SUP'COM
Ecole Supérieure des Communications de Tunis
Higher School of Communication of Tunis

P2M Project Report

Burned area detection based on satellite data

Realized By

Tarek Bouchoucha

Alaa Ben Turkia

Supervised by

Prof. Riadh Abdelfatteh

Academic year 2023-2024

Acknowledgment

We would like to express our heartfelt gratitude to our supervisor, **Mr. Riadh Abdelfatteh**, for his invaluable assistance and unwavering support throughout the duration of this project. His insightful guidance and constructive feedback have been instrumental, particularly during the initial phase when laying the groundwork was crucial. Mr. Abdelfatteh's continuous encouragement and expertise have greatly contributed to the development and refinement of our work.

General introduction

Wildfires have become an increasingly severe global issue, impacting natural ecosystems, human settlements, and overall environmental stability. The unpredictable nature and rapid spread of these fires present significant challenges for effective environmental management and disaster response. Traditional ground-based methods for detecting and mapping burned areas, while useful, are often inadequate. They are labor-intensive, time-consuming, and restricted to accessible regions, failing to provide the timely data required for effective response and management.

In contrast, satellite remote sensing has revolutionized the ability to monitor and assess burned areas. Satellites offer extensive spatial coverage and frequent temporal updates, making them invaluable for capturing the dynamic nature of wildfires. By leveraging advanced sensors and sophisticated image processing techniques, satellite data can be utilized to detect and map burned regions with high precision.

This report focuses on the use of satellite data, particularly from Sentinel-2, to develop an efficient, accurate, and scalable solution for burned area detection. By applying image processing and computer vision models to satellite imagery, this approach enhances the speed and accuracy of detecting burned areas, providing critical insights into the severity and extent of wildfires. The following sections will delve into the problem at hand, the proposed solution, and the results of our analysis, contributing to the advancement of remote sensing applications in wildfire management and offering practical tools for better managing fire-affected landscapes.

SUP'COM
Ecole Supérieure des Communications de Tunis
Higher School of Communication of Tunis

Table of Contents

List of Figures

I- Context of the project

1. Problem:

As mentioned in the general introduction wildfires pose significant challenges to environmental management due to their unpredictable nature and rapid spread. Traditional ground-based methods for detecting and mapping burned areas are often inadequate, as they are labor-intensive, time-consuming, and limited to accessible regions. Moreover, these methods do not provide the timely data required for effective response and management. The primary problem addressed in this project is the need for an efficient, accurate, and scalable solution to detect burned areas quickly and over large geographic extents.

2. Solution:

The solution resides in using satellite images from sentinel-2 to detect and map burned areas. Satellite remote sensing offers several advantages, including wide-area coverage, frequent data acquisition, and the ability to capture diverse spectral information. By applying image processing techniques to the data and using computer vision models, we can identify burned areas accurately. This approach not only enhances the speed and efficiency of burned area detection but also provides valuable insights into the severity and extent of wildfires.

3. Required work:

To achieve the objectives of this project, performing the following tasks was required:
- Conducting a comprehensive search and review of existing technologies and methods used in similar projects to get a comprehensive idea and perspective of the project and its requirements.
- Gathering and obtaining the required data and performing the necessary preprocessing to make it suitable for our project.
- Preparing the working environment with the necessary tools and libraries and then performing the training and testing of the models with the finalized version data.
- Interpreting the results and making some kind of simple implementation.

II- Tools and libraries

1. QGIS:

QGIS (Quantum Geographic Information System) is an open-source geographic information system software used for creating, editing, visualizing, analyzing, and publishing geospatial information. It is a powerful tool that supports a wide range of vector, raster, and database formats and functionalities.

In our project, we used this software to make corrections to the sentinel-2 images to be able to use them in our model training.

Fig.1 QGIS logo

2. Visual Studio Code:

Visual Studio Code (VS Code) is a free, open-source code editor developed by Microsoft, known for its versatility, powerful features, and extensive customization options. It is lightweight, fast, and supports multiple programming languages. Key features include an extension marketplace, built-in Git integration, IntelliSense for intelligent code completion, robust debugging tools, and an integrated terminal. It can be customized with settings, key bindings, and themes, and the Live Share extension enables real-time collaboration.

Fig.2 VS Code logo

3. Google Colab:

Google Colab (short for Collaboratory) is a free, cloud-based platform provided by Google that allows users to write and execute Python code through a web browser. It is particularly well-suited for machine learning, data analysis, and educational purposes.

Fig.3 Google Colab logo

III- Development of the project

1. Data acquisition:

Our data acquisition process began with selecting Sentinel-2 imagery due to its high resolution and frequent revisit times, making it ideal for monitoring dynamic events like wildfires. We obtained ground truth data used in the Portuguese wildfire project from the official website of the Nature and Forests Conservation Institute of Portugal. Additionally, we sourced additional data from the Copernicus Open Access Hub, focusing on those same regions affected by wildfires. This combination of datasets ensured we had a robust collection of both imagery and reference data.

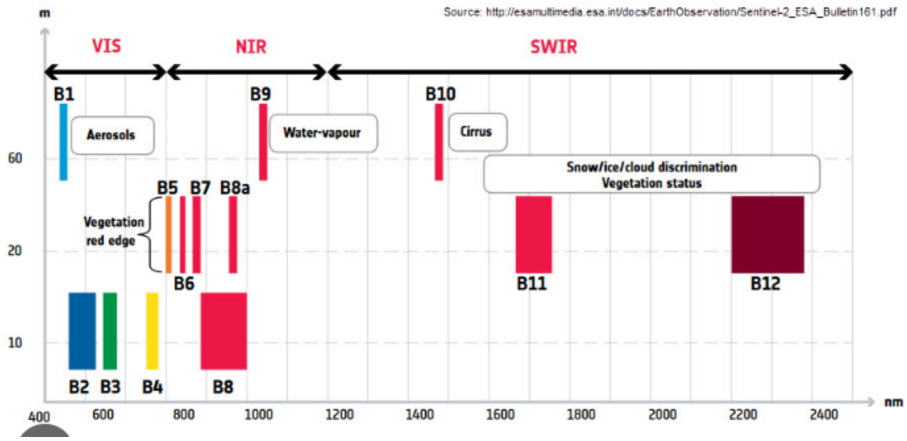

Fig.4 Spectral bands of Sentinel-2 images

2. Data preprocessing:

To ensure the quality and usability of our data, we undertook extensive preprocessing steps. This phase was crucial for correcting and preparing the data for model training.

a. Data correction:

The raw Sentinel-2 images required correction to address atmospheric effects, geometric distortions, and other anomalies. We applied atmospheric and geometric correction techniques using **sen2cor** to align the images spatially, which is essential for accurate analysis and comparison over time.
We chose images with spatial resolution 20 meters per pixel.

b. Data Preparation:

- Extract band: We extracted the Green, NIR, and SWIR bands and then with the help of **QGIS** we transformed them from **jp2** to **.tiff** format which is a format that allows to store high-quality images with their meta-data (localization).
- Merge bands**:** We merged the Green, NIR, and SWIR bands of the atmospherically corrected Sentinel-2 data and created false-color images that emphasize the burned areas by transferring the spectral bands to RGB channels and normalizing the values to the 0-255 color scale.
- Patch extraction: we matched the burned area ground truth data to the false-color images by area and date and extracted small patches of size 128x128 from the false-color images and ground truth data.
  We only took patches that contain burned areas and with are superior to 5ha, we also took an interval of one month for the ground truth of each patch since burned areas take time to recover.
- Patch sorting: we removed empty or broken patches from the datasets.

- Data augmentation: we performed the following data augmentation techniques: Flip (Horizontal and Vertical), 90° Rotate (Clockwise, Counter-Clockwise, Upside Down), Noise (Up to 0.34% of pixels) which allowed us to augment the number of patches from 1862 patches to 3012 patches.
- Ground truth type converting we transformed the ground truth from the json format to the coco format which will help us later for training.
- Data Split: 70% for training, 20% for validation and 10% for testing.

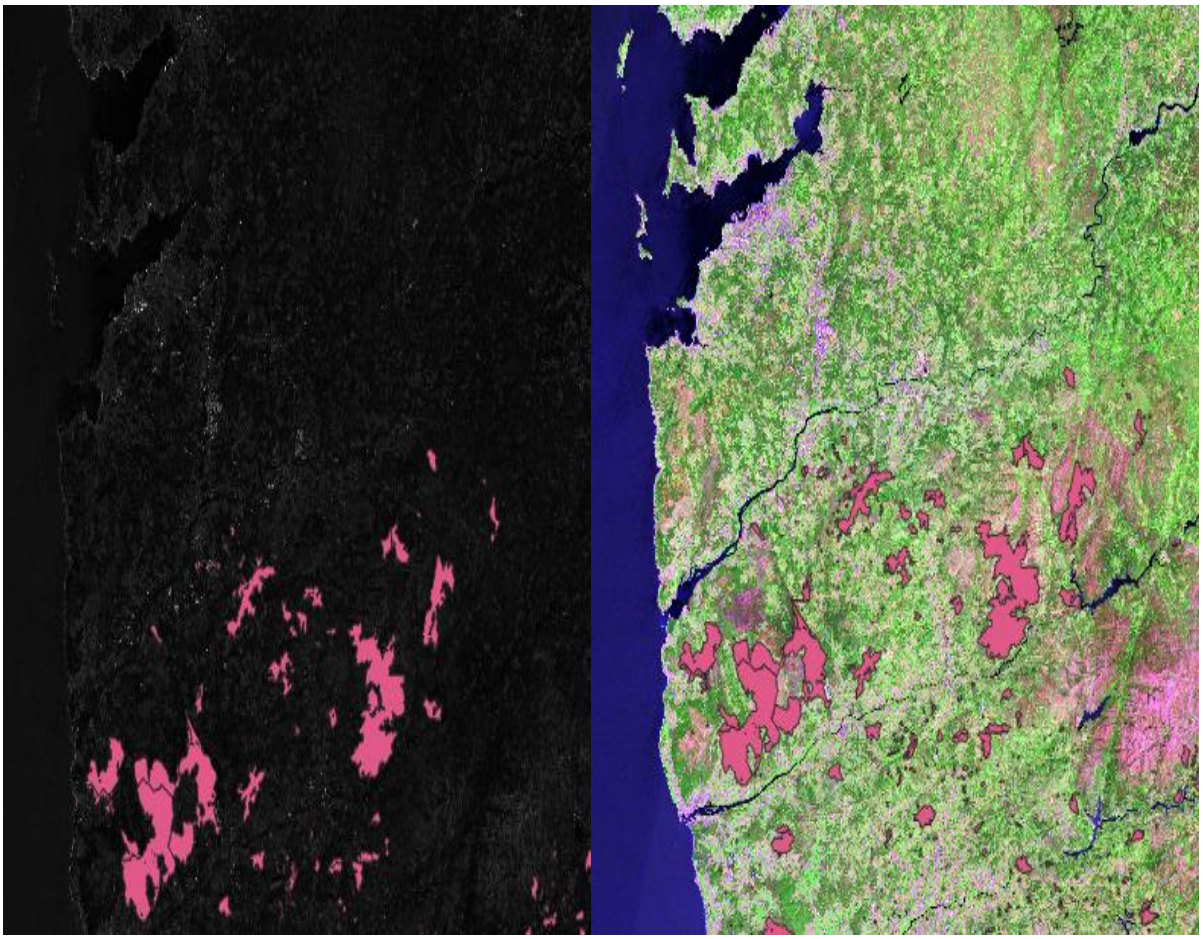

Fig.5 Band merging visualization

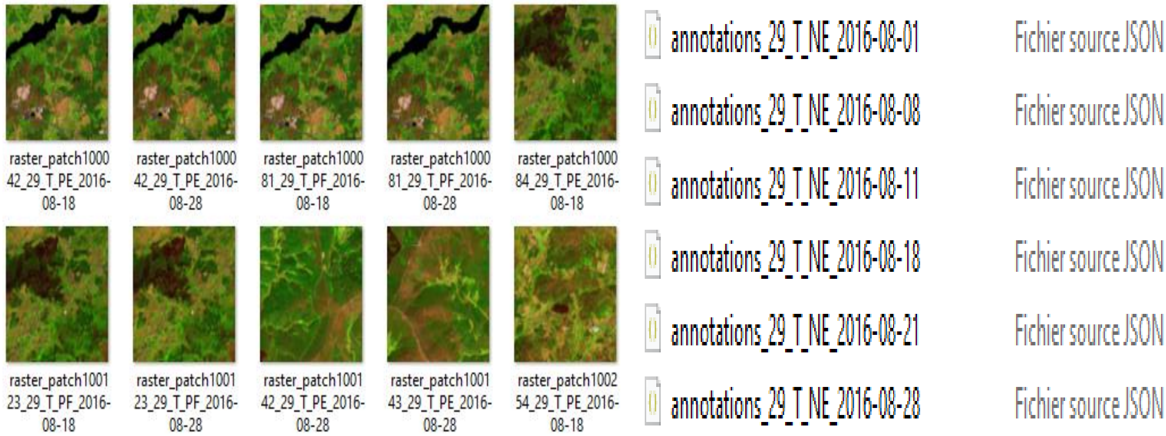

Fig.6 Patches and annotations after preprocessing

3. Training:

For the training phase, we selected the YOLOv8 (You Only Look Once, version 8) model due to its balance between speed and accuracy in object detection tasks. We first structured the data into 3 repositories (train, valid and test) with yaml which will help us train our model, we then trained the model using the prepared patches and the annotated ground truth data. The training process involved several iterations of adjusting hyperparameters and optimizing the model to improve its performance.
The training dataset comprised 70% of the image set and we took 50 epochs for training.

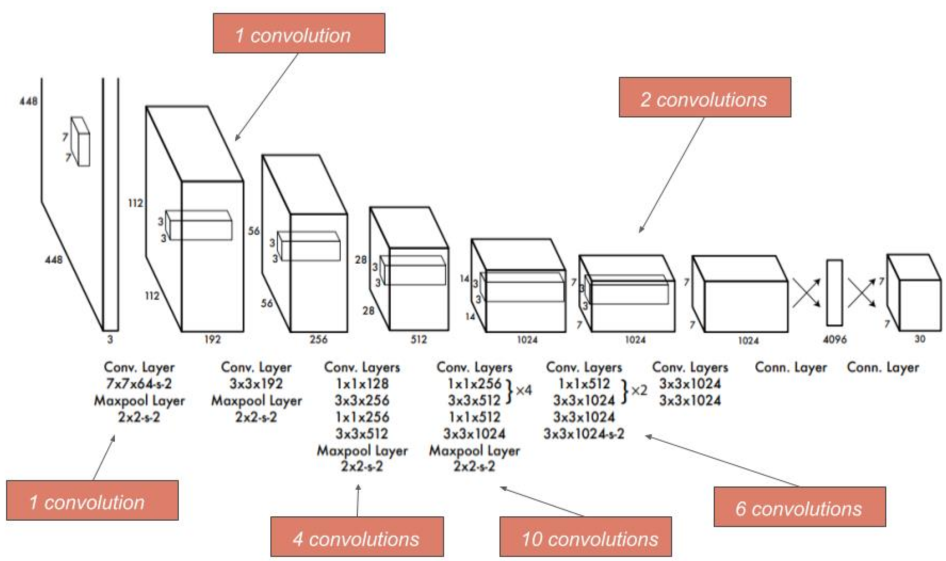

Fig.7 YOLO v.8 model architecture

- <u>Training results:</u>

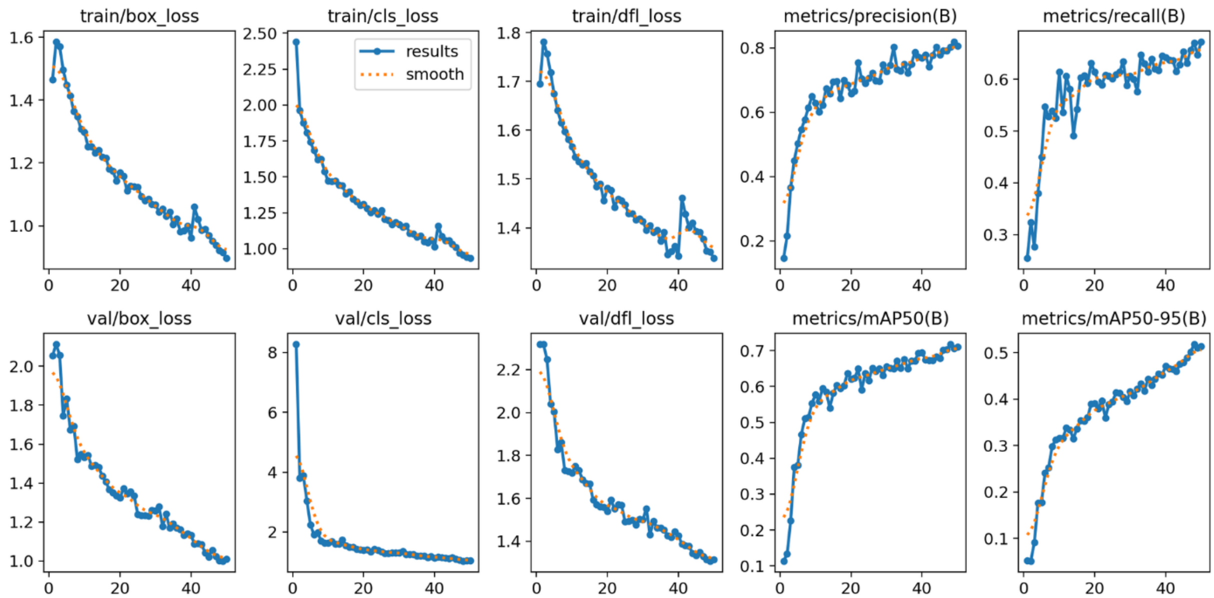

Fig.8 Training metrics

<u>Decreasing Losses</u>: The consistent decrease in training and validation losses (box loss, class loss, and distribution focal loss) indicates that the model is learning effectively and not overfitting.
<u>Increasing Precision and Recall</u>: The increasing trends in precision and recall metrics signify that the model is becoming better at identifying true positives while minimizing false negatives and false positives.
<u>Improving mAP Scores</u>: The improvement in both mAP50 and mAP50-95 metrics suggests that the model's performance in terms of object detection accuracy is getting better over a range of IoU thresholds.

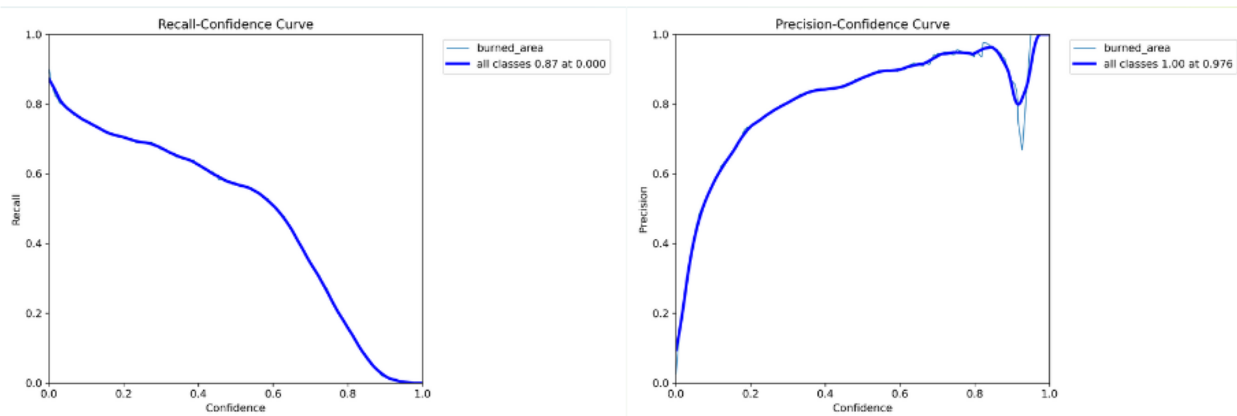

Fig.9 Recall-Precision-Confidence relationship graphs

On the recall-confidence curve, we notice a steep initial drop in recall with increasing confidence suggesting that the model is initially very sensitive, but this sensitivity drops

quickly as confidence requirements rise.
On the precision-confidence curve, we notice a noticeable dip in precision around the 0.8 confidence threshold. This could indicate that the model has specific confidence ranges where its predictions are less reliable, possibly due to overfitting or other issues in the model's training.

4. Testing and validation:

Testing the model was a critical step to evaluate its accuracy and effectiveness in detecting burned areas. We used a separate set of images that were not included in the training set to test the model. This validation process involved measuring the model's performance using metrics such as precision, recall, and model confidence. The results of these tests provided insights into how well the model could generalize to new, unseen data.

- Results:

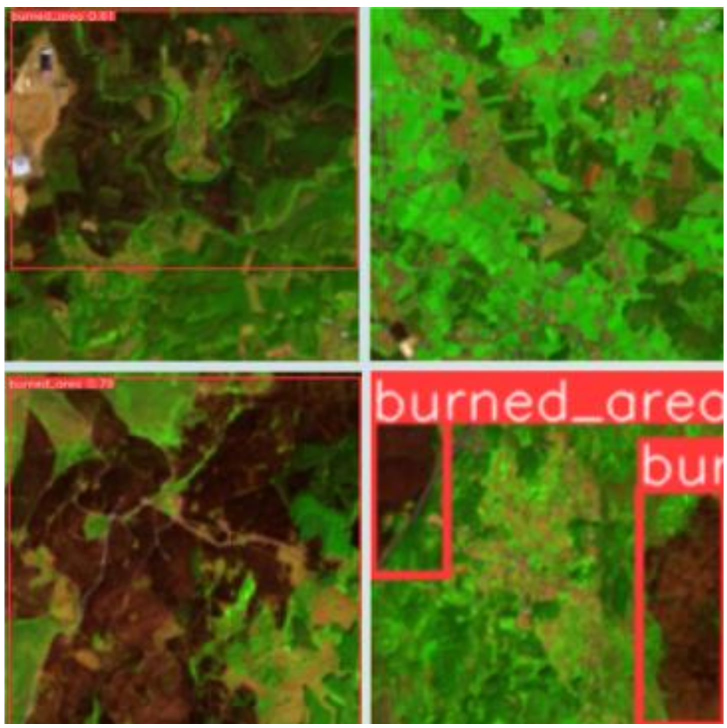

Fig.10 Example of testing results

5. Deployment:

To demonstrate the practical application of our trained model, we developed a simple interface implementation using Streamlit, a Python library for creating interactive web applications. This implementation allows users to upload patches and receive real-time detections of burned areas.

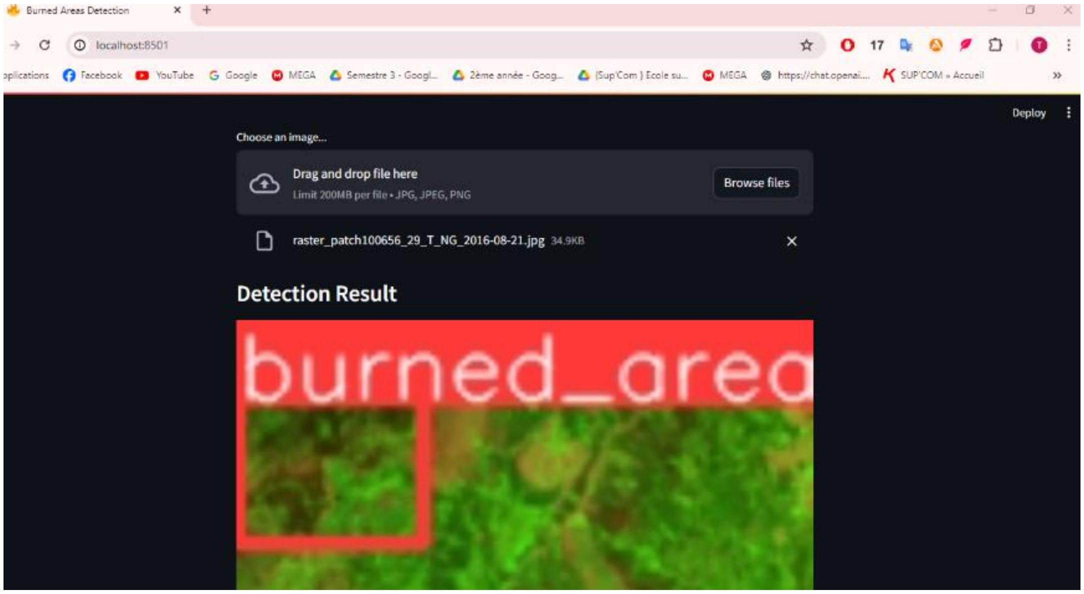

Fig.11 Deployment interface

6. Testing on Tunisian data:

In the final phase of our project, we tested our model on recent images of burned areas in Tunisia more specifically the region of Malloula, Jendouba governate, that witnessed a great deal of wildfires in the summer of 2023. This testing was essential to assess the model's applicability in different geographic and climatic conditions. We sourced the latest Sentinel-2 images covering this region. By applying our trained model to this new data, we evaluated its performance and adaptability.
The steps included extracting images from sentinel-2 correcting them and then annotating them manually using QGIS and proceeding with the same steps of extracting patches and sorting them.

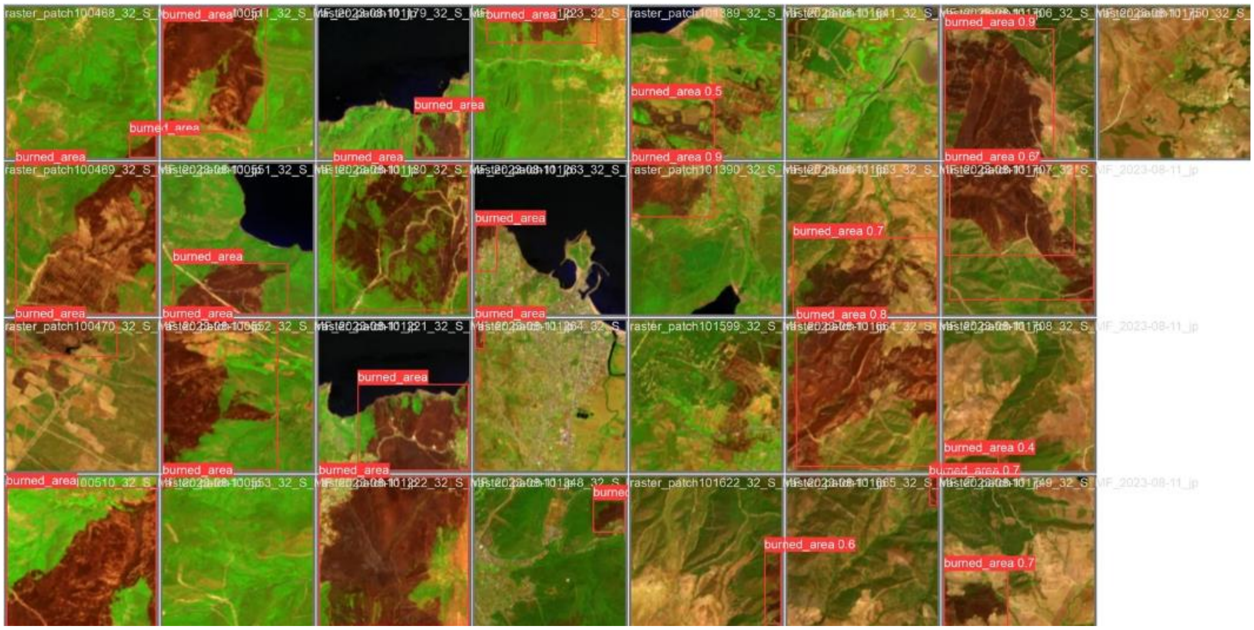

Fig.12 Results of testing on Tunisian data

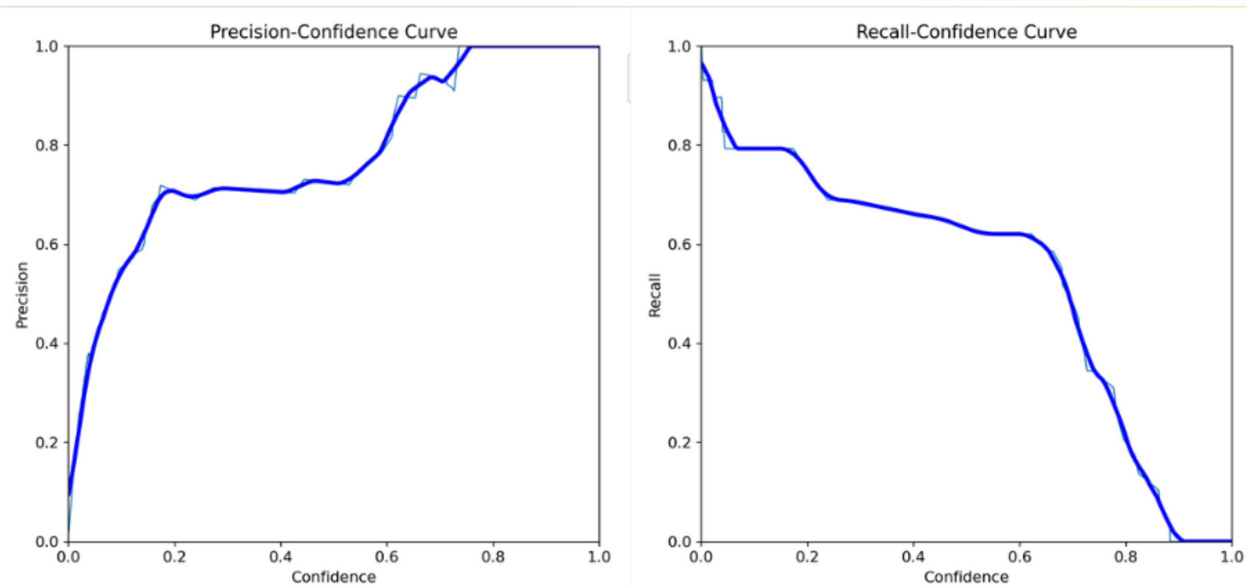

Fig.13 Precision-Confidence-Recall graphs

The results indicated the model's potential for broader applications such as predicting the evolution of forest fires in the next summer and assessing the damage caused by these phenomena both on an ecological level and human level and helping prevent or at least decrease them.

IV- Conclusion

In summary, this report underscores the critical need for advanced methods in wildfire detection and management, given the limitations of traditional ground-based approaches. By leveraging the capabilities of satellite remote sensing, particularly through the use of Sentinel-2 imagery, we have demonstrated a significant improvement in the accuracy, efficiency, and scalability of burned area detection.

The application of image processing techniques and computer vision models to satellite data enables timely and precise identification of burned regions, addressing the challenges posed by the unpredictable and rapid spread of wildfires. The comprehensive review of existing technologies, meticulous data preprocessing, and the strategic use of the computer vision model Yolo v8 culminated in a robust solution.

V- Bibliography

- Sentinel-2 [link 1]
- Copernicus data browser [link 2]
- Instituto da concervação da natureza e das florestas (ICNF) [link 3]
- YOLO v8 documentation [link 4]