M.Sc. Geoinformatics Engineering

Geoinformatics Project

Project Technical Report

COMPARISION OF SPATIO-TEMPORAL BURNED AREA DISTRIBUTION USING GOOGLE EARTH ENGINE

Tutor: Prof. Venuti Giovanna
(Chair of Geoinformatics Engineering Department, Politecnico di Milano)

Co-Tutor: Dr. Daniela Stroppiana
(Senior Researcher, CNR-IREA)

Authors: Rishikesh Miriyala (10814272)
rishikesh.miriyala@mail.polimi.it
Sriram Seenivasan Krishnasamy (10888750)
sriram.krishnasamy@mail.polimi.it
Ghulam Abbas Zafari (10944305)
ghulamabbas.zafari@mail.polimi.it

GitHub Link: https://github.com/Sri603/Comparision-of-spatio-temporal-BAD-using-GEE.git

Table of Contents

ABSTRACT

The Spatio-Temporal distribution of burned areas (BA) is critical for understanding fire dynamics and their ecological, climatic, and socio-economic impacts. This study provides a comprehensive comparison of burned areas estimates derived from major global datasets, including MCD64 and VIIRS. Using multi-annual datasets in Google Earth Engine (GEE), we analyze spatial and temporal patterns of burned areas at global and regional scales. Our findings highlight the variability and land use practices over Australia, contributing to improved monitoring methodologies. A custom-developed algorithm implemented in GEE facilitates the comparison of products, ensuring consistent and accurate assessment.

We propose two different steps for extracting information on wildfires from multi-spectral satellite imagery to identify the factors affecting wildfire distribution across Australia. First, an image incorporating all relevant factors will be created using Google Earth Engine (GEE). We will visualize burned areas, forest loss, and land use/land cover through machine learning classification tools. Active fires, burned areas, and other contributing factors will be extracted by filtering data based on date, region of interest (Australia), with suitable band and scale, and further visualized using Python.

The second step includes creating a burn mask and predictor image using soil moisture, precipitation data, elevation, etc. Further stratified sampling of the burn mask is done to create a dataset of points that represent burned and unburned areas. This method ensures that both classes are proportionally represented in the sample. The predictor image is trained over the sample using a Gradient Boost Tree classifier, allowing us to achieve higher accuracy.

Our work sets the scene and material for further analysis of these factors while providing preliminary experimentation with the data.

KEYWORDS:

Burned Area (BA), Spatio-Temporal Analysis, Google Earth Engine (GEE), MCD64, VIIRS, Land Use Land Cover, Global Forest Cover, Multi-Annual Analysis, Gradient Boost Tree.

Ⅰ. INTRODUCTION:

Australia experiences a significant number of bushfires, particularly during dry summer months. These fires have devastating impacts on the environment, human communities, and the economy. The distribution of burned areas across Australia is not uniform and varies considerably depending on factors such as climate, vegetation, and human activities. Understanding the trends in burned area distribution, the impacts of fires, and effective management strategies is crucial for mitigating the risk and promoting sustainable land management [1].

The distribution of burned areas in Australia is influenced by a complex interplay of factors. Hotter and drier conditions increase the risk of fire ignition and spread. Key meteorological variables include temperature, precipitation, humidity, wind speed, and wind direction. For instance, the 2019–2020 bushfire season was exacerbated by ongoing drought, low soil moisture, and heatwaves. Human activities, such as land clearing, agricultural practices, and the presence of infrastructure, also influence fire regimes and the distribution of burned areas [1].

The impacts of bushfires are far-reaching. Fires can result in the loss of life, damage to property, and disruption to essential services. They also have significant environmental consequences, including loss of biodiversity, habitat destruction, soil erosion, and reduced water quality. Bushfires release substantial amounts of greenhouse gases into the atmosphere, exacerbating climate change. Furthermore, they have socioeconomic impacts, affecting tourism, agriculture, and timber production.

Managing bushfires effectively requires a multi-pronged approach. Prevention strategies, such as reducing fuel loads, controlling ignition sources, and implementing fire-resistant building practices, are critical for minimizing the risk of fires. Early detection and rapid response are essential for containing fires and reducing their impact. This involves investing in advanced monitoring systems, deploying firefighting resources, and ensuring the effectiveness of community fire management programs [2].

Australia has a long history of fire management, and the country's approach has evolved over time. The focus has shifted from suppression to a more holistic approach that incorporates ecological considerations. This involves managing landscapes to reduce the intensity and spread of fires, promoting fire-tolerant species, and allowing for natural fire regimes in certain ecosystems [3].

Ⅱ. SPECIFICATIONS:

A. Objectives

The primary objectives of this study are:

1. **Dataset Comparison:** Evaluate the spatial and temporal discrepancies between MCD64 and VIIRS datasets to assess their reliability and consistency.
2. **Environmental Impact Assessment:** Analyze the influence of environmental factors such as Land Use Land Cover (LULC), temperature, soil moisture, and wind direction on the distribution of burned area across Australia.
3. **Visualization and Interpretation:** Develop interactive maps and charts for better understanding and interpretation of fire patterns across various regions of Australia.
4. **Automation and Scalability:** Utilize GEE’s cloud computing capabilities to ensure scalability and automation of the analysis across multiple timeframes and regions.
5. **Classifier Training and Evaluation:** Train and evaluate a classifier using the Gradient Boost Tree method, perform cross-validation, and generate a confusion matrix to assess classification accuracy.

B. Functionalities Used

i. Software and Tools Used

- **Google Earth Engine (GEE):** A cloud-based platform for global-scale geospatial analysis that processes a variety of geographical data at scale and handles large geographical datasets.
- **API:** GEE's Python API allows access from a Python environment such as Jupyter Notebook, which was used for the final analysis.
- **QGIS:** An open-source cross-platform GIS application that supports viewing, editing, printing, and analyzing geospatial data.

ii. Input Products Collection

The analysis relies on multiple geospatial datasets available in GEE:

Dataset	Description	Purpose
MCD64	Modis Burned Area	
SUOMI VIIRS	Active Fire Data	Fire Intensity Classes
HANSEN	Deforestation Data	Forest Fire Impact
MODIS LULC	Land Use Land Cover	Land Types Analysis
MODIS LST	Land Surface Temperature	Temperature Correlation
MODIS NDVI	Normalize Difference Vegetation Index	Pre/Post fire assessment

Table 1: Global Datasets

C. Analysis and Approach

This section describes the step-by-step approach used in implementing the burned area analysis and key methodologies employed.

i. Clipping the Region of Interest (ROI)

- **Importing Global Datasets:** Utilize GEE's API to access global datasets pertinent to burned area analysis, such as Landsat or Sentinel-2 imagery.
- **Filtering Data by Date and Clip to ROI:** Apply temporal filters to select data within specific date ranges (2019 to 2023) and spatially clip the dataset to focus on the ROI, e.g., Australia [4].
- **Select Relevant Bands:** Choose specific spectral bands necessary for the analysis, such as those used in calculating indices like BurnDate.

```
var Modis = ee.ImageCollection("MODIS/061/MCD64A1")
              .filter(ee.Filter.date('2021-01-01','2021-12-31'))
              .filterBounds(Aus);

var burned = Modis.select('BurnDate');
```

By clipping datasets to the ROI, the volume of data processed is reduced, mitigating the risk of exceeding GEE's memory limits. This approach aligns with methodologies discussed in the study [5].

Detecting and classifying burned areas is essential for understanding fire dynamics and their ecological impacts. Two prominent satellite-derived products used in the context are the MODIS MCD64A1 and the Suomi VIIRS Collection 2 enhances the accuracy and efficiency of the process [6].

ii . Classification of Fire Severity Using MCD64A1 and Suomi VIIRS C2

The MODIS MCD64A1 product provides global burned area mapping at a 500-meter resolution. It employs an improved algorithm that reduces omission errors and enhances the detection of smaller burns. The product includes data layers such as Burn Date, Burn Date Uncertainty, and Quality Assurance, which are instrumental in assessing fire severity [7].

Similarly, the VIIRS VNP64A1 burned area product continues the MODIS record into the Suomi NPP and JPSS era. It is generated from time series of daily 750-meter VIIRS land surface reflectance data, resampled to 500 meters to maintain consistency with MODIS products. This product aims to provide a consistent burned area Earth System data record, facilitating the classification of fire severity over time [8].

iii . Generating Burned Area Masking Using Threshold Values

By applying specific threshold values to active fire data, it is possible to generate burned area masks that delineate the extent of fire-affected regions (Figure 1). This approach involves analyzing spectral indices sensitive to fire effects and setting thresholds to classify burned and unburned areas [9].

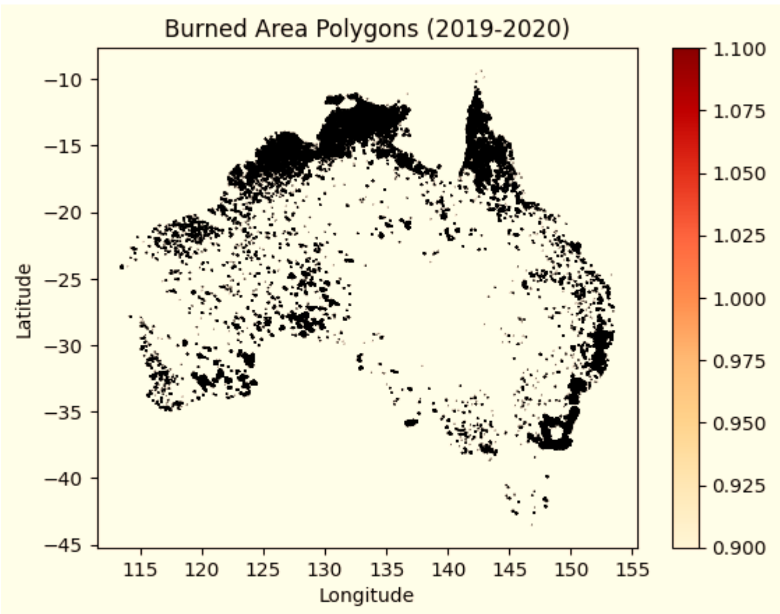

Figure 1: Burned Area Polygons Australia

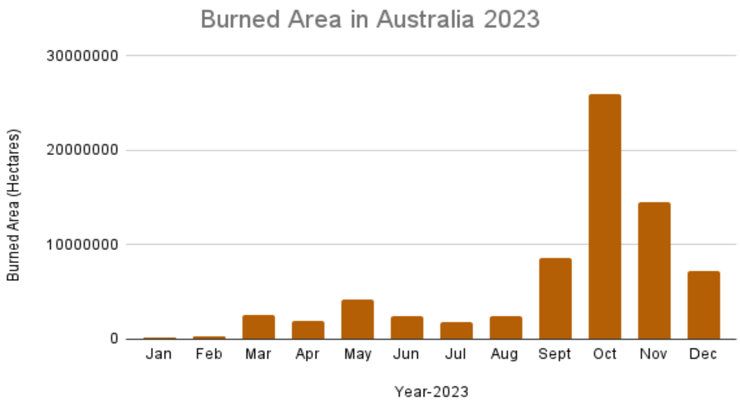

Figure 2: Monthly Burned Area in Australia 2023

iv . Spatial Analysis

- **Evaluating Land Cover Classes Affected by Fires:** Accessing the impact of fires on various lands covers classes requires the integration of burned area data with Land Use/ Land Cover (LULC) datasets. By overlaying these datasets, one can identify which land cover types are most affected by fires. For instance, a study in Europe analysed transitions between LULC classes and burned areas over two decades, revealing significant losses in forests and arable lands due to wildfires (Figure 3) [10].
- **Calculating Areas of Each Land Cover Trend During the Year:** Quantifying annual changes in land cover classes involves calculating the area of each class before and after fire events. This can be achieved using Geographic Information System (GIS) tools to process LULC datasets and detect changes over time. Such analyses help in understanding the dynamics of land cover changes and the extent of areas affected by fires (Figure 4) [10].

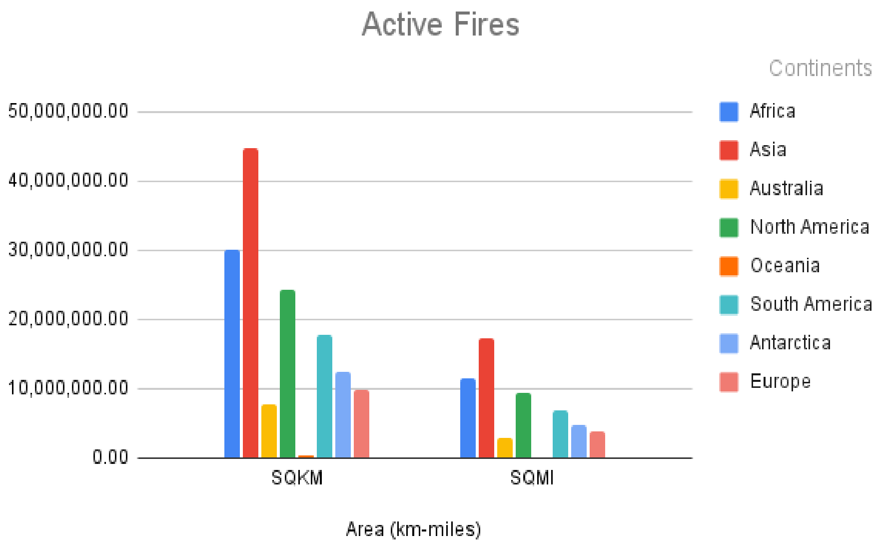

Figure 3: Active Fires Across Globe

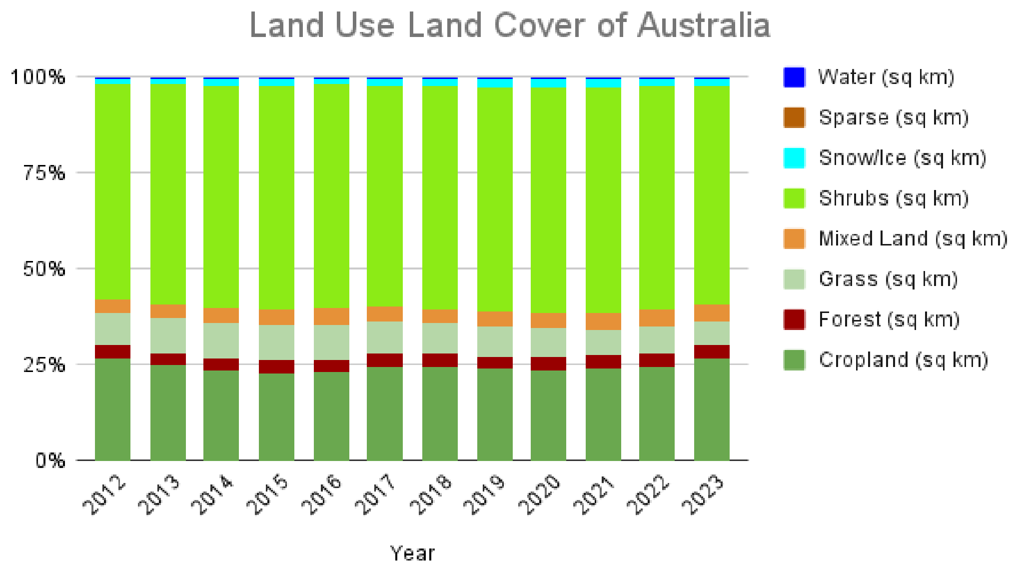

Figure 4: Land Use and Land Cover Area of Australia

Conducting a comprehensive analysis of burned areas involves spatial and temporal assessments, as well as the application of machine learning classifiers. Below are the methodologies for each component

V . Temporal Analysis

- **Creating Time Series for NDVI Patterns:** The Normalized Difference Vegetation Index (NDVI) is a key indicator of vegetation health. By generating time series of NDVI data, researchers can monitor vegetation dynamics and assess the impact of fires on vegetation health over time. This involves processing satellite imagery to extract NDVI values at different time intervals (Figure 5) [10].

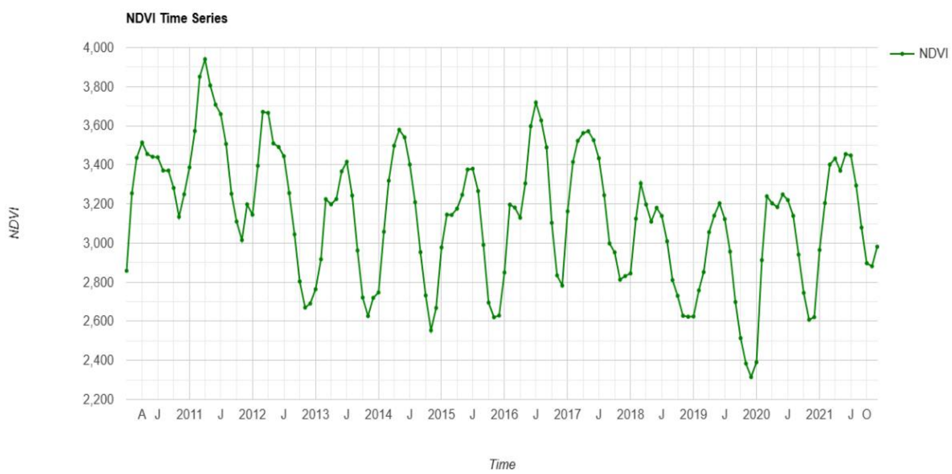

Figure 5: NDVI Time Series

- **Evaluating During the Period 2019 to 2023:** Focusing on specific time frame such as 2019 to 2023, allows for the assessment of the recent trends in the vegetation health and fire impacts. Analysing NDVI time series can reveal patterns of recovery or degradation of post fire events [10].

- **Analysing Fire Severity Patterns:** Fire severity refers to the extent of environmental change caused by fire. By examining post-fire NDVI reductions, researchers can infer the severity of fires across different regions and time periods. This analysis aids in identifying areas that require restoration or further monitoring [10].

vi. Classifier Training and Evaluation

- **Training a Gradient Boosted Tree Classifier:** Machine Learning models such as Gradient Boosted Tree are effective in predicting burned areas by learning patterns from various predictor ancillary layers. For example, GBT algorithm is utilized to predict wildfires occurrences, incorporating variables like landcover, vegetation indices, soil moisture, and climatic factors [11].

- **Dataset Split for Validation:** To evaluate the performance of the classifier, the dataset is typically divided into training and testing subsets. A common practice to allocate 70% of the data for training the model and 30% for testing its predictive accuracy. This approach ensures that the model is assessed the unseen data, providing an unbiased evaluation of its performance.

- **Generating a Confusion Matrix and Calculating Accuracy Metrics:** After making predictions on the test dataset, a confusion matrix is generated to compare predicted and actual outcomes. This matrix helps in calculating various accuracy metrics such as precision, recall and overall accuracy, which are essential for assessing the model's effectiveness in predicting burned areas [10].
- By integrating spatial and temporal analysis with machine learning techniques, we develop a robust model for predicting and understanding the dynamics of burned areas, thereby aiding in effective wildfire management and mitigation strategies.

Ⅲ. IMPLEMENTATION

In the implementation phase, we apply techniques to analyse burned areas using geospatial dataset and machine learning algorithms. focuses on applying methodologies and techniques required to analyse. This section encompasses data acquisition, processing burned area classification, spatial and temporal analysis.

A. Data Acquisition and Processing

ⅰ. Data Collection

- **Datasets:** The study utilizes datasets from Google Earth Engine (GEE), including MODIS MCD64A1, MODIS NDVI, LULC, Land Surface Temperature (LST), and SUOMI VIIRS. These datasets provide comprehensive information on burned areas, vegetation health, land cover types, and temperature variations.

- **Region of Interest (ROI):** The study focuses on Australia, defined using administrative boundary layers from the Global Administrative Unit Layers (GAUL) dataset.

- **Analysis Period:** In the implementation phase, we apply techniques to analyse burned areas using geospatial dataset and machine learning algorithms. focuses on applying methodologies and techniques required to analyse. This section encompasses data acquisition, processing burned area classification, spatial and temporal analysis.

ii . Data Cleaning and Processing

- **Alignment:** Datasets are reprojected and rescaled to ensure uniform spatial resolution and coordinate systems, facilitating accurate overlay and analysis.
- **Temporal Adjustment:** The date ranges of datasets are adjusted to align with the study period, ensuring consistency across temporal analysis.
- **Calculations:** Burned area extents are computed, LULC types are classified, and NDVI values are calculated to assess vegetation health.
- **Band Extraction:** Relevant spectral bands are extracted for visualization and analysis.

iii . Feature Extraction

- **Burn Metrics:** Derive burn dates, assess fire severity, and delineate burned areas using the MCD64A1 dataset.
- **LULC Analysis:** Calculate areas of different land cover types affected by fires.
- **Burn Mask Creation:** A burn mask generated from predictor images and monthly data to identify and isolate burned areas.

B. Burned Area Classification Approach

i . Burn Mask Creation: A burn mask is generated using the MCD64A1 dataset by selecting pixels corresponding to burned areas. Data is reprojected to a common reference system and clipped to the ROI.

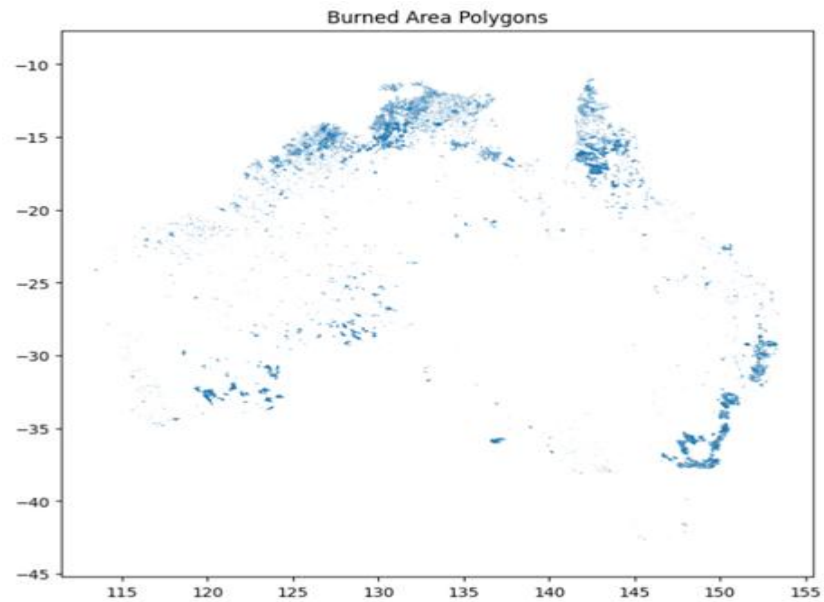

Figure 6: Burned Area Polygons Australia 2019

ii . Predictor Image Compilation: A predictor image is created by integrating datasets such as NDVI, MODIS LULC, soil moisture, elevation, precipitation, and temperature. These ancillary factors influence fire dynamics.

iii . Gradient Boost Tree Classifier Implementation: The Gradient Boost Tree (GBT) classifier is selected for its ability to handle non-linearity and achieve high predictive accuracy. Key parameters include:

- **Number of Trees:** 3000 trees are used to reduce bias and capture complex patterns, acknowledging the trade off with increased training time.
- **Learning Rate:** A low learning rate of 0.001 ensures gradual learning and enhance precision.
- **Sampling Rate:** A sampling rate of 0.8 introduces randomness and minimizes overfitting.
- **Maximum Nodes:** A maximum of 512 nodes per tree capture complex relationships in the data.

C. Spatial and Temporal Analysis

i . Spatial Analysis:

- **Burned Area Distribution:** Thematic maps are created to visualize the spatial distribution of burned areas across Australia.
- **Active Fires Severity:** Active fires are categorized by severity levels to assess immediate impact (Figure 7).
- **Land Cover Impact:** Analysis of MODIS LULC data determines the extent of land cover types affected by fires.
- **Temperature Variation:** The LST data monitors temperature variations during forest fire events, providing insights into fire behaviour.
- **Forest Loss Assessment:** Global Forest Change (GFC) data is used to calculate forest loss, enhancing understanding of fire effects on forest patterns [12].

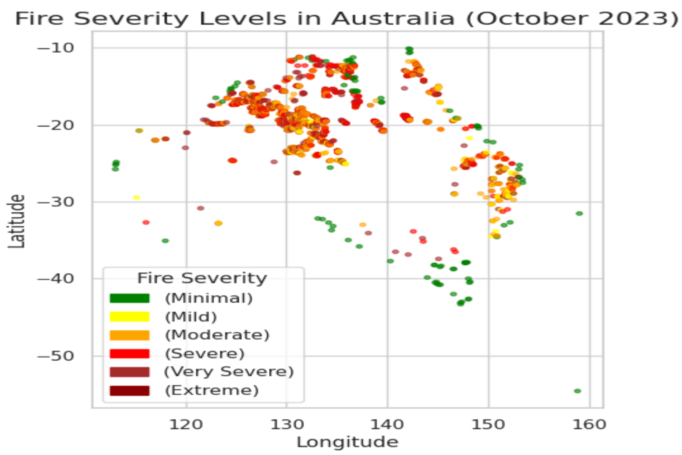

Figure 7: Active Fire Severity Levels Australia

ii . Temporal Analysis

- **Burned Area Trends:** Trends in burned area extents are plotted over the study period to identify temporal patterns.
- **NDVI Time Series:** Monthly NDVI time series are extracted to assess vegetation dynamics and post-fire recovery (Figure 8 & Figure 9).

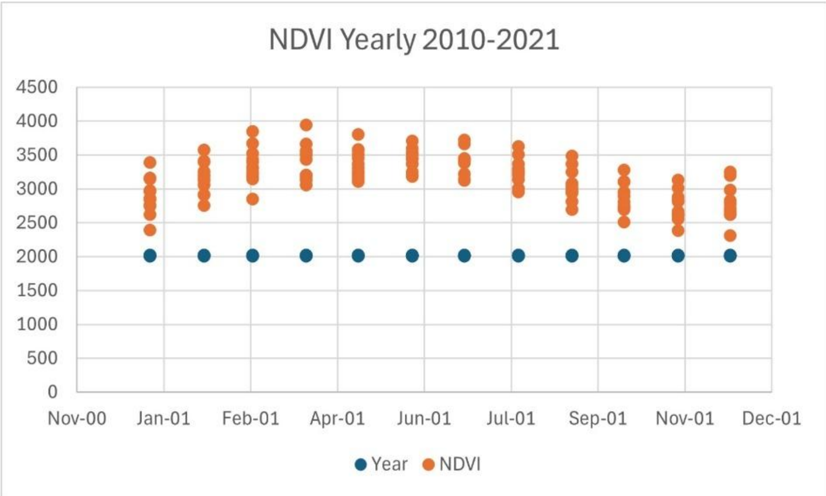

Figure 8: Monthly NDVI - Australia

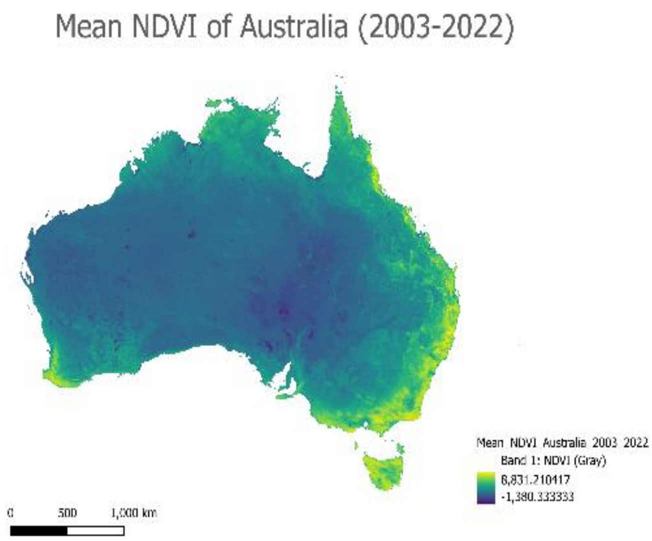

Figure 9: Mean NDVI Australia (2019-2023)

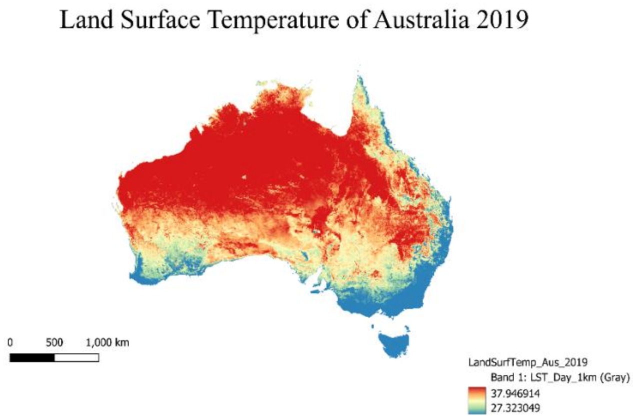

Figure 10: Land Surface Temperature Australia

Ⅳ. TESTING AND CONCLUSION:

This section evaluates the classifier's performance and discusses the conclusions drawn from the analysis.

A. Model Evaluation Techniques

ⅰ. Confusion Matrix

The confusion matrix assesses the classification model's performance by comparing predicted labels to actual labels [13]. Key metrics include:

- **True Positive (TP):** Correctly predicted burned areas.
- **False Positive (FP):** Incorrectly predicted burned areas.
- **True Negative (TN):** Correctly predicted unburned areas.
- **False Negative (FN):** Incorrectly predicted unburned areas.

ⅱ. Accuracy Metrics

- **Overall Accuracy:** Measures the proportion of total correct predictions (both burned and unburned) out of all predictions made by the model.

$$\boldsymbol{OA = \frac{TP + TN}{TP + TN + FP + FN}}$$

- **Kappa Coefficient:** Evaluates agreement between predicted and actual values, accounting for the possibility of agreement occurring by chance. It is particularly useful in assessing the performance of classifiers on imbalanced datasets.
- **Producer's Accuracy (Recall):** Measures the ability to correctly identify actual burned areas.

$$PA = \frac{TP}{TP + FN}$$

- **Consumer's Accuracy (Precision):** Measures the likelihood that predicted burned areas are actually burned.

$$CA = \frac{TP}{TP + FP}$$

B. Evaluation Results

Metrics	Percentage
Overall Accuracy	85%
Kappa Coefficient	0.70
Producer's Accuracy	84%
Consumer's Accuracy	86%

Table 2: Evaluation Results

Take Aways:

- The classifier achieved an overall accuracy of 85%, indicating good predictive performance.
- The kappa coefficient of 0.7 suggests strong agreement between predictions and actual values, though refinement is possible.
- The model performed well in detecting burned areas but showed some false positives, requiring potential refinement of input features.

C. Challenges Faced

1. **Data Acquisition Issues:** Selecting an appropriate period for analysis to obtain accurate results was challenging, as data availability and quality varied over time.
2. **Diverse Datasets:** Integrating various datasets with different resolutions, formats, and temporal coverages required significant preprocessing to ensure compatibility and relevance to the project objectives.
3. **Computational Limitations:** Processing large datasets demanded substantial computational resources, leading to increased processing times and occasional system crashes. Efficient data handling and optimization techniques were necessary to manage these limitations.
4. **Classifier Optimization:** Determining optional parameters for Gradient Boosting Classifier was time-consuming and required multiple iterations to balance bias and variance effectively.
   Addressing these challenges involved implementing robust data preprocessing pipelines, optimizing computational workflows, and conducting extensive hyperparameter tuning to enhance model performance.

D. Conclusions and Recommendations

Conclusions:

- **Effectiveness of MCD64 and VIIRS in Burned Area Detection:** The MCD64 product, derived from MODIS data, and the VIIRS sensor have been instrumental in detecting and mapping burned areas. The MCD64 product utilizes specific MODIS bands to identify burned regions, providing valuable insights into fire-affected areas. Similarly, VIIRS, with its 375-meter spatial resolution, enhances the detection of active fires and burned areas, offering improved spatial detail.

- **Efficacy of Gradient Boosting classifier:** The Gradient Boosting classifier has demonstrated effectiveness in classifying fire-affected regions. Studies have shown that ensemble learning methods, such as Gradient Boosting, can produce accurate fire susceptibility maps, outperforming traditional methods like logistic regression and random forest in terms of classification accuracy and area under the curve (AUC).

- **Scalability of Burn Mask:** The generated burn mask, developed using the methodologies, is scalable and can be applied to larger regions beyond the initial area of interest. This scalability ensures that the approach can be adapted for broader applications in different geographical contexts.

- **Workflow Validation in Diverse Regions:** The project workflow has been successfully tested in various countries across Africa and South America, indicating its robustness and adaptability to different environmental conditions and datasets.

Recommendations:

1. **Incorporate Additional Datasets:** To enhance prediction capabilities, it is recommended to integrate additional datasets such as soil moisture, wind patterns, and socio-economic factors. Incorporating these variables can provide a more comprehensive understanding of fire dynamics and improve model accuracy.

2. **Develop Multi-Temporal Analysis:** Implementing multi-temporal analysis at a higher scale can offer insights into temporal patterns of burned areas. This approach allows for the assessment of fire progression and recovery over time, contributing to more effective fire management strategies.

3. **Explore Deep Learning Methods:** Exploring deep learning-based methods, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), may further improve classifier performance. These advanced techniques have the potential to capture complex patterns in the data, leading to more accurate predictions.

4. **Refine Burn Mask:** Enhancing the burn mask through the integration of higher-resolution data and advanced image processing techniques can improve its accuracy. Refinements may include the use of object-based image analysis and the incorporation of spectral indices sensitive to burned areas.

REFERENCES:

1. Proloy Deb, “Causes of the Widespread 2019–2020 Australian Bushfire Season.” (https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020EF001671)

2. Taylor, “New Spatial Analyses of Australian Wildfires.” (https://www.pnas.org/doi/10.1073/pnas.2002269117)

3. B. Jones, “Averting the Flames: How to Prevent Bushfires in Australia.” (https://www.dryad.net/post/bushfire-prevention-in-australia)

4. G. S. Exchange, “Clipping imagery to regions of interest with Google Earth Engine before exporting,” 23 Sept 2020. (https://gis.stackexchange.com/questions/364901/clipping-imagery-to-regions-of-interest-with-google-earth-engine-before-exportin)

5. Ekhi Roteta, “A Preliminary Global Automatic Burned-Area Algorithm in GEE.” (https://www.mdpi.com/2072-4292/13/21/4298)

6. Josep G Canadell, “Multi-Decadal Increase of Forest Burned Area in Australia.” (https://www.nature.com/articles/s41467-021-27225-4)

7. USGS Earth Data, “MODIS Burned Area Data Product.” (https://lpdaac.usgs.gov/products/mcd64a1v006/)

8. NASA, “VIIRS Burned Area Product.” (https://viirsland.gsfc.nasa.gov/Products/NASA/BurnedAreaESDR.html)

9. Quan Duan, Burned area detection from a single satellite image using an adaptive thresholds algorithm,” (https://www.tandfonline.com/doi/full/10.1080/17538947.2024.2376275)

10. Joana Parente, “Quantitative Assessment of LULC Changes and Wildfires.” (https://www.mdpi.com/2571-6255/6/5/198)

11. Yang Zhang, “Identifying Key Drivers of Wildfires Using Machine Learning.” (https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020EF001910)

12. Ana Militino, “Detection of burned areas with extreme gradient,” *Research Square,* 2022. (v1_covered.pdf)

13. Understanding the Confusion Matrix in Machine Learning. (https://www.geeksforgeeks.org/confusion-matrix-machine-learning/)