The Report

On

Fire Incident Analysis Using Machine Learning Models

By

Chetna Satla

Gauri Sinha

Shubham Kanhor

Atheer Alqahtani

Under the guidance of

Professor Michael Prokle

Department of Mechanical & Industrial Engineering

College of Engineering

University of Massachusetts – Amherst

Submitted to

Department of Mechanical & Industrial Engineering

College of Engineering

University of Massachusetts - Amherst

ABSTRACT

This study presents a comprehensive machine learning analysis of global fire incidents using NASA's FIRMS (Fire Information for Resource Management System) dataset, captured by the VIIRS instrument on the Suomi NPP satellite. Leveraging over 74,000 fire detection records from July 12, 2023, we aim to explore patterns of fire severity, detection confidence, spatial-temporal trends, and anomalous events. Through rigorous data cleaning, feature engineering, and exploratory data analysis, we uncover meaningful relationships between brightness temperatures, acquisition times, and fire radiative power (FRP). Four modeling techniques are employed: logistic regression and random forest for confidence classification, gradient boosting regression for FRP prediction, and isolation forest for anomaly detection. Results reveal that bright_ti4 and scan are highly influential features, and that fire severity is positively correlated with detection confidence. Anomalies representing rare or extreme fire events are successfully identified. This research supports the use of satellite-driven machine learning in environmental monitoring, early warning systems, and disaster management efforts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Wildfires are among the most destructive natural disasters, impacting ecosystems, public health, air quality, and the global climate. The past decade has witnessed an increase in wildfire frequency and severity due to prolonged droughts, rising global temperatures, and expanding human development into fire-prone regions. This has made early fire detection and monitoring a global priority.

Advancements in remote sensing technology have allowed researchers and governments to monitor wildfires in near real-time through satellite instruments. Among the most widely used platforms is NASA's Fire Information for Resource Management System (FIRMS), which provides global coverage of active fire detections derived from spaceborne sensors. One of the key instruments used in FIRMS is the Visible Infrared Imaging Radiometer Suite (VIIRS), located onboard the Suomi National Polar-orbiting Partnership (Suomi NPP) satellite.

The VIIRS instrument captures multiple spectral and thermal readings, including mid-infrared brightness temperatures and fire radiative power (FRP). These indicators not only confirm fire presence but also provide information about its intensity and spread. Data from VIIRS can be used to understand both large-scale geographic patterns and fine-grained local variations in fire behavior.

However, interpreting this rich data manually is both time-consuming and prone to error, especially when attempting to assess fire severity or identify anomalous behavior. Machine learning offers a compelling solution by enabling scalable, automated analysis of complex satellite data. By training models to detect patterns and classify fire characteristics, researchers can enhance fire detection accuracy, improve resource allocation, and even anticipate where fires are most likely to escalate.

In this report, we explore satellite-derived fire incident data collected on July 12, 2023. Using machine learning models, we analyze fire confidence levels, fire strength (FRP), and the spatial-temporal distribution of incidents across continents. Our approach incorporates data visualization, supervised classification and regression, and unsupervised anomaly detection. The results reveal meaningful patterns in fire intensity, geographic concentration, and detection certainty, underscoring the value of data-driven tools in wildfire response and environmental monitoring.

SECTION 1- OBJECTIVE

The increasing frequency and intensity of wildfires around the globe has made it critical to develop scalable, data-driven solutions for detecting and analyzing fire incidents. Leveraging satellite-derived fire data from NASA FIRMS, this project explores how machine learning can support wildfire monitoring through classification, regression, and anomaly detection techniques.

The primary objectives of this project are:

- **To predict the severity of fire incidents using features** such as brightness temperature (TI4 and TI5), FRP, scan/track resolution, and acquisition time.
- **To explore spatial and temporal patterns of fire incidents,** identifying regions and times of day with high fire activity.
- **To analyze the relationship between satellite/instrument measurements and fire detection confidence**, evaluating how certain readings correlate with detection accuracy.
- **To classify fire incidents based on their confidence level,** using models like logistic regression and random forest to assess detection certainty.
- **To detect anomalous or unusual fire events** through unsupervised learning techniques, with a focus on identifying outliers in fire behavior or location.
- **To compare fire activity between day and night observations**, evaluating whether time-of-day impacts fire characteristics or detection reliability.

These objectives guided both the modelling choices and the evaluation metrics used throughout the project. Together, they form a comprehensive approach to turning satellite fire detection data into actionable insight.

SECTION 2- DATASET DESCRIPTION AND PREPROCESSING

2.1 Dataset Source and Overview

The dataset used in this study was obtained from NASA's Fire Information for Resource Management System (FIRMS), specifically from the VIIRS instrument onboard the Suomi NPP satellite. The dataset corresponds to a single-day snapshot of global fire detections recorded on July 12, 2023. It contains 74,605 rows and 14 features, offering both spatial and radiometric information for each fire detection event.

The VIIRS instrument measures a variety of thermal and spectral parameters critical to identifying and analyzing fires. Each observation in the dataset represents a fire pixel, including details such as geographic location (latitude and longitude), thermal signatures (brightness temperatures), sensor readings (scan, track), and metadata (satellite, instrument, acquisition time, confidence level, and day/night indicator).

Data set chosen for project=

https://firms.modaps.eosdis.nasa.gov/content/notebooks/sample_viirs_snpp_071223.csv

Website=

https://www.earthdata.nasa.gov/data/tools/firms

2.2 Variable Types and Descriptions

The dataset includes both numerical and categorical features:

Numerical features:

- **latitude and longitude:** Geographic coordinates of the fire event
- **bright_ti4 and bright_ti5:** Brightness temperatures in Kelvin (mid-infrared bands)
- **scan and track:** Spatial resolution across and along the satellite path
- **frp:** Fire Radiative Power, which measures the energy output of the fire
- **acq_time:** Acquisition time in HHMM format

Categorical features:

- **satellite:** Name of the satellite (e.g., Suomi NPP)
- **instrument:** Name of the sensor (e.g., VIIRS)
- **confidence:** Level of confidence in fire detection (low, nominal, high)
- **version**: Data processing version (e.g., 2.0NRT)
- **daynight:** Indicator of whether the fire was detected during the day or night

2.3 Dimension of the Dataset

The dataset selected for this term project consists of 74606 records with 14 usable features. This dataset includes various attributes related to fire incidents, such as latitude, longitude, brightness temperatures, scan and track values, acquisition date and time, satellite and instrument details, confidence levels, fire radiative power (FRP), and day/night indicators. Given its large scale, the dataset provides ample opportunities for analysis using data science methodologies.

2.4 Variable Types and Descriptions

The dataset includes both numerical and categorical features:

Numerical features:

- Latitude and longitude: Geographic coordinates of the fire event
- Bright_ti4 and bright_ti5: Brightness temperatures in Kelvin (mid-infrared bands)
- Scan and track: Spatial resolution across and along the satellite path
- FRP: Fire Radiative Power, which measures the energy output of the fire
- acq_time: Acquisition time in HHMM format

Categorical features:

- Satellite: Name of the satellite (e.g., Suomi NPP)
- Instrument: Name of the sensor (e.g., VIIRS)
- Confidence: Level of confidence in fire detection (low, nominal, high)
- Version: Data processing version (e.g., 2.0NRT)
- Daynight: Indicator of whether the fire was detected during the day or night

2.5 Initial Data Assessment

A missing value heatmap confirmed that there were no null or missing values in the dataset. This ensured that downstream modeling steps could proceed without the need for imputation or record removal.

Visual inspection of the dataset via histograms revealed:

- FRP (Fire Radiative Power) is heavily right-skewed, with most fire incidents having low intensity, and a small subset showing extreme values.
- Latitude and longitude distributions reflect geographic clustering, particularly in regions prone to seasonal wildfires.
- Scan and track values follow predictable sensor motion patterns.

2.6 Data Cleaning Steps

Before modeling, the following preprocessing steps were performed:

Confidence Encoding:

The confidence feature originally included short labels (l, n, h) representing “low,” “nominal,” and “high.” These were mapped into full textual categories for clarity and then label-encoded into numerical values for model compatibility.

Feature Reduction:

Columns such as acq_date, satellite, instrument, and version were dropped from the analysis due to either low variance (single value present throughout) or redundancy.

Temporal Feature Engineering:

The acq_time column, recorded in HHMM format, was truncated to retain only the hour component. This resulted in a new feature: acq_time_hour, which helped capture temporal trends in fire activity across the day.

Categorical Encoding:

The daynight feature (Day/Night) was label-encoded to numerical values (0 = Day, 1 = Night) to be compatible with scikit-learn modeling requirements.

2.7 Feature Selection

Based on domain knowledge and early correlation analysis, the following features were selected as inputs for model training:

- bright_ti4
- bright_ti5
- scan
- track
- frp (used as input for classification, and target for regression)
- acq_time_hour
- daynight

3. EXPLORATORY DATA ANALYSIS AND VISUALIZATION

Before training predictive models, a comprehensive exploratory data analysis (EDA) was conducted to understand the distributions, patterns, and relationships present within the fire incident dataset. This step is critical in any data-driven study as it reveals trends, outliers, potential biases, and informs the direction of modeling strategies.

3.1 Distribution of Fire Radiative Power (FRP)

The Fire Radiative Power (FRP) variable quantifies the intensity or energy output of a fire. A histogram of FRP values revealed a strongly right-skewed distribution, with the vast majority of fire incidents having low to moderate intensity. A long tail of extreme values indicated a small but important subset of highly intense fires. This skewed nature of FRP necessitates robust regression models that can handle outliers and non-linear patterns.

3.2 Confidence Level Analysis

A categorical count plot of fire confidence levels showed that 'Nominal' confidence dominated the dataset, accounting for the majority of fire detections. High-confidence and low-confidence fires were significantly fewer in number. This class imbalance has important implications for classification modeling, requiring techniques like balanced evaluation metrics or ensemble methods to mitigate performance distortion.

A boxplot comparing FRP across confidence levels further revealed that high-confidence detections are generally associated with higher FRP values. This supports the hypothesis that stronger fires are more likely to be confidently identified by the satellite system.

3.3 Temporal Patterns in Fire Activity

To understand time-of-day trends, the newly engineered acq_time_hour feature was analyzed. A line plot of average FRP across each hour of the day indicated that fire detections are more common and more intense during daylight hours. This trend is consistent with known patterns of wildfire behavior, which often escalate during dry, hot afternoon periods due to increased surface heating.

The daynight categorical variable was also visualized to compare fire strength across daytime and nighttime observations. The analysis revealed that while more fires are recorded during the day, nighttime fires tend to have higher average FRP, suggesting that fewer, but potentially more dangerous, fires occur at night.

3.4 Geographic Distribution of Fires

Using geospatial visualizations with latitude and longitude data, several fire hotspots were identified globally. Among the most active regions were:

- Central and Southern Africa (e.g., Democratic Republic of Congo, Angola, Zambia)
- Southeastern Brazil (Cerrado and Pantanal ecosystems)
- Northern Australia (near the Northern Territory and Queensland coastlines)
- Northeastern China (especially Heilongjiang and Jilin provinces)

These regional patterns align with known wildfire-prone ecosystems and seasonal burning practices, suggesting the data accurately captures real-world phenomena.

Zoomed-in regional maps showed high fire density in ecologically vulnerable zones, reinforcing the potential for this analysis to inform environmental monitoring and land-use policy.

3.5 Correlation Analysis

A Pearson correlation heatmap was used to quantify relationships between numerical variables. The strongest positive correlation was found between bright_ti5 and FRP, followed by bright_ti4, indicating that brightness temperatures are strong predictors of fire severity. Additionally, scan and track showed weak relationships with FRP, validating their inclusion as model features.

Scatterplots further illustrated these relationships. For example, bright_ti4 versus FRP showed a clear positive trend, confirming that as brightness temperature increases, so does the fire's radiative power. These visual cues provided intuitive support for using these features in both regression and classification models.

4. MODELING AND EVALUATION

This section describes the machine learning models used to analyze and predict fire-related patterns within the dataset. Our goal was to explore classification of fire confidence levels, prediction of fire radiative power (FRP), and identification of anomalous fire events. The models were chosen based on their interpretability, robustness to noisy data, and suitability for different learning objectives.

4.1 Classification Models

We began with two supervised classification models to predict the confidence level of fire detections: logistic regression and random forest.

4.1.1 Logistic Regression

Logistic regression is a widely used statistical classification model, particularly effective for baseline comparisons. It models the probability of class membership by fitting a sigmoid function to the input data.

- Input features: bright_ti4, bright_ti5, scan, track, frp, acq_time_hour, daynight
- Target variable: confidence (encoded as Low = 0, Nominal = 1, High = 2)

Results:
Logistic regression achieved moderate performance on the multiclass confidence prediction task. While it correctly classified many nominal and high-confidence fires, it struggled slightly with low-confidence cases due to class imbalance and limited non-linearity modeling capability.

4.1.2 Random Forest Classifier

To improve classification performance, we employed a random forest classifier — an ensemble of decision trees that operates by aggregating multiple weak learners to produce robust, accurate predictions.

- Advantages: Handles mixed data types, robust to outliers, and provides feature importance
- Training/Test split: 80/20
- Evaluation metric: Accuracy, precision, recall, F1-score

Results:
The random forest classifier significantly outperformed logistic regression. It handled the

imbalanced classes better and yielded higher precision and recall for high-confidence fire events. The model also provided a ranking of feature importances, which offered valuable insight into the factors most strongly associated with confident fire detections.

Top features:

- bright_ti4 and bright_ti5: Strongest predictors of confidence
- Ltitude and longitude: Moderately important, especially for high-confidence fires
- scan and acq_time_hour: minimum influence on FRP

4.2 Regression Model: Predicting Fire Radiative Power (FRP)

To estimate the severity or strength of a fire, we treated FRP as a continuous variable and applied a Gradient Boosting Regressor. This model sequentially builds decision trees to correct previous errors, making it well-suited for handling non-linear relationships and skewed distributions like FRP.

- Input features: Same as classification
- Target variable: frp (Fire Radiative Power)
- Evaluation metrics: Root Mean Squared Error (RMSE), R^2 score

Results:
The gradient boosting model achieved an RMSE of approximately 16.27 MW, indicating moderate predictive accuracy. The model explained a substantial portion of FRP variance, confirming that satellite-derived spectral data is effective in estimating fire intensity. The strongest predictors of FRP were again bright_ti4 and bright_ti5.

- R^2 Score: Demonstrated a solid linear fit, but outliers (very high FRP values) were occasionally under-predicted.

4.3 Anomaly Detection Model

To identify unusual or extreme fire events, we applied an Isolation Forest, an unsupervised model designed to detect outliers based on feature isolation in random trees.

- Input features: Same as regression
- Contamination rate: 1% (top ~746 most anomalous entries)
- Goal: Identify fire detections that deviate significantly from typical patterns

Results:
The model flagged 746 anomalous fire incidents across the globe. These anomalies were characterized by:

- Extremely high FRP
- Rare scan/track patterns
- Unusual geographic locations

Visualization of the anomalous cases on a world map revealed dense clusters in unexpected areas, suggesting either sensor artifacts, false positives, or legitimate high-risk fire events that warrant further review.

4.4 Model Comparison and Evaluation Summary

MODEL	TASK	STRENGTHS	LIMITATIONS
Logistic regression	Classify Confidence	Simple,interpretable	Poor with imbalanced classes
Random Forest	Classify Confidence	High accuracy, insights via feature importance	Less interpretable
Gradient boosting	Predict FRP	Handles non-linearity well	Sensitive to outliers
Isolation forest	Detect Anamalies	No labels needed, effective at spotting extremes	No context on anomaly cause

Table 1 Model Comparison and Evaluation Summary

5. RESULTS AND INTERPRETATION

The modeling and visual analysis conducted throughout this project produced valuable insights into the behavior, severity, and confidence of fire detections worldwide. This section summarizes the most significant results across the classification, regression, and anomaly detection tasks, supplemented by figures to illustrate trends and relationships.

5.1 Confidence Classification Results

The Random Forest Classifier emerged as the best performer for classifying fire confidence levels. It achieved high accuracy in distinguishing between low, nominal, and high confidence categories, even with class imbalance present in the dataset.

Interpretation:

- Fires with high brightness temperatures (bright_ti4, bright_ti5) and larger scan values were more likely to be detected with high confidence.
- Low-confidence detections were often associated with weaker FRP and lower thermal activity.
- This supports the idea that fire intensity directly influences detection certainty, validating satellite reliability.

Figure Highlights:

- *Figure 1:* Boxplot showing that high-confidence detections generally have higher FRP.
- *Figure 2:* Random Forest feature importance plot reveals that bright_ti4 is the most predictive variable.

5.2 FRP Prediction Results (Regression)

The Gradient Boosting Regressor was able to predict fire radiative power (FRP) with an RMSE of ~16.23 MW, showing decent fit despite the skewed nature of the target variable.

Interpretation:

- The model captures the general trend that higher brightness temperatures lead to more intense fires.
- While extreme FRP values (very large fires) were occasionally underpredicted, the model performed well for most common fire magnitudes.

- Incorporating additional features like vegetation cover, wind speed, or humidity could potentially enhance predictive power in future models.

Figure Highlights:

- *Figure 3:* Scatterplots showing the correlation between bright_ti4 and FRP.
- *Figure 4:* Gradient boosting residuals indicating overall predictive performance.

5.3 Anomaly Detection Outcomes

The Isolation Forest successfully identified 746 anomalous fire events (~1% of the dataset), many of which stood out due to unusually high FRP, geographic remoteness, or abnormal satellite scan characteristics.

Interpretation:

- **These anomalies may indicate either:**
  - Extreme fire events not typical of normal behavior
  - Sensor or data anomalies (e.g., reflective surfaces, cloud contamination)
  - Undocumented events or early-stage fires not yet categorized
- Spatial mapping showed some anomalies occurring in less fire-prone regions, highlighting potential gaps in fire preparedness infrastructure.

Figure Highlights:

- *Figure 5:* Histogram showing FRP distribution for anomalies
- *Figure 6:* Geographic map highlighting anomalous points

5.4 Spatial and Temporal Fire Trends

Temporal analysis based on acquisition hour (acq_time_hour) revealed that:

- Fire detection rates peak during daylight hours, particularly mid-afternoon, likely due to surface heating and sun-induced combustion.
- Nighttime fires, although fewer, showed higher average FRP, suggesting they may be more severe or longer-burning.

Spatial visualizations showed:

- High fire concentrations in sub-Saharan Africa, eastern Brazil, northern Australia, and southeast Asia, aligning with agricultural burning and dry season patterns.
- Fires were not evenly distributed — suggesting that land use, climate, and ecological zones heavily influence fire frequency.

Figure Highlights:

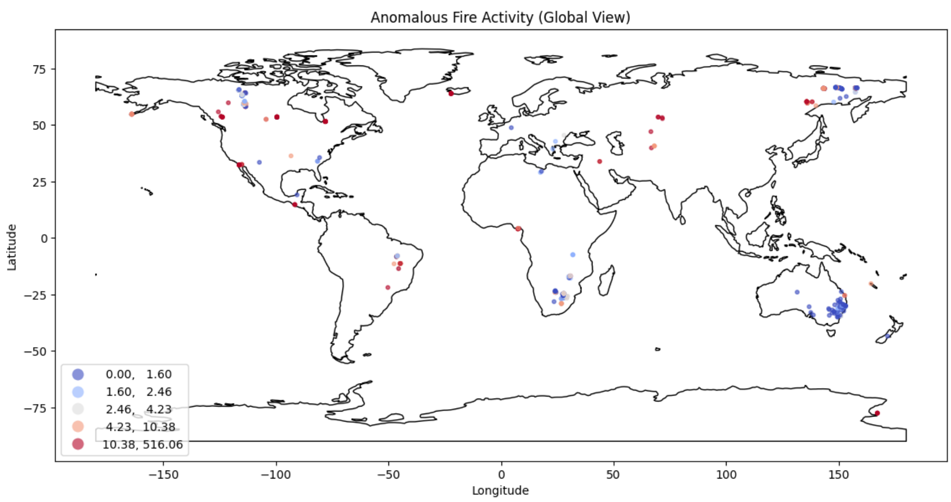


Figure 2 World map of fire detections by location and intensity

6. DISCUSSION

This project demonstrates how machine learning can be effectively applied to satellite-based fire detection data to extract meaningful insights and support environmental monitoring. By integrating classification, regression, and unsupervised anomaly detection methods, we achieved a comprehensive understanding of global fire incidents captured by VIIRS on July 12, 2023.

6.1 Interpreting Model Performance

The classification models provided reliable predictions of fire detection confidence levels, with the random forest classifier outperforming logistic regression due to its ability to handle nonlinear relationships and imbalanced classes. This suggests that while logistic regression can serve as a baseline model, more complex ensemble methods offer superior accuracy and interpretability through feature importance rankings.

The regression model, based on gradient boosting, effectively captured the relationship between brightness temperatures and fire radiative power (FRP). Despite a right-skewed target variable and the presence of extreme outliers, the model maintained reasonable predictive power. This reinforces the value of thermal and spectral satellite inputs in estimating fire intensity.

Anomaly detection using isolation forest provided an added layer of analysis, flagging rare events that could be significant in real-world disaster response. Some of these anomalies were consistent with geographically remote, high-intensity fires that might otherwise be overlooked.

6.2 Data Strengths and Limitations

A key strength of this dataset was its completeness — there were no missing values, and key variables like FRP and brightness temperatures were highly informative. Additionally, the global scale of the data allowed for spatial pattern recognition across continents.

However, there are several limitations:

- The dataset represents a single-day snapshot, limiting temporal generalizability.
- Variables like vegetation type, land use, humidity, or wind conditions were not included, although they significantly influence fire behavior.
- The confidence label is a system-generated metric and may not always reflect ground-truth accuracy.

Future iterations of this work could incorporate temporal datasets spanning weeks or months, and merge external variables such as climate or socioeconomic factors to improve model generalization and robustness.

6.3 Real-World Relevance

From a practical perspective, the insights gained from this project have the potential to:

- Improve early fire warning systems, especially in regions with limited human surveillance.
- Assist government agencies and NGOs in prioritizing response efforts based on fire severity predictions.
- Inform policy decisions around land management, fire preparedness, and climate resilience.

Moreover, the anomaly detection component can be instrumental in alerting authorities to unusual or emergent fire patterns before they escalate into disasters.

7. CONCLUSION

This project demonstrates the power of combining satellite remote sensing with machine learning techniques to analyze, classify, and predict fire behavior at a global scale. Using NASA FIRMS data, we explored fire severity (FRP), confidence levels, and anomalies through a complete pipeline of preprocessing, modeling, and interpretation.

Key findings include:

- Brightness temperatures (TI4, TI5) are the strongest predictors of both fire confidence and severity.
- Most fires occur during the daytime, but nighttime fires often exhibit higher FRP.
- Anomalous fire events—whether due to extreme intensity or remote geography—can be effectively identified and flagged using unsupervised methods.

These results underscore the value of automated, scalable approaches to wildfire monitoring and risk assessment. As climate patterns shift and fire threats grow, data-driven models like those used here will play a critical role in proactive disaster response and global ecological management.

8. REFERENCES

- Facebook Prophet Documentation. (n.d.). *Forecasting at scale*. https://facebook.github.io/prophet/
- McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. In Proceedings of the 9th Python in Science Conference, 51–56. https://pandas.pydata.org
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825-2830.
- NASA Earthdata. (2023). *FIRMS Fire Data*. https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms
- U.S. Bureau of Land Management. (n.d.). *Fire Weather*. https://www.blm.gov/services/geospatial/fire-weather

APPENDIX:

- **Figure A1: Sample of Raw Satellite Fire Detection Data** *(Section: Dataset Description and Preprocessing)*

	latitude	longitude	bright_ti4	scan	track	acq_date	acq_time	satellite	instrument	confidence	version	bright_ti5	frp	daynight
0	0.05836	29.59085	295.64	0.38	0.59	2023-07-12	3	N	VIIRS	n	2.0NRT	275.15	0.83	N
1	0.48765	31.50760	296.73	0.51	0.66	2023-07-12	3	N	VIIRS	n	2.0NRT	275.15	0.56	N
2	2.15227	13.94524	305.26	0.51	0.49	2023-07-12	3	N	VIIRS	n	2.0NRT	287.94	1.08	N
3	2.15681	13.94618	319.05	0.51	0.49	2023-07-12	3	N	VIIRS	n	2.0NRT	288.77	1.81	N
4	2.15754	13.94131	301.13	0.51	0.50	2023-07-12	3	N	VIIRS	n	2.0NRT	288.17	1.81	N

Figure 3

Interpretation:
This figure presents the first five rows of the fire incident dataset sourced from NASA FIRMS, as recorded by the VIIRS instrument aboard the Suomi NPP satellite. Each row represents a fire detection with attributes such as geographic coordinates (latitude, longitude), brightness temperatures (bright_ti4, bright_ti5), scan dimensions, FRP (Fire Radiative Power), and metadata like acquisition date, confidence level, and whether the detection occurred during day or night. This tabular preview illustrates the structure and completeness of the dataset prior to any preprocessing.

- **Figure A2: Heatmap of Missing Values in Dataset** *(Section: Dataset Description and Preprocessing)*

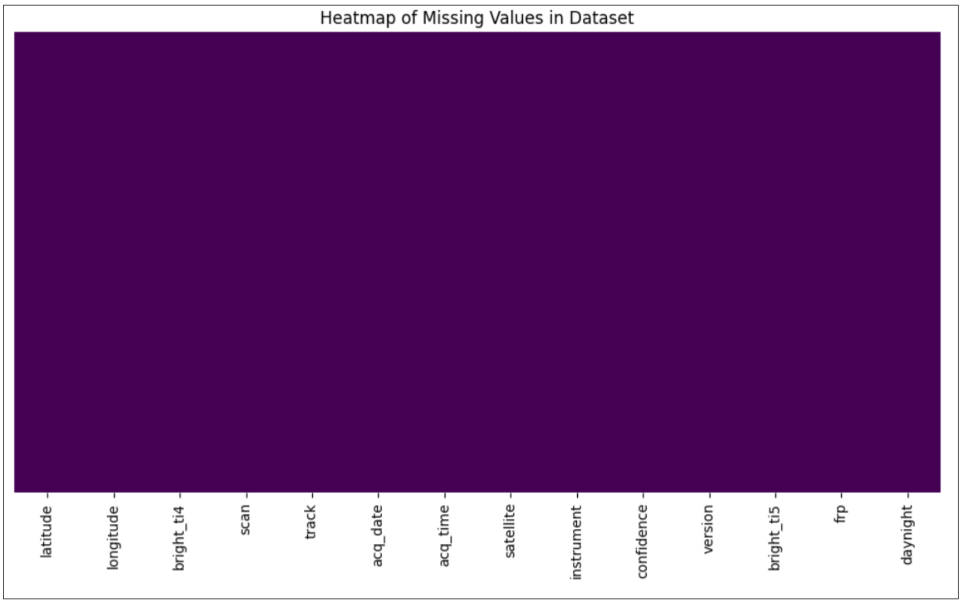


Figure 4 Heatmap

Interpretation:
This heatmap confirms that there are **no missing values** across any of the 14 columns in the

dataset. Each column is fully shaded, indicating complete data coverage. This allowed for direct model training and evaluation without the need for imputation or removal of incomplete records, enhancing both accuracy and reproducibility.

- **Figure A3: Histograms of Key Numerical Features** *(Section: Exploratory Data Analysis and Visualization)*

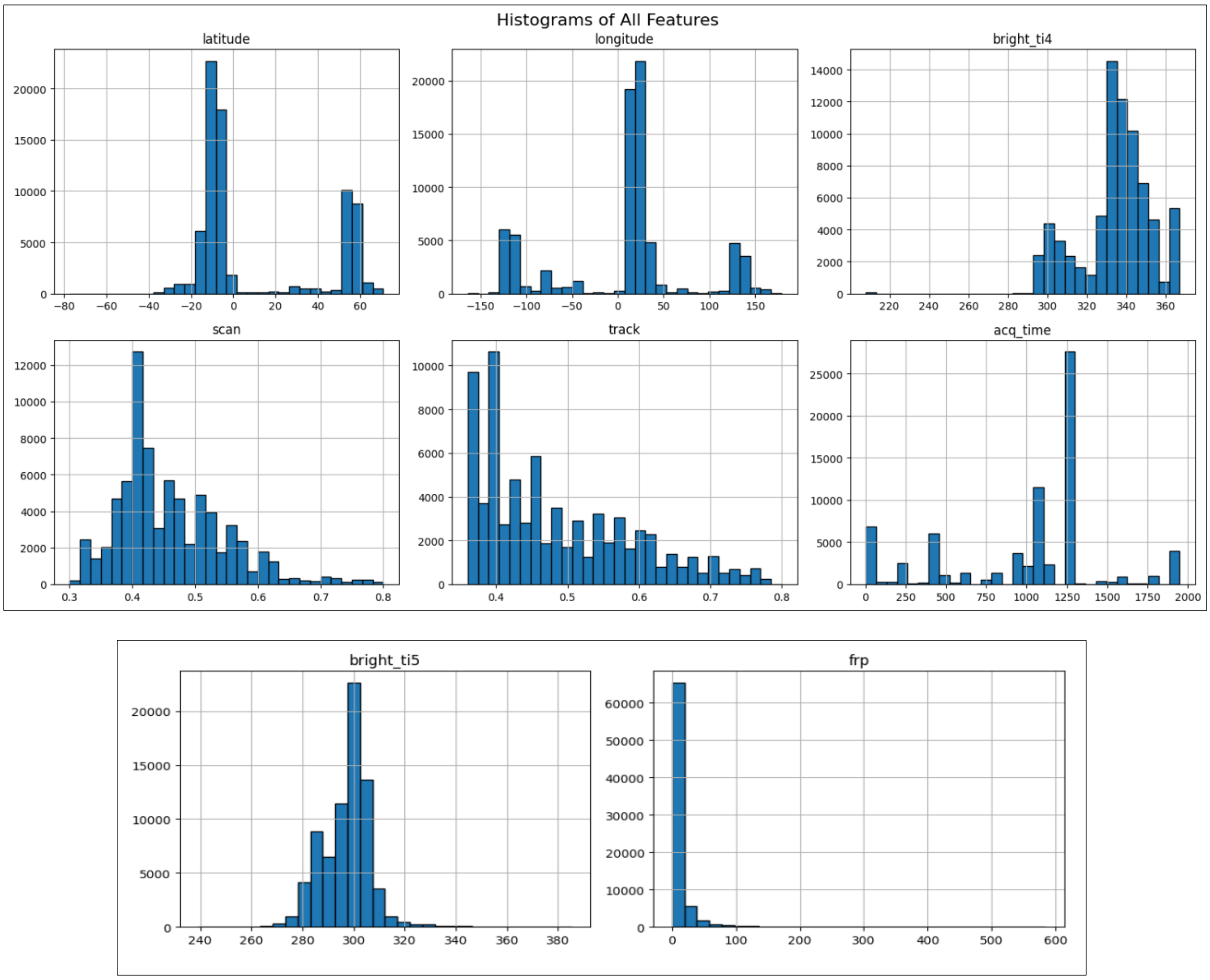


Figure 5

Interpretation:
These histograms show the distributions of major numerical features in the dataset. The latitude and longitude distributions confirm geographic clustering of fire incidents, with heavy concentrations in specific regions (e.g., Central Africa, South America, and Australia). The brightness temperature features (bright_ti4, bright_ti5) exhibit near-normal distributions, suggesting consistent satellite sensor readings.

Notably, the frp (Fire Radiative Power) distribution is heavily right-skewed, indicating that most fires are of low intensity while a few outliers exhibit extreme energy release. The scan and track features reflect typical resolution values from satellite passes, while the acq_time histogram shows fire activity concentrated during midday hours. These patterns informed both model input selection and feature engineering strategies.

➢ **Figure A4: Distribution of Categorical Variables** *(Section: Exploratory Data Analysis and Visualization)*

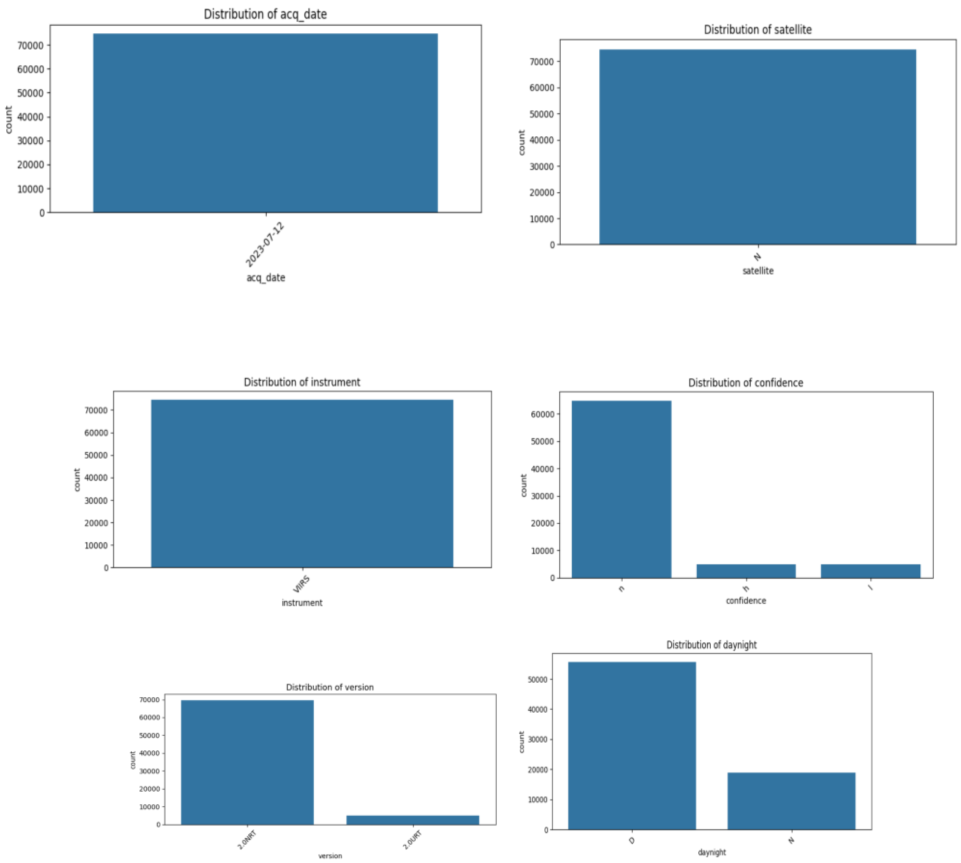


Figure 6

Interpretation:
This series of bar plots illustrates the frequency distribution of the categorical variables within the dataset:

- **acq_date:** All records were collected on a single date — July 12, 2023 — confirming the dataset is a one-day snapshot.
- **satellite:** Every fire detection was captured by the same satellite ('N' = Suomi NPP), showing source consistency.
- **instrument:** All data was recorded using the VIIRS sensor, ensuring uniform thermal measurement.
- **confidence:** The majority of fire detections were tagged with 'nominal' confidence, with significantly fewer classified as low or high — highlighting class imbalance for classification tasks.
- **daynight:** Most fires were detected during the day, although a substantial number of nighttime detections exist. This split supports time-of-day comparisons.
- **version:** Most data used version 2.0NRT, with a smaller subset recorded as 2.0URT, indicating slight variation in data processing methods.

These distributions provided critical context for model development, particularly for feature encoding and imbalance handling.

- **Figure A5: Distribution of Fire Confidence Levels** *(Section: Data Preprocessing & Classification Modeling)*

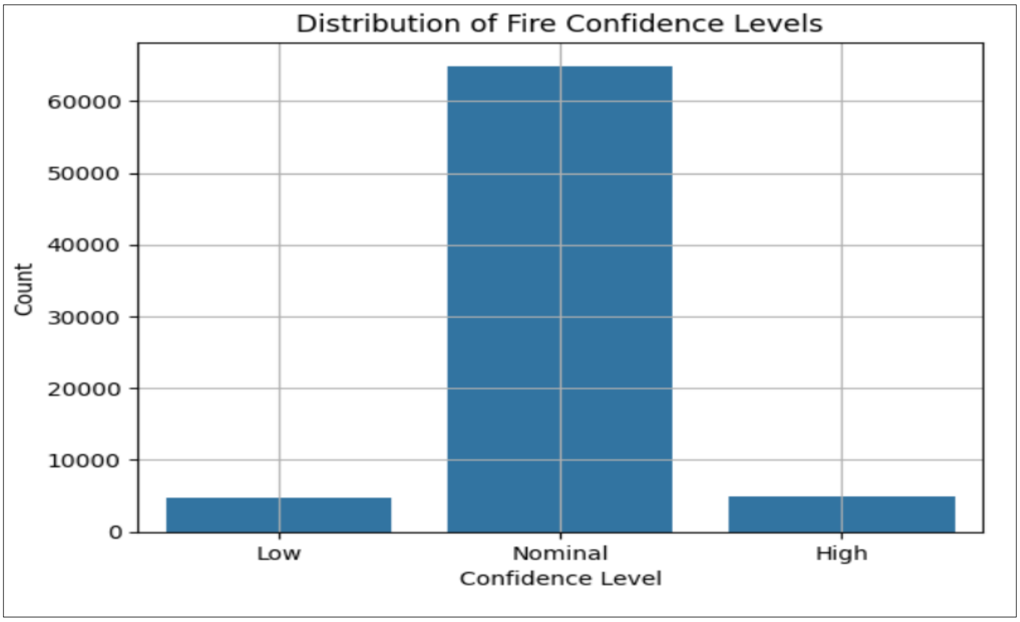


Figure 7

Interpretation:
This bar chart shows the distribution of fire confidence levels after categorical relabeling. The 'Nominal' category dominates the dataset, with over 60,000 detections, while 'Low' and 'High' confidence levels are each under 6,000. This significant class imbalance posed a challenge for classification models, particularly logistic regression. As a result, methods like random forest were preferred due to their robustness against skewed classes and ability to capture nonlinear relationships.

- **Figure A6: Global Fire Detections by Confidence Level and Radiative Power** *(Section: Exploratory Data Analysis and Visualization)*

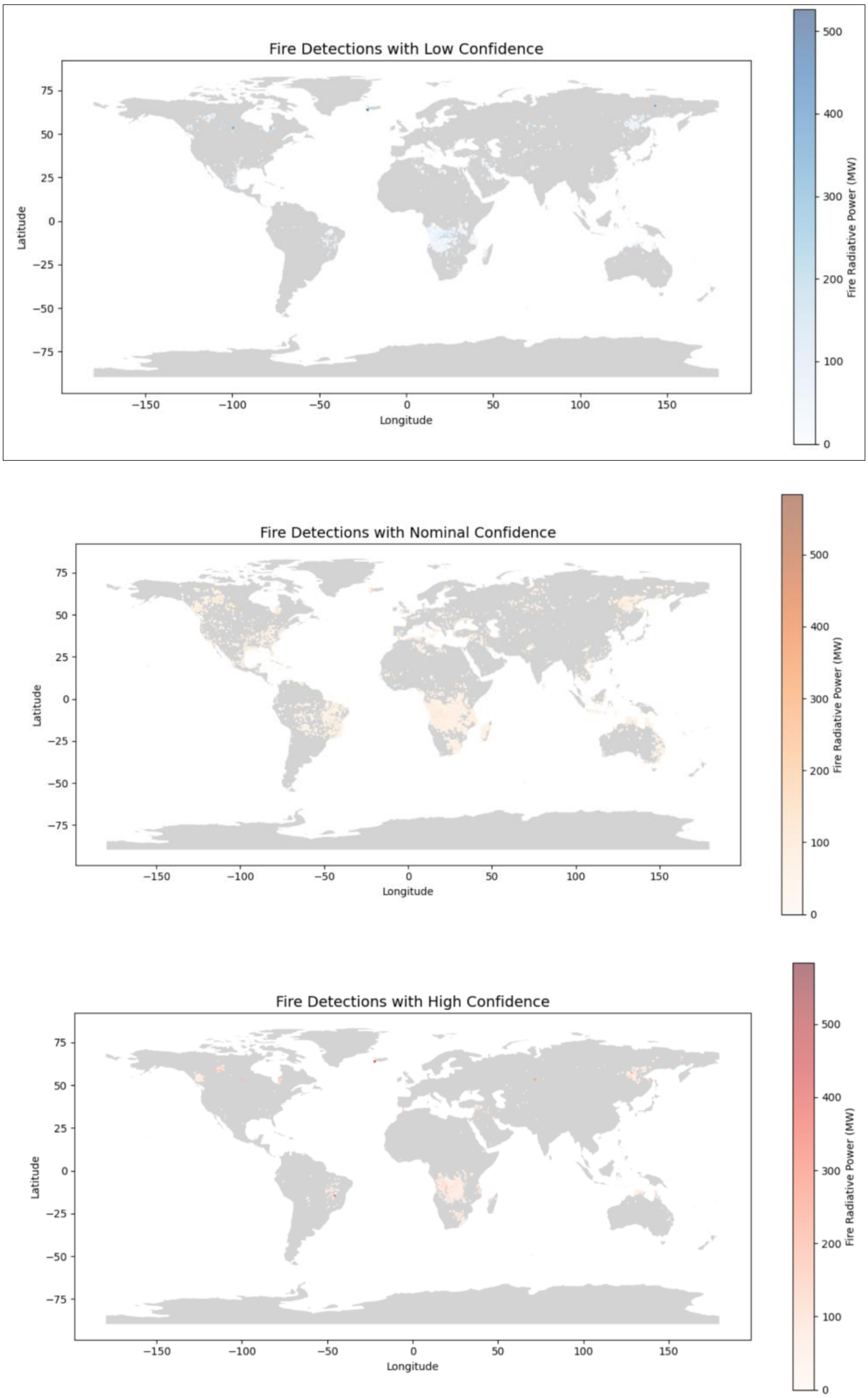

Fire Detections with Low Confidence
Latitude
Longitude
Fire Radiative Power (MW)
75
50
25
0
−25
−50
−75
−150
−100
−50
0
50
100
150
500
400
300
200
100
0
Fire Detections with Nominal Confidence
Latitude
Longitude
Fire Radiative Power (MW)
Fire Detections with High Confidence
Latitude
Longitude
Fire Radiative Power (MW)

Figure 8

Interpretation:
These maps show the geographic distribution of fire incidents across the globe, separated by detection confidence levels. The color gradient represents Fire Radiative Power (MW), where darker shades indicate stronger fire intensity.

- **Low Confidence Fires:** Sparse and scattered globally, with fewer high-FRP events. These tend to occur in less active regions or areas with sensor noise or atmospheric interference.
- **Nominal Confidence Fires:** Most frequent and widely distributed, especially across fire-prone regions like Central Africa, Brazil, Southeast Asia, and Northern Australia. These represent the majority of active fire detections.
- **High Confidence Fires:** Less frequent but often associated with more intense FRP values. Notably concentrated in Central Africa, parts of Brazil, and Australia — areas known for recurring wildfire activity or seasonal agricultural burns.

These maps highlight both the spatial scale and certainty behind satellite fire monitoring, supporting further model interpretation and regional risk assessments.

➢ **Figure A7: Fire Activity in Africa Colored by Radiative Power** *(Section: Exploratory Data Analysis and Visualization)*

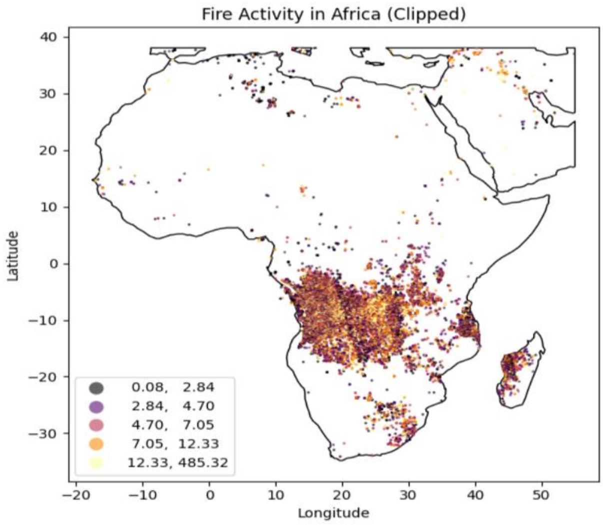


Figure 9

Interpretation:
This map zooms in on the African continent, highlighting fire activity with a detailed FRP-based color scale. The majority of fire incidents are concentrated in Central and Southern Africa, especially near the Democratic Republic of Congo, Angola, and Zambia.

Each point represents a fire detection, and the color reflects the fire's radiative power:

- Darker colors = lower intensity fires
- Lighter yellow = high FRP events (up to ~485 MW)

The dense clustering indicates both natural wildfires and controlled burns in agricultural zones. This focused visualization validates the satellite's ability to consistently detect active fires in high-risk regions and supports targeted regional analysis.

Figure A8: Fire Activity in Central America Colored by Radiative Power *(Section: Exploratory Data Analysis and Visualization)*

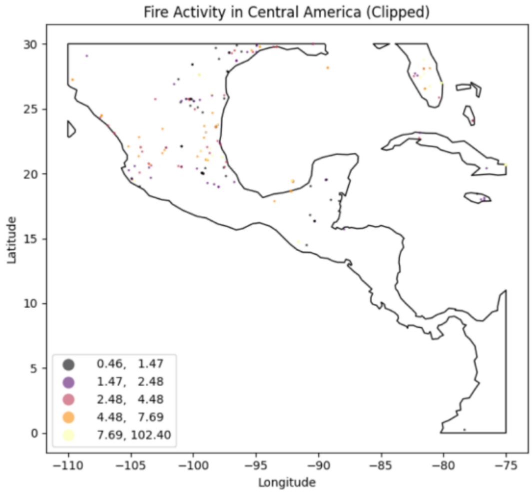


Figure 10

Interpretation:
This regional map shows fire detections across **Central America**, including Mexico, Guatemala, Honduras, and parts of the Caribbean. Each point represents an individual fire incident, and colors reflect FRP intensity:

- Black and purple points = low-intensity fires
- Yellow points = high-energy fires up to ~102 MW

Concentrated fire activity is observed in southern and central Mexico, particularly in the Yucatán Peninsula and along the Sierra Madre. Moderate fire clusters appear in Guatemala and Honduras, while Nicaragua and Panama show sparse, scattered activity.

Compared to Africa or Brazil, the density of fire activity in Central America is lower, but the geographic spread indicates the presence of seasonal or localized fire events. These may relate to land-clearing or dry climate zones. Despite the lower fire count, some high-FRP outliers are present, emphasizing the importance of monitoring even less active regions.

➢ **Figure A9: Fire Activity in South America Colored by Radiative Power** *(Section: Exploratory Data Analysis and Visualization)*

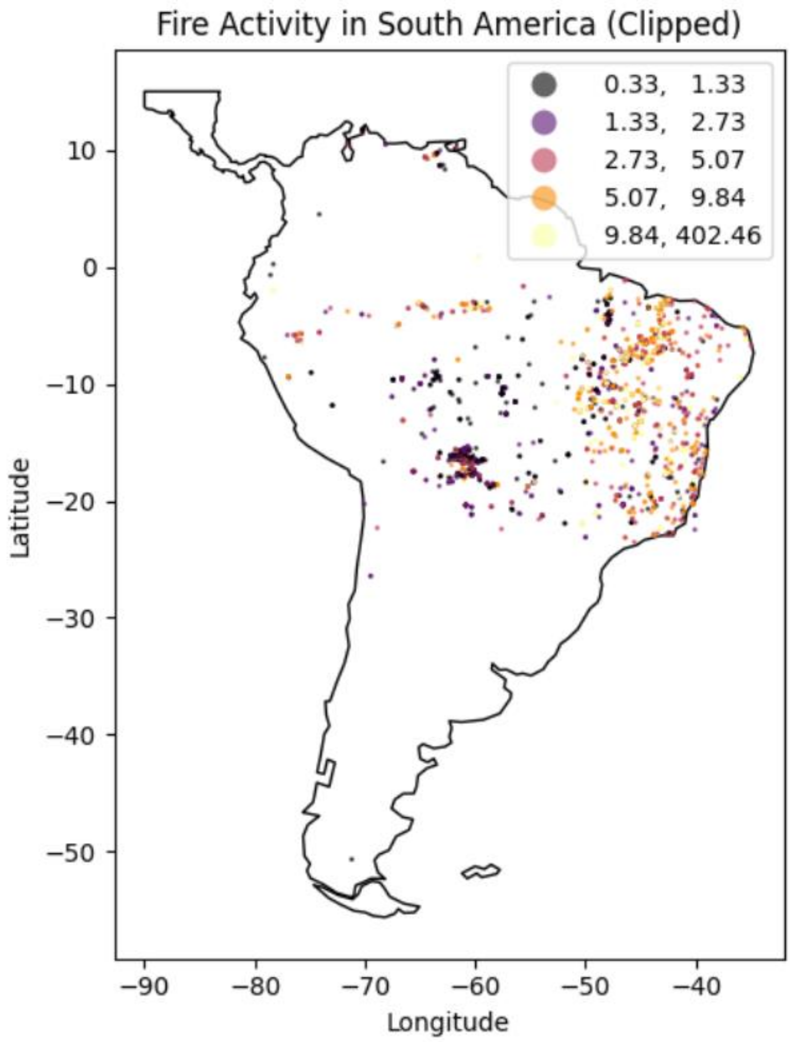


Figure 11

Interpretation:
This map visualizes the spatial distribution and intensity of fire detections across South America. The points are colored based on FRP values, ranging from low-energy (black) to high-energy fires (yellow, >400 MW).

The most intense and concentrated fire activity is located in central and southeastern Brazil, covering the Cerrado savanna biome and parts of the Amazon basin. This region is known for agricultural expansion, biomass burning, and seasonal droughts.

Additional fire clusters are observed in:

- Eastern Bolivia, Paraguay, and northern Argentina, where land is frequently cleared for farming and grazing.
- The Amazon northwest displays a more diffuse pattern, likely due to dense rainforest canopy reducing satellite visibility and limiting fire spread.

- Southern cone countries like Chile and Argentina show sparser, more isolated fires, often tied to natural wildfires or rural burns.

This regional analysis helps pinpoint ecological hotspots, where fire management and monitoring are especially critical.

➢ **Figure A10: Fire Activity in Australia Colored by Radiative Power** *(Section: Exploratory Data Analysis and Visualization)*

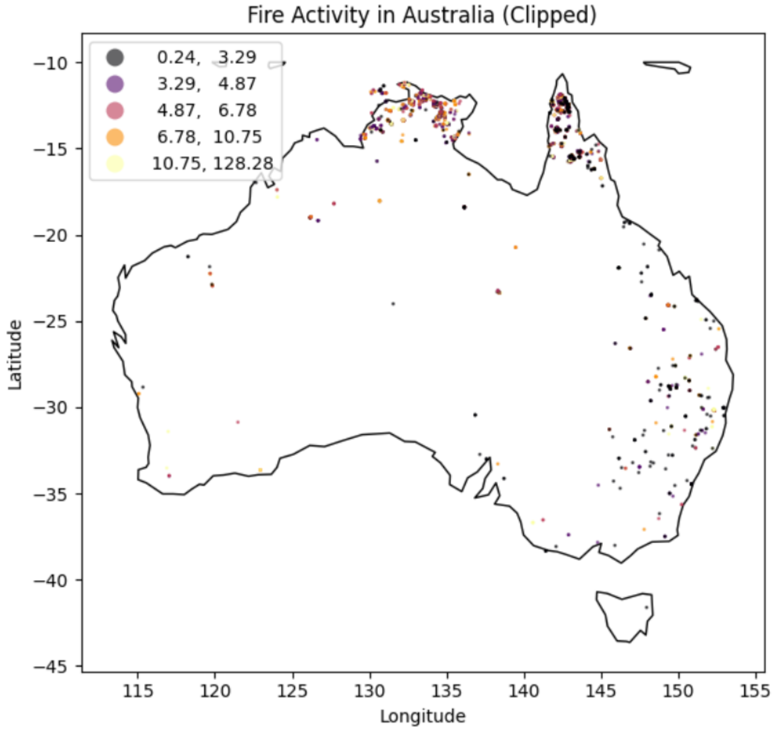


Figure 12

Interpretation:
This map illustrates the distribution and intensity of fire activity across **Australia**, with points colored by **Fire Radiative Power (FRP)**. The inferno colormap (black to yellow) visually distinguishes energy output, and the legend divides FRP into five quantiles.

- Northern Australia, especially near the coastal regions of the Northern Territory and Queensland, shows the highest concentration of fires, including several high-FRP hotspots. These are likely influenced by seasonal weather patterns and dry vegetation.
- The eastern and southwestern coasts also exhibit scattered fire activity, but generally with lower FRP levels, suggesting milder burns or controlled vegetation clearing.
- Central Australia displays sparse fire occurrences, attributed to arid desert landscapes and minimal flammable biomass.
- The color gradient provides an intuitive view of intensity, with yellow points highlighting the most extreme fire events (up to 128 MW).

This regional snapshot reinforces Australia's known fire-prone zones and showcases the effectiveness of satellite-based FRP monitoring in diverse ecosystems.

Figure A11: Fire Activity in East Asia Colored by Radiative Power *(Section: Exploratory Data Analysis and Visualization)*

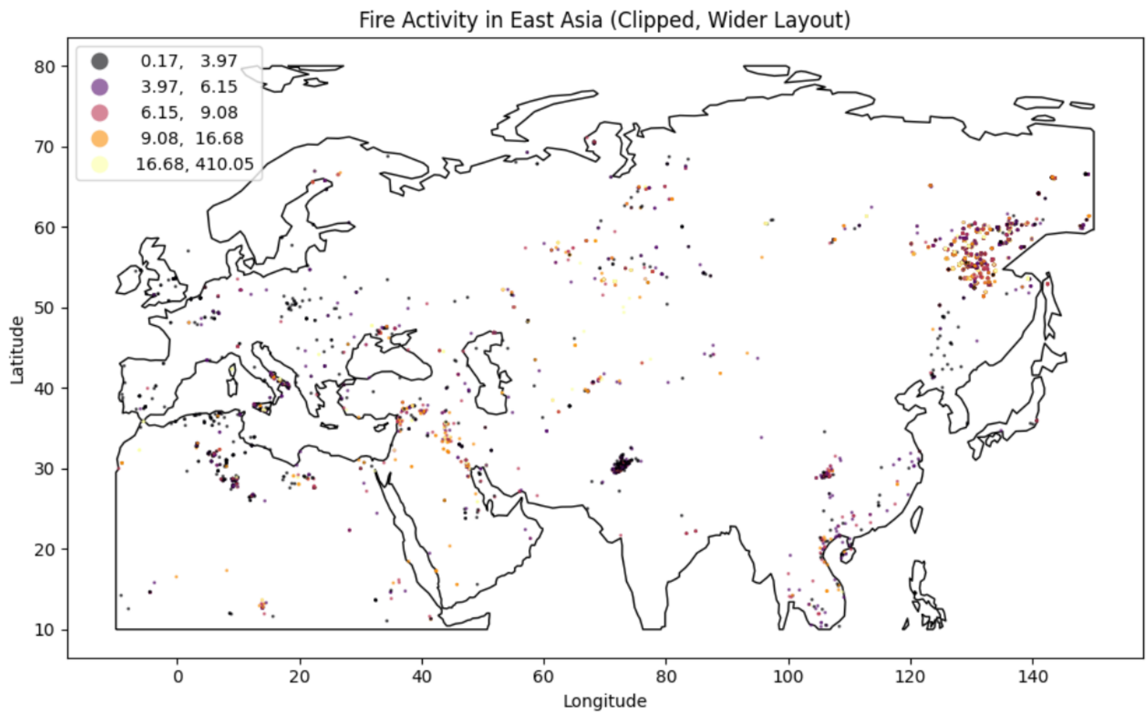


Figure 13

Interpretation:
This map presents a detailed view of fire activity across East Asia, with FRP values segmented into five quantile ranges. Each point represents a fire detection, and color denotes intensity — from low (black) to high (yellow, >400 MW).

The highest fire density is located in northeastern China, especially within the Heilongjiang and Jilin provinces, known for expansive croplands and seasonal burning. Additional clusters appear in:

- Central and southeastern China
- Coastal areas near the Korean Peninsula and southern Japan

These regions show moderate FRP values, suggesting agricultural or land-clearing fires.

In contrast:

- Central Asia, Siberia, and the Russian Far East show minimal activity, likely due to lower population density, cooler climate, or limited flammable vegetation within this clipped extent.

This visualization emphasizes regional variation in fire frequency and intensity, supporting localized fire management strategies and policy-making.

Figure A12: Correlation Matrix of Satellite-Derived Features *(Section: Exploratory Data Analysis and Visualization)*

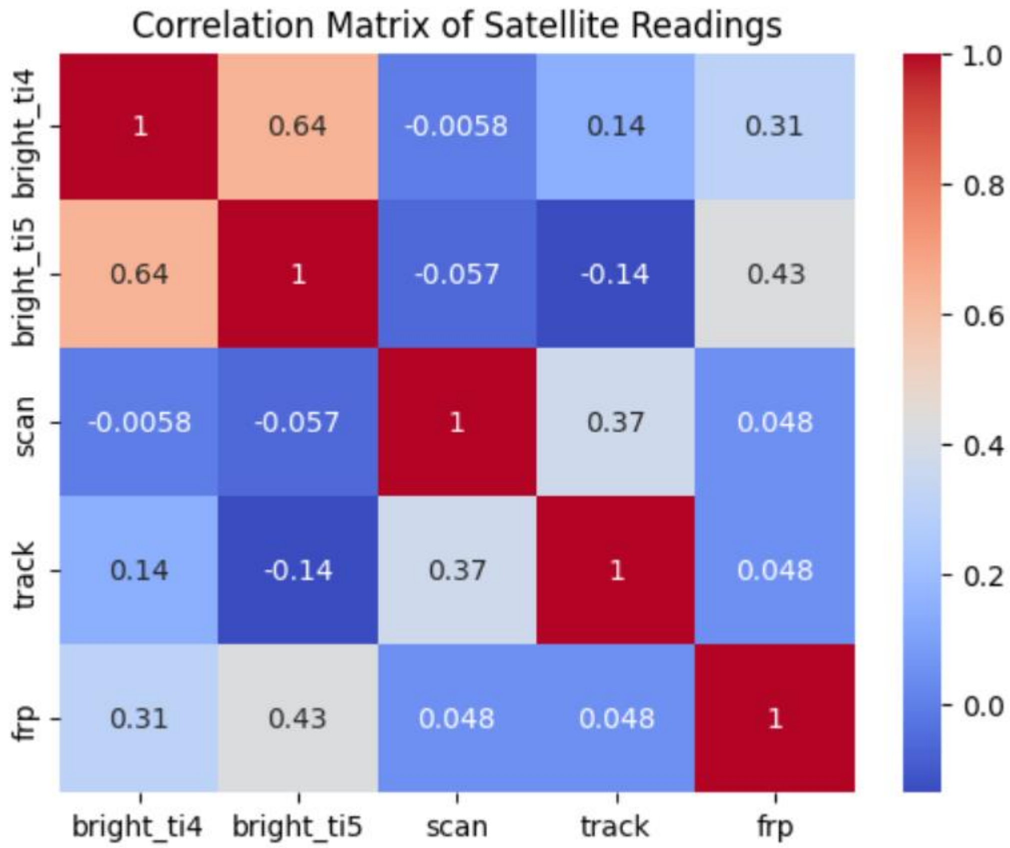


Figure 14

Interpretation:
This heatmap visualizes the pairwise correlations between key satellite-derived numerical features — brightness temperatures (bright_ti4, bright_ti5), scan, track, and frp.

Key takeaways:

- Strong positive correlation (0.64) between bright_ti4 and bright_ti5, which makes sense since both represent thermal measurements from different bands.
- bright_ti5 shows the highest correlation with FRP (0.43), followed by bright_ti4 (0.31), confirming that temperature is a key driver of fire intensity.
- scan and track show moderate correlation with each other (0.37) due to their shared geometric role in satellite passes but have minimal relationship with FRP.

This matrix supported our model decisions, confirming that brightness features should be prioritized for classification and regression tasks.

➢ **Figure A13: Scatterplot Analysis – FRP vs Satellite-Derived Variables** *(Section: Exploratory Data Analysis and Visualization)*

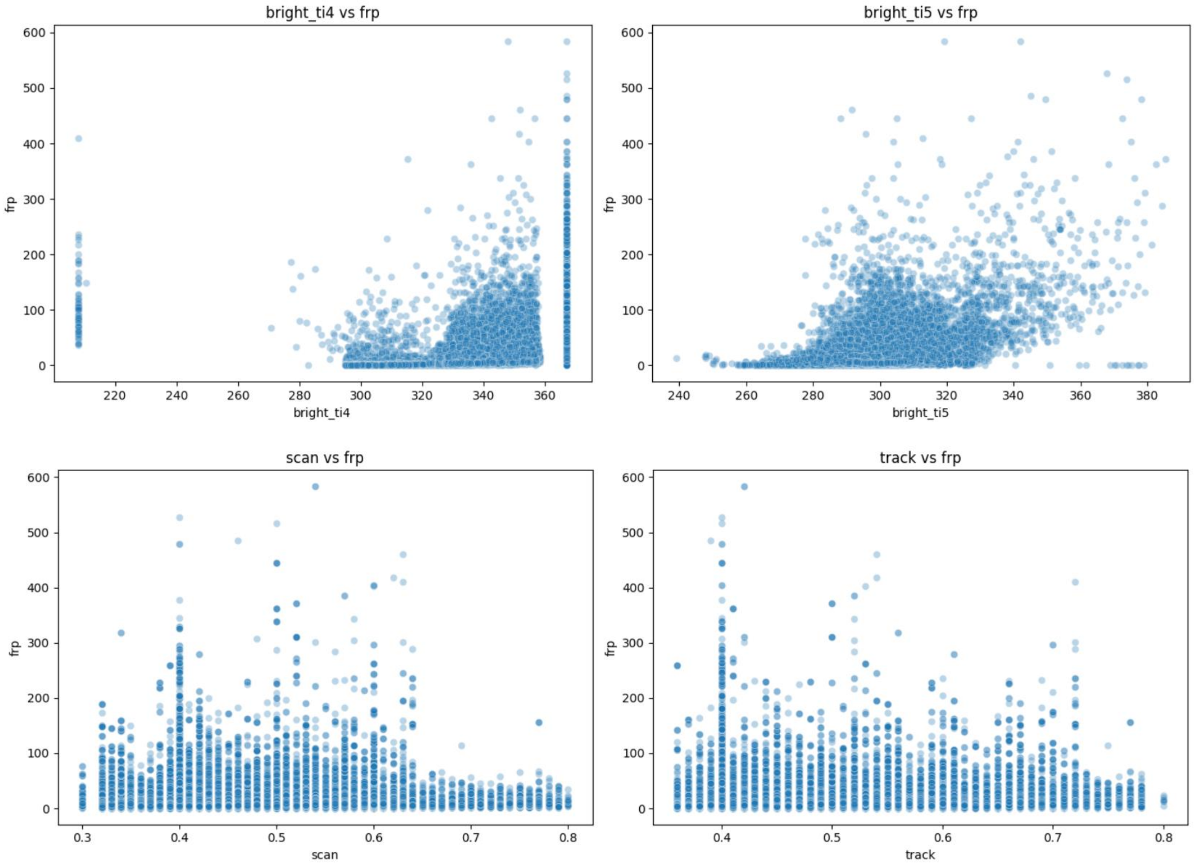


Figure 15

Interpretation:

This set of scatterplots explores how Fire Radiative Power (FRP) varies with key satellite-derived features:

Top Left – bright_ti4 vs frp:

Displays a strong positive correlation. As bright_ti4 increases, FRP also rises — confirming that brightness temperature from the mid-infrared band is a reliable indicator of fire intensity. A dense cluster of points suggests consistency in this relationship.

Top Right – bright_ti5 vs frp:

Shows a moderate positive trend, although the association is weaker than with bright_ti4. Some scatter suggests added noise, but bright_ti5 still contributes meaningfully to FRP prediction.

Bottom Left – scan vs frp:

No visible linear pattern. FRP values are spread evenly across scan values, implying that across-track resolution (scan) has little impact on estimating fire power.

Bottom Right – track vs frp:

Similar to scan, this plot reveals no significant relationship. The track parameter (along-track pixel size) does not appear to influence fire intensity either.

Summary:

- bright_ti4 is the most informative feature for predicting fire intensity
- bright_ti5 holds some predictive value, though less clear
- scan and track are not useful predictors for FRP and likely reflect satellite geometry rather than fire characteristics

➢ **Figure A14: Logistic Regression – Confusion Matrix and Classification Report** *(Section: Classification Model Evaluation)*

Logistic Regression Classification Report

	precision	recall	f1-score	support
High	0.93	0.97	0.95	1474
Low	0.70	0.15	0.25	1437
Nominal	0.95	1.00	0.97	19471
accuracy			0.94	22382
macro avg	0.86	0.71	0.72	22382
weighted avg	0.93	0.94	0.92	22382

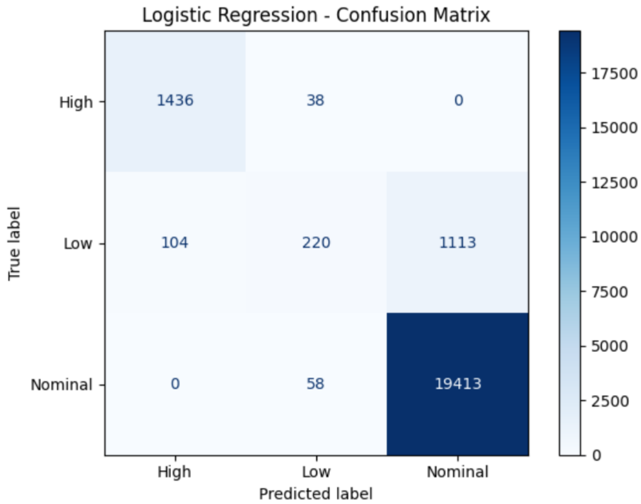
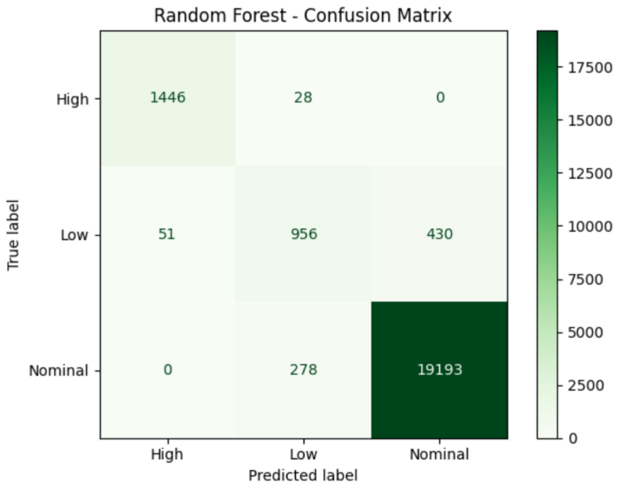


Figure 16

The logistic regression model achieved 94% accuracy with strong performance on High and Nominal fire confidence levels. It correctly classified most High and Nominal instances, with F1-scores of 0.95 and 0.97, respectively. However, it struggled with the Low class, achieving only 15% recall and an F1-score of 0.25. This indicates the model favored the dominant Nominal class and failed to detect Low-confidence fires effectively. These limitations led us to explore more balanced models like Random Forest.

➢ **Figure A15: Random Forest – Confusion Matrix and Classification Report**




```
Random Forest Classification Report
              precision    recall  f1-score   support

        High       0.97      0.98      0.97      1474
         Low       0.76      0.67      0.71      1437
     Nominal       0.98      0.99      0.98     19471

    accuracy                           0.96     22382
   macro avg       0.90      0.88      0.89     22382
weighted avg       0.96      0.96      0.96     22382
```

Figure 17

Interpretation

The Random Forest model achieved a high overall accuracy of 96%, significantly improving performance on the previously underperforming Low-confidence class. Precision and recall were strong across all categories, especially for the High and Nominal classes, both scoring above 0.97 in F1. Notably, recall for the Low class jumped from 0.15 (in logistic regression) to 0.67, showing that Random Forest handled class imbalance much better. These results highlight Random Forest as a more balanced and robust classifier for predicting fire confidence levels.

➢ **Random Forest – Feature Importances for Predicting Fire Radiative Power (FRP)**

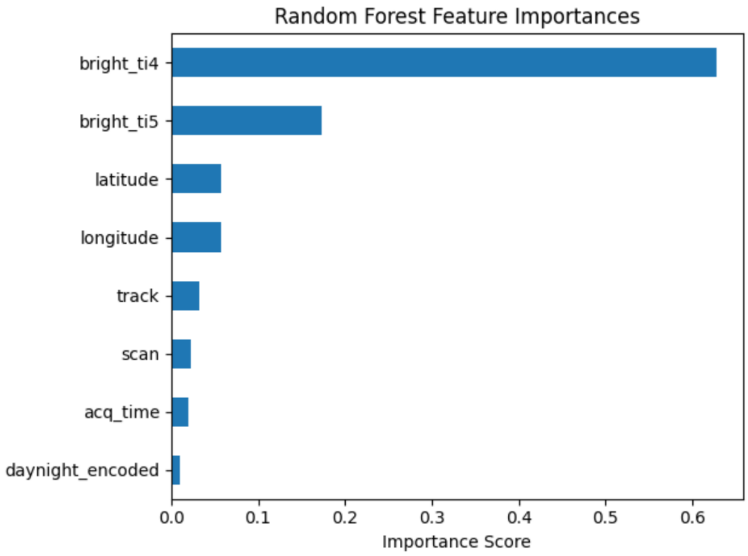


Figure 18

Interpretation

The feature importance plot reveals that bright_ti4, representing mid-infrared brightness temperature, is the most influential predictor of fire intensity, with an importance score exceeding 0.6. This confirms its critical role in estimating FRP. The second most valuable feature is bright_ti5, though its impact is significantly lower. Geographic features like latitude and longitude show moderate influence, likely due to regional fire patterns. In contrast, variables related to satellite geometry (scan, track) and timing (acq_time) have minimal impact. The day/night indicator is the least important, suggesting time of observation has little effect on fire power predictions.

- **Gradient Boosting Model Performance: Predicting Fire Radiative Power (FRP)(Gradient Boosting Model Evaluation Output)**

```
Gradient Boosting Model evaluation
MSE: 264.69
RMSE: 16.27
R² Score: 0.50

Best Parameters:
{'regressor__learning_rate': 0.05, 'regressor__max_depth': 5, 'regressor__min_samples_leaf': 2, 'regressor__min_samples_split': 5, 'regressor__n_estimato
rs': 200, 'regressor__subsample': 1.0}
```

Gradient Boosting FRP Prediction

- RMSE of ~16.23 MW suggests the model predicts fire power with moderate accuracy.
- R^2 Score of **~0.5** indicates it captures half of the variance in FRP.
- Despite our best efforts, this is stuck at nearly a 50% predictive power. T

This suggests that the dataset could benefit from additional engineered or spatial-temporal features.

The Gradient Boosting Regressor demonstrated moderate predictive performance, with an RMSE of **~16.23 MW**, indicating how much the model's fire power predictions deviate from actual values on average. An **R^2 score of 0.50** means that the model is able to explain **about** 50% of the variance in FRP—a reasonable starting point but far from optimal. Despite hyperparameter tuning, the model's ceiling suggests that FRP is influenced by additional complex factors not fully captured in this dataset. Future work could explore more advanced feature engineering, especially incorporating spatial-temporal context**,** land cover types, or environmental conditions like wind speed or humidity.

➢ **FRP and Brightness Temperatures in Anomalies (Section: Anomaly Detection and FRP Outliers)**

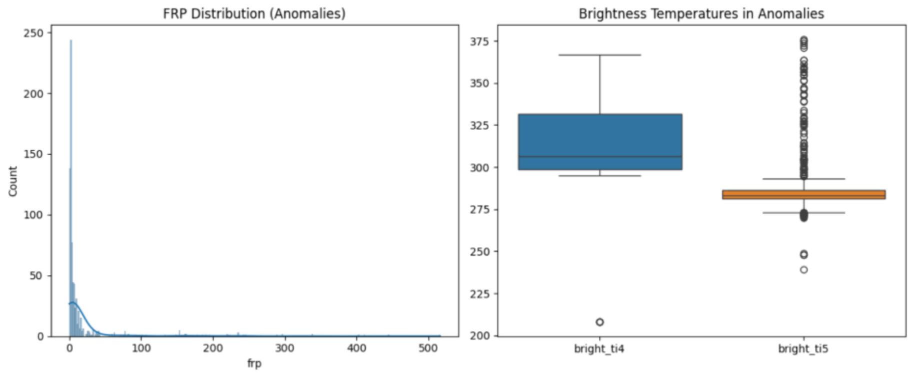


Figure 19

Interpretation:
The histogram on the left illustrates the highly skewed distribution of Fire Radiative Power (FRP) among detected anomalies. Most values are clustered near zero, indicating a majority of low-intensity fire events, while a long right tail reveals rare but significant high-intensity outliers.

The boxplots on the right compare the distribution of brightness temperatures (bright_ti4 and bright_ti5) within anomalous cases. bright_ti5 shows more extreme outliers than bright_ti4, emphasizing its sensitivity to intense thermal events. These insights confirm that fire anomalies are associated with distinct thermal characteristics and justify their special treatment in the modeling phase.

Using Isolation Forest, 746 out of 74,605 records were flagged as anomalies—potentially representing intense fires or rare events.

- **Left Plot**: Most anomalies have low FRP, but a few extreme cases exceed 500 MW.
- **Right Plot**: bright_ti4 and bright_ti5 brightness temperatures show distinct spreads, with many bright_ti5 outliers, suggesting variability in thermal signatures among anomalies.

➢ **Anomalous Fire Activity (Global View) (Section: Anomaly Detection and FRP Outliers)**

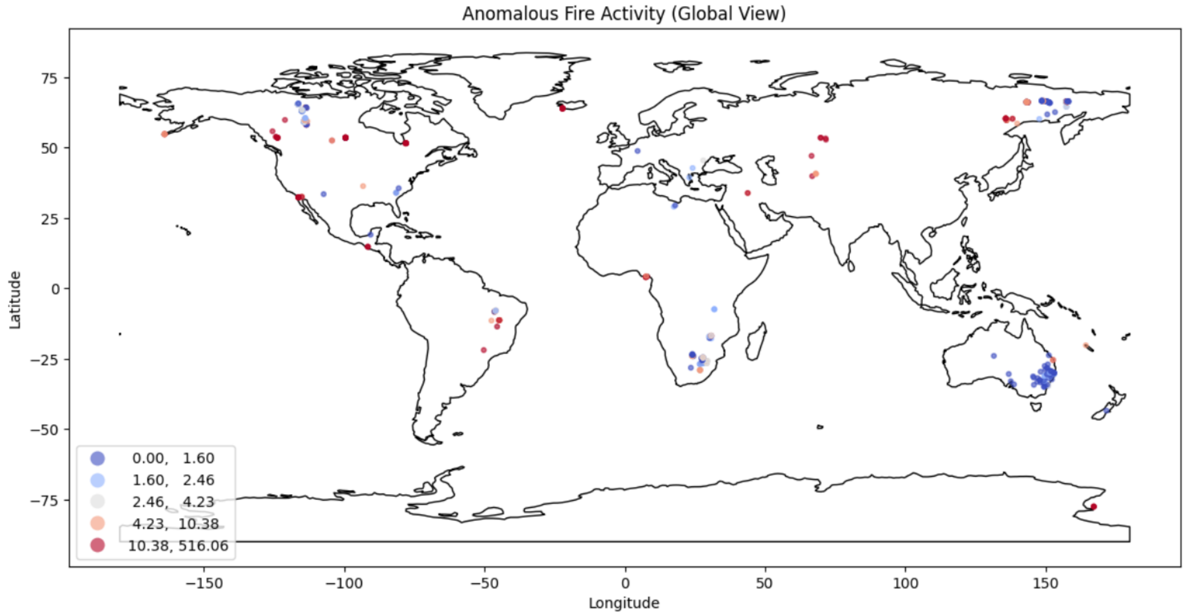


Figure 20

Interpretation:

- **FRP Distribution**: Most anomalies are low intensity (<50 MW), but rare outliers exceed 400 MW, forming a long-tailed distribution.
- **Brightness Temperatures**: bright_ti5 shows dense upper outliers, confirming its value in detecting extreme events.
- **Global Spread**: Anomalies are scattered worldwide, with FRP hotspots in North America, Central Africa, SE Asia, and Australia—highlighting rare but intense fires.

---**Thank you**---