and management up to the end of 2019. The publications were categorized into six broad problem domains: (i) fuels characterization, fire detection, and mapping; (ii) fire weather and climate change; (iii) fire occurrence, susceptibility, and risk; (iv) fire behavior prediction; (v) fire effects; and (vi) fire management. The review highlights the diverse and challenging problems in wildfire science that are amenable to ML approaches; opportunities exist to apply more advanced Machine Learning methods like deep learning, especially for large multivariate datasets. Expertise in wildfire science is necessary to ensure realistic modeling of fire processes across scales, and the wildfire community should play an active role in providing high-quality, freely available data.

2.2 Deep Learning

Diving into the realm of Deep Learning in the context of forest fires detection, a system to detect desert/forest fires[2] was proposed alongside the creation of a new desert fire detection dataset, the Utah Desert Fire dataset, to address the lack of representation in publicly available datasets, and modified transfer learning approaches using Xception and ResNet-50 architectures for improved fire detection performance. The study provides a detailed analysis of the modified Xception and ResNet-50 architectures, including their layer structures, training parameters, and performance, alongside extensive simulations and hyperparameter tuning to optimize the performance of the proposed models.

Forest Guard[7], was a system developed using integrated sensors and AI for the mapping of fire-prone areas in order to aid in early forest fire detection. It utilized a network of low-cost sensors to monitor various parameters, enabling efficient fire prediction and detection, and then integrates sensor data with AI algorithms to predict fire-prone areas and detect wildfires early. It was tested in multiple controlled burn experiments, demonstrating the ability to detect fires within 2-3 minutes of ignition. The sound analysis model was able to detect the fire within 2 minutes, and the confirmatory alarm was raised within 4 minutes. The system was also tested in a real-world controlled burn scenario, where it was able to detect the fire within 2 minutes and 3 seconds, and raise the confirmatory alarm within 5 minutes and 42 seconds.

2.3 Convolutional Neural Networks

In a paper published by IEEE[1], the use of Convolutional Neural Networks has been explored, primarily for the detection of forest fires via image classification. The goal is to verify whether a forest fire is visible by the CNN in a given picture or not. The dataset used to train this model was divided into photos that contained a fire and photos that did not contain a fire. In the methodology for this paper, APIs present in Keras and TensorFlow were used to efficiently develop and train their CNN. In addition to high accuracy, this model also works on low resolutions and frame rates, making it a very suitable pick.

Put forth by a team of researchers from India, a study[3] was done that was focused on early identification of wildfires, employing techniques such as median filtering to remove noise from images containing forest fires, color features extraction to aid in segmentation and classification of fires, and a fire recognition system using a CNN with maximum pooling layers to accurately predict whether a forest fire is present in a given image or not. This data would then be passed on to an alert system which would take action accordingly.

A similar study done by another team of researchers from India[5] proposed an early forest fire detection surveillance system that would help in reacting quickly to it. To this end, the CNN they proposed consisted of nine layers to allow it to classify images with and without a fire in them. The proposed accuracy of their CNN is 96.71%.

3 Methodology

We took four main steps in devising a system for predicting forest fires. These include preprocessing of the dataset, developing a fuzzy rule based system by making fuzzy rules, making an ANFIS and using a Genetic Algorithm on the dataset.

3.1 Preprocessing

We preprocessed the dataset to ensure its suitability for analysis and model development. The dataset comprises various attributes, including temporal and meteorological variables, alongside the target variable indicating fire occurrence.

Prior to analysis, exploratory data analysis (EDA) was conducted. This function facilitated a comprehensive overview of the dataset, including its structure, summary statistics, missing values, and unique values. Additionally, a correlation matrix was generated to assess the relationships between variables, aiding in feature selection and model interpretation. Moreover, categorical attributes such as day and month were encoded appropriately to facilitate their inclusion in subsequent analyses. The target variable, indicating fire occurrence, was encoded as binary, with '0' representing no fire and '1' indicating the presence of a fire event.

3.2 EDA

This section discusses on the techniques we employed in the EDA section of our project. We carried out an extensive Exploratory Data Analysis on our Algerian forest fires dataset.

3.2.1 Histograms

Histograms are particularly useful when analyzing datasets related to forest fires, such as the Algerian forest fire dataset, as they provide a visual representation of the distribution of various parameters like temperature, humidity, wind speed, and other environmental factors. By using histograms, we can gain insights into the frequency and distribution of these parameters, helping us understand their variability and potential impact on forest fire occurrences.

3.2.2 Scatter Plots

Scatter plots are instrumental when examining the relationship between key variables like temperature, relative humidity (RH), rainfall, DMC (Duff Moisture Code), and DC (Drought Code) in forest fire datasets like the Algerian forest fire dataset.

By plotting these variables against each other, we can visually assess any potential correlations or patterns, aiding in the identification of factors that contribute to forest fire occurrences. This graphical approach allows for a comprehensive exploration of how variations in these parameters interact and influence the likelihood of forest fires, enhancing our understanding of their interplay in fire prediction models.

3.2.3 Box Plots

Box plots are valuable tools in exploring the distribution and variability of key variables in forest fire datasets like the Algerian forest fire dataset. These plots provide a concise summary of the data's central tendency, dispersion, and any potential outliers, offering insights into the range and spread of each variable. By visually comparing the box plots of different variables, we can identify variations in their distributions and potential relationships, aiding in the identification of factors contributing to forest fire occurrences and informing predictive modeling efforts.

3.2.4 Outlier Detection

Outlier detection is crucial when analyzing forest fire datasets like the Algerian forest fire dataset because it helps identify data points that deviate significantly from the expected patterns or distributions. By identifying outliers in key variables such as temperature, relative humidity (RH), rainfall, DMC (Duff Moisture Code), and DC (Drought Code), we can gain insights into anomalous environmental conditions or measurement errors that may affect the accuracy of predictive models. Removing or addressing outliers appropriately ensures that the modeling efforts are based on reliable and representative data.

3.2.5 Principle Component Analysis

Principal Component Analysis (PCA) is employed in the analysis of Algerian forest fire datasets to reduce the dimensionality of the data while preserving its essential information, PCA helps in identifying underlying patterns and correlations among the dataset variables. By transforming the dataset into a lower-dimensional space defined by principal components, PCA facilitates visualization, interpretation, and modeling of complex relationships within the dataset. This technique aids in identifying the most influential variables contributing to forest fire occurrences.

3.2.6 K Means Clustering

K-means clustering is applied to the Algerian forest fire dataset to identify distinct groups or clusters within the key variables. This unsupervised learning technique helps in uncovering natural groupings or patterns in the data, which can provide valuable insights into the factors influencing forest fire occurrences. By partitioning the dataset into clusters based on similarity in variable values, k-means clustering aids in understanding the underlying structure of the data and can assist in delineating different risk zones for forest fires based on environmental conditions.

3.2.7 Mean Shift Clustering

Mean shift clustering is employed on the Algerian forest fire dataset to uncover inherent patterns and cluster the key variables. This technique is advantageous for its ability to automatically determine the number of clusters without prior specification, making it suitable for exploring complex and unknown structures in the data. Mean shift clustering is particularly useful for detecting dense regions or peaks in the data distribution, which can help in identifying areas with similar environmental conditions that are prone to forest fires.

3.2.8 Innovation - Birch Clustering

BIRCH clustering is applied to the Algerian forest fire dataset to efficiently handle large volumes of data and identify clusters among the key variables. This technique is advantageous for its scalability and ability to incrementally cluster data points, making it suitable for datasets with high dimensionality and diverse data distributions. BIRCH clustering helps in summarizing the data by creating a hierarchical structure of clusters, which aids in understanding the underlying patterns and spatial distribution of forest fire-prone regions based on environmental conditions.

3.2.9 Innovation - DBSCAN Clustering

DBSCAN clustering is employed on the Algerian forest fire dataset to identify spatial clusters within the key variables. This approach is particularly useful for detecting clusters of varying shapes and sizes, as well as for handling noise and outliers present in the dataset. By leveraging the density-based nature of DBSCAN, it helps in uncovering regions with similar environmental conditions.

3.2.10 Innovation - Fuzzy C Means Clustering

Fuzzy C-means clustering is applied to the Algerian forest fire dataset to capture the inherent uncertainty and fuzziness in the key variables, including temperature, relative humidity (RH), rainfall, DMC (Duff Moisture Code), and DC (Drought Code). Unlike traditional clustering methods, Fuzzy C-means allows data points to belong to multiple clusters simultaneously, reflecting the gradual transition between different environmental conditions conducive to forest fires. This approach is beneficial for identifying overlapping regions of fire risk and for providing a more nuanced understanding of the spatial distribution of fire-susceptible areas in the dataset.

3.3 Fuzzy Rule Based System

We implemented a fuzzy logic-based fire risk prediction system utilizing the Sugeno inference method. The system was designed to assess the likelihood of fire occurrence based on various environmental factors. These factors included Temperature (Temp), Relative Humidity (RH), Wind Speed (Ws), Rainfall (Rain), and indices such as Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI), Buildup Index (BUI), and Fire Weather Index (FWI). Each variable was assigned linguistic terms such as "low," "medium," and "high," and associated with appropriate membership functions to capture the fuzzy nature of the input data.

Furthermore, the fuzzy rule base was established to govern the inference process. These rules defined the relationships between the input variables (antecedents) and the output variable (consequent), specifying conditions such as "if [antecedent], then [consequent]." By combining multiple rules, the system could make predictions regarding fire risk based on the current environmental conditions.

The Sugeno inference method was employed for defuzzification, a process by which fuzzy sets are converted into crisp values. This involved computing the weighted average of the consequent variables based on the degree of membership of the antecedents.

Implementation details encompassed the setup of the fuzzy inference system and the initialization of variables with specific values derived from input data. We utilized the simpful library in Python to facilitate the implementation process, leveraging its capabilities for fuzzy logic modeling and inference.

3.4 Adaptive Neuro Fuzzy Inference System

We extended our fire risk prediction system by integrating an Adaptive Neuro Fuzzy Inference System (ANFIS). ANFIS represents a hybrid approach that combines the strengths of neural networks and fuzzy logic to automatically generate fuzzy rules from the dataset. This integration aimed to enhance the predictive capabilities of our system by leveraging the flexibility of fuzzy logic and the learning ability of neural networks.

The implementation of ANFIS involved several key steps, including data preprocessing, model training, and ANFIS architecture design. Initially, the dataset containing historical environmental variables like FWI Index etc and corresponding fire occurrence data was prepared for input into the ANFIS model. This involved cleaning the data, handling missing values, and partitioning it into training and testing sets to facilitate model training and validation.

Subsequently, the ANFIS model was trained using the training dataset to learn the underlying relationships between the input variables and the output (fire occurrence). During the training process, the model automatically generated fuzzy rules based on the provided dataset, adapting its structure and parameters to minimize prediction errors. The architecture design of ANFIS involved defining the membership functions, and the structure of the hybrid system. Parameters such as the number of input variables, linguistic terms, and the type of fuzzy inference system were determined based on the characteristics of the dataset and domain knowledge.

This hybrid ANFIS system's ability to automatically generate fuzzy rules was highlighted as a significant advantage, eliminating the need for manual rule creation and tuning. This feature enhanced the scalability and adaptability of the system, allowing it to accommodate diverse datasets and environmental conditions without requiring extensive human intervention.

3.5 Genetic Algorithm

We integrated a Genetic Algorithm (GA) into our fire risk prediction system to optimize the fuzzy rules generated by the ANFIS model. This integration aimed to optimize the performance of the fuzzy inference system by iteratively improving the rule set to better capture the complex relationships between input environmental variables and fire occurrence.

The implementation of the Genetic Algorithm involved several key steps, beginning with the initialization of a population of candidate rule sets. Each rule set represented a potential solution, comprising a combination of linguistic terms and membership functions for the input variables, as well as corresponding output fuzzy sets. These models were initialized with random parameters to start the evolutionary process.

We defined a fitness function to evaluate the performance of each model in predicting fire occurrence. This fitness function quantified the goodness-of-fit of each model based on its predictions compared to the actual target values on the training dataset.

We implemented different operators to simulate the evolutionary process within the Genetic Algorithm. Selection mechanisms were used to choose individuals with higher fitness for reproduction favoring those with higher performance, introducing variation and exploration into the population.

We iterated through multiple generations, evaluating the fitness of each individual in the population, selecting individuals for reproduction, and applying genetic operators to create offspring. This process continued until a termination criterion was met, such as reaching a maximum number of generations or achieving a satisfactory level of performance.

During the optimization process, the Genetic Algorithm leveraged the fuzzy rules generated by the ANFIS model as a starting point, refining them through evolutionary iterations to better adapt to the characteristics of the dataset and improve predictive accuracy and other metrics.

4 Experimental Results & Discussion

This section discusses the dataset and the experiment outputs we got.

4.1 Dataset

For our fire risk prediction system, we utilized a dataset sourced from Kaggle, specifically from the Forest Fires Data Set provided by the user "elikplim." This dataset contains valuable information related to forest fires, which is crucial for training and testing our predictive models.

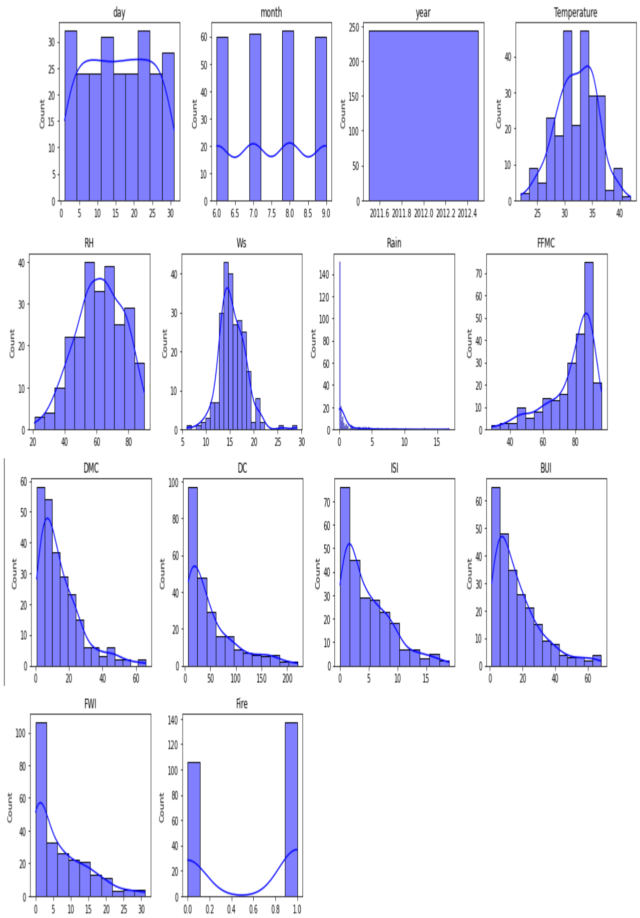
Attribute	Description
X	x-axis coordinate (from 1 to 9)
Y	y-axis coordinate (from 1 to 9)
month	Month of the year (January to December)
day	Day of the week (Monday to Sunday)
FFMC	FFMC code
DMC	DMC code
DC	DC code
ISI	ISI index
temp	Outside temperature (in °C)
RH	Outside relative humidity (in %)
wind	Outside wind speed (in km/h)
rain	Outside rain (in mm/m^2)
area	Total burned area (in ha)

4.2 EDA Results

This section discusses the results we got when we applied several EDA techniques and the innovative clustering approach that we used.

4.2.1 Histogram Results

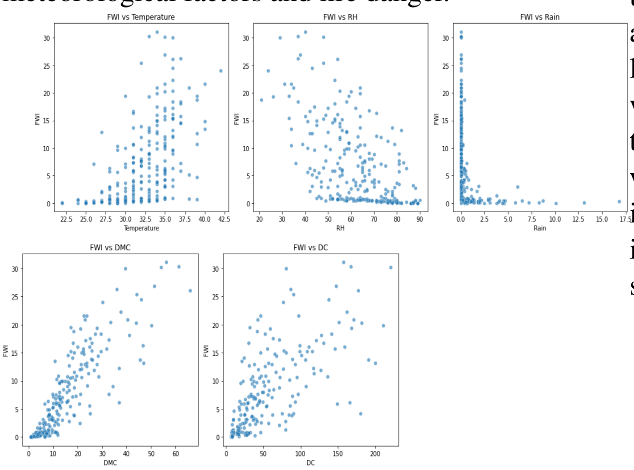
Variables like Temperature, RH, and Ws show relatively normal distributions. Rain is heavily skewed towards lower values, indicating most days have little to no rainfall. Fire indices (FFMC, DMC, DC, ISI, BUI, FWI) display varied distributions, some skewed and others more uniform, reflecting different aspects of fire risk conditions.



Now if we look at the results, we can see that rain negatively impacts Fire Chance and hence the plot is empty at majority of the instances. FFMC against Fire Chance gives us a left skewed graph stating that the distribution of values for FFMC is concentrated toward the higher end of the scale with a tail extending toward the lower values. However, DMC, DC, ISI, BUI and FWI gives us right skewed graphs against Fire Chance which suggests that the distribution of values for these variables are concentrated toward the lower end of the scale with a tail extending toward the higher values.

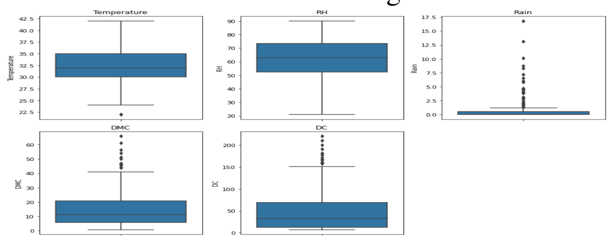
4.2.2 Scatter Plot Results

The scatter plots depict the relationships between the Fire Weather Index (FWI) and key environmental variables. The FWI vs Temperature plot reveals a positive trend, suggesting that higher temperatures tend to coincide with elevated FWI values, indicating an increased risk of fire occurrence. Conversely, FWI vs Relative Humidity (RH) displays a negative correlation, implying that lower humidity levels are associated with higher FWI values, underscoring the role of drier conditions in enhancing fire risk. The FWI vs Rain plot illustrates a clear negative relationship, indicating that higher levels of rainfall are linked to lower FWI values, implying a mitigating effect on fire danger. Moreover, FWI vs Duff Moisture Code (DMC) and FWI vs Drought Code (DC) both exhibit positive correlations, suggesting that as moisture levels decrease (DMC increases) and drought conditions intensify (DC increases), the FWI values and consequently, the fire risk, tend to escalate. These insights underscore the intricate interplay between meteorological factors and fire danger.



4.2.3 Box Plot Results

The box plots provide insights into the distribution of key environmental variables. The median value of Temperature is around 32.5, the median value for RH is around 62 and the median value for DMC is around 11. Temperature and Relative Humidity (RH) exhibit a relatively tight clustering of data points around the median, with a few outliers indicating extreme values. Conversely, the Rain plot depicts a skewed distribution, with the majority of data points clustered at lower values, suggesting a prevalence of days with minimal rainfall, while outliers signify instances of substantial precipitation. Duff Moisture Code (DMC) and Drought Code (DC) reveal wide ranges of values, with notable outliers indicating days characterized by unusually high moisture or drought conditions. These findings underscore the variability and complexity of environmental factors influencing forest fire risk.



4.2.4 Outlier Detection Results

The outlier detection process identified 25 instances of data points that deviated significantly from the typical patterns observed in the dataset. These outliers span various days across different months and years, with notable deviations in key environmental variables such as Temperature, Relative Humidity (RH), Wind speed (Ws), Rainfall, and indices like the Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), and Drought Code (DC). For example, some outliers exhibit unusually low temperatures coupled with high RH values, while others display extreme FFMC values indicative of exceptional fire weather conditions. These outliers underscore the importance of robust outlier detection techniques in identifying anomalous data instances that may skew analyses.

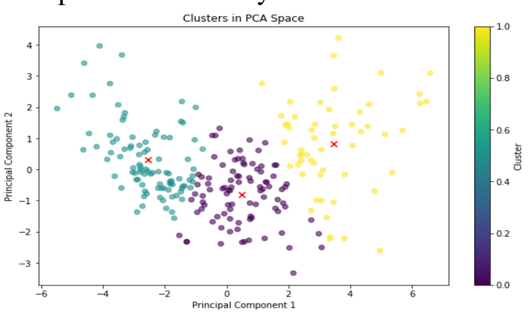
```
Number of outliers detected: 25
   day  month  year  Temperature  RH  Ws  Rain  FFMC   DMC     DC  ISI   BUI  \
2    3      6  2012           26  82  22  13.1  47.1   2.5    7.1  0.3   2.7
3    4      6  2012           25  89  13   2.5  28.6   1.3    6.9  0.0   1.7
28  29      6  2012           32  47  13   0.3  79.9  18.4   84.4  2.2  23.8
86  26      8  2012           31  78  18   0.0  85.8  45.6  190.6  4.7  57.1
87  27      8  2012           33  82  21   0.0  84.9  47.0  200.2  4.4  59.3

     FWI  Fire  outlier
2    0.1     0       -1
3    0.0     0       -1
28   3.9     0       -1
86  13.7     1       -1
87  13.2     1       -1
```

4.2.5 PCA Results

The principal component analysis (PCA) yielded two principal components, which collectively explained approximately 63.6% of the total variance in the dataset. The first principal component accounted for a significant portion of the variance, approximately 49.8%, indicating that it captured the majority of the underlying patterns in the data. The second principal component contributed a smaller but still notable proportion, approximately 13.7%, further elucidating additional variance not fully captured by the first component alone. The cumulative variance explained by both components underscores the effectiveness of PCA in reducing the dimensionality of the dataset while retaining a substantial amount of information. This reduction in dimensionality facilitates a more concise representation of the data while preserving the essential features within the dataset.

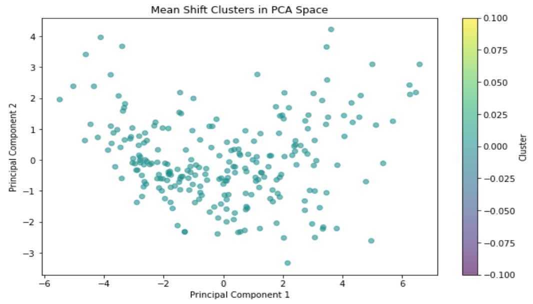
4.2.6 K Means Clustering Results

The clusters are well-separated in the PCA space, indicating distinct groups within the data. The red 'X' marks represent the centroids of each cluster, providing a reference point for the central location of each group. The dataset was partitioned into distinct clusters based on similarity in feature space. The algorithm identified a predetermined number of clusters, optimizing their centroids to minimize the within-cluster sum of squares. The result of the k-means clustering revealed the formation of cohesive clusters, each characterized by similar patterns in the key variables.



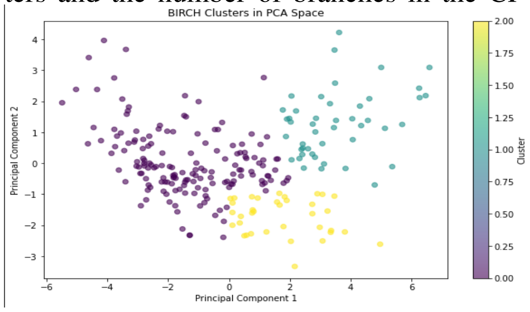
4.2.7 Mean Shift Clustering Results

The color bar indicates density values ranging from -0.100 to 0.100. This suggests that the mean shift algorithm is using a density-based approach to identify clusters. The scatter plot does not show distinct clusters, which could mean that the data points are too spread out or too close together for the mean shift algorithm to identify separate clusters effectively. There is a concentration of data points around the center of the PCA space, but without clear separation into clusters. This could indicate that the data has one large central cluster.



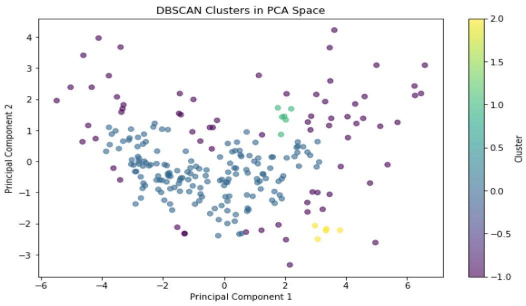
4.2.8 Innovation - BIRCH Clustering Results

The color bar indicates cluster distance, with values ranging from 0.00 to 2.00. This suggests that the plot is showing the distance of each data point from the cluster centroids. The BIRCH algorithm has identified clusters based on the similarity of data points in the PCA space. The scatter plot shows how the data points are grouped together, with each color representing a different cluster. The effectiveness was influenced by the threshold parameters and the number of branches in the CF tree.



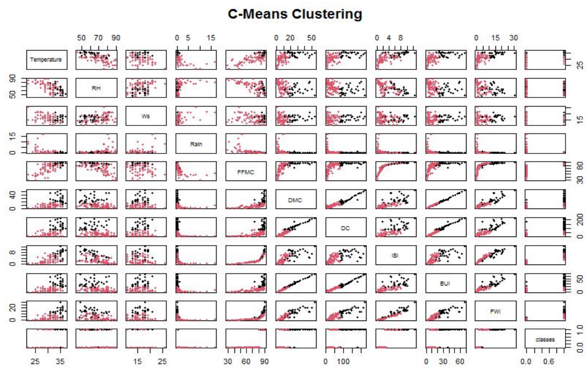
4.2.9 Innovation - DBSCAN Clustering Results

The points are color-coded based on the cluster they belong to. The color bar on the right side of the plot ranges from -1 to 2, indicating the cluster identification number assigned by DBSCAN. Points colored in purple represent noise, typically denoted by a -1 in DBSCAN clustering. These are points that do not belong to any cluster and are considered outliers. It was influenced by eps, which is the maximum distance between two points for one to be considered as in the neighborhood of the other, and min_samples, which is the number of points in a neighborhood for a point to be considered as a core point.



4.2.10 Innovation - Fuzzy C Means Clustering Results

The scatter plots from the fuzzy c-means clustering suggest that Relative Humidity (RH) and Temperature are significant indicators for the presence of fire, as evidenced by the clearer separation between the red and black dots. These variables show distinct groupings, indicating a stronger correlation with fire occurrences. On the other hand, variables like Rain and Initial Spread Index (ISI) displays a more scattered and intermixed pattern of red and black dots, suggesting a weaker predictive power for fire presence. This implies that while they are factors, they may not be as influential or may require the context of other variables to be useful predictors. It is important to remember that the strength of fuzzy c-means clustering is in its ability to use all input variables collectively to discern patterns, which might not be apparent when examining variables in isolation. Therefore, the best-performing variable is RH followed closely by Temperature, and the least effective seems to be Rain and ISI in this specific dataset.

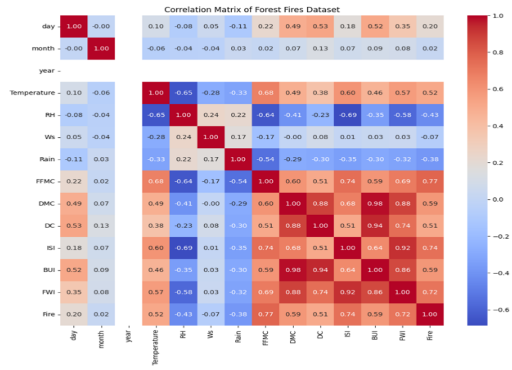


The application of silhouette analysis yielded a silhouette score of 1, indicating a high degree of separation and cohesion among the clusters. This score suggests that the clusters are well-defined and distinct from each other, with data points within clusters being more similar to each other than to those in other clusters. The high silhouette score validates the effectiveness of the fuzzy c-means algorithm in this dataset.

4.3 Fuzzy Rule Based Results

The analysis on the fuzzy rule-based system, particularly the correlation matrix, provides valuable insights into the relationship between different variables and the ”Fire Chance” variable. The correlation matrix serves as a visual representation of these relationships, indicating the strength and direction of correlations between variables.

If we look at the results, we can see that in most of the cases when relative humidity and rain are less, fire chance is the high. These two variables negatively impact fire chance and whenever temperature is low, fire chance is low.

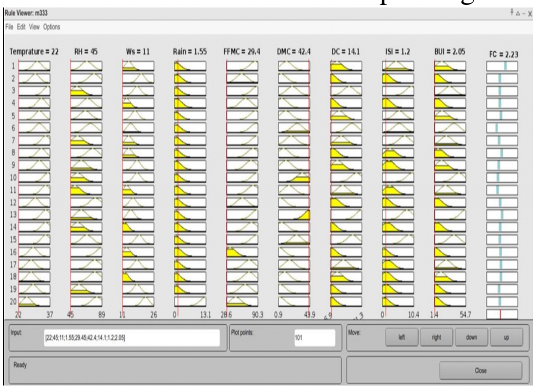


Next, we move on to Sugeno Inference. The outcome of the Sugeno Inference process yielding a fire chance value of 0.95, demonstrates the predictive capability of the fuzzy rule-based system in assessing the likelihood of forest fire occurrence. This numerical output reflects the system's assessment based on the input variables and their respective linguistic fuzzy sets indicating a high likelihood of forest fire occurrence as assessed by the fuzzy rule-based system.

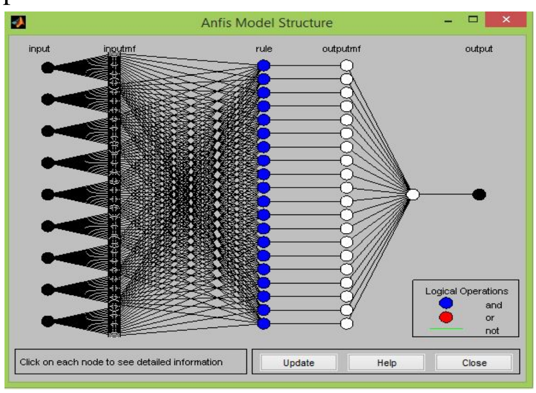
4.4 ANFIS Results

The ANFIS (Adaptive Neuro-Fuzzy Inference System) architecture employed in your model exhibits a sophisticated structure designed to capture the intricate relationships within the forest fire dataset. Comprising 412 nodes, including both linear and nonlinear parameters, the ANFIS model leverages its extensive network to process and analyze the input data effectively. The presence of 200 linear parameters and 360 nonlinear parameters underscores the model's adaptability and flexibility in accommodating the diverse characteristics of the dataset. With a total of 560 parameters, the ANFIS model demonstrates a robust capacity for learning and generalizing from the training data.

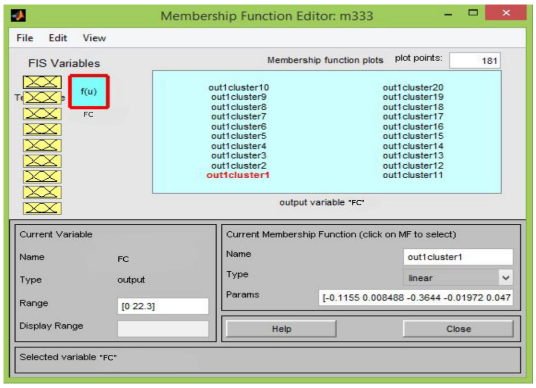
At the core of the ANFIS framework lies the concept of fuzzy logic, which enables the representation of uncertain and imprecise information inherent in environmental datasets such as forest fire records. The model incorporates 20 fuzzy rules, each comprising a set of linguistic variables and membership functions that capture the underlying patterns in the input-output relationships. These fuzzy rules serve as the basis for the inference process, wherein the model evaluates the degree of membership of each input variable to determine the corresponding output.



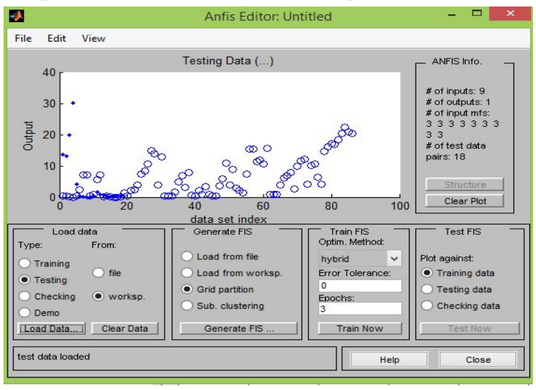
Furthermore, the ANFIS architecture employs a layered structure, consisting of input, fuzzy inference, and output layers. The input layer receives the normalized input data, which is then processed through the fuzzy inference layer, where the fuzzy rules are applied to compute the firing strengths of each rule. These firing strengths are subsequently aggregated to generate the final output through the output layer. This hierarchical arrangement enables the ANFIS model to systematically integrate the fuzzy logic-based reasoning with the computational power of neural networks, resulting in a hybrid system that combines the strengths of both approaches.



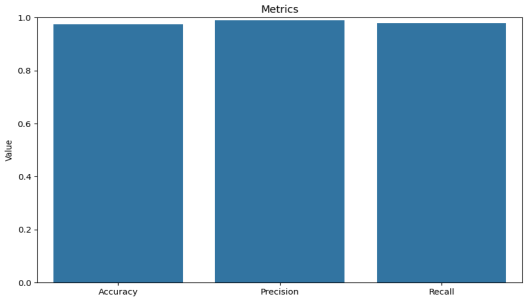
In addition to the ANFIS architecture, the model's ability to cluster the initially formulated 105 rules into a more concise set of 20 rules highlights its capability to distill complex information while maintaining predictive accuracy. This clustering process, driven by the model's learning algorithm, optimizes the rule set by identifying redundant or overlapping rules and consolidating them into more representative and interpretable clusters. By reducing the number of rules without compromising predictive performance, the ANFIS model enhances computational efficiency and facilitates easier interpretation of the underlying decision-making process, thereby empowering stakeholders with actionable insights for effective forest fire management.



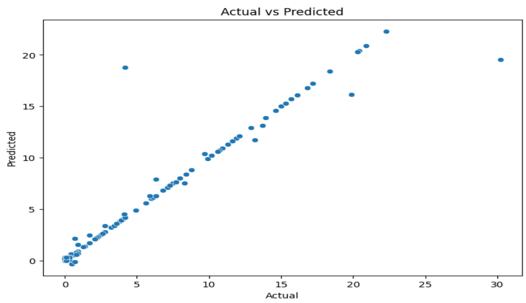
Training the ANFIS model involved the iterative optimization of its parameters to minimize the error between the predicted and actual outputs on the training dataset. The training data, consisting of input-output pairs, serves as the basis for updating the model's parameters. The model made good predictions on our input data, categorizing 93% of the input rules correctly for fire predictions.



With an impressive accuracy and recall of 93% and precision of 97%, respectively, our model demonstrates robust performance in predicting forest fire occurrences. Our model is correctly classifying roughly 93% of all instances across all classes. High accuracy suggests that your model is performing well in making overall correct predictions. Out of all instances that our model predicts as positive, 97% are actually positive.



The straight line trajectory of RMSE signifies minimal deviation between predicted and actual values, indicating the model's high fidelity in capturing the underlying patterns in the dataset. The points lying precisely on the line indicate that for those specific data points, the model's predictions perfectly match the observed values. The scattered points away from the line suggest that for some other data points are outliers.

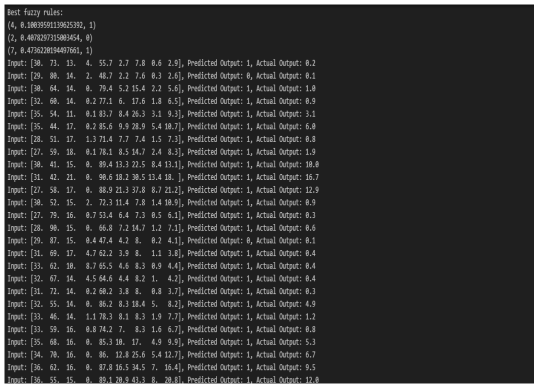


4.5 Genetic Algorithm Results

The output from training the Genetic Algorithm (GA) on the rules produced by ANFIS reveals a set of best fuzzy rules along with their associated parameters. Each rule is characterized by its antecedents, consequents, and firing strength, providing insights into the decision-making process of the hybrid system. Analyzing the provided output, we observe the predictions made by the hybrid system for various input scenarios with the model predicting the likelihood of a forest fire occurrence. A comparison between the predicted output and the actual output indicates the efficacy of the hybrid system in capturing the underlying patterns in the dataset. Notably, the model demonstrates a degree of accuracy in its predictions, with most instances closely aligning with the actual outcomes, while some exhibit variations.

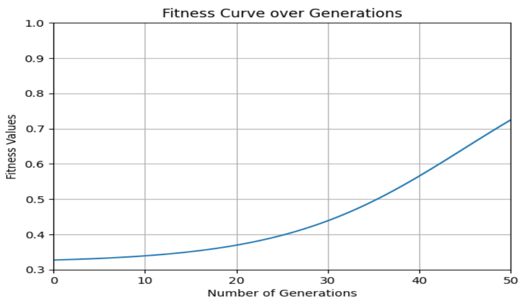
Furthermore, the output showcases the hybrid system's ability to generalize to unseen data, as evidenced by its predictions across a range of input conditions. Despite the inherent complexity and uncertainty associated with forest fire prediction, the integration of ANFIS and GA enables the model to adapt and evolve, refining its rules and parameters to improve predictive performance.

The best rules are represented within parentheses and consists of three components: the rule index, the firing strength, and the consequent. In this case, the input includes temperature, relative humidity, wind speed, and other relevant parameters, with specific values provided within brackets. The predicted output indicates the forest fire occurrence predicted by the hybrid system based on the given input. In the first row, the model predicts a fire occurrence with a value of 1. By analyzing these components, we gain insights into how the hybrid system processes input data, generates predictions, and compares them.



```
Best fuzzy rules:
(4, 0.10039591139625392, 1)
(2, 0.4078297315003454, 0)
(7, 0.4736220194497661, 1)
Input: [30.  73.  13.   4.  55.7  2.7  7.8  0.6  2.9], Predicted Output: 1, Actual Output: 0.2
Input: [29.  80.  14.   2.  48.7  2.2  7.6  0.3  2.6], Predicted Output: 0, Actual Output: 0.1
Input: [30.  64.  14.   0.  79.4  5.2 15.4  2.2  5.6], Predicted Output: 1, Actual Output: 1.0
Input: [32.  60.  14.   0.2 77.1  6.  17.6  1.8  6.5], Predicted Output: 1, Actual Output: 0.9
Input: [35.  54.  11.   0.1 83.7  8.4 26.3  3.1  9.3], Predicted Output: 1, Actual Output: 3.1
Input: [35.  44.  17.   0.2 85.6  9.9 28.9  5.4 10.7], Predicted Output: 1, Actual Output: 6.0
Input: [28.  51.  17.   1.3 71.4  7.7  7.4  1.5  7.3], Predicted Output: 1, Actual Output: 0.8
Input: [27.  59.  18.   0.1 78.1  8.5 14.7  2.4  8.3], Predicted Output: 1, Actual Output: 1.9
Input: [30.  41.  15.   0.  89.4 13.3 22.5  8.4 13.1], Predicted Output: 1, Actual Output: 10.0
Input: [31.  42.  21.   0.  90.6 18.2 30.5 13.4 18. ], Predicted Output: 1, Actual Output: 16.7
Input: [27.  58.  17.   0.  88.9 21.3 37.8  8.7 21.2], Predicted Output: 1, Actual Output: 12.9
Input: [30.  52.  15.   2.  72.3 11.4  7.8  1.4 10.9], Predicted Output: 1, Actual Output: 0.9
Input: [27.  79.  16.   0.7 53.4  6.4  7.3  0.5  6.1], Predicted Output: 1, Actual Output: 0.3
Input: [28.  90.  15.   0.  66.8  7.2 14.7  1.2  7.1], Predicted Output: 1, Actual Output: 0.6
Input: [29.  87.  15.   0.4 47.4  4.2  8.   0.2  4.1], Predicted Output: 0, Actual Output: 0.1
Input: [31.  69.  17.   4.7 62.2  3.9  8.   1.1  3.8], Predicted Output: 1, Actual Output: 0.4
Input: [33.  62.  10.   8.7 65.5  4.6  8.3  0.9  4.4], Predicted Output: 1, Actual Output: 0.4
Input: [32.  67.  14.   4.5 64.6  4.4  8.2  1.   4.2], Predicted Output: 1, Actual Output: 0.4
Input: [31.  72.  14.   0.2 60.2  3.8  8.   0.8  3.7], Predicted Output: 1, Actual Output: 0.3
Input: [32.  55.  14.   0.  86.2  8.3 18.4  5.   8.2], Predicted Output: 1, Actual Output: 4.9
Input: [33.  46.  14.   1.1 78.3  8.1  8.3  1.9  7.7], Predicted Output: 1, Actual Output: 1.2
Input: [33.  59.  16.   0.8 74.2  7.   8.3  1.6  6.7], Predicted Output: 1, Actual Output: 0.8
Input: [35.  68.  16.   0.  85.3 10.  17.   4.9  9.9], Predicted Output: 1, Actual Output: 5.3
Input: [34.  70.  16.   0.  86.  12.8 25.6  5.4 12.7], Predicted Output: 1, Actual Output: 6.7
Input: [36.  62.  16.   0.  87.8 16.5 34.5  7.  16.4], Predicted Output: 1, Actual Output: 9.5
Input: [36.  55.  15.   0.  89.1 20.9 43.3  8.  20.8], Predicted Output: 1, Actual Output: 12.0
```

In analyzing the performance of our genetic algorithm (GA) approach, we plotted the fitness values over 50 generations. The fitness values exhibited a notable trend, starting from 0.35 in the initial generation and gradually increasing to 0.73 by the 50th generation. This upward trend indicates an improvement in the fitness of the population over successive generations, suggesting that the GA successfully optimized the parameters of our hybrid model to better fit the training data. As the generations progressed, the fitness values approached a plateau, indicating that the GA reached a relatively stable state where further iterations did not lead to significant improvements in fitness.



5 Discussion

The fuzzy rule-based system developed for forest fire prediction embodies a meticulous process aimed at capturing the intricate relationships between meteorological variables and fire occurrence. Central to this endeavor was the formulation of comprehensive fuzzy rules, a task that involved careful consideration of relevant input variables and the design of appropriate membership functions.

Through extensive domain expertise and empirical analysis, we identified key meteorological parameters known to influence fire behavior, such as temperature, relative humidity, wind speed. Each input variable was associated with a set of linguistic terms, delineating their respective fuzzy sets and membership functions. The selection of linguistic terms and membership functions was guided by both empirical evidence and expert knowledge, ensuring a nuanced representation of the underlying uncertainties in forest fire dynamics.

The rationale behind adopting a fuzzy rule-based approach is stemmed from its inherent capacity to handle uncertainty and imprecision inherent in forest fire prediction. Unlike traditional statistical models or machine learning algorithms, which may struggle to capture the nonlinearity and vagueness of environmental variables, fuzzy logic offers a flexible framework for encoding qualitative and quantitative knowledge into a coherent decision-making system. By formalizing fuzzy rules based on heuristic principles and empirical observations, we aimed to harness the interpretability and transparency of fuzzy logic, facilitating a deeper understanding of the underlying factors driving fire risk assessments. Furthermore, the adaptive nature of fuzzy systems enables continuous refinement and adaptation to evolving environmental conditions, enhancing their robustness and applicability in dynamic wildfire scenarios.

Despite the inherent advantages of fuzzy rule-based systems, several challenges were encountered during their development and implementation. One notable challenge pertained to the elicitation and validation of fuzzy rules, which necessitated extensive collaboration between domain experts and data scientists. Ensuring the representativeness and generalizability of fuzzy rules across diverse geographic regions and ecosystems posed a significant logistical hurdle, requiring iterative refinement and validation against empirical data.

Additionally, the computational complexity associated with processing large volumes of meteorological data and optimizing fuzzy rule parameters presented technical challenges, necessitating the adoption of efficient algorithms and computational techniques. Through a concerted effort to address these challenges, we were able to develop a robust fuzzy rule-based system for forest fire prediction, offering valuable insights into fire risk assessment and management strategies.

6 Future Directions

In envisioning future trajectories for forest fire prediction research, the integration of advanced machine learning techniques like deep learning and ensemble methods with fuzzy logic-based approaches emerges as a promising avenue. By harnessing convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to extract intricate spatiotemporal patterns from satellite imagery and meteorological data, predictive models can achieve heightened accuracy and interpretability. Complementing this technological advancement, the deployment of unmanned aerial vehicles (UAVs) and Internet of Things (IoT) sensors for real-time monitoring holds potential for early fire detection and swift intervention. Moreover, the integration of geographical information systems (GIS) and remote sensing techniques promises to refine fire risk assessments by mapping vulnerable areas with precision, guiding targeted mitigation efforts.

Beyond technological innovations, interdisciplinary research collaborations are crucial for bridging natural and social sciences in wildfire management. By incorporating socio-economic factors into fire risk models, such as land use practices, demographic trends, and policy interventions, researchers can offer holistic insights into wildfire vulnerability and resilience. Collaborative initiatives that engage local communities and integrate indigenous knowledge systems can further enhance adaptive capacity and resilience in wildfire-prone regions. By embracing a transdisciplinary approach and leveraging diverse stakeholder expertise, researchers can unlock novel insights and solutions to address the multifaceted challenges posed by wildfires in a rapidly changing climate.

7 Conclusion

In conclusion, this research underscores the efficacy of fuzzy logic-based systems in augmenting forest fire prediction and management strategies. Through the development of a fuzzy rule-based system integrated with ANFIS and Genetic Algorithm, this study has demonstrated the potential for accurate and interpretable fire risk assessments. By harnessing the inherent flexibility of fuzzy logic to model complex, uncertain, and imprecise relationships among meteorological variables and fire occurrence, the proposed approach offers a valuable tool for proactive wildfire mitigation. Moreover, the synergy between fuzzy logic and machine learning techniques like ANFIS and Genetic Algorithm opens avenues for hybrid systems capable of automatic rule generation and optimization, thereby enhancing predictive performance and scalability.

Looking ahead, the research highlights the importance of continued interdisciplinary collaboration and innovation in wildfire research and management. By embracing a holistic approach that integrates scientific insights with socio-economic considerations and indigenous knowledge systems, stakeholders can foster adaptive governance frameworks and sustainable practices to mitigate the escalating threats posed by wildfires. Ultimately, by leveraging the power of fuzzy logic and cutting-edge technologies, researchers and practitioners can navigate and foster resilience and safeguard ecosystems and livelihoods for generations to come.

References

K. Gayathri, J.V.D. Prasad, T. Divya Kiran, and V. Mythili. 2022. Forest fire detection using Convolution Neural Networks. *2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC) (December 2022)*, DOI: dx.doi.org/10.1109/icmnwc56175.2022.10031936

Mason Davis and Mohammad Shekaramiz. 2023. Desert/forest fire detection using machine/Deep Learning Techniques. *Fire 6, 11 (October 2023)*, 418. DOI: dx.doi.org/10.3390/fire6110418

Mrs.R. Sujatha, K.M. Gokul, M. Dinesh, T. Gopinath, and C. Dineshkumar. 2023. Alert system for forest fire detection using CNN algorithm. *Recent Trends in Data Science and its Applications (2023)*. DOI: dx.doi.org/10.13052/rp-9788770040723.167

Piyush Jain, Sean C.P. Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D. Flannigan. 2020. A review of Machine Learning Applications in wildfire science and management. *Environmental Reviews 28, 4 (December 2020)*, 478–505. DOI: dx.doi.org/10.1139/er-2020-0019

Prabira Kumar Sethy, A. Geetha Devi, and Santi Kumari Behera. 2023. Forest fire detection using a nine-layer deep convolutional neural network. *Multimedia Data Processing and Computing (September 2023)*, 65–72. DOI: dx.doi.org/10.1201/9781003391272-6

Ramez Alkhatib, Wahib Sahwan, Anas Alkhatieb, and Brigitta Schütt. 2023. A brief review of machine learning algorithms in Forest Fires Science. *Applied Sciences 13, 14 (July 2023)*, 8275. DOI: dx.doi.org/10.3390/app13148275

Rudransh Agnihotri, Utkarsh Mittal, Rakesh Kumar Dhammi, Charu Gaur, Arun Gupta, and Mumtaz Ahmad Khan. 2023. Forest Guard: An integrated sensor cum AI-based fire-prone area mapping and Early Forest Fire Detection System. *2022 (July 2023)*. DOI: dx.doi.org/10.21203/rs.3.rs-3083433/v1