DEEP-LEARNING BASED ANALYSIS OF SATELITES IMAGE TIME SERIES FOR MAPPING FOREST EGENERATION IN AMAZON RAINFOREST

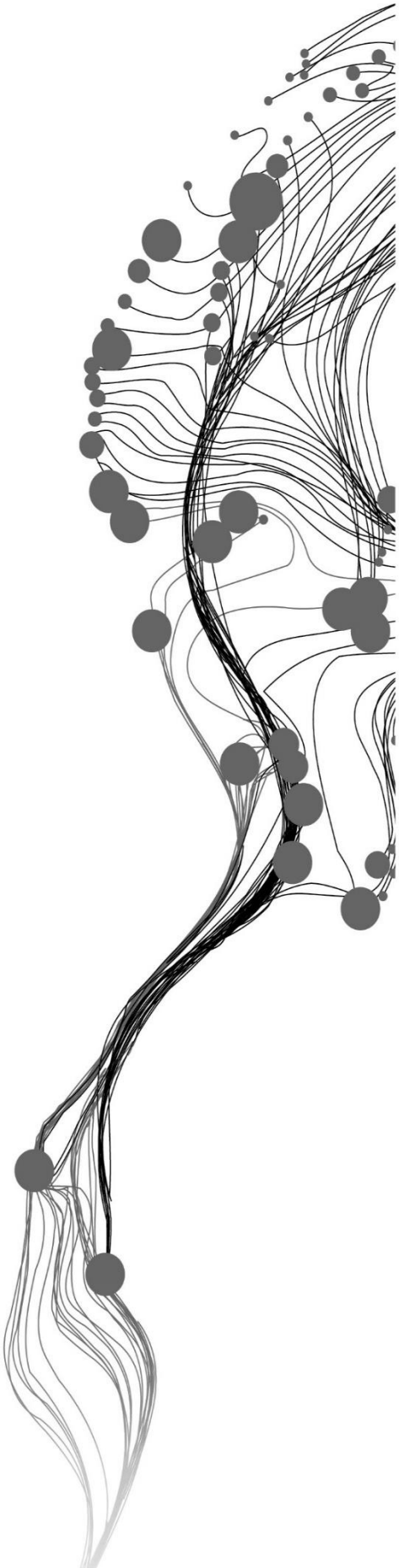
HATANGIMANA FULGENCE

Enschede, The Netherlands, AUGUST, 2024

SUPERVISORS

dr. R. Vargas Maretto

prof.dr.ir. C. Persello



DEEP-LEARNING BASED ANALYSIS OF SATELITES IMAGE TIME SERIES FOR MAPPING FOREST REGENERATION IN AMAZON RAINFOREST

HATANGIMANA FULGENCE

Enschede, The Netherlands, July 2024

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Geoinformatics

SUPERVISORS:

dr. R. Vargas Maretto

Prof.dr.ir. C. Persello

THESIS ASSESSMENT BOARD:

dr. Mariana Belgiu

dr. Marcos Adami

ABSTRACT

The Amazon Rainforest, crucial for global climate regulation and biodiversity, faces significant threats from deforestation and degradation. Traditional monitoring methods often lack precision and scalability, failing to capture the complex temporal dynamics of forest ecosystems. This study addresses these gaps by developing and implementing a deep-learning-based model to map forest regeneration areas in the Amazon Rainforest using satellite image time series. The methodology involved data pre-processing, model development, and performance evaluation using metrics such as accuracy, precision, recall, and F1-score. Precision-recall and ROC curves (AUC-ROC) were also employed to assess model performance.

Time-series analysis is essential for understanding the temporal patterns of forest regeneration. We found that the hybrid transformer architecture outperformed the standard transformer model in distinguishing regenerated areas from other classes. The hybrid transformer model demonstrated superior performance, achieving an overall accuracy of 86.36% compared to the traditional transformer's 85.48%. The model achieved an F1-score of 0.863. When analysing longer periods, the hybrid transformer achieved an overall accuracy of 86.88%, a recall of 0.86, and a precision of 0.86. According to the most accurate model, secondary forest occupies 6.4% of the research area and has a mapping accuracy of 86.38%, which aligns with previous studies.

In conclusion, the hybrid transformer model is a valuable tool for conservation and management, providing precise and reliable maps of forest regeneration. Future research should continue investigating hybrid transformers that utilize both spatial and temporal data and explore more advanced deep-learning architectures, such as Long Short-Term Memory (LSTM) networks. additional features such as the Normalized Difference Vegetation Index (NDVI), which is crucial for differentiating vegetation from other classes Additionally, integrating other additional features such as the Normalized Difference Vegetation Index (NDVI), which is crucial for differentiating vegetation from other classes.

Keywords: Deep-learning models, Forest regeneration, Satellite image time series, Performance metrics.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, **Dr. R. Vargas Maretto** and **Prof. Dr. Ir. C. Persello**, for their invaluable guidance, support, and encouragement throughout my research. Their expertise, insightful feedback, and patience have been instrumental in the completion of this thesis.

I am also profoundly grateful to **Prof. Dr. Ir. A. Stein**, a member of my Thesis Assessment Board, for his constructive criticism and helpful suggestions which significantly contributed to the improvement of my work.

I would like to thank the government of the Netherlands for providing me with financial support through the Orange Knowledge Programme (OKP) to study for my master's degree in science at ITC, the University of Twente.

I would like to extend my heartfelt thanks to my family and friends, whose steady support and understanding have been a source of strength and motivation. Their encouragement has been vital in helping me overcome challenges and stay focused on my goals.

Special thanks to my colleagues and peers who have provided me with a supportive environment and collaborative spirit. Their friendship and shared experiences have made this journey a memorable and enriching one.

Thank you all for your support and belief in me.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AUC-ROC : Area Under the Receiver Operating Characteristic Curve

CNN : Convolutional Neural Network

GEE : Google Earth Engine

IoU : Intersection over Union

LSTM : Long Short-Term Memory

MLP : Multi-Layer Perceptron

NIR : Near-Infrared

NLP : Natural Language Processing

OKP : Orange Knowledge Programme

RNN : Recurrent Neural Network

SAR : Synthetic Aperture Radar

SWIR : Short-Wave Infrared

TPR : True Positive Rate

USGS : United States Geological Survey

ViT : Vision Transformer

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Background

Forests cover approximately one-third of the Earth's surface, or approximately 4 billion hectares, making them a critical component of our planet's ecosystem (Dieterle, 2010; Katila et al., 2014; Lakshmi, 2015; Sukumar, 2008). They play a crucial role in regulating the carbon cycle by absorbing and storing carbon, which helps mitigate high atmospheric carbon dioxide concentrations (Reichstein et al., 2013). Carbon absorption is essential because it reduces the greenhouse effect, which is a major driver of global warming. Therefore, protecting and restoring forests is essential to mitigate climate change and reduce carbon emissions.

Deforestation, driven by agricultural expansion, infrastructure development, and urbanization, is a significant environmental issue that threatens biodiversity and contributes to high levels of CO2 emissions (Suding et al., 2014), leading to loss of forest cover, exacerbation of climate change, and habitat damage. The Amazon Rainforest, the largest tropical rainforest in the world, covering approximately 5.5 million square kilometers, exemplifies this issue. It plays a vital role in global climate regulation and biodiversity. However, it faces severe threats from deforestation and degradation, which impact its capacity to function as a carbon sink (Brando et al., 2014; Lima et al., 2014), which is crucial for maintaining the health of the Amazon and its ability to regulate the global climate.

In the Amazon, forests are severely affected by factors, such as agricultural expansion, illegal logging, and infrastructure development (Shiferaw & Suryabhagavan, 2019)). These activities not only cause local ecological disruptions but also have profound global implications, including climate change, biodiversity damage, carbon release, and diminishing the forest's role as a carbon sink (Coe et al. 2013). Consequently, addressing deforestation is essential for preserving the Amazon Rainforest and mitigating the effects of climate change (Nepstad et al., 2014).

Forest regeneration is a vital process that helps forests recover and grow, enhancing the ability of the Amazon rainforest to absorb and store carbon dioxide from the atmosphere (Chazdon et al., 2016). This process also conserves biodiversity and enhances ecosystem services, which are crucial for maintaining the balanced and stable functioning of ecosystems worldwide. By focusing on forest regeneration, the Amazon Rainforest can sequester carbon and support a diverse array of species, thereby contributing to the stabilization of the global climate.

Traditional machine learning techniques such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forests (RF) have been widely used for land use and land cover classification using satellite imagery (Bharghavi et al., 2023; Loganathan et al., 2021). Although these methods have proven to be effective in various applications, they are typically applied on a small scale and face challenges related to precision, accuracy, and time efficiency. Specifically, the limitations of these techniques become apparent when handling large datasets, resulting in reduced classification performance and increased computational demands. Consequently, although traditional machine-learning methods provide valuable insights, their scalability and efficiency highlight the need for more advanced approaches in large-scale remote-sensing applications.

The integration of deep-learning techniques, specifically transformer and hybrid transformer neural networks, with satellite image time series analysis, represents a significant advancement in environmental monitoring, particularly in forest mapping. These models learn complex temporal patterns from large datasets and are powerful tools for analyzing time series from optical images, such as Landsat 5 and Landsat 8. Using Landsat images, this approach overcomes traditional limitations, such as spectral confusion in optical data, thereby improving the precision, accuracy, and effectiveness of forest regeneration mapping (Hirschmugl et al., 2020).

Previous research on mapping regeneration areas in the Amazon rainforest primarily relied on manual techniques and single-time-point analyses and lacked the precision and automation necessary for large-scale assessments (da Silva et al., 2023). This study aims to bridge this gap by evaluating two advanced deep-learning techniques: a hybrid transformer model and a transformer model. These models automate the process of image classification using satellite image time-series data from 2012 to 2021 instead of relying on single time slots.

The advantage of this approach is its ability to capture temporal dynamics and reduce spectral confusion, which are common challenges in single-time-slot analyses. By improving the precision, accuracy, and effectiveness of forest regeneration mapping, this method not only enhances the reliability of results but also provides a robust tool for future research on secondary forest recovery in the Amazon rainforest.

1.2 Main Objective

The main objective of this research was to develop and implement a deep learning-based model for the analysis of satellite image time series for mapping forest regeneration areas.

1.2.1 Specific Objectives:

1. To explore deep learning-based models, specifically the transformer and hybrid transformer, to accurately identify and distinguish regeneration areas from other classes in the Amazon rainforest.
2. To Assess the model's capability by analyzing time-series data over different time lengths, ensuring improved accuracy and a more realistic representation of the regeneration area.
3. Evaluate the best-performing models for mapping regenerated areas from the developed models.
4. To quantify the area of secondary forest across the entire study area based on the classification generated by the proposed method.

1.3 Research questions

1. Which of the two deep learning architectures, transformer and hybrid transformer, is the most suitable for accurately differentiating regeneration areas from primary forests in the Amazon rainforest?
2. What are the capabilities of the model to effectively manage time-series data across varying time steps to achieve a more accurate and realistic representation of regeneration in the Amazon rainforest?
3. What is the performance of the best-performing model for mapping regenerated areas from the developed models?
4. To what extent is there a secondary forest across the entire study area, based on the classification generated by the proposed method?

1.4 Scope of the study

The scope of this research is to develop and evaluate two deep-learning-based models, specifically the transformer and hybrid transformer, using satellite image time series to identify regenerated areas from other classes in the Amazon Rainforest, particularly in Pará state. These models were evaluated to accurately map the regenerated areas by analyzing temporal information. This study aimed to accurately quantify secondary forests in Pará state and assess model performance using metrics such as accuracy, precision, recall, and AUC, as well as RoC and precision-recall curves. Key aspects of the methodology include model training using satellite temporal series data extracted from satellite image time series with tools such as Python and Google Earth Engine and employing ArcGIS for data visualization.

1.5 Thesis structure

Chapter 1 provides an overview of the research including the study's background, problem statement, research gaps, research objectives, research questions, scope, and thesis structure. Section 2 presents a literature review. Section 3 describes the study area, datasets, and materials used. Chapter 4 discusses the methodology used to answer the research question. Chapter 5 presents an analysis of the results and addresses the research question. The final section concludes this thesis and offers recommendations for future research.

CHAPTER 2: LITERATURE REVIEW

2.1 Overview

This literature review examines the various deep learning architectures used in the mapping of forest regeneration in the Amazon Rainforest. This chapter highlights how these models process temporal data, which are essential for accurately identifying regeneration areas. Understanding the capabilities and limitations of these technologies is crucial to improve forest management and conservation strategies.

The Transformer model, introduced by Vaswani et al. (2017) for natural language processing (NLP) with a novel application of attention mechanisms, allows for efficient processing of sequential data (Dosovitskiy et al., 2021). Unlike recurrent neural networks (RNNs), transformers can capture long-range dependencies without the limitations of sequential processing (Yan et al., 2023), making them faster and more effective for large datasets. Originally developed for natural language processing (NLP), these models have found significant applications in other domains, including image recognition and classification (Dosovitskiy et al., 2021).

This architecture comprises an encoder and decoder, each consisting of multiple layers of self-attention and feed-forward neural networks. The self-attention mechanism computes the relationships between different positions in the input sequence, enabling the model to focus on the relevant parts of the data. This mechanism is particularly useful for tasks that require understanding of the context and relationships within the data. Building on the success of Transformers in NLP, Dosovitskiy et al. (2021) applied transformer architecture to image recognition tasks by introducing a transformer. The transformer processes images as sequences of patches, analogous to tokens in NLP, which allows the model to capture long-range dependencies and global contexts more effectively than traditional convolutional neural networks (CNNs). By treating an image as a sequence of patches, such as words in a sentence, transformers can use their attention mechanisms to capture spatial relationships. This approach has been shown to outperform traditional CNNs for certain image classification and segmentation tasks.

The Transformer model for image processing involves embedding each image patch into a high-dimensional space and adding positional encodings to retain the spatial information. These embeddings are then processed by the transformer encoder, which learns the complex relationships between the patches. The output is used for various downstream tasks such as classification or segmentation.

2.2 Case studies

Table 1 summarizes various case studies involving different models used to analyze forest regeneration patterns and land cover classification. It highlights the specific capabilities of each model, their performance metrics such as overall accuracy, F1-Score, and AUC-ROC, and the relevant references for each study.

Table 1: Case studies related to work.

Case Study	Model Used	Capabilities	Performance Metrics (Accuracy, F1-Score, AUC-ROC)	Reference
Application of LSTM in Predicting Forest Regeneration Patterns in Tropical Forests	LSTM	Multi-temporal analysis across time length s	Accuracy: 82%, F1-Score: 0.81, AUC-ROC: 0.88	(Ball et al., 2022)
GRU-Based Analysis of Secondary Forest Dynamics in the Amazon Basin	GRU	Single time length data input	Accuracy: 85%, F1-Score: 0.84, AUC-ROC: 0.90	(Carvalho et al., 2019)
Temporal Dynamics Analysis with RNN in North American Forests	RNN	Sequential data processing for long-term trends	Accuracy: 76%, F1-Score: 0.74, AUC-ROC: 0.80	(Carvalho et al., 2019)
Hybrid Transformer for Comprehensive Mapping of Forest Regeneration in Southeast Asia	Hybrid Transformer and CNN)	Integration of multispectral and radar data across two-time length s	Accuracy: 89%, F1-Score: 0.88, AUC-ROC: 0.93	(Lucas et al., 2000)
HyFormer: Hybrid Transformer and CNN for Pixel-Level Multispectral Image Land Cover Classification in Changxing County and Nanxun District	Hybrid Transformer	Pixel-level classification of multispectral images	Accuracy: 91%, F1-Score: 0.90, AUC-ROC: 0.94	(Yan et al., 2023)

2.3 Deep-learning Architectures for Time Series Data

This section reviews case studies that apply advanced deep learning models to time-series data to map secondary forests. Table 2 summarizes the different models used along with their descriptions and limitations.

Table 2 : Summary of Deep-learning Architectures for Time Series Data

Model	Description	Limitation	Literature
RNN (Recurrent Neural Network)	Utilizes sequential data processing to maintain state information across inputs, suitable for modelling time-dependent patterns.	Prone to vanishing gradient problem, which can impede learning in long data sequences.	(Park et al., 2020) (Noh, 2021
LSTM (Long Short-Term Memory)	Designed to overcome vanishing gradient issues in RNNs, enabling better performance on long sequence data.	Computationally intensive, requiring more resources which can be a limitation in large-scale deployments.	(Chandra et al., 2021)
GRU (Gated Recurrent Units)	Provides a simplified architecture compared to LSTM, which requires fewer computational resources while achieving similar performance.	Although efficient, may not capture complexities as effectively as LSTM in certain applications.	(Chandra et al., 2021)
Hybrid Transformer	Integrates convolutional neural layers for spatial data processing with transformer mechanisms for handling sequential data.	High computational cost and requires large amounts of data to train effectively.	(Ouyang et al., 2023)

CHAPTER 3: STUDY AREA AND DATASETS

In this section, we provide an overview of our study area, including its location, the datasets we used, and their availability. We explain each component in detail and discuss its significance in our research.

3.1 Study area description

This study was conducted in the Amazon rainforest of Pará State, Brazil. Geographically, the Amazon rainforest lies between the latitudes of 2°37'56.928" and 9°50'44.268" and the longitudes of 58°53'53.7576" and 46°3'5.2308". Approximately 60% of the Amazon rainforest is in Brazil and covers an area of approximately 5.5 million square kilometers. The study area map is shown in the figure1.

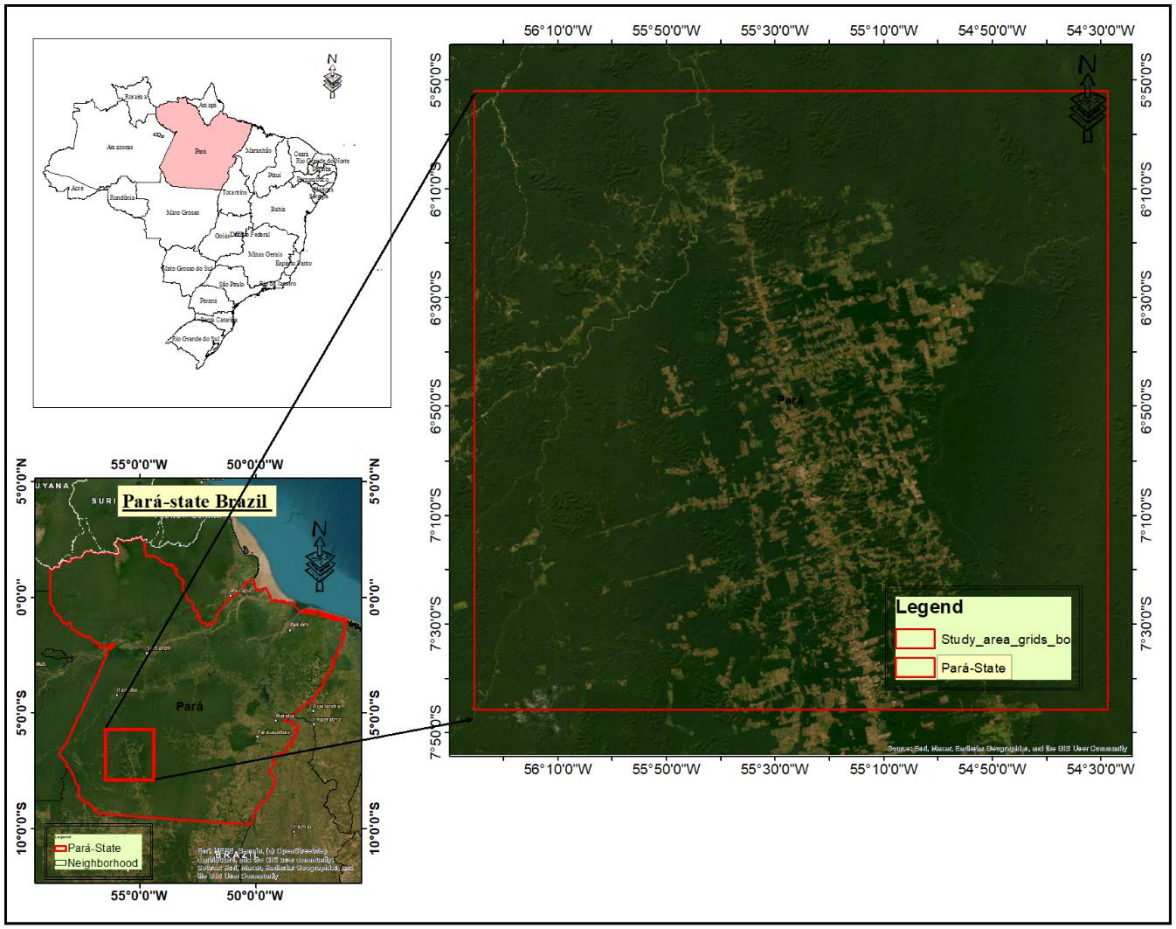


Figure 1: The study area is the Amazon rainforest in Pará State, Brazil. The large map highlights the study grid (red outline).

Pará State is a vast state situated in the Amazon rainforest, making it challenging to simultaneously study the entire region. We chose a smaller area within Pará for our study because it has experienced high rates of deforestation and forest regeneration. Focusing on this specific area makes it easier to simplify the model, reduce processing time, and manage memory usage.

3.2 Datasets

3.2.1 Reference Land Use and Land Cover Data

The Land Use and Land Cover dataset was downloaded from the MapBiomas project (https://brasil.mapbiomas.org/). These maps were downloaded in raster format using the WGS84 spatial reference system (EPSG: 4326). The datasets were generated using machine learning algorithms supplemented by human interventions to improve accuracy. Table 3 provides an overview of these datasets along with their descriptions. Table 3 presents these datasets and their descriptions.

Table 3: Datasets and sources used.

	Dataset	Source	Temporal
1	Land Use and Land Cover Map	https://brasil.mapbiomas.org	2021
2	Deforestation data from PRODES	https://terrabrasilis.dpi.inpe.br	2021
3	Satellite images	USGS Earth Explorer	2012-2021
4	Brazil tiles	Brazill dataCube	2022
5	Administrative boundaries	https://www.diva-gis.org	2022

3.2.2 Satellite image

We obtained images covering the period from 2012 to 2021 from satellites such as Landsat-5 and Landsat-8. These were freely acquired through the Google Earth Engine (GEE) from the European Space Agency (ESA) Earth Explorer or Open Access Hub. Landsat 5 TM has 7 bands, and Landsat 8 OLI has 11 bands, as shown in Table 4.

Table 4: Optical Data characteristics for 5 TM Bands and Landsat 8 OLI/TIRS Bands

Landsat 5 TM Bands				Landsat 8 OLI/TIRS Bands			
Band	Resolution (m)	Color	Wavelength (µm)	Band	Resolution (m)	Color	Wavelength (µm)
				1	30	Coastal/ Aerosol	0.435-0.451
1	30	Blue	0.45 -0.52	2	30	Blue	0.452-0.512
2	30	Green	0.52 - 0.60	3	30	Green	0.533-0.590
3	30	Red	0.63-0.69	4	30	Red	0.636-0.673
4	30	NIR	0.76-0.90	5	30	NIR	1.566-1.651
5	30	SWIR-1	1.55 -1.75	6	30	SWIR-1	1.566-1.651
6	60	TIR	10.4 -12.3	10	100, (for landsat5 120)	TIR-1	10.60-11.19
				11	100	TIR-2	11.50-12.51
7	30	SWIR-2	2.08 - 2.35	7	30	SWIR-2	2.107-2.294
				8	15	Pan	0.503-0.676
				9	30	Cirrus	1.363-1.384

3.3 Availability and Materials

Using Google Earth Engine, we downloaded Landsat-5 and 8 datasets from repositories, such as the USGS Earth Explorer (https://earthexplorer.usgs.gov/). Reference data was also available at https://mapbiomas.org, and we obtained Brazilian administrative boundaries from diva-gis.org. These repositories offer freely accessible data for research. To process and analyze these datasets, we used geospatial software packages, such as Google Earth Engine and ArcGIS Pro, for visualization. For data processing and analysis, we used Python with libraries, including Rasterio and PyTorch. We performed additional processing and analysis using computing resources equipped with 72 vCPU Intel x86-64 processors, 768 GB RAM, and an NVIDIA RTX A4000 GPU.

CHAPTER 4: METHODOLOGY

4.1 Overview of methodology

Figure 2 provides an overview of the methods used in this study. Section 4.2 provides a thorough explanation and discussion of each part of the methodology, including data preprocessing, model architecture development, model training, and the application of these methods to predict forest regeneration in the study area.

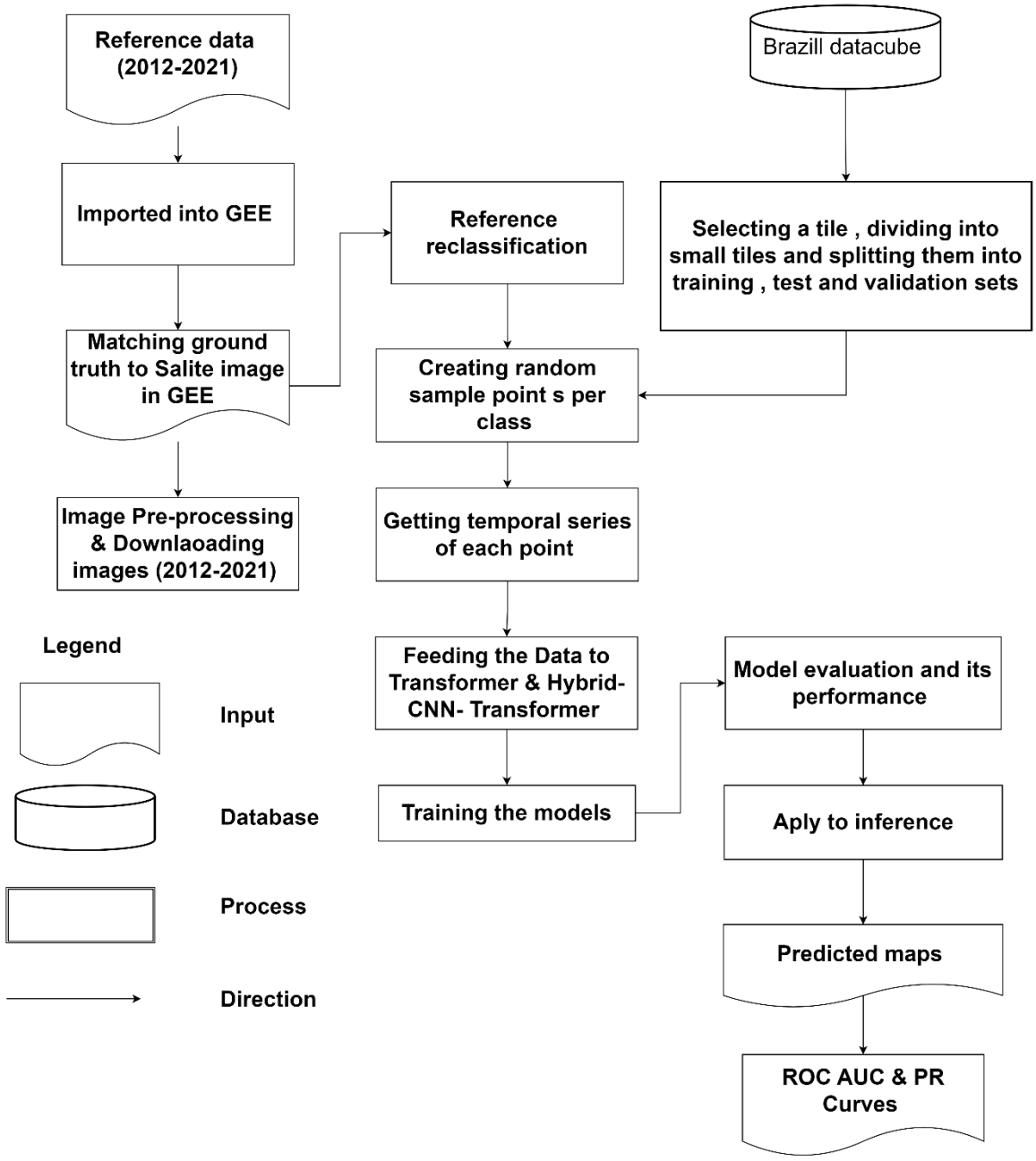


Figure 2: Overview of the methodology

4.2 Data pre-processing

a) Satellite images

In this section, we perform preprocessing tasks using the Google Earth Engine platform before downloading images. These tasks included atmospheric correction, cloud masking, and normalization. We selected the most informative spectral bands, red, green, blue, near-infrared (NIR), and short-wave infrared (SWIR) (Jin & Wang, 2019), to effectively distinguish forest regeneration and reduce the computational costs during model training. SWIR bands, which are sensitive to moisture content in vegetation and soil, are particularly valuable for land cover identification (Chuvieco et al., 2002; Wang et al., 2008). These bands are recognized for their ability to capture critical information about vegetation health, moisture content, and land cover changes (Rukhovich et al., 2022).

b) Reference preparation

Figure 3 illustrates the division of a large tile (211,200 m × 211,200 m) from the Brazil Data Cube into 64 smaller tiles, each measuring 26,400 m × 26,400 m. To make the data more manageable, we split the smaller tiles into three sets: 70% for training, 15% for validation, and 15% for testing. This distribution ensured that the model was trained on a substantial portion of the data, validated to prevent overfitting, and tested to evaluate its performance on unseen data. Consequently, this structured approach to ground-truth preparation enhances the reliability and robustness of the image classification model.

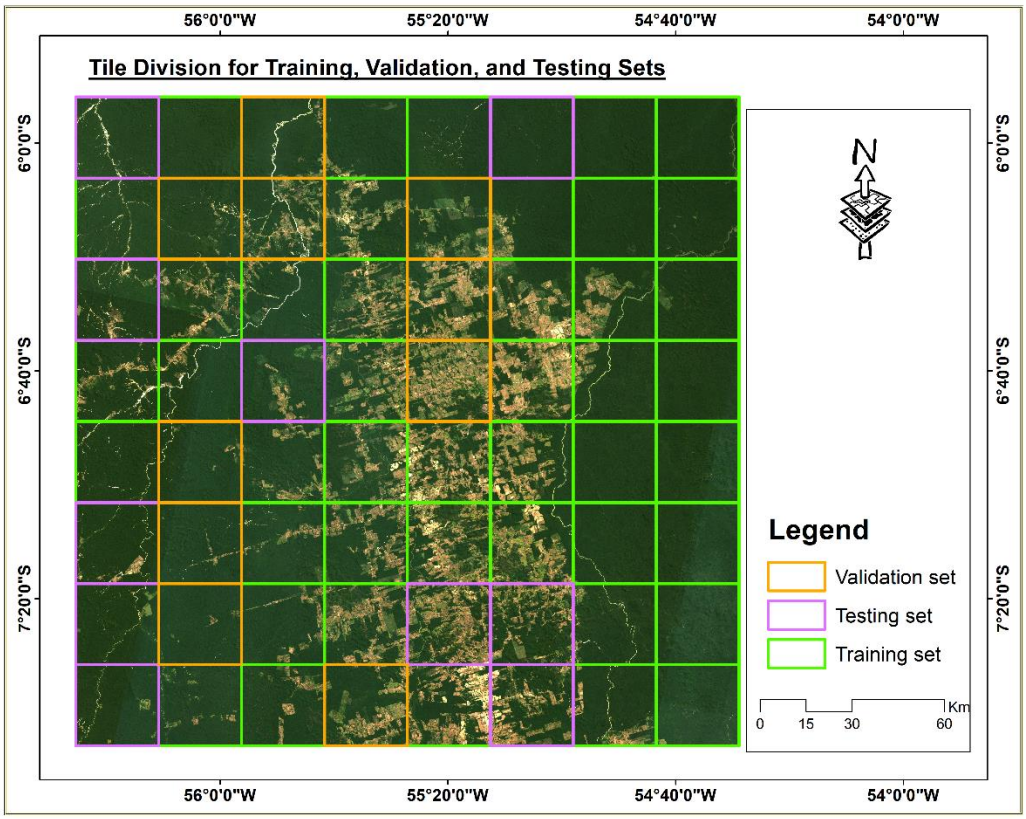


Figure 3: Division of large tiles into 64 smaller tiles and splitting them into training (green), validation (orange), and testing (yellow) sets.

c) LULC maps and reclassification

The maps included several classes, such as anthropic areas, primary vegetation, secondary forests, suppression of primary forests, recovery of secondary forests, and suppression of secondary forests. We grouped these classes into five new categories: anthropic, deforestation, primary forest, secondary forest, and water (see Figure 4).

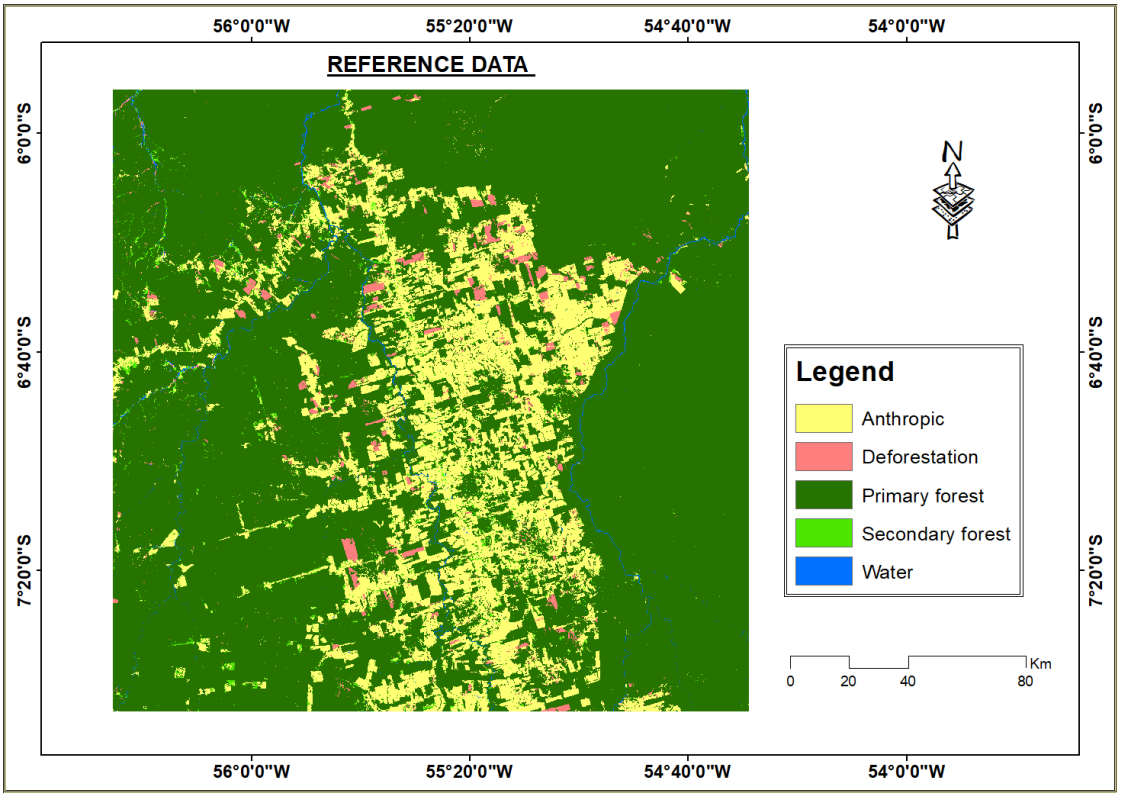


Figure 4: Reclassified LULC map from Mapbiomas data, showing anthropic areas (yellow), deforestation (red), primary forests (dark green), secondary forests (light green), and water (blue).

After reclassification, we generated 2000 random points per class for the training set and 1000 random points per class for both the testing and validation sets. We then extracted a temporal series of spectral features from five bands: blue, green, red, near-infrared (NIR), and short-wave infrared (SWIR). Before feeding the data to the model, we applied one-hot encoding to the categorical class labels. This technique converts each class label into a binary vector, making it easier for the model to read and classify data by indicating the presence or absence of each class. This one-hot encoding ensures that the model can reliably distinguish between different classes, thereby enhancing classification performance. Figure 5 shows the distributions of the training, validation, and testing points.

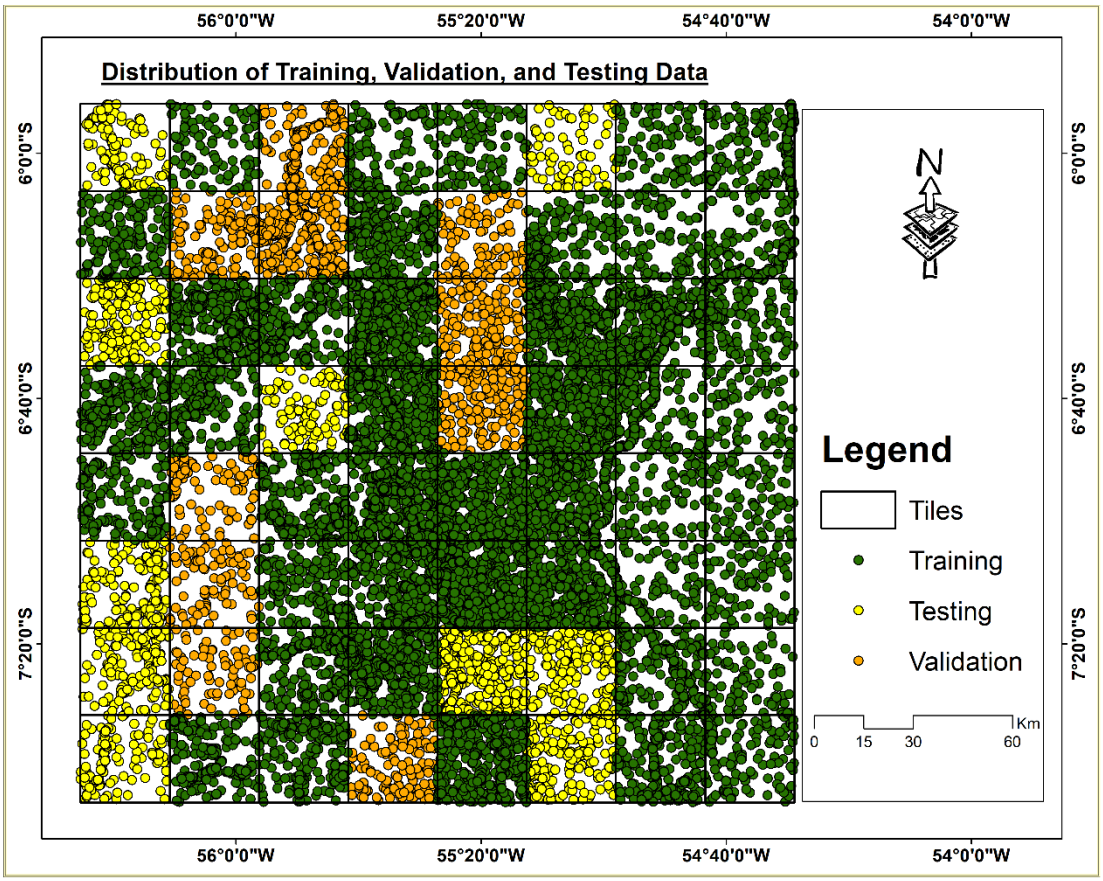


Figure 5: Distribution of random points for the training (green), testing (yellow), and validation (orange) sets after reclassification.

4.3 Model Development

In this section, we describe two deep-learning architectures, Transformer and Hybrid Transformer, which were explored and evaluated for their effectiveness in distinguishing regenerated areas from other classes. Once the predictions were made, the extent of the regenerated area was quantified.

4.3.1 Transformer

The Transformer model, originally developed for natural language processing (NLP) by (Dosovitskiy et al. (2020), was adapted for various image processing tasks. In this study, we employed a transformer to predict forest regeneration using satellite images.

The model processes a one-dimensional sequence of pixels, with each sequence containing five spectral bands over varying time series lengths: a single image (t1), three time lengths (t3), five time lengths(t5), and ten time lengths(t10). The input data are first converted into dense vector embeddings and then projected into a higher-dimensional space, specifically, a multi-channel feature map. Positional information was added to maintain the sequence order.

Temporal patterns were extracted using the multihead attention mechanism (Shih et al., 2018), and these combined patterns and positional embeddings were fed into the transformer encoders. Finally, the MLP (Multilayer Perceptron) head converts the embedded features into classification results.

Table 5 outlines the key components and layers of our transformer model, highlighting the transformation of the input data through convolutional operations, positional embeddings, multiread attention mechanisms, and final classification.

Table 5: Transformer Model Architecture for Spectral Data that adapted from a study by Yan et al. (2023).

Layer	**Formula**	**Description**
input embeddings	$X \in \mathbb{R}^{L*5}$	Input 1 D sequence of pixels across five bands over different time lengths.
linear projection	$E_{i=} W.X_i + b$ for i=1,….. L	Each pixel sequence (of five spectral bands) is linearly projected into high-dimensional space.
Positional embeddings	$P \in \mathbb{R}^{L*D}$	Positional embeddings are added to the input sequence to encode positional information
Adding Positional Information:	$Z = E + P$	Combines linearly projected embeddings with positional embeddings to form the input for the multi-head attention mechanism.
Multi-head attention	$(Q, K, V) = \text{SoftMax}($	Applies multi-head attention to the input embeddings to focus on different relevant parts of the input sequence for making predictions
Layer Normalization and Residual Connections	FFN(x) =max (0, xW1+b1) W2+b2	Applies multi-head attention to the input embeddings to focus on different relevant parts of the input sequence for making predictions
MLP	Dense (5, activation='SoftMax')	A multi-layer perceptron (MLP) with a dense layer of 5 units, using SoftMax activation to classify the input into one of five categories based on the learned features.

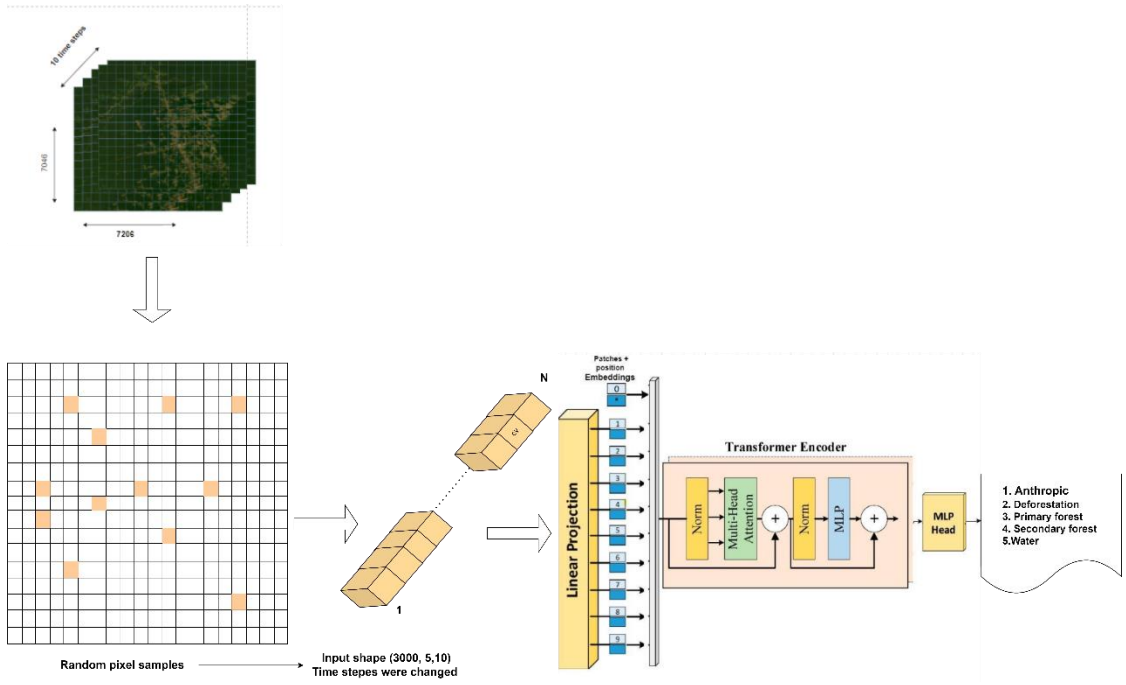


Figure 6: Workflow of the Transformer model for forest classification across different time lengths. This methodology was partly inspired by Dosovitskiy et al**.** (2020).

Figure 6: Steps of using a transformer model to analyze land cover changes over different time periods. The input raster maps were converted into a 1D sequence of pixels, which were then projected linearly and embedded in positional information. These embeddings are fed into the Transformer encoder, which processes the data to classify each pixel into one of five categories: anthropic, deforestation, primary forest, secondary forest, and water

4.3.2 Hybrid Transformer

Hybrid transformer architecture involves the integration of CNNs and transformers. Figure 7 depicts the hybrid transformer architecture used in this project to predict the one-dimensional pixel sequence of the five classes. This architecture focuses on temporal information using a CNN layer to capture temporal patterns and a transformer to capture temporal dependencies.

We processed a one-dimensional pixel sequence of five classes over different time lengths, including t1, t3, t5, and t10. Each pixel was represented as a sequence of five spectral values. The input data are first fed into the CNN. Temporal patterns were extracted using a 2D CNN layer and embedded into a higher-dimensional space, specifically, a multi-channel feature map for further analysis using transformer blocks. The transformer combines the CNN-extracted temporal patterns with positional encoding, which are then fed into the transformer encoders. Finally, the MLP (Multilayer Perceptron) head converts the embedded features into classification results.

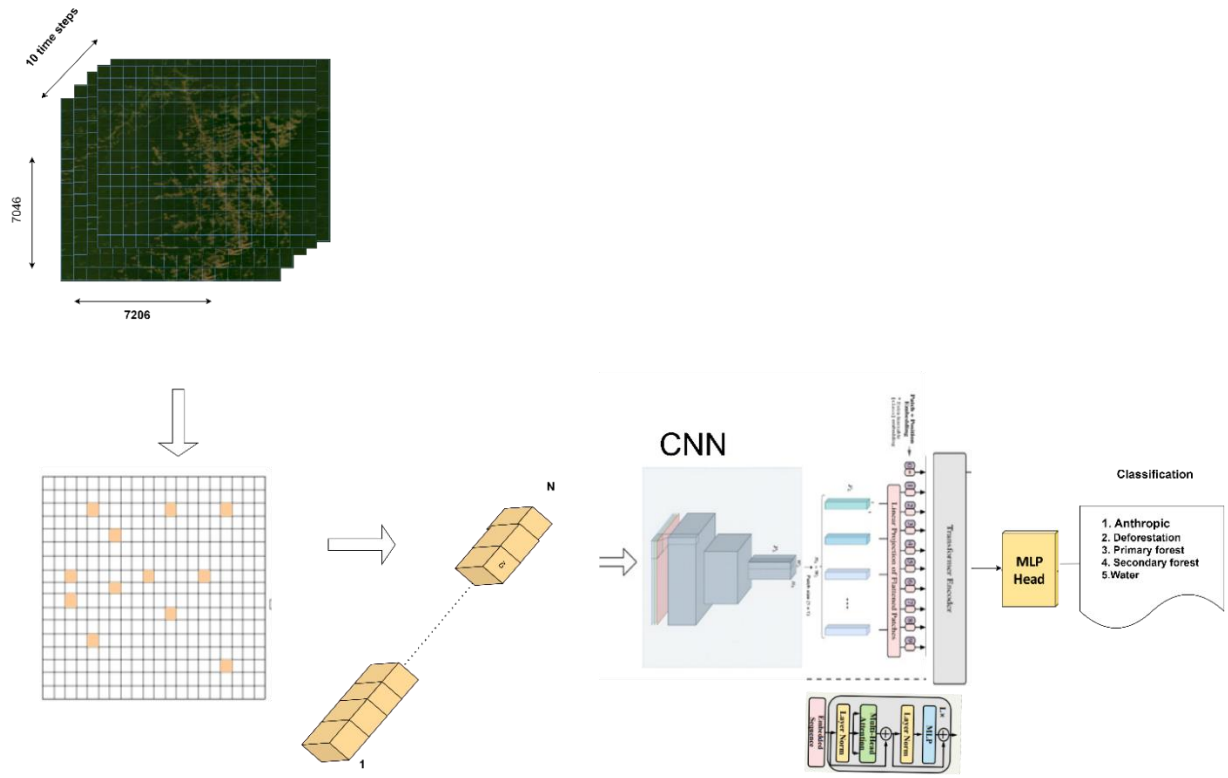


Figure 7: Hybrid Transformer architecture for predicting one-dimensional pixel sequences across five classes. The methodology was partly inspired by that adapted from Yan et al. (2023)

Layer Details

Table 6 presents a detailed list of the layers used in the hybrid transformer model. Each layer is described using its corresponding formula and functionality within the model.

Table 6: Transformer Model Architecture for Spectral data. This table was adapted from Yan et al. (2023).

Layer	Formula	Description
CNN Block		
2D convolution	Y = Conv2D (X, W)	Applies a 2D convolution over the input image to extract temporal information.
Pixel-wise nonlinear activation	Y = ReLU(X)	Applies ReLU activation function to introduce non-linearity.
Max-pooling	Y = MaxPool(X)	Downsamples the input along the temporal dimension by taking the maximum value over an input window (pooling window) for each channel.
Transformer Encoder		
Adding Positional Information:	Z = E + P	Combines linearly projected embeddings with positional embeddings to form the input for the multi-head attention mechanism.
Multi-head attention	$Z=E+P$	Applies multi-head attention to the input embeddings to focus on different relevant parts of the input sequence for making predictions.
Layer Normalization and Residual Connections	(Q, K, V) = SoftMax $(QK^T/\sqrt{\boldsymbol{dk}})V$	Applies layer normalization and adds the residual connections to stabilize and accelerate the training.

MLP	Dense (5, activation='SoftMax')	A multi-layer perceptron (MLP) with a dense layer of 5 units, using SoftMax activation to classify the input into one of five categories based on the learned features.

4.4 Loss function

Cross-entropy is commonly used for multiclass image classification because of its ability to manage multiple categories (Andreieva & Shvai, 2021). It measures the model performance of the predicted class probabilities compared with the actual target classes for each pixel. Cross-entropy (L) was mathematically calculated as follows:

$$L = -\frac{1}{N}\sum_{i=1}^{M} yic \ \log(pic) \quad (4.1)$$

Where N is samples, M denotes the number of classes and yic is the sign function (0 or 1). If the true category of sample *i* is equal to c, take 1, otherwise take 0; pic is the predicted probability that the observed sample *i* belongs to category c. This formula quantifies the discrepancy between the predicted probabilities and actual class labels, thereby effectively guiding the model to improve its predictions.

4.5 MODEL PERFORMANCE METRICS

4.5.1 Overall Accuracy

Overall accuracy is a fundamental metric used to evaluate the performance of classification models (Novakovic et al., 2017). This represents the proportion of correctly classified instances (both positive and negative) out of the total number of instances. Simply put, it measures how often the model makes correct predictions across all classes (Bharghavi et al., 2023). The formula for calculating overall accuracy is as follows

$$\boldsymbol{Overall\ accuracy} = \frac{\boldsymbol{TP+FN}}{\boldsymbol{TP+TN+FP+FN}} \quad (4.2)$$

4.5.2 Precision, Recall, and F1-Score

In addition to the overall accuracy, we analyzed the confusion matrix to derive other critical metrics, such as precision, recall, and F1-score. These metrics provide more detailed insights into the performance of the model, particularly in the context of imbalanced datasets. Using Eq. (4.2), we calculated the precision and recall for each class, as follows:

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}}, \boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \quad (4.3)$$

Where TP, FP, and FN are the total number of true positives, number of false positives, and number of false negatives, respectively.

Precision indicates the proportion of correct positive predictions, whereas recall indicates the proportion of actual positive cases that were correctly identified by the classifier (Alvarez, 2002). The F1-score, which is the harmonic mean of the precision and recall, provides a single measure that balances the two metrics.

4.5.3 Receiver Operating Characteristic (ROC) Curve and Precision-Recall Curve

To further evaluate the performance of our model, we generated a Receiver Operating Characteristic (ROC) curve and precision-recall curve. These curves provide additional insights into the ability of the model to distinguish between different classes, which is crucial when dealing with imbalanced datasets.

$$\boldsymbol{True\ positive\ Rate\ (TPR)} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \quad \textbf{(4.4)}$$

$$\boldsymbol{False\ positive\ Rate} = \frac{\boldsymbol{FP}}{\boldsymbol{TP+FP}} \quad \textbf{(4.5)}$$

The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) across various thresholds, helping to assess the model's overall discriminatory power. The Area Under the ROC Curve (AUC) was used as the summary statistic to quantify the overall performance of the model.

Similarly, the precision-recall curve plots precision against recall for different thresholds, offering a more informative perspective, particularly in scenarios where the dataset is imbalanced. These curves, along with the calculated metrics, provide a comprehensive evaluation of our model's performance, allowing us to identify both strengths and areas for potential improvement.

4.6 Application to Full Image

The trained models were applied to the entire Landsat image, which was preprocessed before making predictions for each pixel. The results were visualized as classified images to provide a detailed visual representation of the predicted land cover types across the landscape. This step demonstrated the practical application of the models in real-world scenarios.

4.7 Post-processing and Analysis

In the post-processing phase, we quantified the secondary forest by calculating the class statistics. We counted the pixels for each land cover class, calculated the area in square kilometers (km^2), and determined the percentage of each land cover type. We then plotted a pie chart to visualize these proportions.

$$\boldsymbol{Area\ (km2)} = pixelcount\ X\ pixel\ size \quad (4.6)$$

CHAPTER 5: RESULTS AND DISCUSSIONS

5.1 Introduction

This chapter presents the results obtained from the study and provides detailed findings, along with the research objectives and questions. The performance of the deep learning models, their accuracy, and the implications of these results for mapping forest regeneration were analyzed.

The main objective of this research was to develop and implement a deep learning-based model for the analysis of satellite image time series to map forest regeneration areas. The results were organized according to specific objectives and research questions.

5.2 Exploration of deep learning models

To explore the effectiveness of deep learning models in identifying and distinguishing regenerated areas, we developed and tested two models: a transformer model and a hybrid transformer model. The transformer model uses self-attention mechanisms to capture global dependencies within the satellite image time series data, whereas the hybrid transformer model combines convolutional neural networks (CNN) with transformer layers to integrate both local and global feature extraction capabilities. The CNN layer was used to capture temporal patterns, and a transformer was used to capture temporal dependencies. Both models were trained on a dataset consisting of a 1D sequence of pixels across five bands over different time lengths, which were labelled into various classes such as anthropic, deforestation, primary forest, secondary forest, and water bodies.

The evaluation metrics further demonstrated the superiority of the hybrid model, with an overall accuracy of 86.38%, precision of 0.86, recall of 0.86, and F1-score of 0.86. In comparison, the traditional transformer achieved an overall accuracy of 85.48% with a precision of 0.86, recall of 0.85, and F1-score of 0.85. The visual results shown in section 5.3.1 indicated the qualitative analysis of the predicted maps.

5.3 Model capability with time series data over different time lengths

In this study, we evaluated the performance of two models: the Transformer and the Hybrid Transformer in predicting regenerated areas in the Amazon rainforest using time-series data of varying lengths: a single time step (t1), three time steps (t3), five time steps (t5), and ten time steps (t10). We assessed each model's effectiveness through precision, recall, F1-score, and overall accuracy (see Table 7).

At t1, the Transformer model achieved an F1-score of 0.71 with an overall accuracy of 70.80%, indicating its ability to identify some patterns but with limitations due to the short temporal context. The Hybrid Transformer performed slightly better, with an F1-score of 0.74 and an overall accuracy of 73.88%, likely due to its CNN layer, which captures local temporal patterns.

As the time steps increased to t3, both models improved; the Transformer achieved an F1-score of 0.80 and an accuracy of 80.74%, while the Hybrid Transformer reached 0.88 and 81.76%, respectively. The Hybrid model's architecture, combining CNN and Transformer layers, allowed for more effective utilization of temporal data, resulting in more accurate predictions.

By t5, the Transformer's F1-score rose to 0.83 with 83.70% accuracy, but the Hybrid Transformer continued to outperform it, achieving an F1-score of 0.84 and 85.50% accuracy. At t10, the Transformer peaked with

an F1-score of 0.85 and 85.48% accuracy, while the Hybrid Transformer maintained its lead with 0.86 and 86.36%.

Therefore, the Hybrid Transformer consistently outperformed the standard Transformer across all time steps due to its superior architecture, which effectively captures both local and long-term temporal patterns. This makes the Hybrid Transformer a more accurate and reliable model for predicting forest regeneration in the Amazon rainforest, especially when using extended time-series data.

Table 7: Model capabilities with time series data on performance metrics over different time lengths (Using testing **points)**

Time length (t_i)	Model	Precision	Recall	F1-score	Overall accuracy (%)
1	Transformer	0.71	0.71	0.71	70.80
	Hybrid Transformer	0.74	0.74	0.74	73.88
3	Transformer	0.81	0.81	0.8	80.74
	Hybrid Transformer	0.83	0.82	0.88	81.76
5	Transformer	0.84	0.83	0.83	83.70
	Hybrid Transformer	0.85	0.84	0.84	85.50
10	Transformer	0.86	0.85	0.85	85.48
	Hybrid Transformer	0.86	0.86	0.86	86.36

5.4 The performance of the best-performing model for mapping regenerated areas

Once the model capabilities were tested and assessed, it became evident that both the Hybrid Transformer and the Traditional Transformer performed well at larger time steps. Given their strong performance at these extended time steps, particularly in classes with distinct spectral signatures, these two models were subsequently applied to predict forest regeneration across the Amazon rainforest. The following sections present a detailed evaluation of the models' performance in mapping regenerated areas using testing sample points and by assessing the accuracy using all pixels from predicted images corresponding to testing set tiles.

5.4.1 Evaluation using testing sample points

The performance of the Hybrid Transformer and Traditional Transformer models for mapping regenerated areas is summarized in Tables 8 and 11. Overall, the Hybrid Transformer outperformed the Traditional Transformer across most land-cover classes, demonstrating higher accuracy, precision, recall, and F1 scores.

In Table 8, the Hybrid Transformer exhibited strong performance in the water class, achieving a precision of 97.80% and an F1 score of 97.85%. This indicates the model's effectiveness in identifying water bodies, a land-cover class with distinct and consistent spectral signals. Similarly, in the primary forest class, the Hybrid Transformer excelled, with a precision of 79.94% and an F1 score of 86.5%. The model's integration of CNN layers likely contributed to this success by enhancing its ability to capture and process local temporal patterns, which are crucial for accurately identifying stable land-cover types like primary forests.

Despite its overall strong performance, the Hybrid Transformer encountered challenges in accurately classifying more complex classes, such as secondary forests and deforestation areas. In Table 8, the model's precision for the deforestation class was 84.79%, with an F1 score of 81.58%, reflecting difficulties in distinguishing deforested areas. The Traditional Transformer struggled even more in this class, achieving a precision of 76.22% and an F1 score of 79.49%.

Additionally, both models showed weaker performance in the secondary forest class, where the Hybrid Transformer achieved a precision of 81.65% and an F1 score of 81.32%, while the Traditional Transformer had a precision of 76.22% and an F1 score of 80.77%. These results underscore the complexities involved in classifying secondary forests, which often exhibit mixed or transitional spectral characteristics, making them challenging for both models.

The Hybrid Transformer consistently demonstrated superior performance, particularly in classes with distinct and stable spectral signatures, such as water bodies and primary forests. However, both models showed limitations in more challenging classes like deforestation and secondary forests, where spectral variability and complexity are higher. These findings suggest that while the Hybrid Transformer is generally more reliable, further improvements are needed to enhance its accuracy in these difficult-to-classify areas.

Table 8: Performance of the hybrid model to map regenerated areas using test sample points.

	hybrid transformer				Transformer			
Class	**Precision (User)**	**Recall (Producer)**	**F1-Score**	**Overall Accuracy**	**Precision**	**Recall (Producer)**	**F1-Score**	**Overall Accuracy**
Anthropic	87.98	87.8	87.89		84.41	88.8	86.55	
Deforestation	84.79	78.6	81.58		94.49	68.6	79.49	
Primary Forest	79.94	86.5	86.5		78.56	85.4	81.84	
Secondary Forest	81.65	81	81.32		76.22	85.9	80.77	
Water	97.8	97.9	97.85		97.92	98.7	98.31	
Overall Accuracy				**86.36**				**85.48**

5.4.2 Assessing the accuracy using all pixels from predicted images corresponding to testing set tiles

The classification capabilities of the Hybrid Transformer and Traditional Transformer models were evaluated using precision, recall, and F1 scores across various land-cover classes, as detailed in Table 9. The Hybrid Transformer generally outperformed the Traditional Transformer, achieving an overall accuracy of 81.087% compared to 79.346%.

The Hybrid Transformer excelled in the anthropic class, with a precision of 99.42% and an F1 score of 93.02%, and in the primary forest class, where it achieved a recall of 94.98% and an F1 score of 94.89%. This strong performance can be attributed to the model's ability to capture clear and consistent spectral patterns through its CNN layers, making it particularly effective in classes with well-defined characteristics.

However, the Hybrid Transformer showed limitations in the water class, where it recorded a precision of 55.42% and an F1 score of 69.85%, and in the deforestation class, with a precision of 51.21% and an F1 score of 59.75%. These lower scores suggest challenges in classifying land-cover types with less distinct spectral signatures, likely due to variability and mixed signals within these categories.

To improve performance in challenging classes such as water and deforestation, several strategies could be implemented. Enhancing the model's training data with more representative samples, particularly in areas with high spectral variability, could help the model better distinguish these complex classes. Incorporating post-classification smoothing techniques, such as spatial filtering methods like majority filtering, may reduce the impact of mixed pixels and improve the consistency of the model's predictions. Additionally, integrating features such as NDVI, critical for vegetation analysis, could further enhance classification accuracy.

While the Traditional Transformer showed higher precision in the primary forest class (95.48%) with a higher F1 score of 97.61%, it suffered from significantly lower recall in the deforestation class (30.74%), suggesting a potential underestimation of this category. In secondary forests, the Hybrid Transformer's balanced precision (64.95%), recall (67.73%), and F1 score (66.32%) demonstrate its robustness in handling complex, transitional land-cover types. Conversely, the Traditional Transformer struggled more with the spectral variability in these areas.

Thus, while the Hybrid Transformer demonstrates strong performance in classes with distinct spectral signatures, both models face challenges in more complex environments. Addressing these limitations through improved data, post-classification techniques, and multi-temporal data integration could enhance their accuracy and robustness in diverse real-world applications.

Table 9: Performance of the hybrid model for mapping regenerated areas.

	Hybrid transformer				**Transformer**			
Class	**Precision (User)**	**Recall (Producer)**	**F1-Score**	**Overall Accuracy (%)**	**Precision (User)**	**Recall (Producer)**	**F1-Score**	**Overall Accuracy (%)**
Anthropic	99.42	87.56	93.02		89.98	89.12	90	
Deforestation	51.21	71.82	59.75		38.97	30.74	51.45	
Primary Forest	94.8	94.98	94.89		95.48	98.63	97.61	
Secondary Forest	64.95	67.73	66.32		61.24	59.69	61.91	
Water	55.42	94.29	69.85		88.81	54.1	67.82	
Overall Accuracy				**81.087**				**79.346**

5.4.3 ROC curve and precision-recall curves

I. Using testing points

The evaluation of the Hybrid Transformer and Traditional Transformer models trained using sample points, through ROC and precision-recall curves provides valuable insights into their classification capabilities. The ROC curve's AUC values, ranging from 0 to 1, serve as a key metric for assessing model performance, where an AUC closer to 1 indicates excellent predictive ability, while values around 0.5 suggest random guessing, and below 0.5 indicate poor performance. In this study, the ROC curves demonstrated the superiority of the Hybrid Transformer, which achieved impressive AUC values of 0.97 for anthropic, 0.95 for deforestation, 0.967 for primary forest, 0.96 for secondary forest, and 1.00 for water. These results reflect the model's robust classification capabilities, particularly in distinguishing between different land-cover types.

However, a critical analysis reveals that while the Hybrid Transformer performs exceptionally well in most classes, its performance in the deforestation class, although strong with an AUC of 0.95, still indicates areas for improvement. The inherent complexity and spectral variability within deforested areas might introduce challenges in maintaining high classification accuracy. The Traditional Transformer, in contrast, showed generally lower AUC values, particularly in the deforestation class (0.85), which emphasizes the advantage of the hybrid approach but also highlights the necessity for enhanced methodologies to handle difficult classes like deforestation and secondary forests.

The precision-recall curves offer additional insights into the models' effectiveness. For the Hybrid Transformer, the anthropic class displayed a high AUC of 0.92, indicating excellent precision across various recall levels, underscoring the model's strength in handling this relatively straightforward class. However, the deforestation class, with an AUC of 0.92, although balanced, showed some precision loss at higher recall levels. This suggests that while the model is generally robust, the nuanced characteristics of deforested areas require further refinement, possibly through more sophisticated feature extraction or enhanced training data that better represents these complex environments.

In contrast, the Traditional Transformer’s performance, particularly in the deforestation class with an AUC of just 0.56, reveals significant limitations. The sharp drop in precision at higher recall levels suggests that the model struggles considerably with the spectral complexity and variability inherent in deforestation areas, likely leading to a higher rate of misclassifications. This highlights a crucial weakness that must be addressed if the model is to be effectively applied in more complex, real-world scenarios.

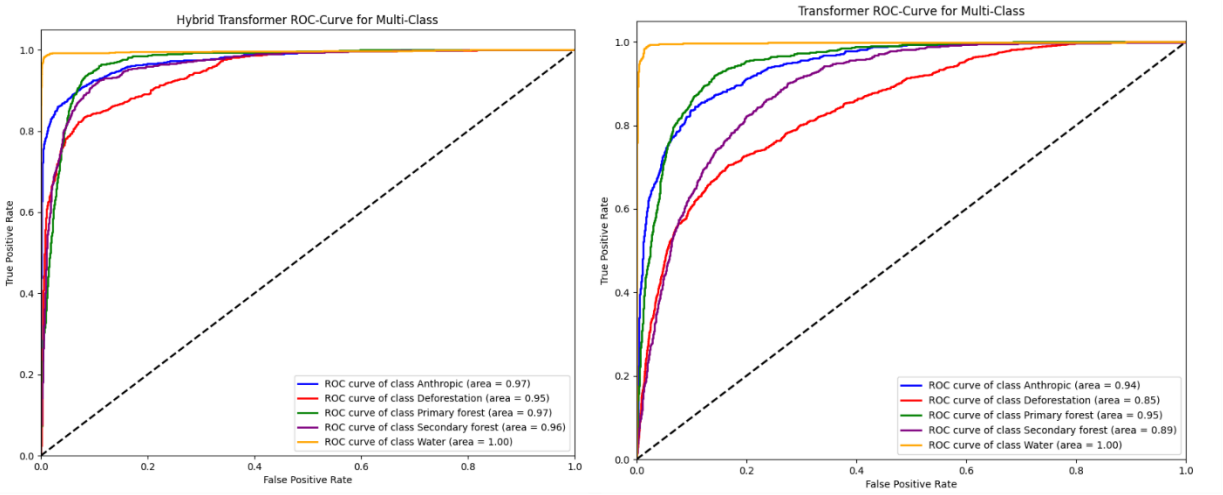


Figure 8: ROC curves for Transformer and Hybrid Transformer

Precision-recall curves were also used to represent high-performance models in forest regeneration mapping. The blue curve (Anthropic, AUC = 0.92) showed high precision across different recall levels. The red curve (Deforestation, AUC = 0.92) indicates balanced precision and recall. The green curve (Primary Forest, AUC = 0.88) showed a good performance, although there was a slight drop in precision at higher recall levels. The purple curve (Secondary Forest, AUC = 0.86) shows lower performance with challenges in maintaining high precision at higher recall.

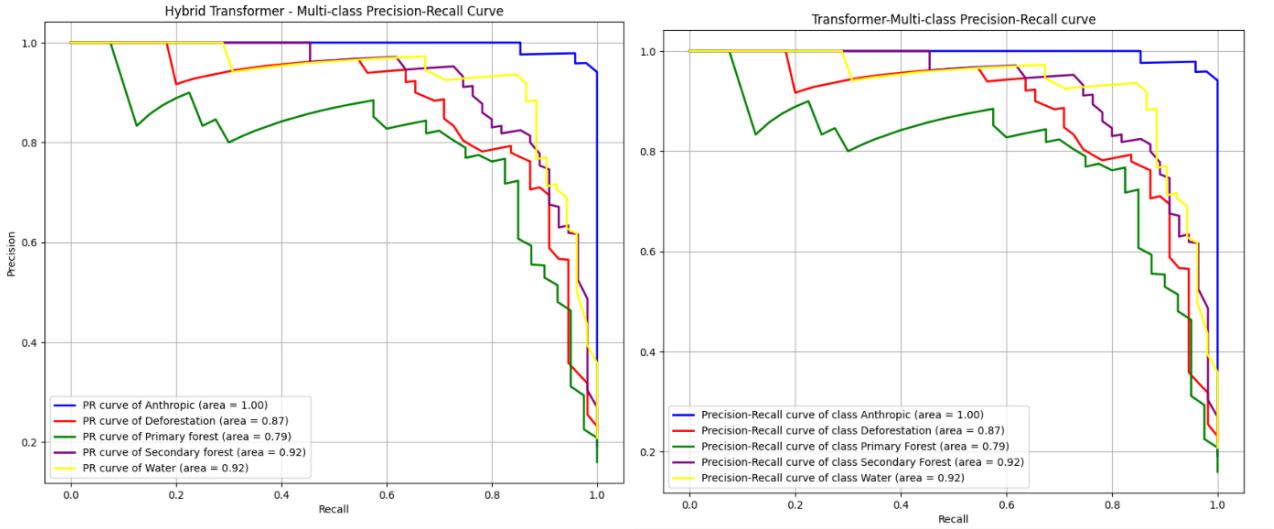


Figure 9: Precision-Recall curves for Transformer and Hybrid Transformer

II. Precision-Recall Curves

For the hybrid transformer model using testing points, the precision-recall curve for the anthropic class showed an AUC of 0.92, indicating excellent precision across various recall levels. The deforestation class had an AUC of 0.87, showing balanced precision and recall, with some decreases at higher recall levels. The

primary forest class exhibited an AUC of 0.79, reflecting a good performance with a slight drop in precision at higher recall levels. The secondary forest class showed strong performance, with an AUC of 0.92, but faced challenges in maintaining high precision at higher recall levels. The water class demonstrated excellent precision and recall with an AUC of 0.92.

In comparison, the transformer model using the testing points yielded different results. The anthropic class had a precision-recall curve AUC of 0.90, indicating high precision but slightly lower than that of the hybrid model. The deforestation class exhibited poor performance, with an AUC of 0.56, indicating a significant drop in precision at higher recall levels. The primary forest class performed very well, with an AUC of 0.97, indicating very high precision and recall. However, the secondary forest class reflects poor performance, with an AUC of 0.45. The water class showed a moderate performance, with an AUC of 0.71.

III. ROC-AUC Curves

For the hybrid transformer model using testing points, the ROC curves revealed high AUC values across all classes: 0.97, 0.95, 0.97 for primary forest, 0.96 for secondary forest, and 1.00. These values indicate excellent classification performance, particularly for water, primary forest, and anthropic classes.

In comparison, the transformer model using testing points showed lower AUC values: 0.94 for anthropic, 0.85 for deforestation, 0.95 for primary forest, 0.89 for secondary forest, and 1.00 for water. Although the performance is good, it is noticeably lower for deforestation and secondary forest classes compared to the hybrid transformer model.

IV. Comparative Evaluation of Model Performance Using Accuracy Metrics, precision-recall and ROC-AUC Analysis

The relationship between the evaluation techniques in sections 5.3.1 and 5.3.3 is evident, as both assess the performance of the Hybrid Transformer and Traditional Transformer models. Section 5.3.1 evaluates the models using traditional accuracy metrics such as precision, recall, and F1 scores based on testing sample points. In contrast, section 5.3.3 offers a more detailed assessment through ROC-AUC and precision-recall curves. These complementary approaches provide a comprehensive understanding of each model's performance across various land-cover classes.

Both sections highlight the Hybrid Transformer's strong performance in the water class. In section 5.3.1, the Hybrid Transformer achieved a precision of 97.80% and an F1 score of 97.85%, which aligns with the perfect AUC of 1.00 observed in section 5.3.3, underscoring the model's reliability in handling distinct and consistent spectral signatures.

The deforestation class presents more complexity. In section 5.3.1, the Hybrid Transformer's precision was 84.79%, with an F1 score of 81.58%, consistent with an AUC of 0.95 in section 5.3.3. These results suggest areas for refinement, particularly in classes with high spectral variability.

Similarly, for the primary forest class, section 5.3.1 reports a precision of 79.94% and an F1 score of 86.5%, which aligns with an AUC of 0.967 in section 5.3.3, indicating robust performance. Challenges in classifying secondary forests are evident in both sections, where precision, F1 scores, and AUC values suggest the need for further improvement. Overall, the findings in both sections align, reinforcing that while the Hybrid Transformer is generally reliable, there is room for enhancement, especially in more challenging land-cover classes.

5.5 Qualitative analysis of the results

I. Predicted maps made by Hybrid and Transformer Models

Figure 10 provides a visual comparison between the reference data (a) and the predictions made by two different models: the transformer (c) and the Hybrid Transformer (b). Both models utilised temporal information and classified the data into five distinct categories: anthropic, deforestation, primary forest, secondary forest, and water. A close inspection reveals that the Hybrid Transformer model more accurately maps the regenerated areas than the traditional transformer model by capturing finer details and showing fewer misclassifications, particularly in areas transitioning from deforestation to secondary forests.

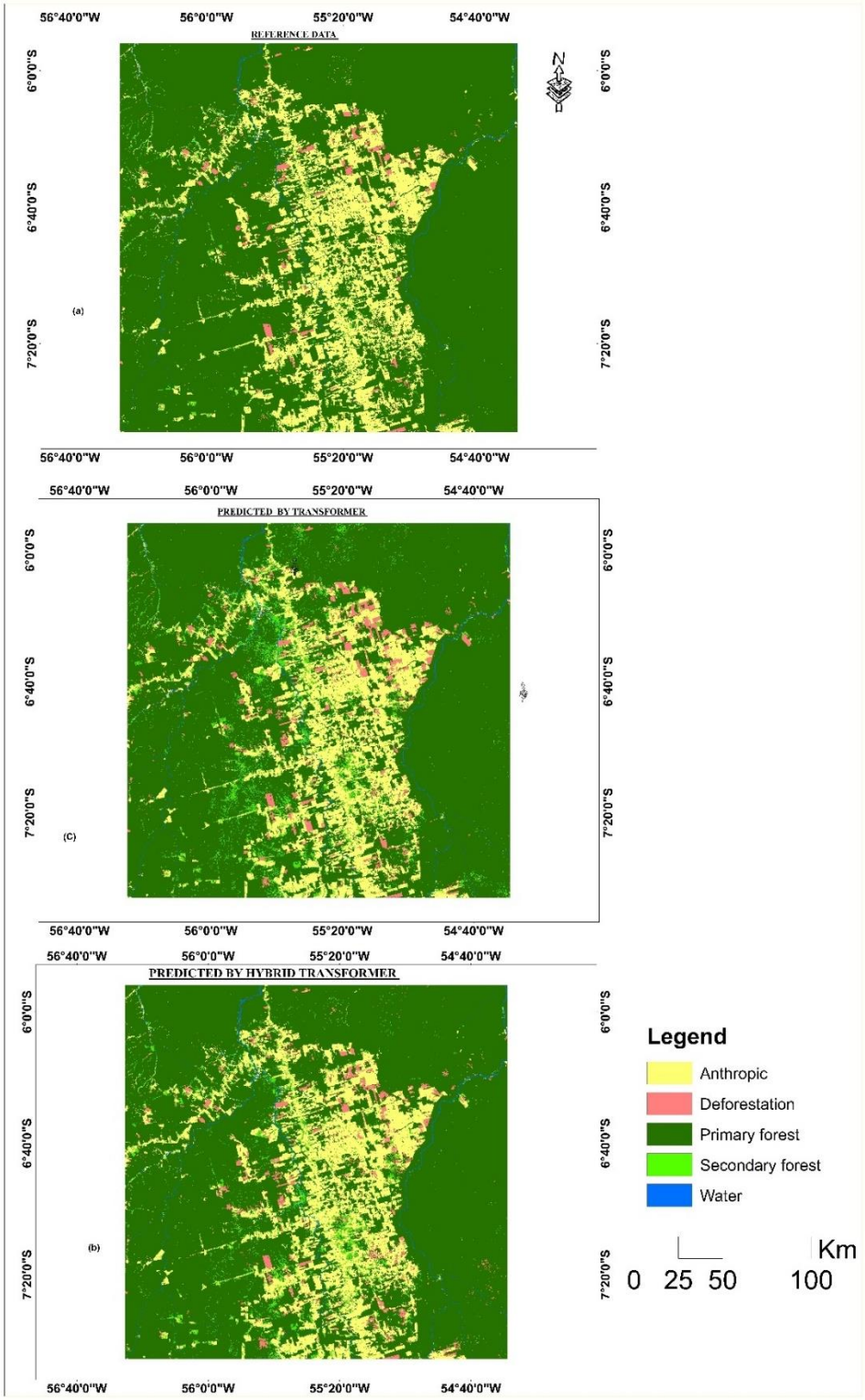


Figure 10: Depicts predicted and reference maps for forest regeneration

II. Identifying areas of accurate prediction

To evaluate the classification performance of the Hybrid Transformer and Standard Transformer models, we conducted an in-depth analysis using historical Landsat images from 2012, recent images from 2021, and predicted maps for 2021. The analysis involved visually comparing reference data with predictions from both models, identifying areas of accurate and inaccurate classification, as illustrated in Figures 11 and 12.

The Hybrid Transformer model strongly aligned with reference data, particularly in regions classified as primary and secondary forests (see Figures 11 and 12). This accuracy is attributed to the model's integration of Convolutional Neural Networks (CNNs), which effectively process local temporal patterns. The CNN layers enable the model to capture subtle variations in vegetation growth, crucial for distinguishing between primary and secondary forests.

In contrast, the Standard Transformer, which lacks CNN layers and relies on processing sequential temporal dependencies, often overestimates the extent of secondary forests (see Figure 12). The Standard Transformer also faces challenges in distinguishing between secondary and primary forests due to their spectral similarities, resulting in higher rates of misclassification.

The Hybrid Transformer, by combining CNNs for local temporal analysis with transformers for sequential dependency processing, proves more ability for managing these challenges. This dual approach allows the Hybrid Transformer to provide more accurate classifications, especially in complex and transitional areas like secondary forests and deforestation zones. While both models performed reasonably well in identifying water bodies, the Hybrid Transformer had a slight edge, consistently capturing temporal variations across seasons, resulting in more precise classifications.

Therefore, the Hybrid Transformer outperforms the Standard Transformer in overall classification accuracy, particularly in complex areas like secondary forests. Its superior performance stems from the integration of CNN layers for enhanced local temporal pattern recognition and the transformer's ability to process sequential data, leading to more accurate and reliable land cover classifications.

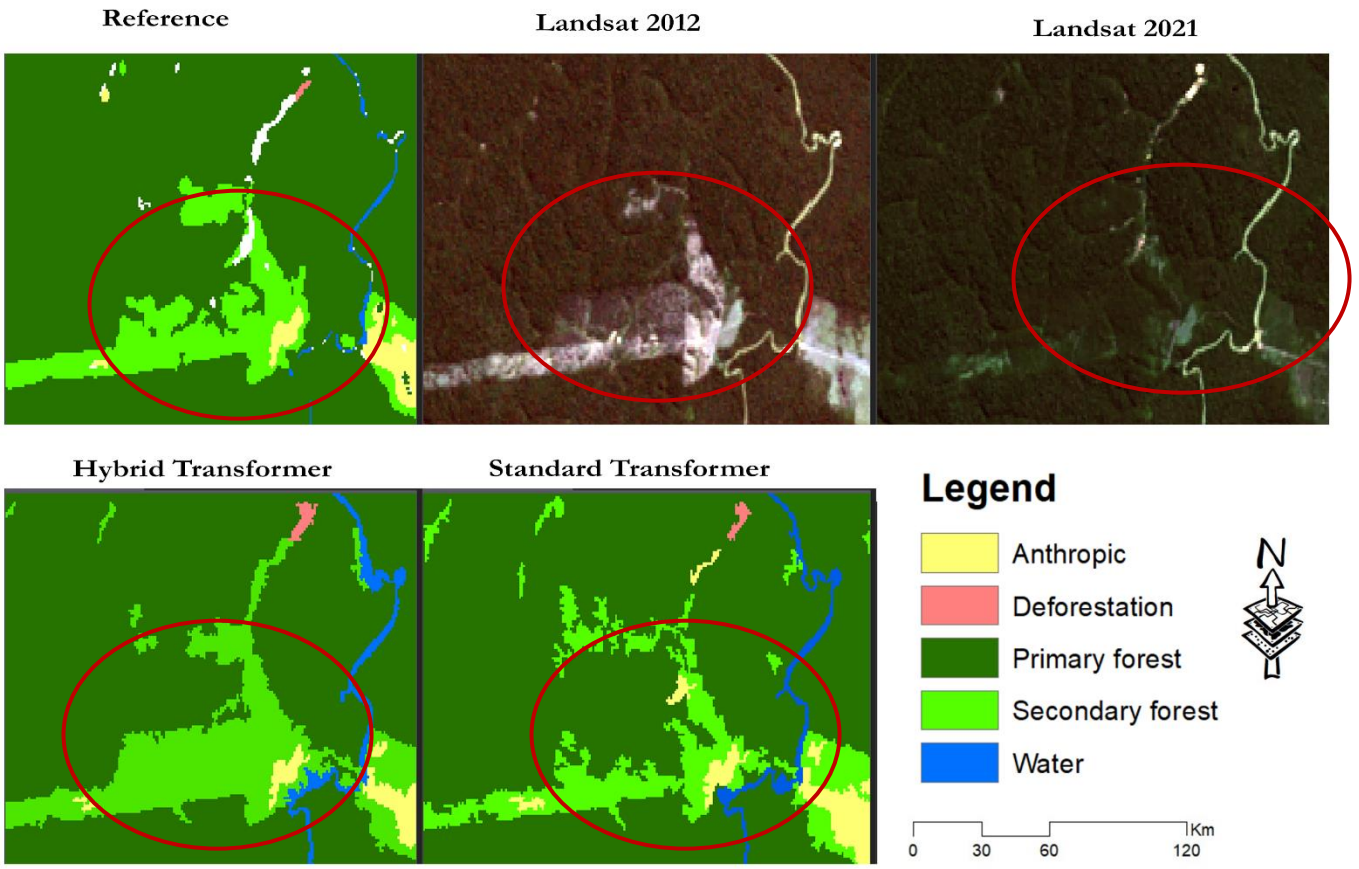


Figure 11: Comparison of model predictions with reference data and historical Landsat images (a)

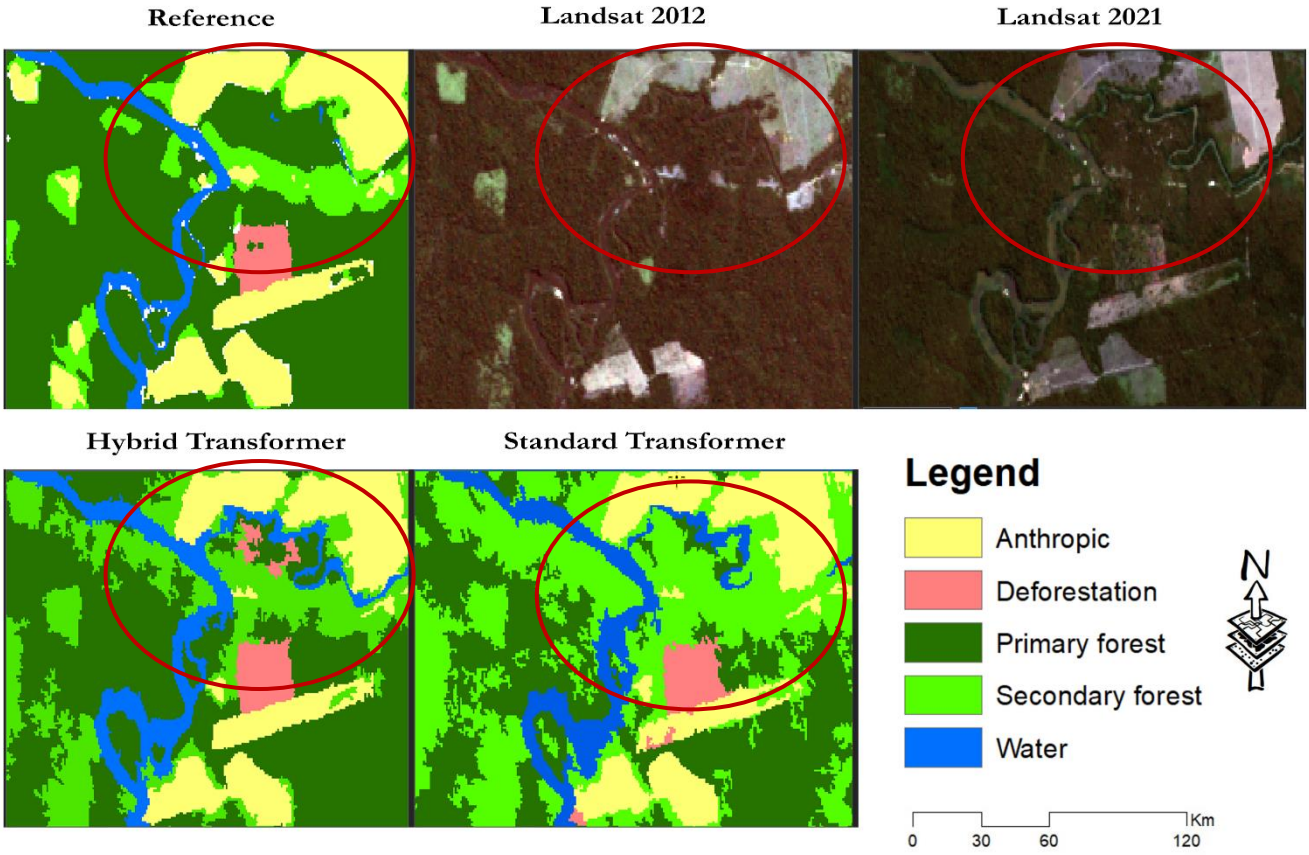


Figure 12: Comparison of model predictions with reference data and historical Landsat images (b).

I. Inaccurate predictions

Despite the overall strong performance of the Hybrid Transformer model, some regions exhibited poor predictions. Figure 9 highlights areas where both the Hybrid Transformer and Standard Transformer models did not align well with the reference data.

These inaccuracies are largely due to uncertainties in the reference data (see Figure 13 and 14), which can confuse the classifiers and negatively impact the models' performance. While the Hybrid transformer generally performed better, it still encountered challenges in areas where reference data noise was more pronounced. In these cases, the model's predictions deviated from the reference data, leading to both overestimations and underestimations.

The Standard Transformer exhibited even greater difficulties, particularly in handling complex temporal patterns. Its reliance on sequential temporal processing, without the ability to process localized temporal variations effectively, made it more prone to misclassifications. This often resulted in a mismatch between the model's predictions and the reference data, further highlighting the limitations of the Standard Transformer in accurately capturing intricate land-cover changes.

To address these issues, future work should focus on improving the quality and accuracy of reference maps, particularly by filtering out transition noise and ensuring more consistent classification of similar land-cover types. By refining the reference data, the models could achieve greater alignment with actual land cover, reducing misclassifications and improving overall predictive accuracy.

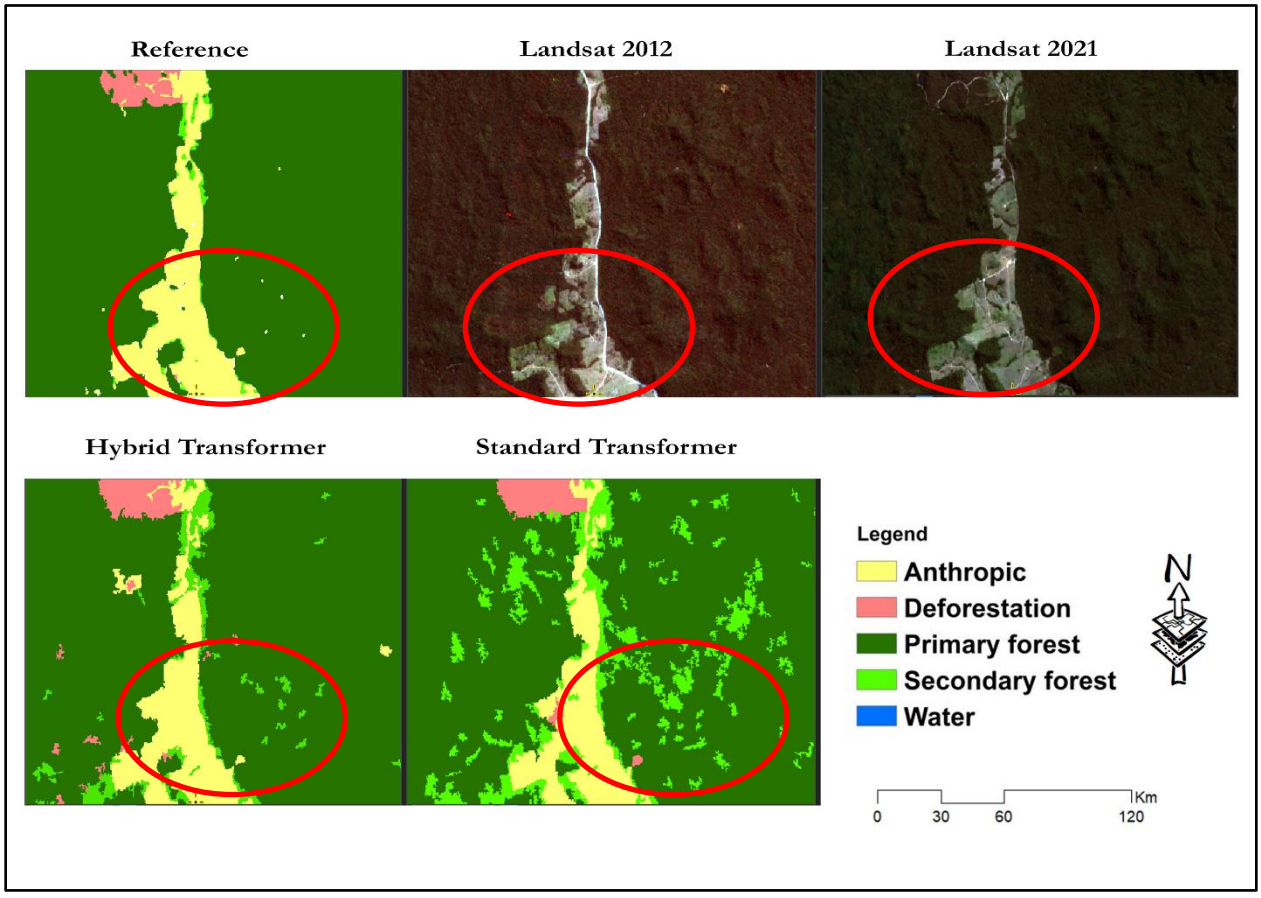


Figure 13: Visual comparison of model predictions with reference data, showing better alignment by the Hybrid Transformer over the Standard Transformer (a)

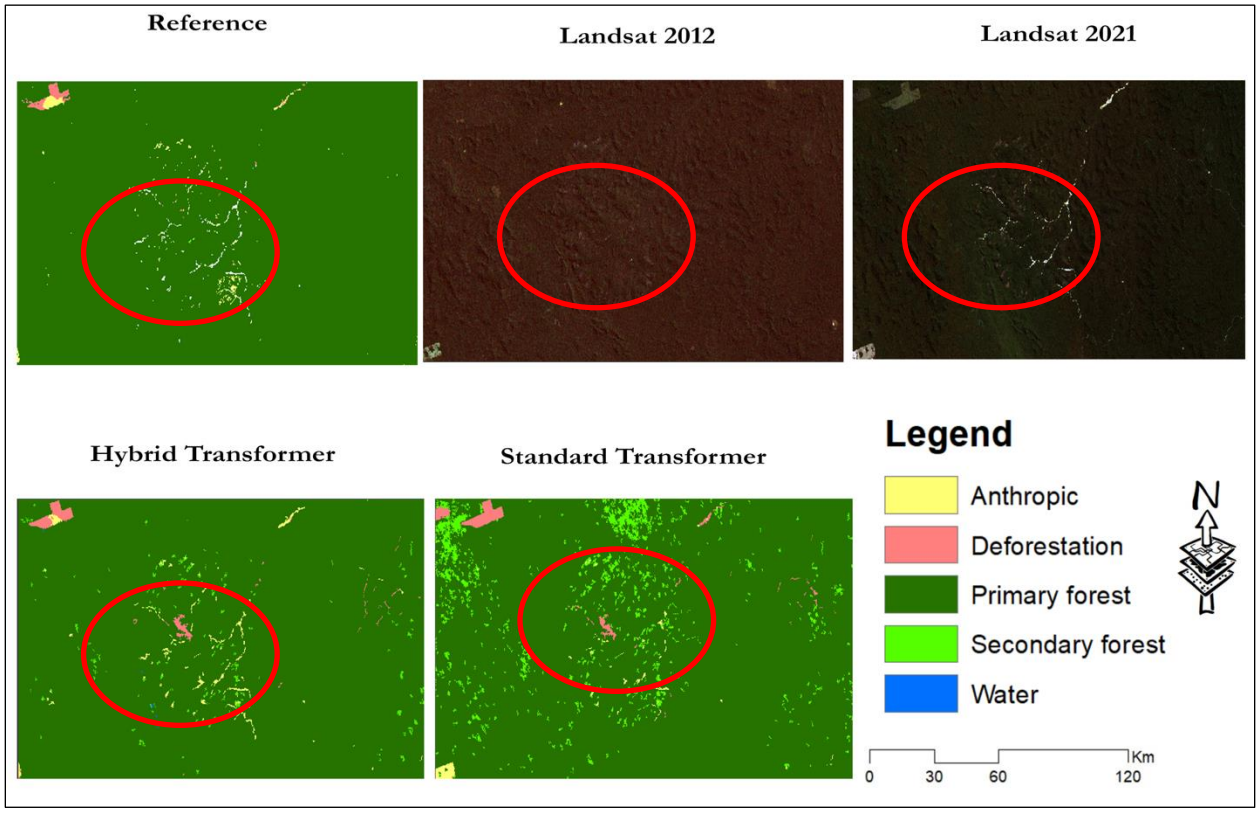


Figure 14: Visual comparison of model predictions with reference data, showing better alignment by the Hybrid Transformer over the Standard Transformer (b)

5.6 Objective 4: Quantify the Extent of Secondary Forest

The extent of the secondary forest within the study area was quantified using predictions from the best-performing model, which demonstrated a high overall accuracy of 86.38%. The results indicated that secondary forest covered 6.4% of the total study area. Other land cover classes were distributed as follows: primary forest (73.2%), water bodies (1.2%), anthropogenic zones (11.5%), and areas of deforestation (7.6%). These findings underscore the model's capability in accurately predicting and quantifying various land cover types, particularly secondary forests, providing critical insights for environmental monitoring and forest management.

However, it is important to acknowledge that these predictions come with inherent uncertainties. While the model's high accuracy indicates reliable performance, the 13.62% error rate suggests potential misclassifications that might impact the accuracy of predicted secondary forest and other land cover extents. This is especially relevant in areas where land cover classes share similar spectral characteristics, like primary and secondary forests or deforested areas, where the model could overestimate or underestimate actual reference data.

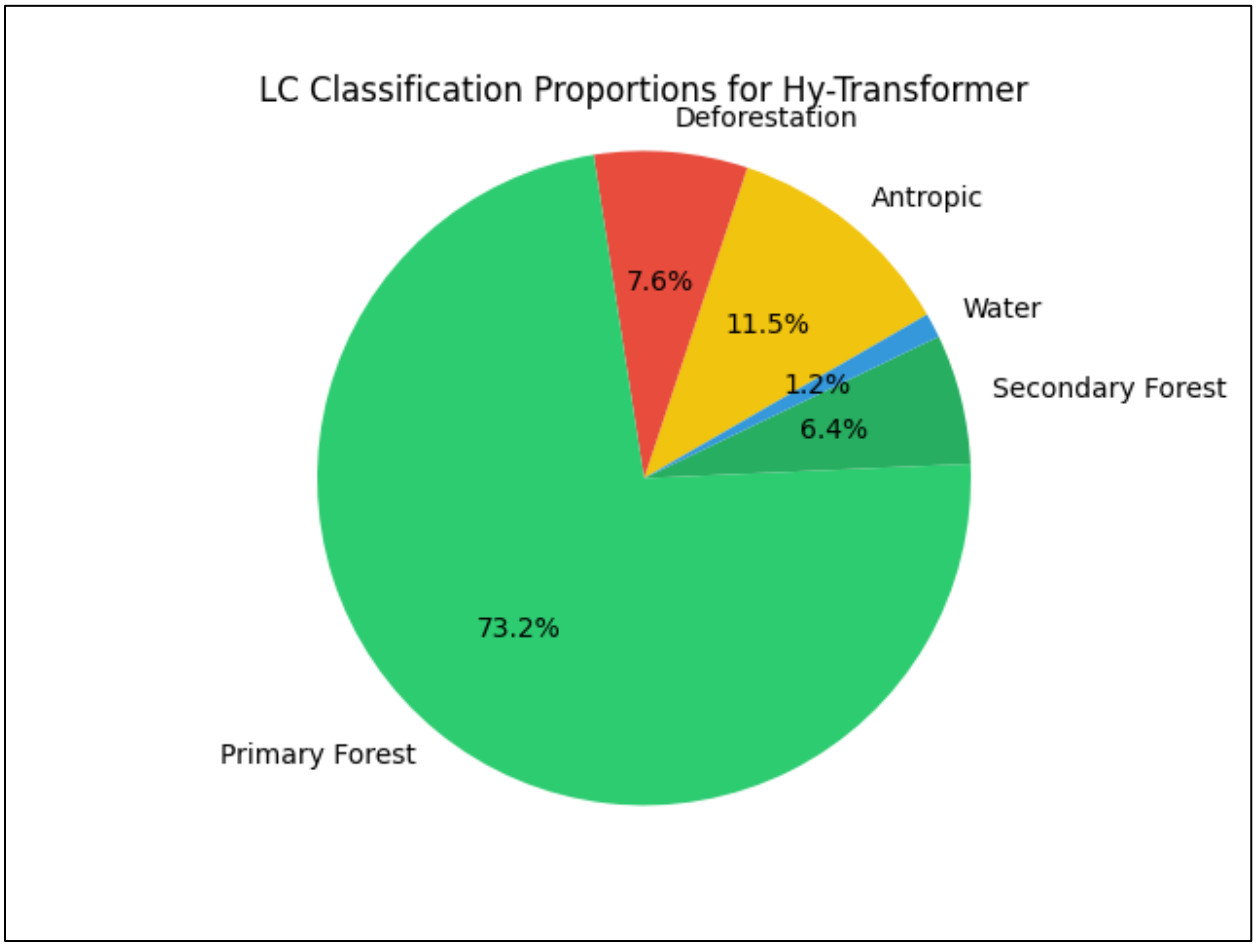


Figure 15: **:** Extent of secondary forest compared with other classes

CHAPTER 6: DISCUSSIONS AND IMPLICATIONS

DISCUSSIONS

Research Question 1: Comparative Performance of Transformer and Hybrid Transformer Models.

Here, the main objective was to compare the performance of two deep-learning architectures, the transformer and the hybrid transformer, to determine which is more suitable for accurately differentiating regeneration areas from other classes in the Amazon rainforest.

The results indicated that the hybrid transformer consistently outperformed the traditional transformer across all evaluated metrics, particularly in its ability to integrate temporal and spectral information effectively. On the one hand, the hybrid transformer achieved an overall accuracy of 86.36%, compared to 85.48% for the traditional transformer. This performance advantage is largely attributed to the integration of CNN layers within the hybrid model, which allowed it to capture local temporal patterns more efficiently.

On the other hand, the traditional transformer indicated its strength in capturing broader temporal dependencies without the need for detailed local pattern extraction, which is why its accuracy is low compared to a hybrid transformer. These findings align with previous research conducted by Li et al. (2022); Liang et al. (2023) and Ma et al. (2022), which has shown that hybrid models often yield better results across various domains by combining different modelling techniques.

Research Question 2: Effectiveness of Handling Time-Series Data Over Different Time Steps

The purpose of this question was to assess how well the models handle time-series data over different time steps and how this affects their accuracy and representation of regeneration in the Amazon rainforest.

Here, the findings revealed that the hybrid transformer is highly effective in utilizing long time-series data, leading to more accurate predictions of forest regeneration in the Amazon rainforest. As the period increased from single-time steps (t1) to ten-time steps (t10), the model's performance consistently improved, indicating its ability to capture long-term trends and subtle changes over time. This capability to model complex temporal dynamics is essential for accurately representing the gradual and nuanced processes involved in forest regeneration.

In terms of performance metrics, the hybrid transformer achieved an overall accuracy of 86.36%, with precision, recall, and F1-scores all at 0.86. In contrast, the standard transformer recorded an overall accuracy of 85.48%, with a precision of 0.86, recall of 0.85, and F1-score of 0.85. This improvement is attributed to the hybrid model's effective integration of both temporal and spectral information across different periods, significantly enhancing its prediction accuracy.

The reason the hybrid transformer outperforms the standard transformer is due to the integration of Convolutional Neural Network (CNN) layers within its architecture. These CNN layers enable the model to capture local temporal patterns more effectively, while the transformer layers focus on processing and understanding local temporal dependencies. This approach aligns with previous studies by Li et al. (2020) that emphasize the advantages of combining CNN and transformer technologies for time-series analysis.

Although the hybrid transformer is very effective for mapping forest regeneration, it performs poorly when dealing with a small number of images or short time lengths. For instance, when processing single time steps (t1), the hybrid transformer, while still better than the traditional transformer, showed a noticeable decline in performance. This decline is attributed to the lack of adequate temporal dependencies when analyzing changes over time. Consequently, the hybrid transformer achieved a precision, recall, and F1-score of 0.74, with an overall accuracy of 73.88%, compared to the traditional transformer's precision, recall, and F1-score of 0.71 and an overall accuracy of 70.80%.

Despite the clear advantages of the hybrid transformer in handling long time-series data, there are significant challenges related to its computational demands. The integration of CNN and transformer layers, especially over extensive time steps, increases the computational burden, which can be a major obstacle in resource-limited settings. This high computational requirement may restrict the model's application in real-time environments or in areas with limited technological infrastructure, where rapid data processing is essential.

Research Question 3: The Best-Performing Model in Mapping Regenerated Areas

The purpose of this question was to evaluate the performance of the hybrid transformer, identified as the best model from previous analyses (RQ1 and RQ2), in mapping secondary forests and other land-cover classes within the Amazon rainforest. The findings from our study demonstrate that the hybrid transformer model outperforms the traditional transformer in this task, particularly in distinguishing secondary forests from other land-cover types.

The hybrid transformer's performance was assessed using both testing points and test tiles, and in both scenarios, the model consistently demonstrated superior accuracy and F1 scores compared to the traditional transformer. These results align with earlier research that emphasized the advantages of combining CNN layers with transformer architectures for complex environmental tasks (Li et al., 2022; Liang et al., 2023). In our study, the hybrid transformer achieved an overall accuracy (OA) of 86.36% with testing points, surpassing the traditional transformer's OA of 85.48%. This result aligns with earlier work by Ouyang et al. (2023),where advanced hybrid models demonstrated improved accuracy in remote sensing tasks, attributed to their ability to capture both local and global features in time-series data.

Notably, in the critical category of secondary forests, the hybrid transformer recorded a precision of 81.65%, a recall of 81.0%, and an F1 score of 81.32%, outperforming the traditional transformer. This performance improvement is consistent with earlier studies such as that of (Ouyang et al., 2023) which highlighted the hybrid transformer's superior capability in handling complex, multi-temporal data. The model's ability to integrate both CNN and transformer layers allowed it to better differentiate secondary forests, which are often difficult to classify due to their spectral similarities with other land-cover types.

Similarly, when evaluated using test tiles, the hybrid transformer continued to show superior performance with an overall accuracy of 81.087%, compared to 79.346% for the traditional transformer. In secondary forests, it achieved a precision of 64.95% and an F1 score of 66.32%, again reflecting the model's enhanced ability to process and analyze complex temporal patterns. This aligns with previous research by Zhou et al. (2019), who demonstrated that models incorporating both temporal features tend to perform better in detecting subtle changes in land cover.

The hybrid transformer excelled in well-defined classes like water and anthropic areas, achieving high precision and F1 scores. This success can be attributed to the distinct spectral signatures of these classes, which the model effectively captures through its CNN layers. However, the model's performance was more variable in complex classes like deforestation and secondary forests. These areas are challenging due to variability in spectral characteristics, driven by factors such as vegetation regrowth and human activities, issues similarly noted in earlier studies by Liu et al., (2021).

In fact, misclassification became particularly evident when secondary forests shared similar spectral features with other land-cover types, leading to overlaps and false positives. This challenge poses a limitation of the current model. To address this, one potential improvement could involve integrating additional features such as the Normalized Difference Vegetation Index (NDVI), which is crucial for differentiating vegetation from other classes. As demonstrated by Kwan et al. (2020), the integration of NDVI into models significantly improves their ability to differentiate between vegetative and non-vegetative areas, thereby reducing the risk of misclassification.

Research Question 4: To what extent is there secondary forest in the entire study area?

In addressing Research Question 4, our study quantified the extent of secondary forest within the entire study area using predictions from the best-performing model, which demonstrated a high overall accuracy of 86.38%. The analysis revealed that secondary forests covered 6.4% of the total study area. The distribution of other land cover classes was as follows: primary forest (73.2%), water bodies (1.2%), anthropogenic zones (11.5%), and areas of deforestation (7.6%). These findings highlight the effectiveness of the hybrid transformer model in accurately identifying and quantifying various land cover classes, offering valuable insights for environmental monitoring and forest management.

The accurate quantification of secondary forest cover is essential for understanding forest recovery dynamics and guiding reforestation efforts. Secondary forests play a crucial role in the ecological landscape, as they represent areas where natural regeneration is occurring, contributing to biodiversity conservation, carbon sequestration, and the maintenance of hydrological cycles. By providing a precise estimate of secondary forest extent, our model supports the evaluation of the success of natural regeneration and reforestation initiatives, which is vital for the effective monitoring of forest dynamics.

The ability to accurately map and quantify secondary forests also informs the development of conservation strategies and policies. Understanding the spatial distribution and extent of secondary forests allows policymakers and conservationists to prioritize areas for protection, restoration, or further intervention. For instance, areas identified as secondary forests may require different management approaches compared to primary forests or deforested regions, depending on their stage of regeneration.

IMPLICATIONS

The high performance of the hybrid transformer model in distinguishing secondary forests from other land-cover classes highlights its potential as a valuable tool for environmental monitoring, forest

management, and sustainability efforts. Its accuracy in mapping regenerated areas supports the continuous monitoring of forest dynamics in the Amazon rainforest, which is essential for informed decision-making regarding ecosystem protection and restoration.

In the Amazon, where forest recovery is crucial for biodiversity, carbon sequestration, and hydrological cycles, the hybrid transformer model's data can be instrumental in evaluating reforestation projects and guiding conservation strategies. Its ability to differentiate between various stages of forest regeneration enhances conservation planning by identifying areas that are successfully regenerating and those requiring targeted intervention or changes in land management practices.

Moreover, the model's application in mapping regenerated areas has broader implications for sustainable forest management. By providing accurate and timely data on forest recovery, the model can inform policy decisions aimed at achieving sustainability goals. This is particularly relevant for climate change mitigation, where understanding the extent of secondary forests is vital for assessing carbon sequestration potential and setting realistic carbon offset targets.

The hybrid transformer's consistent performance across various land-cover classes suggests it could be a valuable tool for monitoring other critical ecosystems. Its ability to provide detailed and accurate mapping of complex landscapes makes it a powerful asset for global conservation efforts, particularly in regions experiencing rapid environmental changes where timely and precise data is crucial.

CHAPTER 7: CONCLUSION AND RECOMMENDATION

7.1 Conclusion

This study aimed to evaluate the effectiveness of transformer-based deep learning architectures, specifically the hybrid transformer model, in distinguishing regeneration areas from other land-cover classes in the Amazon rainforest. The research focused on comparing the traditional transformer model and the hybrid transformer model and assessing their performance across various time lengths (t1, t3, t5, and t10) to predict forest regeneration accurately.

The hybrid transformer model demonstrated superior performance compared with the traditional transformer model. The integration of CNN layers in the hybrid model allowed for more efficient capture of local temporal patterns, enhancing feature extraction and overall accuracy. The hybrid transformer consistently outperformed the traditional transformer across all metrics, achieving an overall accuracy of 86.36% compared with 85.48% for the transformer. The precision, recall, and F1-score metrics also showed superior performance for the hybrid transformer, particularly at longer time lengths.

The study also highlighted the capability of the hybrid transformer to handle extended time-series data, providing a more accurate and realistic representation of regeneration dynamics. The ability of this model to integrate temporal information across multiple time lengths significantly improved its predictive performance, in agreement with previous studies that combined CNNs and transformers for better results in time-series analysis tasks.

In terms of mapping regenerated areas, the hybrid transformer's higher performance metrics indicate its effectiveness. The model achieved a precision of 64.95%, recall of 67.73%, and F1 score of 66.32% for secondary forest mapping, supporting its application in large-scale environmental monitoring and deforestation assessment. Accurate quantification of secondary forest cover is crucial for understanding forest recovery and for guiding reforestation efforts, particularly in the Amazon rainforest.

The extent of secondary forest within the study area was quantified using the hybrid transformer model, which revealed that secondary forest covers 6.4% of the total study area. Accurate quantification is vital for assessing the success of natural regeneration and reforestation initiatives, providing essential data for the effective monitoring of forest dynamics and formulating conservation strategies.

7.2 Recommendations for Future Research

Exploration of other deep-learning architectures: Future research should focus on models like CNN-transformer and LSTM networks that explicitly incorporate spatial dimensions, improving the accuracy and efficiency of forest regeneration mapping.

Extended Time Series Data Analysis: Future studies should continue to explore the potential of hybrid transformer models for handling extended time-series data. This approach provides a more realistic representation of forest dynamics and can significantly improve the predictive accuracy.

Integration with Additional Data Sources: To improve the model's predictive accuracy and robustness, combining satellite imagery with features such as Digital Elevation Model (DEM) data is recommended. This integration can help to reduce the effects of topography on data acquisition, leading to more accurate predictions.

Inclusion of Comprehensive Field Surveys: Owing to the current limitations imposed by inaccuracies in the reference data, it is advised that future research includes comprehensive field surveys. These surveys should aim to collect more precise and reliable data, which will help validate and refine the model's predictions, ensuring better alignment with real-world conditions.

7.3 Ethical Considerations

This study strictly followed ethical standards in accordance with the Research Ethics Policy of the University of Twente. The following points were considered in this study.

Data Privacy and Security: This study utilized open-source datasets in which satellite images such as Landsat 5 and 8 were freely and publicly accessible. Data acquisition methods were scrutinized to ensure that the data were representative, unbiased, and comprehensive, encompassing all relevant variables, without any undue exclusions. Efforts were made to verify the accuracy and completeness of the data through rigorous pre-processing steps, including data cleaning, handling missing values, and removing outliers that could distort the performance of the model.

Moreover, the ethical use of data sources was considered, ensuring that all data were obtained and utilized in compliance with legal and ethical standards, including obtaining the necessary permissions and ensuring the privacy and confidentiality of any sensitive information. Regular audits and validation checks were performed to maintain data integrity and reliability throughout the data-handling process.

Acknowledgement of Previous Research: All previous research related to this study has been properly acknowledged and referenced in the body of the text and the list of references. This aspect respects the contributions of other researchers and maintains their academic integrity. By recognizing the work of others, this study upholds the ethical standards of academic honesty and integrity, fostering a culture of respect and collaboration within the scientific community. This approach ensured that the study was conducted with a high level of ethical responsibility, recognizing the value of both the data and the scholarly work preceding it.

Use of AI Tools: AI tools were used to improve text clarity, grammar, and formatting, and to assist in code development. Their use was limited to improving language precision and coding efficiency, without influencing research findings or content generation. All AI applications adhered to ethical guidelines, ensuring academic integrity.

LIST OF REFERENCES

Alvarez, S. A. (2002). *An exact analytical relation among recall, precision, and classification accuracy in information retrieval*.

Andreieva, V., & Shvai, N. (2021). Generalization of Cross-Entropy Loss Function for Image Classification. *Mohyla Mathematical Journal*, *3*, 3–10. https://doi.org/10.18523/2617-7080320203-10

Ball, J. G. C., Petrova, K., Coomes, D. A., & Flaxman, S. (2022). Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation. *Methods in Ecology and Evolution*, *13*(11), 2622–2634. https://doi.org/10.1111/2041-210X.13953

Bharghavi, Y. M., Pavan Kumar, C. S., Lakshmi, Y. H., & Sri Vyshnavi, K. P. (2023). Agriculture Land Image Classification Using Machine Learning Algorithms and Deep Learning Techniques. *International Conference on Frontiers in Intelligent Computing: Theory and Applications*, *370*, 235–246. https://doi.org/10.1007/978-981-99-6702-5_19

Brando, P. M., Balch, J. K., Nepstad, D. C., Morton, D. C., Putz, F. E., Coe, M. T., Silvério, D., Macedo, M. N., Davidson, E. A., Nóbrega, C. C., Alencar, A., & Soares-Filho, B. S. (2014). Abrupt increases in Amazonian tree mortality due to drought-fire interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(17), 6347–6352. https://doi.org/10.1073/pnas.1305499111

Carvalho, R., Adami, M., Amaral, S., Bezerra, F. G., & de Aguiar, A. P. D. (2019). Changes in secondary vegetation dynamics in a context of decreasing deforestation rates in Pará Brazilian Amazon. *Applied Geography*, *106*(March), 40–49. https://doi.org/10.1016/j.apgeog.2019.03.001

Chandra, N., Ahuja, L., Khatri, S. K., & Monga, H. (2021). Utilizing Gated Recurrent Units to Retain Long Term Dependencies with Recurrent Neural Network in Text Classification. *Journal of Information Systems and Telecommunication*, *9*(34), 89–102. https://doi.org/10.29252/jist.9.34.89

Chazdon, R. L., Broadbent, E. N., Rozendaal, D. M. A., Bongers, F., Zambrano, A. M. A., Aide, T. M., Balvanera, P., Becknell, J. M., Boukili, V., Brancalion, P. H. S., Craven, D., Almeida-Cortez, J. S., Cabral, G. A. L., De Jong, B., Denslow, J. S., Dent, D. H., DeWalt, S. J., Dupuy, J. M., Durán, S. M., … Poorter, L. (2016). Carbon sequestration potential of second-growth forest regeneration in the Latin American tropics. *Science Advances*, *2*(5). https://doi.org/10.1126/SCIADV.1501639

Chuvieco, E., Riaño, D., Aguado, I., & Cocero, D. (2002). Estimation of fuel moisture content from multitemporal analysis of Landsat Thematic Mapper reflectance data: Applications in fire danger assessment. *International Journal of Remote Sensing*, *23*(11), 2145–2162. https://doi.org/10.1080/01431160110069818

Coe, M. T., Marthews, T. R., Costa, M. H., Galbraith, D. R., Greenglass, N. L., Imbuzeiro, H. M. A., Levine, N. M., Malhi, Y., Moorcroft, P. R., Muza, M. N., Powell, T. L., Saleska, S. R., Solorzano, L. A., & Wang, J. (2013). Deforestation and climate feedbacks threaten the ecological integrity of south-southeastern Amazonia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1619). https://doi.org/10.1098/rstb.2012.0155

Dieterle, G. (2010). *Sustaining the World's Forests: Managing Competing Demands for a Vital Resource – The Role of the World Bank*. 9–32. https://doi.org/10.1007/978-90-481-3301-7_2

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR 2021 - 9th International Conference on Learning Representations*. https://arxiv.org/abs/2010.11929v2

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale. *ICLR 2021 - 9th International Conference on Learning Representations*.

Hirschmugl, M., Deutscher, J., Sobe, C., Bouvet, A., Mermoz, S., & Schardt, M. (2020). Use of SAR and Optical Time Series for Tropical Forest Disturbance Mapping. *Remote Sensing*, *12*(4). https://doi.org/10.3390/RS12040727

Jin, J., & Wang, Q. (2019). Selection of Informative Spectral Bands for PLS Models to Estimate Foliar Chlorophyll Content Using Hyperspectral Reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(5), 3064–3072. https://doi.org/10.1109/TGRS.2018.2880193

Katila, P., Galloway, G., Jong, W. de, & Pacheco, P. (2014). *Forests under pressure*.

Kwan, C., Gribben, D., Ayhan, B., Li, J., Bernabe, S., & Plaza, A. (2020). An Accurate Vegetation and Non-Vegetation Differentiation Approach Based on Land Cover Classification. *Remote Sensing*, *12*(23), 1–29. https://doi.org/10.3390/RS12233880

Lakshmi, P. (2015). Prevention and Usage of Forest. *Journal of Pharmacognosy and Phytochemistry*.

Li, S., Wu, C., & Xiong, N. (2022). Hybrid Architecture Based on CNN and Transformer for Strip Steel Surface Defect Classification. *Electronics*, *11*(8). https://doi.org/10.3390/ELECTRONICS11081200

Liang, S., Hua, Z., & Li, J. (2023). Hybrid transformer-CNN networks using superpixel segmentation for remote sensing building change detection. *International Journal of Remote Sensing*, *44*(8), 2754–2780. https://doi.org/10.1080/01431161.2023.2208711

Loganathan, A., Koushmitha, S., & Arun, Y. N. K. (2021). Land Use/Land Cover Classification Using Machine Learning and Deep Learning Algorithms for EuroSAT Dataset - A Review. *International Conference on Intelligent Systems Design and Applications*, *418 LNNS*, 1363–1374. https://doi.org/10.1007/978-3-030-96308-8_126

Lucas, R. M., Honzák, M., Curran, P. J., Foody, G. M., Milne, R., Brown, T., & Amaral, S. (2000). Mapping the regional extent of tropical forest regeneration stages in the Brazilian Legal Amazon using NOAA AVHRR data. *International Journal of Remote Sensing*, *21*(15), 2855–2881. https://doi.org/10.1080/01431160050121285

Nepstad, D., Mcgrath, D., Stickler, C., Alencar, A., Azevedo, A., Swette, B., Bezerra, T., Digiano, M., Shimada, J., Seroa, R., Armijo, E., Castello, L., Brando, P., Hansen, M. C., Mcgrath-horn, M., Carvalho, O., & Hess, L. (2014). and Soy Supply Chains. *Science*, *344*(6188), 1118–1123.

Noh, S. H. (2021). Analysis of gradient vanishing of RNNs and performance comparison. *Information (Switzerland)*, *12*(11). https://doi.org/10.3390/info12110442

Novakovic, J., Veljovic, A., Ilić, S., Papic, Ž., & Milica, T. (2017). Evaluation of Classification Models in Machine Learning. *Theory and Applications of Mathematics & Computer Science*.

Ouyang, E., Li, B., Hu, W., Zhang, G., Zhao, L., & Wu, J. (2023). When Multigranularity Meets Spatial–Spectral Attention: A Hybrid Transformer for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *61*. https://doi.org/10.1109/TGRS.2023.3242978

Park, J., Yi, D., & Ji, S. (2020). Analysis of recurrent neural network and predictions. *Symmetry*, *12*(4), 615. https://doi.org/10.3390/SYM12040615

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D. C., Papale, D., Rammig, A., Smith, P., Thonicke, K., Van Der Velde, M., Vicca, S., Walz, A., & Wattenbach, M. (2013). Climate extremes and the carbon cycle. *Nature*, *500*(7462), 287–295. https://doi.org/10.1038/nature12350

Shih, S. Y., Sun, F. K., & Lee, H. yi. (2018). Temporal pattern attention for multivariate time series forecasting. *Machine-Mediated Learning*, *108*(8–9), 1421–1441. https://doi.org/10.1007/S10994-019-05815-0

Suding, P. H., Coca, A., Navarrete, C., Arango, D., Jarvis, A., Watkins, G. G., & Reymondin, L. (2014). *Potential Impact of Road Projects on Habitat Loss and Greenhouse Gas Emissions in Guyana from 2012 to 2022*. https://doi.org/10.18235/0009189

Sukumar, R. (2008). Forest Research for the 21st Century. *Science*, *320*(5882), 1395. https://doi.org/10.1126/SCIENCE.1160329

Wang, L., Qu, J. J., Hao, X., & Zhu, Q. (2008). Sensitivity studies of the moisture effects on MODIS SWIR reflectance and vegetation water indices. *International Journal of Remote Sensing*, *29*(24), 7065–7075. https://doi.org/10.1080/01431160802226034

Yan, C., Fan, X., Fan, J., Yu, L., Wang, N., Chen, L., & Li, X. (2023). HyFormer: Hybrid Transformer and CNN for Pixel-Level Multispectral Image Land Cover Classification. *International Journal of Environmental Research and Public Health*, *20*(4). https://doi.org/10.3390/ijerph20043059

Appendix 1: Confusion matrices

Hybrid transformer for single time length

	Anthropic	Deforestation	Primary forest	Secondary forest	Water	Total	Precision (user)
Anthropic	**733**	169	14	82	2	1000	**77.57**
Deforestation	154	**536**	140	163	7	1000	**60.56**
Primary forest	0	12	**838**	149	1	1000	**72.49**
Secondary forest	55	144	157	**631**	13	1000	**60.97**
Water	3	24	7	10	**956**	1000	**97.65**
Total	945	885	1156	1035	979	5000	
Recall (producer)	**73.3**	**53.6**	**83.8**	**63.1**	**95.6**		
Overall Accuracy							**73.88**

Hybrid transformer for Three-time length

	Anthropic	Deforestation	Primary forest	Secondary forest	Water	Total	Precision (user)
Anthropic	**880**	55	15	40	10	1000	76.92

Deforestation	75	**705**	172	47	1	1000	**84.23**
Primary forest	6	17	**907**	69	1	1000	**70.31**
Secondary forest	182	58	187	**563**	10	1000	**77.66**
Water	1	2	9	6	**982**	1000	**97.81**
Total	1144	837	1290	725	1004	5000	
Recall (producer)	**88**	**70.5**	**90.7**	**56.3**	**98.2**		
Overall Accuracy							**80.74**

Hybrid transfomer for Five-time length s

	Anthropic	Deforestation	Primary forest	Secondary forest	Water	Total	Precision (user)
Anthropic	**825**	59	16	100	0	1000	**91.16**
Deforestation	37	**712**	178	73	0	1000	**86.3**
Primary forest	2	21	**894**	82	1	1000	**74.44**
Secondary forest	40	32	109	**813**	6	1000	**74.59**
Water	1	1	4	22	**972**	1000	**99.28**
Total	905	825	1201	1090	979	5000	
Recall (producer)	**82.5**	**71.2**	**89.4**	**81.3**	**97.2**		

Overall Accuracy	**84.32**

Transfomer for single time length

	Anthropic	Deforestation	Primary forest	Secondary forest	Water	Total	Precision (user)
Anthropic	**674**	240	13	68	5	1000	**82.1**
Deforestation	116	**565**	142	145	32	1000	**54.27**
Primary forest	0	16	**848**	133	3	1000	**72.23**
Secondary forest	29	216	165	**575**	15	1000	**61.96**
Water	2	4	6	7	**981**	1000	**94.69**
Total	821	1041	1174	928	1036	5000	
Recall (producer)	**67.4**	**56.5**	**84.8**	**57.5**	**98.1**		
Overall Accuracy							**72.86**

Transfomer for Three-time length s

	Anthropic	Deforestation	Primary forest	Secondary forest	Water	Total	Precision (user)
Anthropic	**787**	106	21	82	4	1000	91.62
Deforestation	15	**722**	159	102	2	1000	**79.87**

Primary forest	2	15	**861**	121	1	1000	**72.66**
Secondary forest	48	57	135	**747**	13	1000	**70.41**
Water	7	4	9	9	**971**	1000	**97.98**
Total	859	904	1185	1061	991	5000	
Recall (producer)	**78.7**	**72.2**	**86.1**	**74.7**	**97.1**		
Overall Accuracy							**82.06**

Transfomer for Five-time length s

	Anthropic	Deforestation	Primary forest	Secondary forest	Water	Total	Precision (user)
Anthropic	**866**	63	13	55	3	1000	85.49
Deforestation	40	**752**	174	34	0	1000	**84.4**
Primary forest	3	24	**928**	44	1	1000	**71.94**
Secondary forest	101	51	168	**673**	7	1000	**81.18**
Water	3	1	7	23	**966**	1000	**98.87**
Total	1013	891	1290	829	977	5000	
Recall (producer)	**86.6**	**75.2**	**92.8**	**67.3**	**96.6**		
Overall Accuracy							**83.7**

Link for code : https://github.com/Fulgence8595/Hybrid_transfomer-