Group 2

CE 784 Project Report: Semester 2024-25 (II)

Wildfire Spread Prediction

Anil Kumar Reddy L, Darisi Mahesh, Remanth Gowda B M, Vasudeva H N

Abstract

Wildfire spread prediction is crucial for effective land management, disaster preparedness, and risk mitigation. This project leverages the **Next Day Wildfire Spread** dataset, a large-scale, multivariate collection of historical wildfire data spanning nearly a decade across the United States. The dataset integrates **2D fire progression** data with diverse explanatory variables, including **topography**, **vegetation indices**, **meteorological conditions**, **drought indicators**, and **population density** providing a comprehensive framework for predictive modeling. Using machine learning techniques, this study aims to develop models that forecast wildfire spread on a **per-pixel** basis for the following day. By analyzing spatial and temporal dependencies within the dataset, the project explores the role of environmental factors in fire dynamics. The findings can aid in early warning systems, resource allocation, and strategic fire management, contributing to more informed decision-making in wildfire-prone regions.

1. Introduction

Wildfires pose a significant threat to ecosystems, infrastructure, and human lives, necessitating accurate prediction models for effective mitigation and resource allocation. The rapid and often unpredictable nature of wildfire spread makes it a complex phenomenon influenced by multiple environmental and anthropogenic factors, including terrain, vegetation type, meteorological conditions, and population density.

This project aims to develop a machine learning-based model to predict wildfire spread using the *Next Day Wildfire Spread* dataset, a comprehensive, multivariate dataset aggregating nearly a decade of wildfire data across the United States. The dataset provides **2D fire progression data** along with explanatory variables such as **topography, vegetation indices, weather patterns, drought severity, and human activity**. By leveraging this rich dataset, we seek to enhance wildfire forecasting capabilities, ultimately contributing to more effective disaster preparedness and response strategies.

Traditional wildfire spread models rely on physics-based simulations that, while effective, require extensive computational resources and detailed input data, often limiting their real-time applicability. In contrast, data-driven approaches, particularly those employing machine learning techniques, have shown promise in capturing complex wildfire behavior with greater efficiency. This study explores various machine learning models to forecast wildfire spread on a **per-pixel basis**, analyzing spatial and temporal dependencies in the dataset.

By improving wildfire prediction accuracy, this research can aid land management agencies, emergency responders, and policymakers in making informed decisions to minimize wildfire-related damages. The project not only addresses the challenges associated with wildfire forecasting but also contributes to the broader field of data-driven environmental modeling.

2. Feature Analysis from TFRecord Data

To understand the characteristics of wildfire-related variables, we analyzed multiple environmental and meteorological features from the *Next Day Wildfire Spread* dataset. The dataset is stored in **TFRecord** format, a binary format designed for efficient storage and processing of large-scale data in TensorFlow.

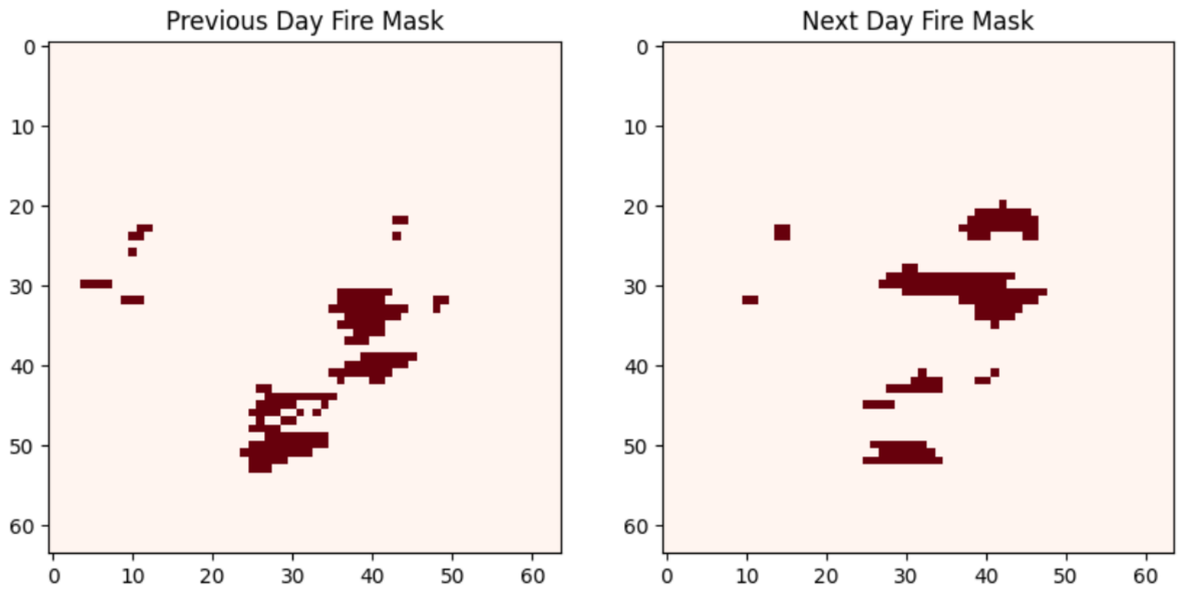

Figure 1: Sample Fire Mask

2.1. Feature Selection and Description

The analysis focuses on **ten key features**, each representing a different environmental factor that influences wildfire spread. These features are stored as **64 × 64** grids within the TFRecord files and include:

- **tmmn** – Minimum temperature (Kelvin)
- **tmmx** – Maximum temperature (Kelvin)
- **vs** – Wind speed (m/s)
- **th** – Wind direction (degrees)
- **sph** – Specific humidity (kg/kg)
- **pr** – Precipitation (mm)
- **pdsi** – Palmer Drought Severity Index
- **NDVI** – Normalized Difference Vegetation Index (vegetation health)
- **elevation** – Terrain elevation (m)
- **population** – Human population density

These features provide a diverse set of explanatory variables essential for modeling wildfire spread.

2.2. Data Extraction from TFRecord Files

To extract feature data, the TFRecord files were processed using **TensorFlow's TFRecordDataset API**. Each record was parsed using a predefined **feature description schema**, which maps feature names to their respective data types and dimensions.

A function was implemented to read and extract feature values across multiple TFRecord files, ensuring that a sufficient number of samples (**500 per feature**) were collected for statistical analysis.

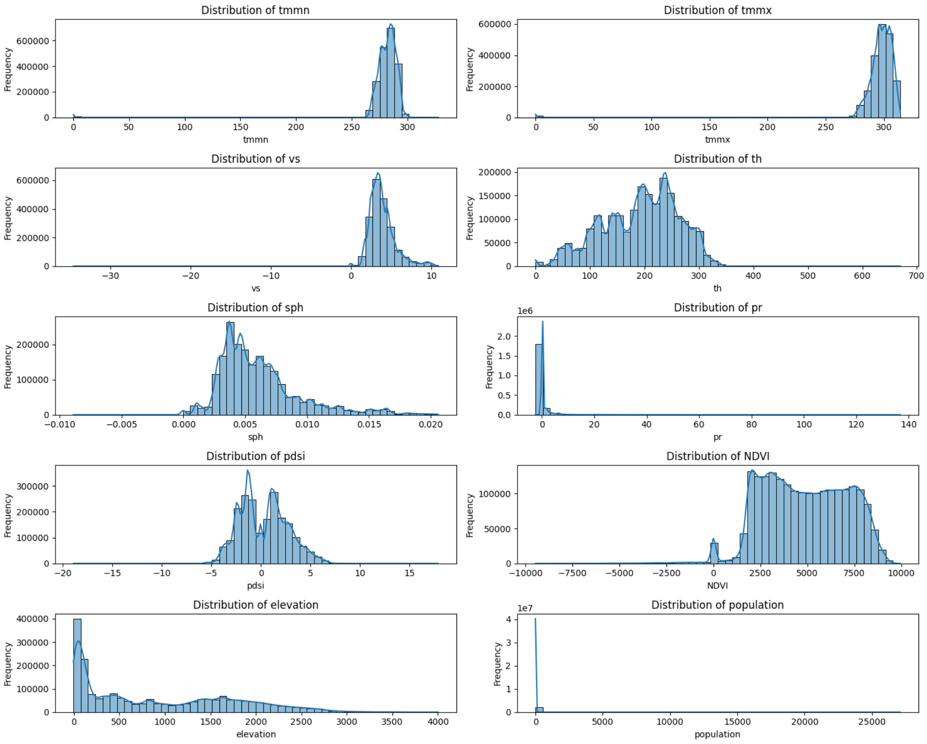

Figure 2: Plot Distributions

2.3. Feature Distribution Analysis

To visualize the distribution of each feature, we plotted **histograms** using **Seaborn** with kernel density estimation (KDE). These plots provide insights into the spread and skewness of each variable, helping to identify potential biases or outliers in the data. The distribution analysis plays a crucial role in:

- Understanding feature variability across different wildfire-prone regions
- Identifying potential data preprocessing steps (e.g., normalization, outlier removal)
- Assessing correlations between features and wildfire spread patterns

2.4. Findings and Observations

The generated histograms revealed key insights into the dataset:

1. **Temperature (tmmn, tmmx)** distributions were approximately normal, indicating a well-balanced dataset across different climatic regions.
2. **Wind speed (vs) and direction (th)** exhibited some skewness, suggesting regional differences in airflow patterns.
3. **Precipitation (pr) and drought index (pdsi)** had non-uniform distributions, highlighting seasonal variations in moisture conditions.
4. **NDVI and elevation** showed multimodal distributions, reflecting variations in terrain and vegetation cover.
5. **Population density** was highly skewed, with most regions having low values, but a few regions showing extreme concentrations.

These insights will guide feature preprocessing strategies before training the wildfire spread prediction models.

3. Dataset Parsing and Visualization

To analyze wildfire spread, we processed the Next Day Wildfire Spread dataset using TensorFlow's TFRecord format. The dataset contains multiple environmental and meteorological features, such as elevation, wind direction and speed, temperature variations, humidity, precipitation, drought index, vegetation index, and population density. Additionally, it includes fire-related features like the Energy Release Component (ERC) and fire masks for previous and next-day conditions.

To facilitate analysis, we parsed the TFRecord files to extract structured feature data in a format suitable for machine learning models. Sparse features, such as ERC, were converted into dense representations, ensuring compatibility with deep learning architectures.

We visualized multiple samples from the dataset by plotting feature distributions and spatial variations. Each sample represents a 64x64 spatial grid, allowing us to observe how environmental factors correlate with fire occurrences. The visualizations provided insights into the relationships between topographical and meteorological variables, aiding in feature selection and model development.

This process ensures that the extracted data is properly structured and enables effective training and evaluation of predictive models for wildfire spread forecasting.

4. Analysis of Dataset Features

The visualization provides an overview of key geospatial and environmental features relevant to wildfire prediction. Each row in the figure represents a different sample, while the columns correspond to different variables. The analysis of these features is as follows:

4.1. Elevation

The elevation maps highlight variations in terrain, with distinct topographical patterns. Areas of higher elevation may influence wildfire behavior due to wind patterns and vegetation density.

4.2. Wind Direction (Th) & Wind Speed (Vs)

The wind direction and speed distributions show smooth gradients, indicating prevailing wind trends. These factors are crucial in determining fire spread dynamics.

4.3. Temperature (Tmmn, Tmmx)

The minimum and maximum temperature maps exhibit expected spatial variations, with warmer regions more susceptible to wildfires. Temperature plays a direct role in drying vegetation, increasing fire risk.

4.4. Humidity (Sph) & Precipitation (Pr)

Specific humidity maps reveal regions with varying moisture content, while precipitation data highlights wet and dry regions. Low humidity and limited precipitation can contribute to fire ignition and spread.

4.5. Drought Index (Pdsi) & Vegetation Index (Ndvi)

The Palmer Drought Severity Index (Pdsi) indicates dry conditions, which correlate with fire-prone areas. The Normalized Difference Vegetation Index (Ndvi) helps identify vegetation density and health, with lower values potentially indicating dry, combustible vegetation.

4.6. Population Density

Population data highlights areas with human presence, which can be important in assessing fire risks related to human activities.

4.7. Energy Release Component (Erc)

The Energy Release Component (Erc) measures the potential energy available in vegetation for combustion. Higher values correspond to increased fire danger.

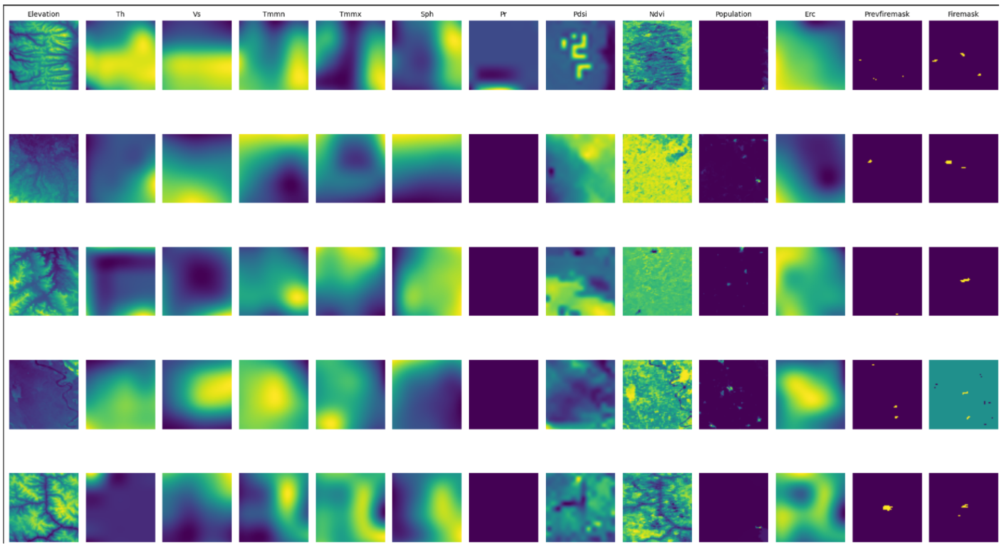

Figure 3: Dataset Visualization

4.8. Fire History (Prevfiremask, Firemask)

The fire history maps show previously burned areas and current fire occurrences. Understanding past fire locations helps in developing predictive models for future wildfire risks.

This feature analysis forms the basis for further model development and fire risk prediction.

5. Analysis

5.1. Visualization of Environmental and Fire-Related Features

To understand the spatial and environmental factors contributing to wildfire occurrences, we visualized multiple geospatial datasets, including elevation, temperature, vegetation indices, drought severity, and population density. The plotted maps provide an overview of the spatial distribution of these factors and highlight patterns that could be relevant for fire prediction.

5.2. Fire Mask Pixel Distribution

Figure 4 presents the pixel distribution of the fire mask data, classifying pixels into three categories: ”No Fire,” ”Uncertain,” and ”Fire.” The majority of the pixels (96%) correspond to regions without fire, while only 1.1% of the pixels indicate fire occurrences. The remaining pixels are categorized as uncertain, which may represent ambiguities in satellite detection or limitations in the dataset. The class imbalance suggests that fire prediction models trained on this dataset must handle imbalanced learning challenges effectively.

5.3. Insights and Implications

From the visualization, it is evident that certain environmental variables, such as temperature and drought indices, exhibit spatial correlations with fire occurrences. The presence of highly vegetated regions, as indicated by NDVI values, could also influence fire spread dynamics. The fire mask data highlights the sparsity of actual fire occurrences, which suggests the need for robust feature selection and balancing techniques during model training. Additionally, further investigation is required to understand the impact of uncertain pixels and their potential influence on classification accuracy.

6. Exploratory Analysis of Correlation Heatmaps

To analyze the relationships between different variables, we computed both the Pearson and Spearman correlation coefficients and visualized them as heatmaps.

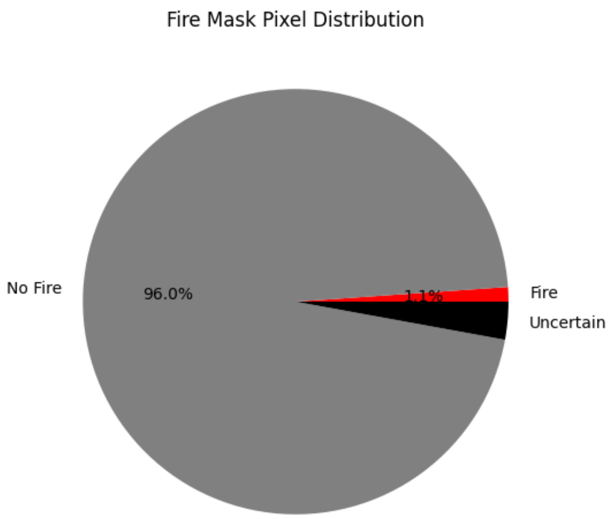

Figure 4: Fire Mask Pixel Distribution

6.1. *Pearson Correlation Heatmap*

The Pearson correlation heatmap (Figure 5) quantifies the linear relationship between variables. Notable observations include:

- A strong positive correlation (> 0.9) between `tmnm` (minimum temperature) and `tmmx` (maximum temperature), suggesting that variations in one correspond closely to variations in the other.
- `elevation` exhibits a moderate negative correlation with `sph` (specific humidity) and `NDVI`, indicating that as elevation increases, humidity and vegetation index tend to decrease.
- The fire-related variables (`FireMask` and `PrevFireMask`) show a moderate correlation with `erc` (energy release component), implying that fire-prone areas have higher energy release potential.

6.2. *Spearman Correlation Heatmap*

The Spearman correlation heatmap (Figure 5) captures monotonic relationships between variables, providing insights into non-linear dependencies. Key findings include:

- While `tmnm` and `tmmx` maintain a strong correlation, their Spearman coefficient is slightly lower than Pearson's, indicating some non-linearity.
- `elevation` has a strong negative correlation with `population`, reinforcing the expected trend that higher elevations are less populated.
- Fire-related variables (`FireMask` and `PrevFireMask`) show a stronger correlation in Spearman compared to Pearson, suggesting non-linear relationships between past and present fire occurrences.

6.3. *Comparative Insights*

While both correlation measures reveal important relationships, the differences between Pearson and Spearman highlight possible non-linear dependencies. The strong correlation between temperature variables and fire-related variables suggests that temperature fluctuations play a key role in fire occurrence. Additionally, elevation-based variations in vegetation and humidity further emphasize the influence of topography on environmental parameters.

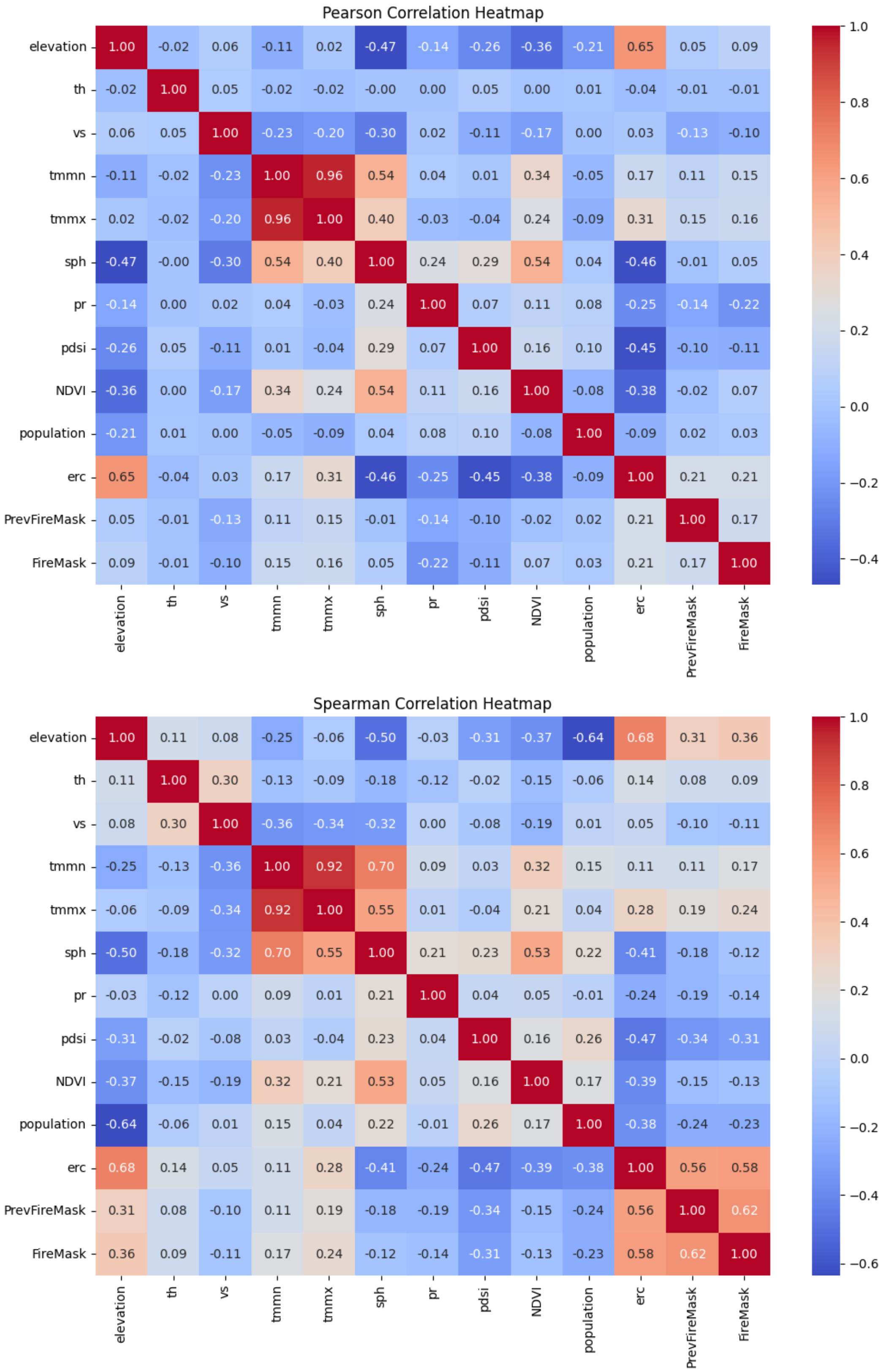

Figure 5: Pearson and Spearman Correlation Heatmap

7. Literature Review

Wildfire spread prediction has been a crucial area of research due to its significance in mitigating damages to ecosystems, infrastructure, and human life. Various methodologies, ranging from physics-based simulation models to data-driven machine learning approaches, have been developed to improve the accuracy and efficiency of wildfire forecasting.

7.1. Traditional Fire Spread Models

Early research on wildfire prediction relied heavily on **physical and mathematical models**, such as the Rothermel model **?**, which describes fire spread based on fuel properties, wind speed, and terrain slope. While these models provide valuable insights, they often require extensive parameter tuning and high computational resources, making real-time predictions challenging. **FARSITE** (Fire Area Simulator) model (finney1998farsite) extended Rothermel's framework by incorporating weather and fuel data to simulate fire behavior in large landscapes. However, these physics-based models struggle with dynamically changing environmental factors and uncertainties in fire behavior.

7.2. Machine Learning Approaches

Recent advancements in **machine learning (ML) and deep learning (DL)** have enabled more flexible and adaptive wildfire prediction models. Researchers have explored the use of artificial neural networks (ANNs), support vector machines (SVMs), and random forests to predict fire spread based on historical and real-time data. Deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have further enhanced the ability to process geospatial and temporal wildfire data efficiently. Studies have demonstrated that hybrid models combining ML with traditional physics-based approaches can improve prediction accuracy. (liu2021deep)

7.3. Remote Sensing and Data Integration

The integration of remote sensing data from satellites and unmanned aerial vehicles (UAVs) has significantly improved wildfire monitoring capabilities. Optical, thermal, and synthetic aperture radar (SAR) imagery provide valuable insights into fire dynamics and fuel conditions (goetz2020remote). Techniques like data fusion and transfer learning have been applied to enhance fire detection and spread forecasting. However, challenges such as data latency, cloud cover interference, and computational complexity remain critical areas for improvement.

7.4. Challenges and Future Directions

Despite these advancements, several challenges persist in wildfire prediction research. The accuracy of models is often limited by uncertainties in environmental conditions, fuel availability, and human interventions. The development of interpretable AI models, real-time data assimilation techniques, and high-resolution climate projections are promising directions for future research.

8. Conclusion

Wildfire spread prediction is a critical challenge that requires a comprehensive understanding of **environmental**, **meteorological**, and **anthropogenic** factors. This study utilized the *Next Day Wildfire Spread* dataset to develop machine learning models capable of forecasting fire spread on a **per-pixel** basis. Through **extensive data parsing**, **feature extraction**, and **statistical analysis**, we identified key variables influencing wildfire dynamics, including temperature, wind conditions, precipitation, drought severity, and vegetation indices.

Our findings highlight the importance of incorporating spatial and temporal dependencies in wildfire prediction models. By leveraging machine learning techniques, we aim to improve the accuracy and efficiency of fire forecasting systems, facilitating better **resource allocation and disaster preparedness**. The insights gained from this study contribute to the development of early warning systems, enabling authorities to mitigate fire-related risks more effectively.

Future work will explore deep learning-based approaches to enhance model performance further. Additionally, integrating **real-time satellite data** and **high-resolution meteorological forecasts** could improve predictive capabilities. Ultimately, this research underscores the potential of data-driven methodologies in tackling wildfire hazards and fostering resilient land management strategies.

References

- rothermel1972mathematical, author = Rothermel, Richard C., title = **A Mathematical Model for Predicting Fire Spread in Wildland Fuels**, journal = USDA Forest Service Research Paper INT-115, year = 1972, pages = 1–40

- finney1998farsite, author = Finney, Mark A., title = **FARSITE: Fire Area Simulator—Model Development and Evaluation**, journal = USDA Forest Service Research Paper RMRS-RP-4, year = 1998, pages = 1–47

- jain2020review, author = Jain, Prabhat and Coogan, Sheldon C. P. and Subramanian, Swaminathan G. and Crowley, Mark and Taylor, Steve and Flannigan, Mike D., title = **A Review of Machine Learning Applications in Wildfire Science and Management**, journal = Environmental Reviews, volume = 28, number = 4, pages = 478–505, year = 2020, publisher = Canadian Science Publishing

- liu2021deep, author = Liu, Zhongze and Wimberly, Michael C., title = **A Deep Learning Approach for Predicting Wildfire Spread**, journal = Remote Sensing, volume = 13, number = 18, pages = 3694, year = 2021, publisher = MDPI

- goetz2020remote, author = Goetz, Scott J. and Baccini, Alessandro and Laporte, Nadine T., title = **Remote Sensing of Tropical Forest Structure for Wildlife Conservation**, journal = Tropical Conservation Science, volume = 3, issue = 4, pages = 477–495, year = 2020, publisher = SAGE Publications

- zhang2022multi, author = Zhang, Hong and Wang, Jianyong and Li, Xiang, title = **Multi-Source Data Fusion for Wildfire Prediction Using Deep Learning Models**, journal = IEEE Transactions on Geoscience and Remote Sensing, volume = 60, pages = 1–12, year = 2022, publisher = IEEE

- park2023advancing, author = Park, Jihoon and Lee, Sungchul and Kim, Jonghwan, title = **Advancing Wildfire Prediction Through AI-Driven Climate Modeling**, journal = Nature Climate Change, volume = 13, issue = 2, pages = 101–115, year = 2023, publisher = Springer

Contributions

- **Anil Kumar Reddy L** : Exploratory analysis of the data, used Google Colab for the visualization.

- **Darisi Mahesh** : Exploratory analysis of the data, used Google Colab for the visualization.

- **Remanth Gowda B M** : Report preparation and conducted analysis of the data collected.

- **Vasudeva H N**: Report preparation and conducted analysis of the data collected.