1. INTRODUCTION

Tropical forests, often described as the "lungs of the planet," are critical reservoirs of biodiversity and essential regulators of the global climate (FAO and UNEP, 2020). Among the nations gifted with this vital resource, Colombia stands out as one of the most biodiverse countries per square kilometre, hosting nearly 10% of the world's species within its varied ecosystems, where nearly 1,900 species of birds, 67 mammals, and more than 130,000 plant species are found (Procolombia, 2023). The Colombian Amazon, with a size of 483,119 km2 and covering 42% of the national territory, is a cornerstone of this natural heritage, containing an extraordinary wealth of life, and serving as the ancestral home for numerous Indigenous communities (SIAT, 2023).

Despite their ecological and cultural significance, these forests are under severe and escalating threat. According to data from Colombia's Institute of Hydrology, Meteorology and Environmental Studies (IDEAM, 2024) the nation has been grappling with alarming deforestation rates, with 123,517 hectares lost in 2022 alone, an area roughly equivalent to the size of Los Angeles or 1,200 square kilometres. This clearing, driven by the expansion of agriculture, cattle ranching, illegal mining, and illicit crops, not only decimates biodiversity but also jeopardises the stability of the Amazonian biome and its role in hemispheric climate systems (Molina, 2024). Effective and timely monitoring of these vast and often inaccessible regions is therefore not just a national priority but a global imperative.

The conventional approach to forest monitoring, reliant on map interpretation, manual ground surveys and ancillary data analysis, is incapable of addressing the scale and urgency of this crisis. Such methods are labour-intensive, and prohibitively expensive (Alshehri, Ouadou and Scott, 2024). In response, the scientific and conservation communities have turned to remote sensing, using satellite imagery to observe and quantify land cover change across broad regions (James, 2025). The paradigm shifted in 2008 when the entire Landsat archive was made available to the public at no cost, a move that unleashed the potential of the data for science and monitoring (Wulder et al., 2016). This was complemented by the Copernicus programme's Sentinel satellites, which were designed specifically to provide systematic, operational data for services like environmental protection (Drusch et al., 2012). The resulting petabyte-scale archives necessitated new cloud-based platforms like Google Earth Engine to make analysis possible (Gorelick et al., 2017). Consequently, this data now underpins global-scale research, from mapping worldwide deforestation (Hansen et al., 2013) to tracking long-term changes in surface water (Pekel et al., 2016), solidifying its role as an essential tool for understanding our planet.

However, the availability of data is only one part of the problem; the primary challenge lies in its interpretation. Early remote-sensing techniques often relied on pixel-based analyses of spectral indices like the Normalized Difference Vegetation Index (NDVI) to assess vegetation health (James, 2025). While useful for assessing vegetation health, these pixel-based indices are fundamentally limited. Their spectral calculations are easily distorted by atmospheric effects like haze and aerosols, which can make a healthy forest appear less vigorous (Rech, 2023). Furthermore, as they analyse each pixel in isolation, they cannot leverage spatial context, often leading to misclassifications, such as confusing a patch of dark, moist soil with the shaded canopy of a dense forest.

The subsequent rise of machine learning (ML) models, such as Random Forest (RF) and Support Vector Machines (SVMs), offered an improvement by classifying pixels based on a range of spectral and textural features (Pal, 2005; Mountrakis, Im and Ogole, 2011; Belgiu and Drăguţ, 2016). These models, however, depend heavily on a process of manual "feature engineering," which is not only time-consuming but also requires substantial domain expertise to select the optimal features for a given task.

The limitations of traditional ML have been largely superseded by the paradigm shift brought about by deep learning (DL), particularly the application of Convolutional Neural Networks (CNNs) to image analysis (Lecun et al., 1998; Zhu et al., 2017). Unlike their predecessors, deep learning models, particularly CNNs, have revolutionised the field by automatically learning hierarchical feature representations directly from raw image data, thus eliminating the feature engineering challenge and enabling the detection of intricate spatial patterns frequently achieving unprecedented levels of accuracy. (Zhang et al., 2016; Zhu et al., 2017).

Within this domain, deep learning models excel at semantic segmentation, the task of classifying every pixel in an image to produce a detailed, wall-to-wall map of land cover categories. Architectures like U-Net, known for its encoder-decoder structure (Ronneberger et al., 2015), have become state-of-the-art for segmenting remote-sensing imagery (Lv et al., 2023). Recent research has further pushed the boundaries with Transformer-based (Panboonyuen et al., 2021) and attention mechanisms (Wang et al., 2022), to perform semantic segmentation in remote sensing with high accuracy.

Despite the power of these models, their application in real-world conservation efforts is constrained by a critical bottleneck: the scarcity of large-scale, accurately labelled training data. Supervised deep learning models are data-hungry, requiring thousands of labelled examples to achieve high performance. The process of manually annotating satellite imagery is demanding, time-consuming, and subjective. While regional datasets have been developed, such as the BrazilDAM (Ferreira et al., 2020) dataset for detecting mining tailings dams, they are often tailored to specific geographic areas or land cover types, limiting their transferability to ecologically distinct regions like the Colombian Amazon.

This project directly confronts this data-scarcity challenge by developing and implementing a complete and scalable deep learning pipeline for deforestation analysis in Colombia. The core of this work's innovation lies in its novel data generation strategy, which leverages two powerful, cloud-based platforms. Firstly, it uses Google Earth Engine (GEE) (Gorelick et al., 2017), a planetary-scale platform for geospatial analysis, to access and pre-process years of Sentinel-2 satellite imagery (Drusch et al., 2012), creating cloud-free composites of the study area without the need for prohibitive local computation. Secondly, to overcome the lack of ground truth labels, the pipeline employs Google's Dynamic World dataset (Brown et al., 2022). Dynamic World provides a reliable and consistently updated source of global land-cover classifications that can serve as "pseudo ground truth" labels for training at a large scale.

By pairing the processed Sentinel-2 imagery with labels from Dynamic World, this project automates the creation of a high-quality dataset. This dataset is then used to train and compare two state-of-the-art semantic segmentation models: the classic U-Net and the more recent Attention U-Net architecture. The final output is a high-resolution land-cover map which, when compared across different years, can be used to precisely identify and quantify areas of forest loss. This work therefore delivers an actionable tool for stakeholders to monitor deforestation, particularly in hotspots where access and real-time monitoring is arduous, and to better direct conservation and policy interventions.

The primary goals of this work are threefold:

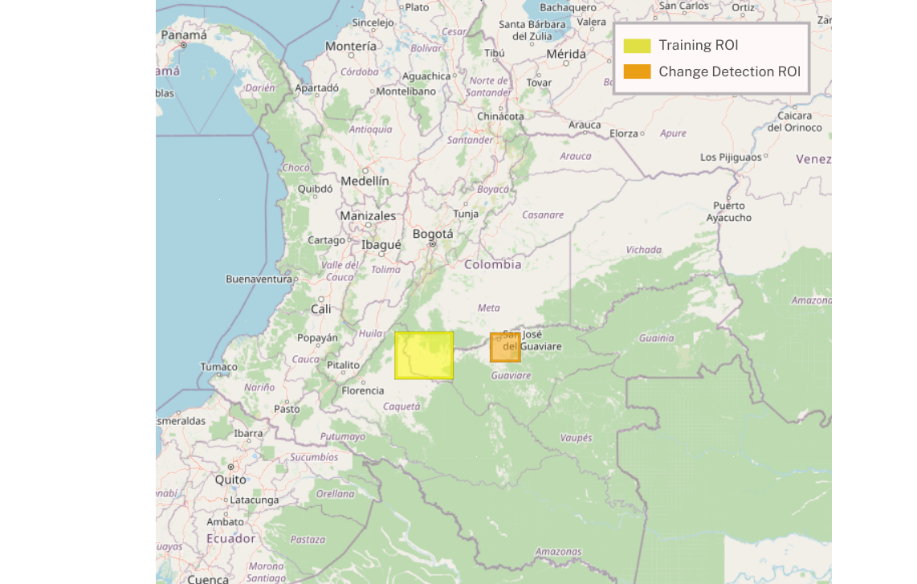
1. The design and implementation of a novel, automated pipeline combining Google Earth Engine and the Dynamic World dataset to generate large-scale training data for land cover classification, addressing a critical data constraint in deforestation monitoring.
2. A comparative analysis of state-of-the-art deep learning architectures (U-Net and Attention U-Net) for semantic segmentation of land cover in the complex ecosystems of Colombia using Sentinel-2 imagery.
3. The production of a high-resolution land-cover map and a quantitative analysis of forest change for a critical deforestation hotspot, providing actionable insights for local conservation efforts.

This paper is organised as follows: Chapter 2 details the study area, the datasets used for training, the implemented neural network architectures, and the experimental configuration. Chapter 3 presents the results of the model training and the land cover change analysis. Finally, Chapters 4 and 5 provide a discussion of the results and the overall conclusions of the study, respectively.

1. MATERIALS AND METHODS
   1. *Study Area*

The geographical focus of this study is the northwestern Colombian Amazon, a region critically important for both its immense biodiversity and its alarming rates of forest loss. This area forms a significant portion of the "Arc of Deforestation," a frontier where agricultural expansion, infrastructure development, and illicit activities exert intense pressure on rainforest ecosystems. According to Colombia's Institute of Hydrology, Meteorology and Environmental Studies (IDEAM), the Amazon region consistently accounts for the majority of national deforestation, with an estimated 68% of the country's total forest loss concentrated there in recent years.

To effectively train and validate the deep learning models, two distinct Regions of Interest (ROIs) were strategically selected within this critical zone, as depicted in Figure 1.

**

*Figure 1. A map showing Colombia with the two ROIs as "Training ROI" and "Change Detection ROI".*

The primary region for training the semantic segmentation models was a large, diverse area of 15,075.31 km² spanning the municipalities of San Vicente del Caguán (Caquetá department) and La Macarena (Meta department). This area was specifically chosen because it encapsulates the complex land-use dynamics characteristic of the deforestation frontier.

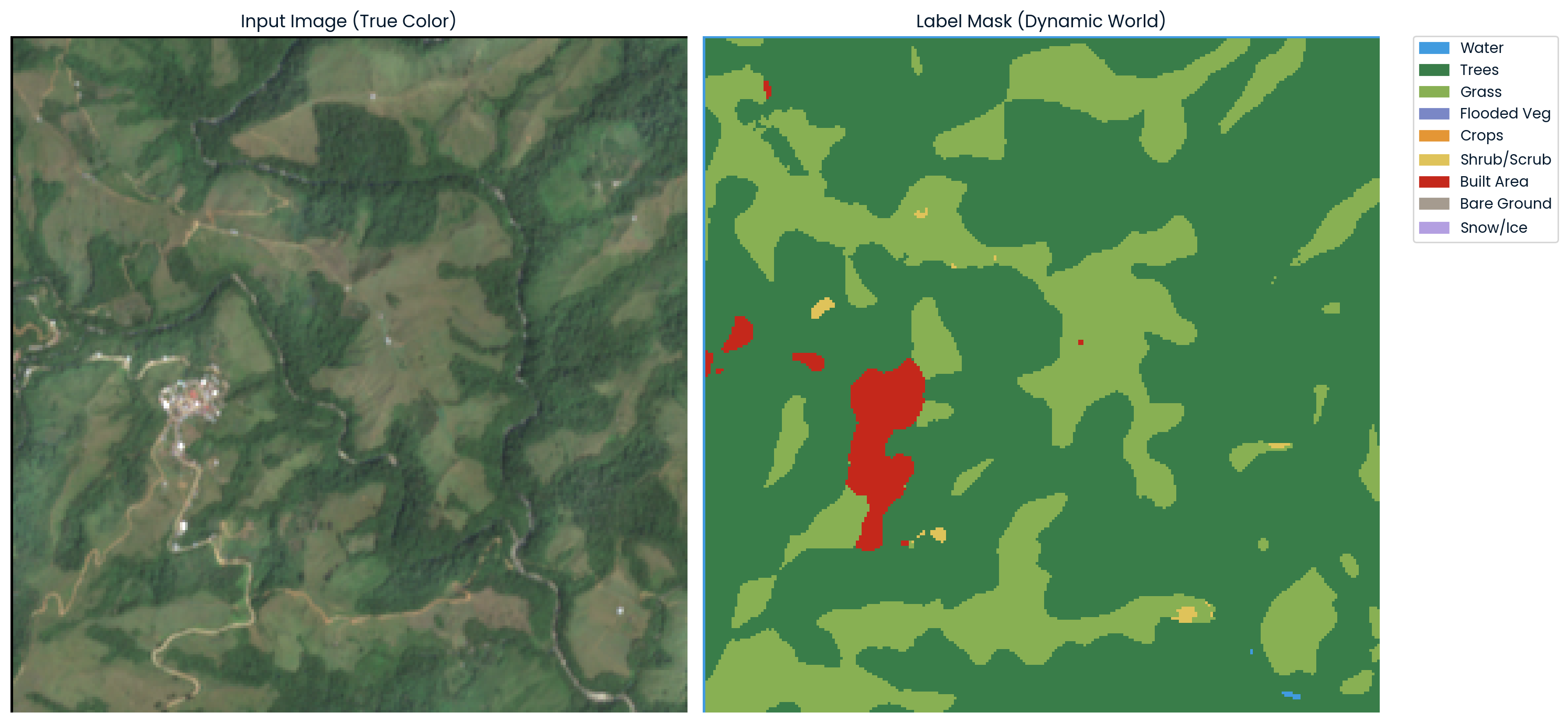
This diversity is essential for training a robust model, forcing it to learn the spectral signatures of various land cover types and the subtle differences between them. Both Meta and Caquetá are consistently ranked by IDEAM as being among the top five departments for deforestation nationwide, making this ROI a representative sample of the problem being addressed (IDEAM, 2024). The bounding box for this training region is defined by the coordinates: 1.722°N to 2.712°N and 74.889°W to 73.656°W.

For the final change detection analysis, a separate and distinct ROI was selected to test the model's generalization capabilities on a new, unseen area. This 4,430.23 km² region is cantered on the municipality of San José del Guaviare (Guaviare department). Guaviare is another critical deforestation hotspot, often cited as a nexus for land grabbing, road construction, and the expansion of illicit crops that drive forest conversion (UNODC, 2021).

This ROI was chosen as it represents an active and rapidly changing landscape, providing an ideal testbed for quantifying the performance of the trained model in a real-world monitoring scenario. The bounding box for this change detection region is defined by the coordinates: 2.1°N to 2.7°N and 72.8°W to 72.2°W.

2.2 *Datasets*

To address the project's objective, two distinct, publicly available satellite data products were utilized, both sourced via the Google Earth Engine (GEE) cloud platform. The selection of these datasets was based on their spatial and temporal resolution, data quality, and suitability for large-scale land cover analysis and deep learning applications.



2.2.1 *Sentinel-2 Multispectral Imagery*

The primary feature data for this study was sourced from the Sentinel-2 satellite mission of the European Union's Copernicus Earth observation program. We utilized the Sentinel-2 Level-2A (L2A) surface reflectance collection, which provides analysis-ready, atmospherically corrected imagery. This ensures spectral consistency across time, which is critical for change detection. The mission's high temporal resolution of approximately five days provides a high density of images, enabling the creation of cloud-free composites even in persistently cloudy regions like the Colombian Amazon.

A key advantage of Sentinel-2 is its combination of high spatial resolution and rich spectral detail. The model was trained using a selection of nine bands at 10 and 20-meter resolution. This included the visible (Red, Green, Blue), Near-Infrared (NIR), and Short-Wave Infrared (SWIR) bands. This spectral range is highly effective for land cover classification (Drusch et al., 2012); the visible bands help identify basic landscape features, while the NIR and SWIR bands are particularly sensitive to vegetation health, structure, and moisture content. This combination provides the model with the comprehensive information needed to accurately differentiate between key classes such as forests, grasslands, water bodies, and urban areas.

*2.2.2 Dynamic World Land Cover Data*

The ground-truth labels required for model training and validation were derived from the Google Dynamic World V1 dataset. This near-real-time product offers a global, 10-meter resolution land cover classification, generated by applying a deep learning model to the entire Sentinel-2 satellite archive (Brown et al., 2022). A key advantage of this dataset is its perfect spatial and temporal alignment with the Sentinel-2 feature imagery, eliminating the need for complex data registration or resampling. For every Sentinel-2 scene, Dynamic World provides a per-pixel probability distribution across nine distinct land cover classes: water, trees, grass, flooded vegetation, crops, shrub/scrub, built area, bare ground, and snow/ice.

* 1. *Network Architectures*

To perform the semantic segmentation of land cover types, two deep learning architectures based on the U-Net framework were implemented and compared. The U-Net was chosen as the foundational model due to its proven effectiveness in biomedical and satellite image segmentation, where precise localization of features is paramount. The second model, an Attention U-Net, was selected to test a specific architectural enhancement designed to improve feature relevance.

* + 1. *U-Net*

The U-Net, originally proposed by Ronneberger et al. (2015), is a fully convolutional neural network designed for end-to-end image segmentation. Its distinctive U-shaped architecture consists of two symmetric paths: a contracting (encoder) path to capture context and an expansive (decoder) path for precise localization. The encoder follows the structure of a typical convolutional network, composed of sequential blocks that each contain two 3x3 convolutional layers with ReLU activations. To mitigate overfitting and improve generalization, Dropout layers are applied within each of these blocks. Each encoder block is followed by a 2x2 max pooling operation for downsampling, which progressively halves the spatial dimensions of the feature maps while doubling the number of feature channels. This process allows the network to build a hierarchical understanding of the image, with deeper layers learning more complex and abstract features.

The decoder's primary function is to upsample these abstract features and recover the original spatial resolution for pixel-wise classification. A key innovation of the U-Net is the use of skip connections, which concatenate the upsampled feature maps from the decoder with the corresponding high-resolution feature maps from the encoder path. This fusion allows the decoder to leverage both the abstract semantic information from the deep layers and the fine-grained spatial details from the shallow layers, which is crucial for accurately delineating boundaries between land cover classes. The final layer of the network is a 1x1 convolution with a softmax activation function, which produces a probability map for each target class for every pixel. For this study, the U-Net was configured with three downsampling and three upsampling stages, creating a network of four distinct resolution levels.

* + 1. *Attention U-Net*

While the standard U-Net is highly effective, its skip connections naively fuse all features from the encoder, potentially propagating redundant or irrelevant information to the decoder. The Attention U-Net, proposed by Oktay et al. (2018), addresses this limitation by integrating attention gates into the skip connections. The attention gate is a sub-network that dynamically generates a weighting mask, or an "attention map," which is applied to the features being passed from the encoder. This map learns to highlight the most salient regions relevant to the segmentation task at a given scale, while suppressing feature responses in irrelevant background areas.

The gate takes two inputs: the high-resolution feature map from the encoder path and a gating signal from the corresponding, deeper decoder path, which provides the necessary contextual information to guide the attention mechanism. By applying this attention map before the concatenation step, the model learns to selectively focus on the most important spatial information, making the feature fusion process more intelligent and efficient. For this project, this allows the model to learn to emphasize the spectral signatures of "Trees" while ignoring nearby, potentially confusing features. The architecture of the Attention U-Net implemented in this study is identical to the baseline U-Net, including the use of Dropout for regularization, with the sole modification being the inclusion of these attention gates at each of the three main skip connection pathways.

* 1. *Experiment Setup*

All model training and experiments were conducted on a high-performance computing system equipped with a 32-core CPU and accelerated by a single NVIDIA Tesla V100-SXM2 GPU with 16 GB of VRAM. The data preparation workflow was executed within the Google Earth Engine (GEE) cloud platform to handle the large-scale geospatial data. For the input features, a cloud-free Sentinel-2 composite was generated for the year 2022 by filtering the COPERNICUS/S2\_SR\_HARMONIZED image collection and applying a per-pixel median() reducer. This technique synthesizes the time-series into a single, high-quality image, effectively removing clouds and transient atmospheric noise. Concurrently, a corresponding label image was created from the GOOGLE/DYNAMICWORLD/V1 collection by applying a mode() reducer to find the most frequent land cover classification for each pixel over the same period.

To prepare this data for the deep learning framework, the Sentinel-2 feature bands and the Dynamic World label band were combined into a single multi-band image stack to ensure perfect pixel-wise alignment. The feature bands were then normalized by dividing their surface reflectance values by 3000, scaling them into a range suitable for neural network processing. The final stacked data was exported from GEE as a set of TFRecord files by partitioning the image into 256x256 pixel patches. This process yielded a final dataset of 2,279 patches, which was subsequently split into 2,051 for training and 228 for validation.

A key methodological decision was to remap the nine original Dynamic World classes into four target categories relevant to deforestation analysis: 'Trees', 'Grass', 'Shrub/Scrub', and a consolidated 'Other' class. This strategy was adopted to address the class imbalance that caused initial models to fail. To further mitigate the remaining imbalance, class weights were calculated based on the inverse frequency of each of the four classes within the training set. This strategy penalizes the model more heavily for misclassifying less frequent classes, encouraging a more balanced predictive performance. The models were trained for a maximum of 50 epochs with a batch size of 5 and a shuffle buffer of 250, parameters selected after an iterative experimentation process to balance performance, training stability, and the available computational resources. To manage memory usage during long runs, validation was performed over 15 steps per epoch. The model's performance was monitored using the 4-class Mean Intersection-over-Union (IoU), with an early stopping patience of 10 epochs to save the best-performing model.

For the final analysis, the model's primary 4-class output was reclassified in a post-processing step to generate a binary Forest vs. Non-Forest map. In this binary scheme, the 'Trees' class was designated as 'Forest', while the 'Grass', 'Shrub/Scrub', and 'Other' classes were aggregated into a 'Non-Forest' category. This two-stage evaluation allows the model to learn from a more nuanced ecological context while providing a final, focused assessment of its efficacy for the ultimate goal of deforestation monitoring. The same data preparation workflow was used to generate two additional inference datasets for the years 2021 and 2023 to serve as the "before" and "after" snapshots for the change detection analysis.

1. RESULTS
2. DISCUSSION
3. CONCLUSIONS
4. REFERENCES

Alshehri, M., Ouadou, A. and Scott, G. J. (2024). ‘Deep Transformer-Based Network Deforestation Detection in the Brazilian Amazon Using Sentinel-2 Imagery’, *IEEE geoscience and remote sensing letters*,21 pp. 1-5. <https://doi.org/10.1109/LGRS.2024.3355104>.

Belgiu, M. and Drăguţ, L. (2016). ‘Random forest in remote sensing: A review of applications and future directions’, *ISPRS journal of photogrammetry and remote sensing*,114 pp. 24-31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.

Brown, C. F., et al. (2022). ‘Dynamic World, Near real-time global 10 m land use land cover mapping’, *Scientific Data*,9 (1), p. 251. <https://doi.org/10.1038/s41597-022-01307-4>.

Drusch, M., et al. (2012). ‘Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services’, *Remote sensing of environment*,120 pp. 25-36. <https://doi.org/10.1016/j.rse.2011.11.026>.

Ferreira, E., et al. (2020). ‘BRAZILDAM: A BENCHMARK DATASET FOR TAILINGS DAM DETECTION’, *International archives of the photogrammetry, remote sensing and spatial information sciences.*,XLII-3/W12-2020 pp. 343-348. <https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-343-2020>.

Gorelick, N., et al. (2017). ‘Google Earth Engine: Planetary-scale geospatial analysis for everyone’, *Remote sensing of environment*,202 pp. 18-27. <https://doi.org/10.1016/j.rse.2017.06.031>.

Hansen, M. C., et al. (2013). ‘High-Resolution Global Maps of 21st-Century Forest Cover Change’, *Science (American Association for the Advancement of Science)*,342 (6160), pp. 850-853. <https://doi.org/10.1126/science.1244693>.

IDEAM (2024). *Monitoreo de la superficie de bosque y la deforestación en Colombia*. Bogotá D.C: Instituto de Hidrología, Meteorología y Estudios Ambientales.

James, F. (2025). ‘Remote Sensing for Deforestation Monitoring and Forest Management’, *Journal of Remote Sensing & GIS*,14 (2).

Lecun, Y., et al. (1998). ‘Gradient-based learning applied to document recognition’, *Proceedings of the IEEE*,86 (11), pp. 2278-2324. <https://doi.org/10.1109/5.726791>.

Lv, J., et al. (2023). ‘Deep learning-based semantic segmentation of remote sensing images: a review’, *Frontiers in Ecology and Evolution*,Volume 11 - 2023. <https://doi.org/10.3389/fevo.2023.1201125>.

Molina, M. (2024). *Drivers of deforestation in the Colombian Amazon*. p. 23.

Mountrakis, G., Im, J. and Ogole, C. (2011). ‘Support vector machines in remote sensing: A review’, *ISPRS journal of photogrammetry and remote sensing*,66 (3), pp. 247-259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>.

Oktay, O., et al. (2018). ‘Attention U-Net: Learning Where to Look for the Pancreas’. <https://doi.org/10.48550/arxiv.1804.03999>.

Pal, M. (2005). ‘Random forest classifier for remote sensing classification’, *International journal of remote sensing*,26 (1), pp. 217-222. <https://doi.org/10.1080/01431160412331269698>.

Panboonyuen, T., et al. (2021). ‘Transformer-Based Decoder Designs for Semantic Segmentation on Remotely Sensed Images’, *Remote sensing (Basel, Switzerland)*,13 (24), p. 5100. <https://doi.org/10.3390/rs13245100>.

Pekel, J.-F., et al. (2016). ‘High-resolution mapping of global surface water and its long-term changes’, *Nature (London)*,540 (7633), pp. 418-422. <https://doi.org/10.1038/nature20584>.

Procolombia (2023). *Amazonia Region: Colombia's green lung*. Available at: <https://colombia.co/en/colombia-country/geography-and-environment/amazonia-region> (Accessed).

Rech, B. (2023). ‘Effects of atmospheric correction on NDVI retrieved from Sentinel-2 imagery over different land cover classes’, *Anais do XX Simpósio Brasileiro de Sensoriamento Remoto.*

Ronneberger, O., et al. (2015). 'U-Net: Convolutional Networks for Biomedical Image Segmentation',  *Lecture Notes in Computer Science*. Switzerland: Springer International Publishing AG, pp. 234-241.

SIAT (2023). *La Amazonia colombiana*: SIAT. Available at: <https://siatac.co/la-amazonia-colombiana/> (Accessed: 7 July 2025 2025).

UNODC (2021). *Colombia: Monitoreo de territorios afectados por cultivos ilícitos 2020*. Vienna, Austria.: United Nations Office on Drugs and Crime.

Wang, Z., et al. (2022). ‘Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3’, *Computers & geosciences*,158 p. 104969. <https://doi.org/10.1016/j.cageo.2021.104969>.

Wulder, M. A., et al. (2016). ‘The global Landsat archive: Status, consolidation, and direction’, *Remote sensing of environment*,185 pp. 271-283. <https://doi.org/10.1016/j.rse.2015.11.032>.

Zhu, X. X., et al. (2017). ‘Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources’, *IEEE geoscience and remote sensing magazine*,5 (4), pp. 8-36. <https://doi.org/10.1109/MGRS.2017.2762307>.