Forestal monitoring & Carbon sequestration analyzer for reforestation projects

Fernando Rojas, Carlos Gernhofer, Juan Azcona, Nicolas Vega.

Universidad Politécnica Taiwan Paraguay

Email: fernando.rojasm@uptp.edu.py, carlosa.gernhofer@uptp.edu.py, juan.azcona@uptp.edu.py, francisco.vega@uptp.edu.py

---

Abstract:. The ongoing deforestation crisis in Paraguay is primarily driven by agricultural expansion and wildfires, leading to a rise in carbon emission levels, causing loss of biodiversity and an increase of the overall temperature in the region. Vital ecosystems are facing alarming rates of deforestation, the loss of which damages wildlife and reduces the region's capacity to capture and store carbon dioxide, further aggravating the global climate crisis.

Previous research made in deforestation monitoring has employed satellite images, geographic information and more to track changes in the forest over time. However, the use of AI and machine learning is a new set of opportunities for more accurate and dynamic tracking of the changes in the environment. This project aims to address the challenges of monitoring deforestation and carbon sequestration potential by developing an artificial intelligence model that is capable of following these issues in order to obtain important data about the damaged zones, enabling the identification of areas where reforestation efforts would yield the maximum carbon capture benefit.

The methodology proposed for this project integrates multiple data sources including historical high-resolution satellite imagery and climate data which will provide a valuable dataset for the training of the model. Primarily employing a supervised learning approach, utilizing Convolutional Neural Networks (CNNs) for image analysis and sequential data and time series forecasting. This enables the identification of areas with the highest carbon capture potential, additionally, the model can detect current deforestation patterns and predict future trends and wildfire risks.

This project intends to provide an accurate deforestation detection system, a predictive model for wildfire risk and a map of carbon sequestration potential across the region. Aiming to contribute a valuable tool to make data-driven decisions for where to implement reforestation programs. Moreover, the project aims to raise awareness of deforestation and wildfire issues in Paraguay. On top of that, it is an effective way to contribute to the development of AI technologies specifically designed to address environmental challenges in Paraguay. The integration of this model is expected to significantly improve forest conservation efforts that will lead to a substantial reduction in CO2 emissions in Paraguay.

Index terms: Artificial Intelligence (AI), Convolutional Neural Network (CNN), carbon emission, machine learning, deforestation, carbon dioxide (CO2).

I. INTRODUCTION

The Chaco is a fragile environment with a wide variety of ecosystems (Madroño). Deforestation is one of the most critical environmental challenges which Paraguay faces, particularly in the Chaco region, which has experienced alarming rates of forest loss in recent decades. This phenomenon, driven by agricultural expansion and frequent wildfires, contributes to significant carbon emissions, biodiversity loss, and regional temperature increase. The deforestation crisis not only threatens vital ecosystems but also reduces the region's capacity to sequester carbon dioxide, thereby exacerbating the global climate analysis.

The forest characterization based on satellite Landsat data and the subsequent change-detection analysis revealed a forest cover loss of 64,700 km^2 between 1987 and 2020, resulting in an annual deforestation rate of 1960 km^2 (Emmanuel & Jennifer, 2022). This highlights the urgent need for innovative tools to monitor deforestation and mitigate its impacts. Traditional satellite-based monitoring methods provide valuable insights but may lack real-time adaptability and predictive capabilities necessary to address the dynamic and multifaceted nature of deforestation.

This project seeks to address this gap by employing artificial intelligence (AI) and machine learning (ML) to enhance deforestation detection and support reforestation efforts in Chaco. By integrating high-resolution satellite imagery with advanced ML techniques such as Convolutional Neural Networks (CNNs) for image analysis, the proposed approach establishes a robust framework for monitoring vegetation changes and predicting deforestation trends.

Additionally, the project focuses on identifying optimal locations for reforestation initiatives based on their carbon sequestration potential, providing policymakers with the tools to prioritize interventions that yield maximum environmental benefits. The use of AI in this context not only enhances the accuracy of monitoring systems but also delivers actionable insights, presenting a scalable and data-driven solution to combat deforestation in Paraguay and similar regions.

By leveraging these methodologies, this work aims to contribute to the preservation of Chaco's ecosystems, mitigate carbon emissions, and offer a practical tool for informed decision-making in conservation and land management. This initiative highlights the potential of AI in addressing critical environmental challenges and aligns with the broader objective of achieving a sustainable balance between development and ecosystem preservation in fragile regions like the Paraguayan Chaco.

II. CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional Neural Networks (CNNs) are essential for predicting near-future deforestation patterns using limited time-series data, leveraging the growing volume of satellite remote sensing imagery to address global deforestation. (Ball, Petrova, Coomes, & Flaxman, 2022). CNNs are pivotal in this project, serving as the backbone for processing and analyzing satellite imagery to detect deforestation patterns and support reforestation recommendations. By automatically and adaptively learning spatial hierarchies of features, CNNs are particularly suitable for high-resolution imagery. These networks extract key features such as vegetation density, land cover patterns, and temporal changes in deforestation trends.

The architecture employed in this project incorporates input images enriched with NDVI index, allowing the model to analyze not only RGB data but also vegetation health indicators. Convolutional layers process spatial patterns, pooling layers reduce feature dimensionality, and fully connected layers classify the land into specific categories such as deforested, forested, or degraded. By leveraging these capabilities, the CNN framework ensures robust deforestation detection and enhances its utility for prioritizing reforestation efforts.

III. USE OF NDVI IN VEGETATION ANALYSIS

The Normalized Difference Vegetation Index (NDVI) is a critical tool in the analysis of vegetation health and density, providing a reliable metric derived from satellite imagery to monitor changes in forested areas. NDVI is calculated using the reflectance values of near-infrared (NIR) and red light, captured by satellite imagery, but since the images are only on RGB the way this problem was tackled was to use the green channel as a placeholder for the NIR channel to later calculate the NDVI values and assign them to their own channel. Healthy vegetation strongly reflects NIR light and absorbs red light during photosynthesis, enabling the computation of NDVI using the formula:

$$NDVI = \frac{(B8-B4)}{(B8+B4)}$$

where B8 represents the near-infrared band, and B4 represents the red band. This index yields values between -1 and 1, where higher values indicate healthier and denser vegetation, while negative or near-zero values typically represent barren land or water.

The integration of NDVI into vegetation analysis was aimed to be given by a two-stage approach. The first step, implemented through Google Earth Engine (GEE), involves calculating NDVI for specific regions of interest. The resulting NDVI maps, saved as GeoTIFF files, highlight spatial patterns of vegetation density and are preprocessed to mask clouds and irrelevant features, ensuring high-quality data for further analysis. These maps are instrumental in classifying land cover changes, monitoring deforestation trends, and identifying areas requiring reforestation. But since the lack of sufficient amount of data for the data set to be enough was a setback for training the model, the investigators opted for using a pre-existent data set of the vegetal coverage of certain regions during different time stamps.

The second stage involves the exploration and preparation of the dataset, including labeled training and testing datasets stored in CSV format. By integrating the NDVI-based metrics, the project enables the development of machine learning models capable of analyzing vegetation patterns and predicting deforestation trends.

The NDVI maps and visualizations play a crucial role in this project. The color scale in these maps represents vegetation density and health, ranging from green (indicating higher forest density and health) to orange (indicating lower vegetation density or stressed vegetation). These maps enable a clear understanding of environmental changes between two time points. For instance, significant drops in NDVI values between paired images indicate deforestation, while areas with stable or increasing NDVI values suggest preservation or recovery of vegetation.

The visualization provided by the NDVI maps allows side-by-side comparisons that clearly demonstrate the changes in vegetation cover and density. This ability to visually inspect temporal differences provides valuable insights into the environmental changes in the study region. NDVI's capability to quantify vegetation health further supports data-driven decision-making for reforestation efforts and conservation strategies. The visualization also serves as a demonstration of the change detection model's performance, effectively communicating the observed environmental transformations.

The combined data pipeline leverages NDVI maps and labeled datasets to extract spatial features such as edges, textures, and vegetation densities, which are fed into Convolutional Neural Networks (CNNs) for deforestation detection and prediction. These features allow the model to identify deforestation patterns, predict future land-use changes, and suggest areas suitable for reforestation based on their carbon sequestration potential. The model's performance is validated on unseen datasets, ensuring its accuracy and scalability for future applications.

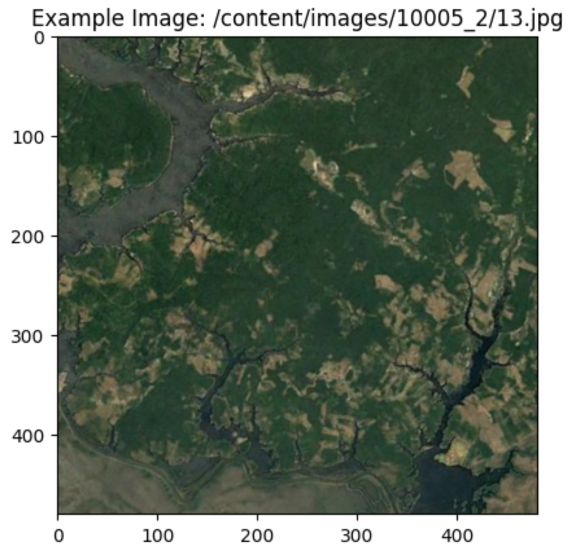

Figure 1: Initial step for validation for NDVI

The datasets

The datasets involved in this analysis consist of dual images, each representing the same location at different time points. This temporal pairing allows for detailed examination of vegetation dynamics, enabling the detection of changes such as deforestation, vegetation recovery, or land-use transformation. Each dataset entry includes metadata specifying file paths, labels, and the corresponding NDVI values for each time point.

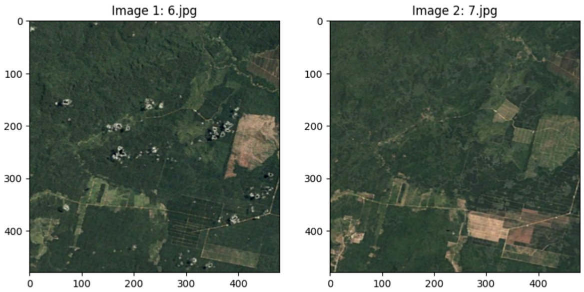

Figure 2: Same location different points in time

By comparing these image pairs, significant insights can be gained regarding temporal changes in vegetation health and density. Each image pair consists of a baseline image taken at an earlier time and a subsequent image taken more recently. This temporal pairing ensures that any detected differences in vegetation or land cover are due to actual environmental changes rather than discrepancies in location or data collection methods.

For this project, these dual images serve as the foundation for detecting deforestation and analyzing its progression. NDVI (Normalized Difference Vegetation Index) is calculated for both images in the pair, providing a quantitative measure of vegetation health for each time frame. The NDVI maps derived from the image pairs are then compared to highlight areas where vegetation has decreased (indicative of deforestation) or increased (potential signs of reforestation or natural recovery).

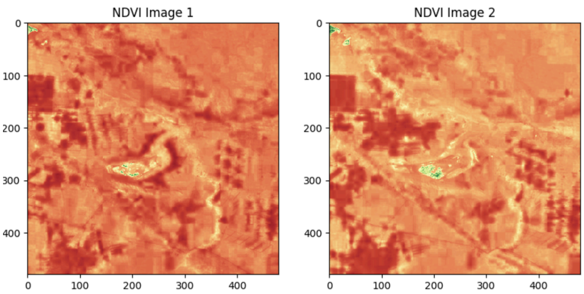

Figure 3: NDVI Comparison Between Two Time Periods

The advantage of using dual images lies in their ability to reveal dynamic changes over time. While a single image can only provide a snapshot of vegetation conditions, dual images enable a temporal analysis, uncovering patterns of change that would otherwise remain hidden. For example, areas with a significant drop in NDVI values between two time points can be flagged as deforested zones, while areas with stable or increasing NDVI values may indicate preserved or recovering ecosystems.

Moreover, the dual image structure is essential for training and validating machine learning models. By pairing images from two points in time, the datasets provide labeled examples of environmental change, which can be used to train models to predict future deforestation patterns or assess the effectiveness of conservation initiatives. This approach combines the power of satellite imagery and temporal analysis to create a robust framework for understanding and addressing deforestation dynamics.

IV. IMAGE COLLECTION AND PROCESSING

The image collection and processing phase is critical to the success of this project, as it establishes the foundation for deforestation detection and analysis. The dataset utilized comprises pairs of images (dual images) representing the same geographical locations but captured at different time points. These dual images allow for a detailed examination of temporal changes in vegetation, enabling the detection of patterns such as deforestation, vegetation recovery, and land-use transformation.

The process begins with data acquisition. Satellite imagery from Sentinel-2 serves as the primary data source, capturing high-resolution images of the Chaco region. These images undergo preprocessing to ensure their usability. The preprocessing includes resizing, normalizing pixel values, and filtering out noise, such as cloud cover, which is managed through a custom cloud masking algorithm. This step ensures that only clear and meaningful data are passed to the analysis pipeline.

Following data acquisition, the dataset is organized into structured files in CSV format. These files, such as train.csv and test.csv, include metadata for each image pair, such as file paths and labels. Using Python's pandas library, these files are loaded and inspected to verify their integrity and structure. For example, each entry in the dataset points to a pair of images: one representing an earlier time point (baseline) and the other representing a later time point (comparison). Labels associated with these pairs describe whether deforestation or other environmental changes occurred.

The preprocessing stage involves calculating the Normalized Difference Vegetation Index (NDVI) for each image in the dataset. For this project, NDVI is computed for both images in a pair, providing a temporal perspective of vegetation health. To enhance the model's understanding of vegetation patterns, NDVI values are appended as additional channels to the original image data, effectively expanding the data's feature set.

Once the NDVI values are computed, the images and NDVI channels are combined into a unified format suitable for machine learning. Each image pair is represented as an array with eight channels: three RGB channels for each image and two additional channels for the NDVI values. This comprehensive representation ensures that the model has access to both visual and vegetation-specific data, maximizing its ability to detect deforestation.

The next step involves preparing the data for model training. The dataset is split into training, validation, and testing subsets to ensure robust evaluation of the model. Labels are extracted and encoded into numerical values, enabling compatibility with machine learning algorithms. Additionally, class weights are computed to address any imbalances in the dataset, ensuring that underrepresented classes (e.g., areas with minimal deforestation) are not overlooked during training.

The processed data is fed into a Convolutional Neural Network (CNN) designed to handle the eight-channel input. This architecture includes convolutional layers for feature extraction, pooling layers for dimensionality reduction, and dense layers for classification. The model leverages the dual image structure and NDVI channels to identify deforestation patterns with high precision. Training is conducted using a weighted loss function, where the computed class weights ensure balanced learning across all classes.

The processed NDVI data, combined with the CNN's capability to learn hierarchical spatial features, results in a robust framework for vegetation analysis. For instance, comparison of NDVI maps across time points enables the generation of "change maps," which visually highlight areas of significant vegetation loss or gain. These maps are invaluable for identifying regions affected by deforestation, quantifying the extent of vegetation loss, and prioritizing conservation efforts.

V. DEFINITION OF THE STUDY AREA

The Chaco region of Paraguay serves as the ideal focal area for this study (further showcases of the model performance are found in the video presentation of this project). This region, characterized by its fragile ecosystems and rich biodiversity, is facing critical environmental challenges due to alarming rates of deforestation. The study area was selected based on its significance as one of the largest deforestation frontiers in South America and its importance in global carbon sequestration efforts. By focusing on the Paraguayan Chaco, the project aims to address the urgent need for sustainable land management and conservation strategies in this vulnerable region.

Geographically, the study area spans a section of the Chaco defined by coordinates that encapsulate the boundaries of significant environmental change. Specifically, the selected area includes regions of diverse land cover types, ranging from dense forests to grasslands, providing a comprehensive view of vegetation dynamics. This spatial diversity is essential for building a robust model capable of detecting deforestation patterns across various types of terrain and vegetation cover.

Satellite imagery from Sentinel-2 was used to capture high-resolution data of the study area. Sentinel-2 provides multispectral images with a 10-meter resolution, making it ideal for detecting changes in vegetation cover and identifying areas affected by deforestation. To ensure the quality of the data, images were filtered based on cloud cover, with only those containing less than 10% cloud cover included in the analysis. This preprocessing step ensured that the data used for modeling and analysis was both clear and representative of the region's actual conditions. All of this imagery served for testing the model performance and ended up being excluded from the training set.

VI. SIMULATIONS AND EXPERIMENTAL RESULTS

The proposed AI model successfully predicts deforestation patterns using satellite imagery and NDVI data, achieving 98% validation accuracy in detecting vegetation changes and classifying deforestation likelihood. This percentage does not apply for urban areas or areas with bodies of water. The validation accuracy graph highlights the model's strong performance and convergence during training.

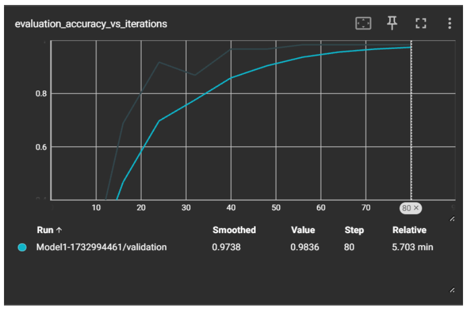

Figure 4: Validation accuracy graph

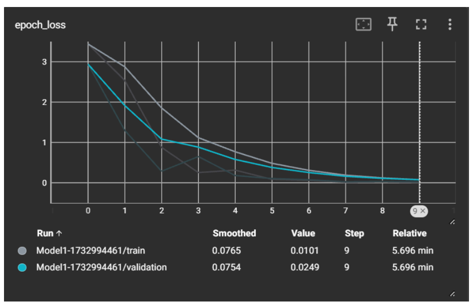

Figure 5: Validation loss graph

To confirm the model is not overfitted, the validation loss graph demonstrates that both training and validation loss decrease consistently, without significant divergence. This indicates that the model generalizes well to unseen data, ensuring reliability for predicting high-risk areas for future deforestation and wildfires. NDVI comparisons further validate its ability to identify critical zones for carbon sequestration and reforestation efforts.

VII. CONCLUSION

This study developed an AI-based model capable of predicting deforestation trends and identifying optimal reforestation zones. By leveraging satellite imagery and NDVI metrics, the model offers a scalable tool for proactive land management and carbon sequestration planning. Future work should focus on real-world testing in regions like the Paraguayan Chaco to validate and refine the model's practical application.

ACKNOWLEDGMENT

The authors would like to express their deepest gratitude to Professor Shan-Hsiang Shen for his invaluable guidance, expertise, and support throughout the course of this research. His insights and explanations on neural networks were essential in shaping the direction and quality of this study.

Special thanks are also extended to teacher assistant Nicolas Maldonado for their contributions and support during the course. Their assistance played a crucial role in the completion of this research.

Data available statement:

The dataset used for tree classification was obtained from the 'Trees in Satellite Imagery' dataset, which provides geospatial satellite images categorized into 'tree' and 'notree' classes (Aksoy, 2022)[1]. The libraries used during the project are cited up next.[5]

REFERENCES

[1] Aksoy, M. Ç. (2022). *Trees in Satellite Imagery [Dataset]*.Retrieved from: https://www.kaggle.com/datasets/mcagriaksoy/trees-in-satellite-imagery.

[2] Ball, J., Petrova, K., Coomes, D. A., & Flaxman, S. (2022). Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation. *British Ecological Society*.

[3] Emmanuel, D. P., & Jennifer, K. (2022). Understanding 34 Years of Forest Cover Dynamics across the Paraguayan Chaco. *MDPI*.

[4] Madroño, A. (s.f.). El Chaco Paraguayo: ambientes naturales, sus aves y problemas de conservación. *Cotinga 4*.

[5] Abadi, M., Barham, P., Chen, J., et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems (Version 2.3) [Software]. TensorFlow.org. https://www.tensorflow.org/

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). Array programming with NumPy. Nature, 585(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

McKinney, W. (2010). Data structures for statistical computing in Python. Proceedings of the 9th Python in Science Conference, 51-56. https://doi.org/10.25080/Majora-92bf1922-00a

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830. http://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

Chollet, F. (2015). Keras [Software]. https://github.com/fchollet/keras