Assessing the Prithvi-100m Foundation Model for Forest Disturbance Detection

Colm Keyes

April 2024

Name student: Colm Keyes
Registration number: 1160524
Period of Internship: 12-11-2023
Date final report: 22-04-2024
Telephone number student: +353 87 9740571
Emergency contact person (name, address & telephone number):
Vera Keyes, Lisanisk, Carrickmacross, Co.Monaghan, Ireland. +353 87 9740571
Contact at internship provider (name, address & telephone number):

Internship Supervisor Name	Xenia Ivashkovych
Name of Company Institution	VITO Remote Sensing
Address	Boertang 280 (TAP), Mol, Belgium
Telephone number	+32 14 33 59 29
MGI Supervisor Name	Johannes Reiche

Contents

List of Figures

List of Tables

Abstract

Foundation models, although widely utilized in geospatial applications, remain underexplored in the context of forest disturbance detection. This research aimed to investigate the potential of one such foundation model, the Prithvi-100m model, to enhance the monitoring and analysis of forest disturbances, a critical area given the increasing anthropogenic pressures on tropical rainforests. The core objective was to evaluate the Prithvi model's effectiveness in identifying forest disturbances, by creating a dataset based on Sentinel-1 RADD alerts and Sentinel-2 HLS data for Borneo, which is an area of significant ecological value and environmental concern. The Prithvi model's performance was assessed against that of a U-Net model, with further examination conducted on the adaptability of the Prithvi model to new data types by replacing the Sentinel-2 SWIR bands with VV and VH polarisations of SAR and InSAR data during the fine-tuning process. Given the limitations of the dataset created in this research, with only 58 and 30 images per dataset, the Prithvi model demonstrated good performance in detecting forest disturbances with an Avg. IoU of 26%, outperforming the U-Net model with a 25% Avg. IoU, when both were fine-tuned on SWIR data. Notably, the U-Net model with InSAR coherence data achieved the best overall performance with Avg. IoU of 27.8%.

In this research, the Prithvi model's effectiveness was demonstrated in forest disturbance detection, with InSAR data also showing significant potential in this domain. However, the study's impact is moderated by limitations of dataset quality and size. Suggested future directions may include the application to forest disturbance detection of foundation models pre-trained on data from tropical regions and potential exploration of the use of InSAR coherence in foundation models.

1 Introduction

Internship Organisation Background
The internship is hosted by VITO, the Flemish Institute for Technological Research, a premier European research organization specializing in cleantech and sustainable development (VITO, 2023). Within VITO, work will be conducted with the Remote Sensing group, which is dedicated to transforming Earth observation data into actionable insights. The group's mission aligns with the broader goals of VITO, aiming to provide innovative and high-quality solutions that contribute to a more sustainable future. The Remote Sensing group has a diverse portfolio, encompassing projects related to climate change adaptation, land use planning, and natural resource management (VITORemoteSensing, 2023).

1.1 Context and Justification

Forests, as biodiversity hotspots and vast sources of carbon, are crucial to global climate change resilience (Mittermeier et al., 2011). They are the lungs of our planet, absorbing carbon dioxide, releasing oxygen, and playing an indispensable role in sustaining life on Earth.
They also provide livelihoods for many people, especially in developing countries (Sunderlin et al., 2005). Forests contribute to the balance of oxygen, carbon dioxide, and humidity in the air (Başkent et al., 2010). They protect watersheds, which supply fresh water to rivers and ultimately to towns and cities (Chung et al., 2021). However, the health and integrity of these forests are under threat due to various anthropogenic activities, necessitating urgent and effective monitoring and management strategies (Morris, 2010; Randhir and Erol, 2013).

Tropical Forests
Tropical forests, in particular, are of immense importance. They are biodiversity hotspots, hosting a vast array of species, many of which are not found anywhere else in the world (Pillay et al., 2021). These forests also play a significant role in global carbon cycling and climate regulation (Cramer et al., 2004).
Tropical forests are not only threatened by deforestation but also by various disturbance events. These disturbances can be natural, such as storms, fires, pests, and diseases, or human-induced, such as logging and mining (Lindenmayer & McCarthy, 2002). Disturbance events can have significant impacts on the structure and function of tropical forests, affecting their biodiversity, carbon storage capacity, and ecosystem services (Cole et al., 2014).
Traditional methods of monitoring these disturbances, such as on-foot surveys, are fraught with challenges. The dense and often inaccessible nature of these forests makes it difficult to detect and quantify disturbances accurately and timely (Misiukas et al., 2021). Moreover, these methods are labor-intensive, time-consuming, and often unable to provide a comprehensive picture of the forest's health and disturbances. This limitation underscores the need for more advanced and efficient methods of monitoring forest disturbances such as drone, aerial, and satellite imagery.
In response to the challenges posed by anthropogenic pressures, international initiatives such as Reducing Emissions from Deforestation and Forest Degradation (REDD+) have been established (Corbera & Schroeder, 2011). REDD+ aims to incentivise developing

countries to reduce emissions from deforestation and forest degradation through financial mechanisms (Angelsen, 2016). One notable example of this initiative in action is the partnership between Norway and Indonesia. Norway has pledged to provide up to $1 billion to Indonesia under the REDD+ framework to support the country's efforts to reduce deforestation and forest degradation (McNeill, 2015). This partnership underscores the critical role of accurate and timely forest disturbance monitoring in achieving REDD+ objectives. It also highlights the potential societal and environmental benefits that can be gained from improving our ability to monitor tropical forest disturbances. However, to fully realize these benefits, there is a need for more advanced and efficient methods of monitoring forest disturbances.

Neural Networks

In recent years, deep learning has emerged as a transformative approach in machine learning, offering unparalleled performance across a range of tasks (Olier et al., 2018). These methods leverage deep neural networks to learn from raw data through multiple layers of abstraction, providing significant improvements in accuracy and efficiency over traditional machine learning algorithms (Schmidhuber, 2015). This is particularly evident in the realm of computer vision, where Convolutional Neural Networks (CNNs) have become the standard for tasks such as image classification and object detection (Tuli et al., 2021).

The utility of Neural Networks has been further enhanced by specialized architectures like U-Nets. Particularly effective in semantic and instance segmentation tasks, U-Nets have proven invaluable for remote sensing applications such as land cover classification (Shafique et al., 2022). Despite the effectiveness of Neural Networks, a significant challenge remains, the requirement for large, well-labeled datasets for effective training. This data scarcity often serves as a bottleneck in the successful deployment of deep neural networks (Tarasiou, 2021). To mitigate this data limitation, various strategies like data augmentation have been employed (Sobien et al., 2022). However, these often require extensive pre-processing by the user to achieve the desired level and diversity of data. An alternative and increasingly popular approach is the use of pre-trained foundation models.

Foundation Models

Foundation models serve as a versatile backbone for a wide array of machine-learning tasks. These models are pre-trained on extensive, diverse datasets, enabling them to capture a broad range of features and patterns (Qin et al., 2023). The architecture often consists of a base model, which is the encoder, and a task-specific decoder, or "head", that can be fine-tuned for specialized applications. This modular approach allows for rapid adaptation to new tasks without the need for training a model from scratch, thereby significantly reducing computational costs and time (T. Zhang et al., 2022).

There are numerous advantages of using foundation models over traditional CNNs, which are trained exclusively for single tasks. First, the pre-trained base model acts as a feature extractor that has already learned a rich set of features from its extensive training data. This pre-training enables the model to generalize better to new tasks, often resulting in superior performance over task-specific networks (Cha et al., 2023). Second, the need for large, well-labeled datasets is mitigated, as the base model can effectively leverage its pre-learned features for fine-tuning on smaller, task-specific datasets. This is particularly beneficial in fields like remote sensing, where acquiring large, high-quality labeled datasets

can be challenging (X. Wang et al., 2023). Lastly, the modular nature of foundation models allows for the leveraging of these complex, high parameter models in situations of low processing capacity, such as those present at many universities and small scale research firms, furthermore opening new avenues for research to be conducted.
While the pretraining of models on large datasets is a fundamental advantage of foundation models, it does not come without limitations, especially when adapting these models to specialized datasets. A core strength of foundation models lies in their ability to generalize to new, unseen data. However, multimodal adaptation is a critical area of underperformance in foundation models. This limitation underscores the need for further exploration into the adaptability of foundation models to new types of data, beyond variations in geographical extent and downstream task.

SAR Data
Synthetic Aperture Radar (SAR) operates in the microwave portion of the electromagnetic spectrum, allowing it to collect data independent of light conditions or cloud cover (Woodhouse, 2017). SAR's capability to function reliably across all weather conditions and times of day make it an indispensable tool for monitoring regions with persistent cloud cover or variable atmospheric conditions, such as those found in Borneo (Keydel, 1992; T. Zhang et al., 2023).
Numerous studies have validated the effectiveness of SAR data in detecting forest disturbances (Ballère et al., 2021; Durieux et al., 2019; Bouvet et al., 2018). For instance, the RADD alert detection system leverages SAR data to provide an alert-based system that offers near real-time detections with high accuracy (Reiche et al., 2021). This alert system has emerged as a useful tool for monitoring disturbances across tropical regions. Additionally, the efficacy of Interferometric SAR (InSAR) data in monitoring subtle environmental changes has been well-documented (Akbari and Solberg, 2022; Pulella et al., 2020; Singh et al., 2020). These applications underscore SAR's potential as a robust data source to evaluate the adaptability of models to new, diverse data types, providing a solid foundation for testing the flexibility of advanced modeling approaches like foundation models.

1.2 Significance of the Topic

As large-scale foundation models continue to be developed by major corporations, their transformative impact across various sectors is becoming increasingly evident. For instance, the recent launch of large language models such as Chat-GPT and BERT in early 2023 has had a profound influence on many industries as well as society as a whole. Their widespread applicability stems from the model's ability to generalize effectively to new tasks, a feature enabled by its training on vast and diverse datasets (Zhou et al., 2023).
In a recent collaboration, NASA and IBM have developed the Prithvi-100m model, a foundation model specifically designed for remote sensing applications ("Prithvi-100M", 2023). The architecture used in this model draws heavily from research conducted by He et al. (2021), with this original architecture expanded into the time domain. This model employs a self-supervised encoder that enhances data capacity by randomly masking portions of the data and predicting the masked pixels. This self-supervised approach, inspired by Natural Language Processing (NLP) models like GPT and BERT, has gained significant traction in the deep learning community. Prithvi-100m stands out as the first

Masked Autoencoder Vision Transformer (MAE-VIT) in its domain, and is pretrained on Harmonized Landsat Sentinel-2 (HLS) data (Claverie et al., 2018). The masked encoder architecture allows the model to focus on relevant features within the data, thereby improving its ability to generalize to new tasks.
The Prithvi-100m model has already demonstrated its efficacy in a range of applications, including burn scar detection, flood mapping, and crop classification. For instance, it achieved an overall accuracy of 60% in crop classification and 96% in both burn scar and flood detection tasks ("Prithvi-100M", 2023). Despite these results, the model is relatively new and its full potential in addressing critical global challenges, such as monitoring forest degradation, remains largely unexplored. As the adoption of foundation models continues to grow, it becomes increasingly crucial to evaluate their capabilities in addressing urgent and complex issues, such as the need for more advanced and efficient methods of monitoring forest disturbances.

1.3 Research Questions

The main goal of this research is to investigate the Prithvi-100m model's capacity to detect forest disturbance events over Borneo and assess its ability to generalize to new data types.

Research Question 1:
How does the Prithvi-100m model perform in identifying forest disturbances over Borneo, when fine-tuned with Sentinel-2 data and RADD alert detection labels?

Research Question 2:
How does the Prithvi model's performance in forest disturbance detection compare against classic transformer models like U-Net?

Research Question 3: What is the impact on the Prithvi-100m model's performance when Sentinel-1 SAR and InSAR data are introduced as new, unseen data during fine-tuning for forest disturbance detection over Borneo?

1.4 Reading Guide

Introduction to the Topic:
This section offers a foundational understanding of the pivotal developments in the field of foundation models within geospatial artificial intelligence, setting the stage for a deeper exploration of their application in forest disturbance detection.
1. (Jakubik et al., 2023): This work introduces the Prithvi model, showcasing its development and application in the geospatial domain. This paper is crucial for understanding the specific context in which the Prithvi model operates and its relevance to the research presented in this report.
2. (He et al., 2021): This paper introduces the integration of the Masked Autoencoder (MAE) and Vision Transformer (ViT) architectures which lead to the development of the Prithvi model. This paper delves into the original MAE VIT architecture, providing insights into the conceptual and technical underpinnings that have influenced its design

and functionality. It is a key reading for grasping the innovations that occurred prior to the creation of the Prithvi model.

Application of Prithvi Model:
3. (Li et al., 2023): provides a comparative analysis between the Prithvi model and the U-Net model for flood mapping, using the Sen1Floods11 dataset. This study demonstrates the application of these models for flood mapping, serving as a reference point for understanding how the Prithvi model's performance in forest disturbance detection may compare to other established models like U-Net in related geospatial tasks.

Comprehensive Survey of Pretrained Foundation Models:
4. (Zhou et al., 2023): offers an extensive review of pretrained foundation models, their theoretical underpinnings, key components, methodologies, and a wide range of applications. This survey provides a thorough understanding of the landscape of pretrained foundation models, covering models like BERT and progressing to advanced systems like ChatGPT. This work contextualizes models within the broader spectrum of foundation models, and underscores their transformative potential in the field of machine learning and artificial intelligence.

Datasets:
5. (Masek, 2023): The HLS Product Guide provides a comprehensive overview of the Harmonized Landsat and Sentinel-2 (HLS) dataset, detailing its specifications and applications. This guide is for readers looking to obtain further information on the HLS data utilized here, and offering insights into the dataset's structure, uses, and nuances.

SAR Data:
6. (Woodhouse, 2017): This book offers a thorough exploration of SAR and InSAR technologies, combining theoretical concepts with practical applications and real-world examples. This book serves as an exhaustive guide, detailing the intricacies of synthetic aperture radar (SAR) and interferometric SAR (InSAR) in remote sensing. This book is crucial for those looking to gain a comprehensive understanding of how SAR and InSAR data are utilized in environmental monitoring.

1.5 Explanation of Technical Terms/Jargon

Foundation Model: A type of deep learning model pre-trained on large datasets to capture a wide array of features, which can then be fine-tuned for specific tasks.
Prithvi-100m Model: A specific foundation model designed for geospatial data analysis, incorporating Vision Transformer (ViT) and Masked Autoencoder (MAE) architectures.
RADD Alerts: Alerts generated from Sentinel-1 SAR data to indicate forest disturbances.
Sentinel-2 HLS Data: Harmonized Landsat and Sentinel-2 data, providing high-resolution multispectral imagery used for environmental monitoring.
SAR and InSAR Data: Synthetic Aperture Radar (SAR) data captures surface characteristics, while Interferometric SAR (InSAR) provides elevation and deformation measurements by comparing phase differences between radar images.
U-Net Model: A convolutional neural network known for its effectiveness in image segmentation tasks, including environmental and medical imaging.
IoU (Intersection over Union): A metric used to quantify the capacity of an object

detector on a particular dataset, calculated by dividing the area of overlap between the predicted and ground truth bounding boxes by the area of their union.

2 Data and Methods

This section details datasets and methods utilized in data pre-processing and analysis of the Prithvi-100m model.

2.1 Test Site

Borneo, the third-largest island in the world, is renowned for its rich biodiversity and extensive tropical rainforests. It contains a wealth of flora and fauna which include many endemic species (Boyce et al., 2010). These species are emblematic of the island's ecological value and the urgent need for effective conservation strategies. Borneo is not only known as a hotspot for its biodiversity but also for its large-scale farming practices, including mega rice projects as well as sprawling palm plantations, which has resulted in extensive deforestation (Hayasaka et al., 2014; Goldstein, 2015). The transformation of forested areas into agricultural land has not only ecological consequences but also socio-economic implications, affecting local communities and indigenous populations who depend on the forest for their livelihoods.
Borneo's diverse landscape, ranging from lowland rainforests to mountainous regions, provides a varied testing ground for assessing the effectiveness of the Prithvi-100m model for forest disturbance detection. This will also contribute insights into the effectiveness of foundation models for remote sensing in supporting conservation efforts and sustainable land management practices. Figure 1 shows several Sentinel-2 scenes utilized for training the Prithvi model.

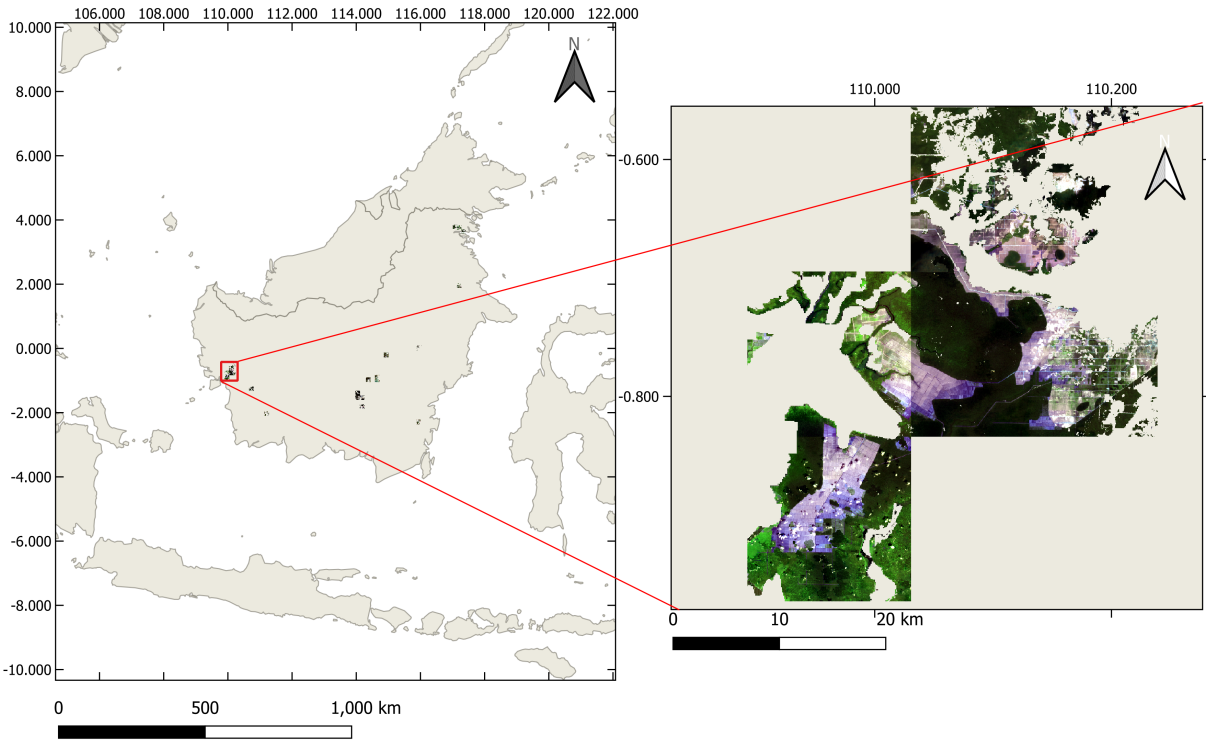


Figure 1: Island of Borneo with an inset showing the specific area within the red rectangle. The right-side image displays several Sentinel-2 scenes used for model training and validation in this study. The study area is contained within this rectangle and corresponds to the regions with visible imagery. Borneo is recognized as a biodiversity hotspot undergoing extensive anthropogenic pressures.

2.2 Datasets

This study used several datasets, including Sentinel-2 Harmonized Landsat Sentinel (HLS) data, Sentinel-1 Single Look Complex (SLC) data, and RADD alert labels. Additionally, several maps were used to filter out unwanted data, such as Above Ground Biomass (AGB) classification, Hansen Forest Loss, and FMask data. The following sections outline the preprocessing and integration of these datasets, explaining how they were combined and utilized in the study.

HLS S30 Data

The main dataset utilized for this research is sourced from the Harmonized Landsat Sentinel-2 (HLS) S30 datasets. The HLS project aims to provide consistent surface reflectance data by harmonizing observations from the Landsat 8 and 9 data and Sentinel-2 MSI sensors (Jakubik and Chu, 2023; Claverie et al., 2018) . The S30 dataset specifically refers to the Sentinel-2 MSI data, which is processed to surface reflectance and gridded to a common tiling system at a 30-meter spatial resolution. Both can be accessed through the NASA Earthdata portal and NASA's Land Processes Distributed Active Archive Center ("LP DAAC 2023", 2023; "NASA Earthdata Search", 2023).

The S30 HLS dataset offers a suite of spectral bands, including visible, near-infrared, and

shortwave infrared, making it highly suitable for a variety of remote sensing applications including forest disturbance mapping. The Prithvi-100m model, which serves as the foundation for this research, is pre-trained on this combined HLS dataset, utilizing 6 of these bands, namely R, G, B, NIR, SWIR 1 and SWIR 2 (Masek, 2023; Jakubik et al., 2023).

Sentinel-1 SAR Data

The Sentinel-1 mission, comprising two polar-orbiting satellites, Sentinel-1A and Sentinel-1B, was launched by the European Space Agency as part of the Copernicus Programme. Both Sentinel-1A and 1B are C-band SAR satellites with an operating wavelength of 5.6 cm. At this wavelength, these satellites directly observe physical phenomena as they occur on the Earth's surface (Geudtner et al., 2014). due to an operational anomaly that occurred on the 23rd of December 2021, power to the Sentinel-1B mission was lost and thus the mission was considered ended. With this in mind, only Sentinel-1A data was utilized for this study, with a revisit time of 12 days. In this research, SAR data is acquired in the raw SLC form. This form of data preserves amplitude information critical for the interferometric data analysis utilized in this study (Kim & van Zyl, 2000). Pre-processing steps occur to obtain InSAR coherence data, which has particular uses in the area of change detection. The processing chain used in this research is designed to optimize the quality and usability of the coherence data in conjunction with S30 HLS data by matching the 30m resolution as closely as possible.

RADD Alerts
The RADD (Radar Alerts for Detecting Deforestation) alert system served as a critical foundation for the fine-tuning of the Prithvi-100m model. The RADD detection system leverages high-resolution Sentinel-1 Ground Range Detected (GRD) images and employs a probabilistic Gaussian Mixture model that triggers forest disturbance alerts based on backscatter observations in VV (Vertical Vertical) and VH (Vertical Horizontal) polarizations (Reiche et al., 2021). This model has consistently achieved high (>80%) accuracy when compared to the Landsat based Global Land Analysis and Discovery (GLAD) forest disturbance detection system (Hansen et al., 2016). RADD alerts are masked by a humid tropical forest mask from Turubanova et al. (2018), annual forest loss by (Hansen et al., 2013) and mangroves are removed utilizing the Global Mangrove Watch baseline by Bunting et al. (2018). RADD alerts contain both low confidence and high confidence alerts. For this research, only high confidence alerts were considered.

2.3 Data Pre-processing

This section outlines the data preprocessing methodology used in this research. Figure 3 provides a visual representation of these steps, and section 2.4 offers a summary of the overall preprocessing process.

Data Labels
Selected HLS images were matched with RADD alerts of a corresponding date, to build a dataset that accurately represents forest disturbance events, as they occur. A preprocessing mechanism was built to accurately match up labels and data. Within this, a time frame was determined within which to match a RADD alert detection to an image.

With considerations for data requirements, it was determined that the maximum time of 2.5 years between the alert and the image would still hold some information on the disturbance events.

Data Masks

Borneo, containing equatorial tropical rainforests, is subject to significant cloud cover throughout the year, which hinders the requisition of cloud-free data. Within the HLS dataset, per-pixel cloud, cloud shadow, snow, and water masks are provided (Masek, 2023). These masks are generated by the Fmask algorithm, reported in Qiu et al. (2019). These masks allow for several mitigation strategies to be explored, including utilizing all images regardless of cloud cover, or adding percentage cover cut-off points such as Zupanc (2023), who suggests cloud cover ranging from 5-90% based on global Sentinel-2 coverage. Here, a cloud cover percentage of 30% was deemed an appropriate balance between the number of images and cloud cover.

To reduce the land use regions considered in this research, geographic sampling splits were considered. One such land classification utilizes Above Ground Biomass (AGB) as its primary indicator (Ferraz et al., 2018). This map has a resolution of 100 m, corresponding to plots 1 ha in size and splits the Indonesian portion of Borneo into several classes: Intact Lowland Forest, Intact Montane Forest, Secondary and Degraded Forest, Peat Swamp Forest, Swamp Scrublands, Scrublands, Crops/Agriculture, Tree Plantation, Urban/Settlement. These different forest classes were utilized to give the data split contextual information across the Island, removing land classifications which did not pertain to primary or secondary forests.

Finally, a forest loss mask by Hansen et al. (2016) was applied for years previous to the first acquisition date of the RADD alert data. This increases the alignment between disturbance events on the ground and the RADD labels used.

Normalization

Per-patch normalization occurs in the attention stage of the model. It was shown that results using normalized tokens were statistically similar to those using normalized pixels (He et al., 2021). Batch normalization(BN) has become a standard technique in deep learning methods, however, small batch sizes may reduce the effectiveness of this normalization. An alternative to BN is Group normalization, which is suggested as a potentially more effective alternative to BN where memory constraints necessitate the use of small batches (Ioffe and Szegedy, 2015; Wu and He, 2018). In this research, a batch size of 20 was utilized, which is deemed sufficiently large to implement batch normalization.

Atmospheric Correction

In addition to cloud masking, the HLS dataset undergoes comprehensive preprocessing to ensure data quality and consistency for various applications. Atmospheric correction is applied to minimize the effects of the atmosphere on the recorded reflectance, enabling an accurate representation of the Earth's surface. Bidirectional Reflectance Distribution Function (BRDF) correction is employed to account for the anisotropic nature of surface reflectance, adjusting for different sun-sensor geometries and ensuring comparability across images taken at different times and angles (D. Roy et al., 2016; D. P. Roy et al., 2017) . Geometric correction is also performed, ensuring that the satellite imagery is

accurately aligned to map coordinates, correcting for potential distortions due to the Earth's shape, satellite orbit, and sensor orientation (Masek, 2023).

Data Augmentation

Data augmentation has been determined as an important step in deep learning methods for remote sensing (Yang et al., 2022). This involves the manipulation of the original dataset to create a more robust and diverse set of training samples. Common augmentation techniques include geometric transformations such as rotation, flipping, and scaling, as well as photometric operations like brightness and contrast adjustments (Perez & Wang, 2017). These methods enrich the dataset by introducing variability is particularly crucial due to the inherent challenges such as varying atmospheric conditions, seasonal changes, and sensor noise. By incorporating these augmentations, the model becomes more resilient to such variations, thereby mitigating the risk of overfitting, particularly when the available labeled data are limited (Oubara et al., 2022). In this research, basic random flipping geometric augmentation was applied with a probability of 50%.

Sentinel-1 SLC Pre-processing

Following the methods utilized in Keyes (2023), Sentinel-1 SLC data was accessed through the ASF Vertex online portal. Images pairs were acquired with a 12-day difference where possible, coordinated as closely as possible with the chosen HLS images for analysis. exact coordination was not always possible, given the difference in orbit times for the Sentinel-2 and Sentinel-1 satellites. Thus SAR images were matched as closely as possible before the HLS image acquisition date, while the interferometric pairs were filtered to be 12 days only, as to prevent temporal decorrelation from occurring. This leads to pairs either occurring some days before, or at times one image before and one image after the acquisition date of the Sentinel-2 image. With this in mind, software was chosen for processing these SLC images. SNAP is a well-known application for processing Sentinel-1 data with a variety of integrated InSAR processing tools. It has been designed to ensure high performance when dealing with very large images such as those from the Copernicus Sentinel-1 mission. It is also open source under a GNU GPL license, making for an ideal platform for InSAR processing (Uppuluri and Jost, n.d.; Foumelis et al., 2018; Rahman et al., 2023). Figure 2 depicts the SNAP Graph Processing Framework (GPF) utilized for InSAR coherence preprocessing, and altered slightly for SAR backscatter processing. The initial data preprocessing steps for InSAR data involved reading both Sentinel-1 SLCs in a pair into the SNAP software. Subswaths and bursts are selected during the 'TOPSAR split' step, followed by the application of orbit files. Orbit files obtained from the satellite are applied to improve the geometric accuracy of the output data (M. Wang et al., 2014). The 'back-geocoding' step involved resampling the data using a Digital Elevation Model (DEM), enabling the preservation of spatial characteristics while reducing coherence bias. Coherence window size, selected during the 'Coherence' step, played a crucial role. In this research, window sizes of 2,8 in azimuth and range directions respectively were processed. This represents a square resolution of 28m in ground range. Debursting and Terrain Corrections were also conducted to further refine the data for the study. Polarizations VV and VH were selected during the 'TOPSAR-Deburst' step while the bilinear interpolation resampling method was selected during 'Terrain Corrections'. This process was then used as a basis for creating a preprocessing pipeline in the Python

programming language.
For SAR backscatter processing, this chain is essentially the same, but for a single SLC rather than a pair. Rather than utilizing the "Coherence" step, a "multilook" step is applied in its stead, where a multilook window size of 2,8 is applied to the data, downsampling the data to a square resolution of 28m in ground range.

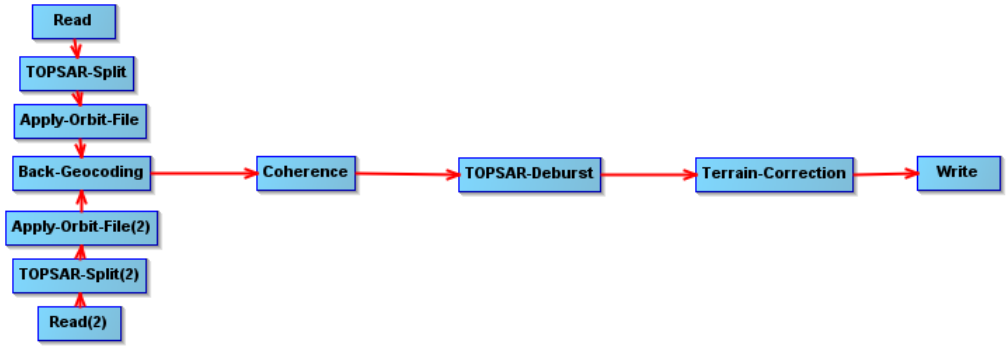


Figure 2: SNAP GPF for InSAR coherence preprocessing.

Train-Test-Validation Split
Jakubik et al. (2023) employs a data split based on temperature and precipitation levels when pre-training the Prithvi-100m based on contiguous U.S. data, where sampling is conducted uniformly across 20 groups based on these statistics. While precipitation levels may be indicative of land classification over Borneo, temperature variations would hold much less useful information due to the island's proximity to the equator and the sea. In this research, a train-test-validation split of 70-20-10 was utilized, with the test set analyzed to ensure no overlap with the training or validation sets.

Data Quality

Due to the difference between the data used by the Sentinel-1 based RADD alert detection system and the Sentinel-2 data utilized in training the Prithvi model, the size of the final dataset needed to be significantly reduced to improve the quality of the testing and training data. In general, the capacity of RADD alerts in detecting small disturbances, only a few pixels in size at 30m resolution, is much greater than the capacity of Sentinel-2 data at the same resolution. Thus, larger disturbed areas are preferred when making a comparison between the Prithvi and U-Net models. With this in mind, a range of minimum disturbance alerts per image were investigated with their effects on model performance analyzed. The minimum number of alerts per image, along with the number of images available at each of these levels are found in Table 1.

Table 1: Minimum number of RADD alerts per image, with the number of images available.

Minimum radd alerts	Number of Images
10000	58
15000	30

2.4 Final Data Pre-processing Chain

Chart 3 visually details the subsequent steps involved in preprocessing the Sentinel-2 and Sentinel-1 data with RADD alert labels, for input into the Prithvi and U-Net models.

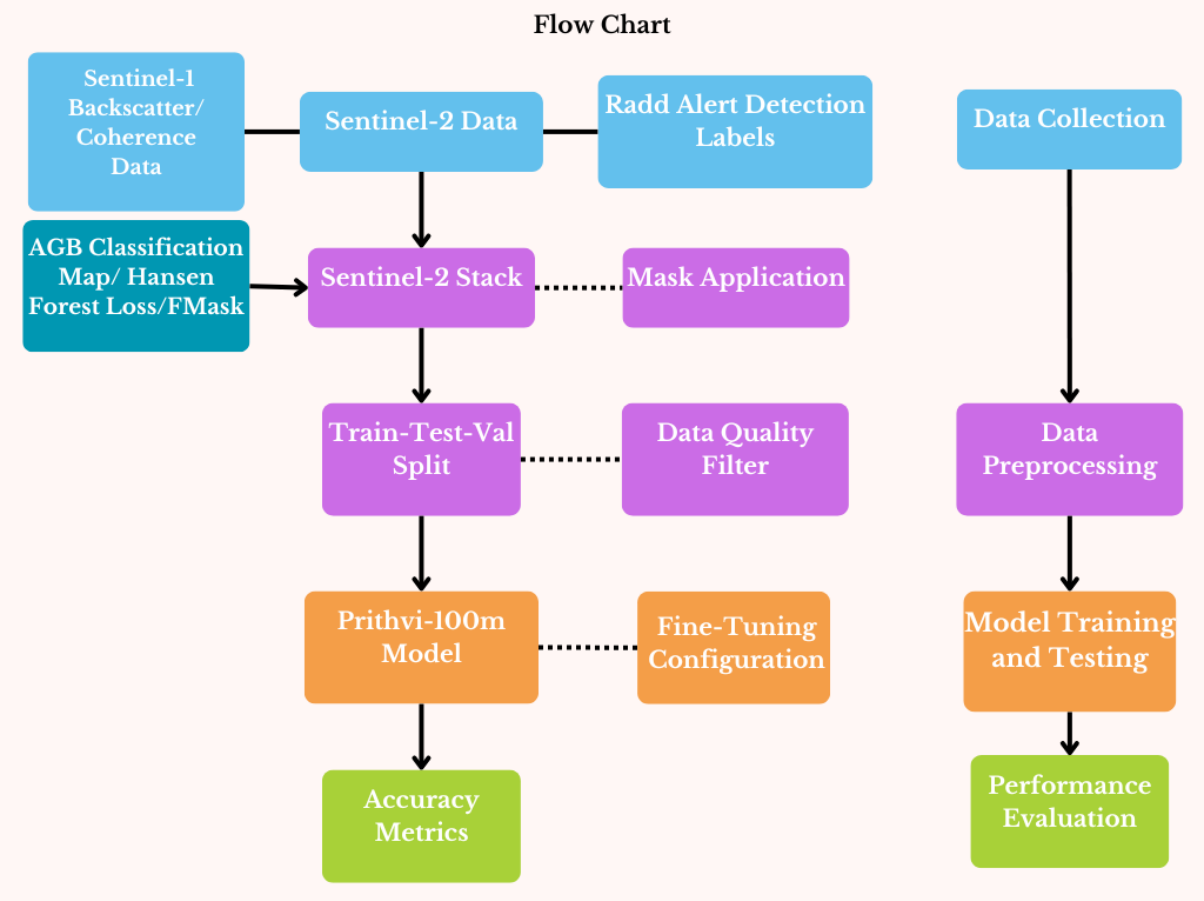


Figure 3: Flow chart, detailing preprocessing steps involved in analyzing the Prithvi-100m model performance.

Data Acquisition
HLS data was obtained from a variety of regions across Borneo, for a time from June of 2021 until December of 2023. Images were filtered to achieve a minimum number of samples, while minimizing cloud cover. Thus, a final selection was made to utilize 625 images obtained at a cloud cover percentage of 30%. This obtained a balance between a high number of images and relatively low cloud cover.

RADD alert Data Matching
The RADD alerts were aligned with the HLS data, involving resampling to a 30m resolution to match the Sentinel-2 data granularity. Subsequent steps included cropping to align with the Sentinel-2 stacks and warping to ensure precise overlay with the HLS raster positions, thereby ensuring that the alert data accurately corresponded to the relevant satellite imagery.

Stacking Data Layers
The HLS data stacks were further processed through a stacking procedure that integrated multiple data layers. This included the cropped Above Ground Biomass classification map, RADD alerts, and the 6-band Sentinel-2 data. Similarly, SAR backscatter and InSAR coherence data were added in at this step, replacing the two SWIR Sentinel-2 bands.

Applying Masks
To refine the quality of the dataset, several masks were applied to the stacked data. The Hansen Global Forest Change dataset provided forest loss events which were used to mask out disturbances occurring before June 2021, ensuring the relevance and recency of the data. Additionally, the FMask algorithm was employed to mask out clouds and cloud shadows, and the AGB classification map was also utilized to mask non-forest areas, focusing the analysis on forested regions.

Filtering Images for Quality
To further ensure the dataset's quality, images were filtered based on the number of RADD alerts, selecting those with alert counts between 10,000 to 15,000 to retain high density alerted areas. This step was crucial to improving the quality of the final dataset, to allow for meaningful analysis of the selected models.

Preparing Data for Model Input
The final stage of the pre-processing involved preparing the data for model ingestion. This included cropping the stacked data to 512x512 pixel tiles, suitable for the models used. Labels were standardized from dates to binary indicators (0 and 1). The train-test-validation split was applied and the dataset underwent a global normalization process to standardize the input values, facilitating more effective model training.

2.5 Model Architecture

Prithvi-100m
Autoencoders are a type of neural network architecture that effectively learn to encode latent representations of data into a network of neurons. they are effective as they only

learn the most important vectors that summarise a dataset, reducing model size. They have a two-step architecture, where an encoder efficiently encodes the information, and a decoder works to reconstruct the input from the latent representation. These networks are particularly effective due to their capacity for unsupervised and self-supervised learning (Bank et al., 2020; Michelucci, 2022).
A vision transformer is a type of data transformation mechanism. A vision transformer flattens patches of an image into a linear sequence of elements. A self-attention or intra-attention mechanism is then applied across this linear set of patches. Between each pair of patches, an attention score is then calculated. This attention score determines how much focus is put on other parts of the input for each position. This information is critical for effectively capturing dependencies across the entire image (Vaswani et al., 2017; Dosovitskiy et al., 2020; Ruan et al., 2022) .
The Prithvi-100m model is based on a Masked Autoencoder Vision Transformer architecture (MAE-VIT). This strategy of masking the autoencoder aims to build a sufficiently difficult learning objective so that the representation of the data cannot be directly learned. This is beneficial in reducing overfitting and increasing the model's capacity to generalize (He et al., 2021). This method operates by masking out a significant portion e.g. 75% of the input image to yield a non-trivial self-supervised learning task. The task given in pre-training, is to recreate the original image. In this way, a model can be trained efficiently on large datasets. Firstly, the high masking ratio significantly reduces processing requirements as only a fraction of the data is actually used for training, as depicted in Figure 4. Secondly, the high masking ratio largely eliminates redundancy, creating a task that cannot be easily solved by extrapolation from features in nearby patches (He et al., 2021).
One of the major steps taken by the Prithvi model over the architecture purposed by He et al. (2021), is the introduction of processing 3D spatiotemporal data transformation, from the original 2D positional and patch embeddings used by He et al. (2021). Here, they "first generate the 1D version of the sine-cosine positional encodings individually for height, width, and time and then combine the individual encodings into a single, 3D positional encoding." (Jakubik et al., 2023,p.9). For patch embeddings, Jakubik et al. (2023) utilize methodology from video processing, where 3D patches with height,width,and time are created. Following this, 3D convolutions are used to process this data.

U-Net

U-Net is characterized by its encoder-decoder structure. The encoder reduces the spatial dimensions of the input image, extracting essential features through convolutional and pooling layers, which is crucial for handling large datasets efficiently. The decoder then reconstructs the image to its original resolution, using upsampling and convolutional layers. A key feature of U-Net is its use of skip connections, which link the encoder and decoder. These connections help in retaining spatial information, crucial for accurate segmentation (Ronneberger et al., 2015).
U-Net's proven effectiveness in segmentation tasks, makes it stand out as a model for comparison. Its architecture offers a contrast to the Prithvi-100m's approach, especially in how it handles spatial information and feature extraction. This comparison will provide insights into the strengths and limitations of the Prithvi-100m model in forest disturbance detection.

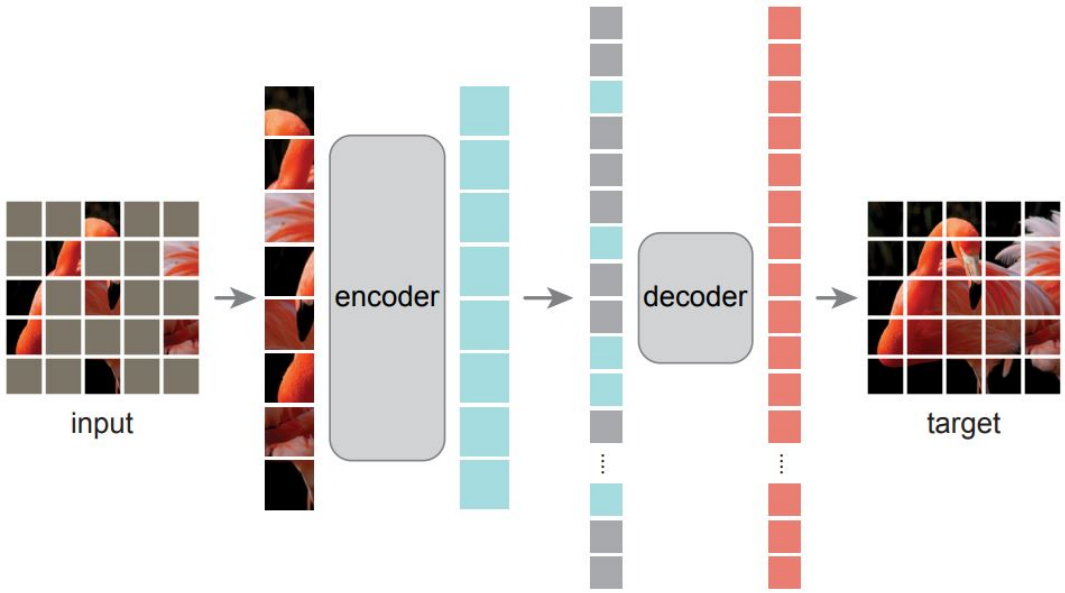


Figure 4: Depiction of Masked Autoencoder architecture proposed by He et al. (2021).

2.6 Model Implementation and Evaluation

Pre-training

Pre-training was conducted on 1TB of multi-spectral HLS data, on up to 64 NVidia A100 GPUs. The AdamW optimizer was used, with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ (Jakubik et al., 2023). The "W" pertains to weight decay, where β_1 and β_2 refer to the decay rate for the first and second moment estimates of the weight gradients (Loshchilov & Hutter, 2017). A one-cycle cosine learning rate scheduler was utilized. This is a learning rate policy that begins at a lower learning rate, increases to a maximum, and decreases using a cosine annealing schedule. This approach can help the model to converge faster and potentially achieve better performance. The paper details that a 5e-04 peak learning is utilized by the scheduler. Input size as 224 x 224 pixels, with a patch size of 1 x 16 x 16 (time x X x Y) (Jakubik et al., 2023).

Fine-tuning and Training

A number of example applications have been detailed along with the Prithvi-100m model, which include pre-trained weights and fine-tuning configurations. The burn scars fine-tuning example was utilized to significantly influence the choice of parameters for fine-tuning the Prithvi-100m model ("Prithvi-100M burn scar", 2023). This burn scars fine-tuned model will also be studied, alongside the Prithvi model without prior fine-tuning. This method is known as sequential fine-tuning, where a model trained for a certain objective is then retrained for another objective. The similarity between forest disturbance detection and burn scar detection lends to the initial use of these parameters, with the burn scar dataset regularly exhibiting disturbed forested areas. The weights within this fine-tuning example were then reset to fine-tune the model on this new application of forest disturbances. Similarly, parameters for UNET training will be based on those reported in Li et al. (2023), who analyzes the Prithvi model for flood inundation and compares it to a UNET model. These parameters are detailed in Table

2.

Table 2: Deep learning model Parameters.

Parameters	**U-Net**	**Prithvi**
Optimizer	AdamW	AdamW
Learning rate	5e-4	6e-5
Weight decay	0.01	0.01
Batch size	8	8
Learning rate scheduler	Poly	Poly
Loss function	Cross-Entropy	Cross-Entropy
Datasets for pre-training	-	HLS

Model Evaluation Metrics

Intersection over Union (IoU): IoU is a crucial metric in segmentation tasks, quantifying the precision of overlap between the predicted and actual target areas. It is defined as the ratio of the intersection area (true positives, TP) to the union area of predicted and ground truth segments. Mathematically, IoU for a class is expressed as:

$$\text{IoU} = \frac{TP}{TP + FP + FN}, \tag{1}$$

where FP represents false positives, and FN denotes false negatives.
Accuracy: This metric provides a general measure of model performance, calculated as the ratio of correctly predicted observations (true positives, TP, and true negatives, TN) to the total observations. The formula for accuracy is:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}. \tag{2}$$

The key difference between IoU and accuracy lies in their treatment of true and false classifications. IoU focuses on the overlap (true positives) in relation to both false positives and false negatives, making it sensitive to the spatial accuracy of segmentation. In contrast, accuracy encompasses all correct predictions (TP and TN) over the dataset, providing a broader view of model performance but potentially less insight into segmentation precision. While both metrics will be utilized to evaluate model runs in this research, particular attention will be given to the IoU metric, as it gives the most information with regard to the desired class of disturbed forest.

3 Problems Encountered, Limitations, and Adaptations

This section describes the various challenges encountered during the project, the strategies employed to overcome them, and the possible impact these adaptations may have had on the study's outcomes.

Deviation between Label Positions and Sentinel-2 Stacks

During the process of aligning RADD alert labels with Sentinel-2 data, a significant challenge was encountered when using Rasterio to crop and stack the labels directly onto the satellite imagery. The primary issue was a noticeable deviation in the spatial alignment of the labels relative to the actual events represented in the Sentinel-2 data. This misalignment is critical in remote sensing, as the precise positioning of labels is essential for accurate model training and subsequent analysis.
Rasterio, while powerful for many geospatial data processing tasks, exhibited limitations in ensuring the exact geo-referencing required for this specific application. The discrepancies likely stemmed from subtle differences in the geospatial metadata interpretation or the resampling methods employed during the stacking process. Such deviations, even if minor, can lead to significant inaccuracies when the model attempts to learn from these misaligned datasets.
To address this challenge, an alternative approach was adopted using GDAL Warp, a tool known for its robust geospatial data manipulation capabilities. GDAL Warp was employed to warp the RADD alert data onto the Sentinel-2 data tiles, ensuring the proper positioning of events relative to the corresponding pixels. This method involves adjusting the RADD alert data's spatial reference to match exactly with that of the Sentinel-2 imagery, thereby aligning the data layers with high precision.
The adoption of GDAL Warp for this task required a more complex processing workflow but ultimately resulted in an accurate alignment of the RADD alert labels with the Sentinel-2 data. This adjustment was crucial for maintaining the integrity of the dataset and ensuring that the subsequent analysis and model training were based on correctly aligned data. The experience highlights the importance of meticulous data preparation in remote sensing and the need for careful selection of tools that suit the specific requirements of the data alignment task.

FMask - Bitvalues, Cloud Shadows, and File Sizes

Understanding and implementing the FMask algorithm presented multiple challenges, particularly in dealing with bit values, cloud shadows, and file sizes.
Bit Values: The interpretation of bit values was a significant hurdle initially. The FMask algorithm used bit values to classify different types of pixels, such as clouds and cloud shadows. The HLS Product Guide referenced bit values ranging from 0 to 7, representing different classifications. However, in the actual raster files, these classifications were embedded in a single 8-bit value, ranging from 0 to 255. This discrepancy required a translation from a single 8-bit value to individual bits, where each bit represented a different condition or classification (Masek, 2023).
For example, cloud presence was indicated by the second bit (bit 1), and cloud shadow by the fourth bit (bit 3). To extract these conditions, bitwise operations were employed, isolating the second bit for clouds and the fourth bit for cloud shadows. When applied to the FMask data, these operations yielded boolean masks indicating the presence of clouds and cloud shadows. Such bitwise manipulations were not straightforward, particularly when one had to consider that many pixels could exhibit multiple conditions, corresponding to an algebraic equation in which the bit values were converted to a range of 0-255 and then summed. This required a precise understanding of how bits corresponded to conditions in the data.

Cloud Shadows: Cloud shadows posed another challenge, as they could alter the spectral signatures of the underlying surface, thereby affecting the quality of the data. Initially, there was consideration to introduce a buffer around cloud and cloud shadows to mitigate this effect. However, given the sparse nature of quality data points in the dataset, it was decided not to implement this buffer to avoid losing valuable data. This decision underscored the trade-off between data purity and data quantity, especially in datasets where quality samples were limited.
File Sizes and Trailing Pixels: Anomalous file sizes, such as images with dimensions of 513x512 for example, were another issue encountered. These trailing pixels in the FMask rasters introduced inconsistencies in data alignment and processing. To rectify this, GDAL Warp was utilized once more to standardize the raster dimensions, ensuring these minor discrepancies did not impact the overall spatial alignment of the data layers. Although the alteration of a single pixel width did not significantly affect data location post-warping, maintaining consistent dimensions was crucial for systematic processing and analysis.
Each of these challenges underscored the intricacies involved in preparing and processing remote sensing data, where attention to detail was crucial in ensuring the accuracy and reliability of subsequent analyses.

Data Sparsity

One of the most difficult challenges encountered in this research was the prevalence of data sparsity within the 512x512 tiles processed. A significant portion of these tiles exhibited minimal or no labels, posing a substantial obstacle to initial training. Data sparsity is a common issue in remote sensing and machine learning, particularly when dealing with the large data requirements regularly needed to train large complex models. The problem is exacerbated when the area of interest, such as specific land cover types or disturbances, occupies a small fraction of the overall landscape, leading to a scarcity of labeled pixels.
To address this challenge, an approach was adopted that involved several steps aimed at mitigating the impact of label disparity. The initial step involved a filtering process, r4educing the dataset based on the number of labeled pixels per image. This filtering was an integral part of the data preprocessing stage, where Sentinel-2 tiles with little to no labels, specifically those with fewer than 1000 labeled pixels per tile, were excluded. This criterion addressed both the issue of data sparsity and also significantly reduced the computational load for future processing. The dataset experienced a sizable reduction, shrinking by a factor of 10, from several thousand 512x512 tiles to a more manageable few hundred, thereby streamlining the subsequent processing stages.
Another strategy implemented to combat data sparsity was the incorporation of the focal loss function during model training. Focal loss, an advanced variation of the traditional cross-entropy loss, is specifically designed to enhance model performance in scenarios characterized by imbalanced datasets. It achieves this by modifying the cross-entropy loss equation to apply a modulating factor to the loss of well-classified examples. As a result, the model focuses more on hard, misclassified examples and less on easy ones, effectively addressing class imbalance and ensuring that the sparse presence of the desired class does not dilute the learning process (Lin et al., 2017).
In addition to focal loss, a class weight mechanism was integrated to further alleviate the imbalance issue, as early attempts were faced with the issue of a decreasing IoU and

Accuracy for the disturbed areas as training continued. Through a series of experiments, an optimal class weight ratio was sought, testing values ranging from 1 to 10000. The objective was to determine a weight that would appropriately scale the loss for the underrepresented class, thereby guiding the model to pay greater attention to these crucial but scarce labels. Ultimately, a class weight ratio of 1:100 was selected, striking a balanced emphasis on the minority class without overshadowing the learning from the majority class. This nuanced approach to dealing with data sparsity and class imbalance underscored the complexity and attention to detail required in handling such challenges in the realm of machine learning with satellite imagery.

OpenMM - Model Inference, Testing, and Training

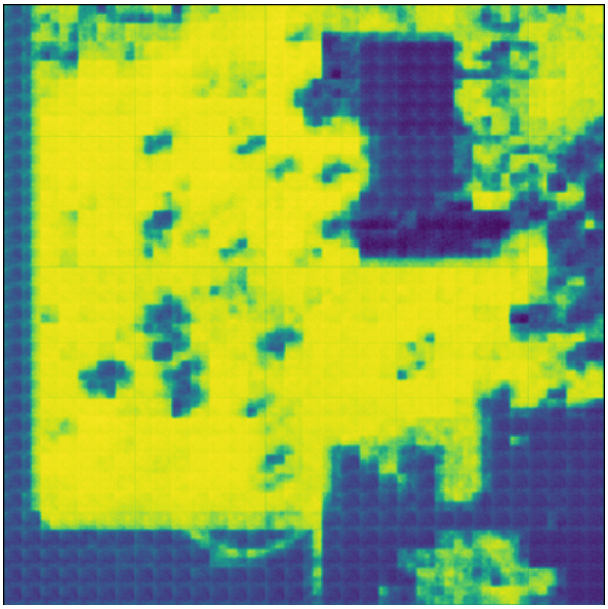
Working with OpenMM, MMseg, and the processes outlined on the HLS Foundation 2023 GitHub page presented several challenges (Jakubik & Chu, 2023). Firstly, the processing workflow utilized by the teams at NASA and IBM was not well-suited for debugging and error analysis. The model initialization using the mim terminal command offers limited debugging capabilities, which restricts data discovery in many Integrated Development Environments (IDEs). Furthermore, OpenMM processes rely on a structure of .JSON configuration files, where parameters are read as tuples and dictionaries into the training regimen. This setup provides no clear back-end tracing for where dictionary values are directed or their roles within the model, leading to a scenario where strings and values are inputted into the model without clear traceability. This configuration-driven model architecture can pose significant challenges for debugging code and a steep learning curve to those unfamiliar with the framework.
While OpenMM offers a user-friendly surface interface compared to frameworks like PyTorch or TensorFlow, any in-depth debugging necessitates delving into the source code of underlying packages, such as MMseg, where most processing limitations are encountered. Specific issues arose when using data stack configurations other than those recommended by the MMseg documentation and the team behind the Prithvi model samples. For example, the PyTorch DataLoader used in the single GPU test function within MMseg testing expects image metadata to contain only "mean" and "std" data and does not accommodate 'no data' values in the form utilized for burn scars. Encountering an error like TypeError: tensor2imgs() got an unexpected keyword argument 'NoData' requires a deep dive through the decoder head, mim testing, and MMseg testing APIs to understand that the 'NoData' value is a dictionary string pair without traceability. Similar challenges arise when working with this JSON configuration-based architecture.
A workaround was developed by transforming terminal training, testing, and inference commands into debuggable Python scripts using the Subprocess package. This approach allows for the execution of command-line instructions within Python, providing greater control and visibility into the processing flow. Significantly, it enables the running of various model configurations by iterating through multiple configuration files and, crucially, allows for the debugging of code, offering a more transparent and manageable way to address issues and refine model performance.

NoData Model Training

The pipelines for testing, training, and inference developed for the Prithvi model were initially designed to operate optimally on well-formatted, preprocessed datasets, such as

the HLS burn scars dataset, which notably does not contain any 'NoData' values and features events centered within the frame. This design presented a significant challenge in this research, as integrating 'NoData' handling became essential across all stages of the processing and model pipelines, appearing in the training and inference stages, as well as the testing stage, as detailed by the TypeError: unexpected keyword argument 'NoData' in the previous section. Compounding the problem, the pipeline creators did indeed include 'NoData' image and label parameter initialization in their config files, but did not integrate these values into the model's functionality, leaving it to the reader to do so.

For instance, as illustrated in Figure 5, 'NoData' values were misinterpreted by the model during inference, causing erroneous outputs where 'NoData' regions were processed as valid data points. This misinterpretation was particularly problematic during the normalization stage of data processing, where 'NoData' values, though flagged, were not treated distinctly within the code, resulting in normalized values that skewed the dataset with large, anomalous negative values, overshadowing valid data.



(a) Example of early inference on Sentinel-2 scene.

(b) Input Sentinel-2 scene.

Figure 5: Incorrect NoData processing by Prithvi hindered initial inference attempts.

This issue was prevalent within much of the initial processing conducted and was not resolved until 'NoData' values were integrated into each stage of the model training, testing, and inference. Custom-built pipeline functions for the Prithvi model, such as TensorNormalise, needed to be redeveloped and tailored to accommodate the data used in this research. This integration proved to be a time-consuming process, requiring extensive debugging and validation. Ensuring that 'NoData' values were appropriately handled at every stage was crucial to maintaining the integrity and accuracy of the model's outputs. The resolution of this issue, while labor-intensive, was critical for the successful application of the Prithvi model to datasets that include 'NoData' values,

highlighting the importance of adaptable and robust preprocessing pipelines in machine learning workflows.

Output Quality

A notable challenge encountered was the production of output images that indicated the Prithvi model might not have fully converged during its training phase. These output images, an example of which is shown in 6 exhibited blockiness and striping patterns during inference, signaling potential issues with model convergence.

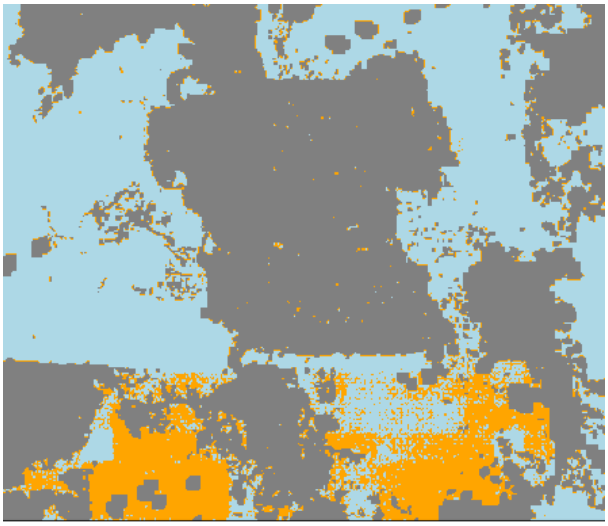


Figure 6: Early Inference for Prithvi 15k alert model showing striping pattern, indicating non-convergence in the model.

Upon investigating the training process, attention was directed towards the optimization settings in the configuration file, particularly for the burn scars scenario on which the training was based. It was discovered that the Adam optimizer's warm-up period in the configuration was initially set to 1500 iterations. This duration was excessively long, especially considering that the training was designed to run for only 500 to 1000 iterations. Such a lengthy warm-up period meant that the optimizer spent too much time adjusting its learning rate, which would impede the model's ability to converge effectively within the allocated training iterations.

To rectify this, the warm-up period was reduced to 100 iterations or 20% of the total iterations utilized here, aligning with research conducted by Izsak et al. (2021) which indicted for NLP models with much larger training periods of several days, that the warm-up phase should constitute approximately 2 to 12% of the total training duration. This adjustment intended to provide the model with an adequate phase to fine-tune the learning rate during the initial stages of training without prolonging the period. Such an optimization was crucial for allowing the model to progress into the main training phase with a better-suited learning rate, thereby facilitating improved convergence and yielding higher-quality output imagery. This change underscored the importance of carefully calibrating the training process, particularly the optimizer's settings, to enhance model performance and output fidelity.

4 Results

In this chapter, the outcomes of this research are presented. These findings help in answering the research questions detailed in Section 1.3 and are analyzed in the following discussion section.

4.1 Model Performance

Table 4 lists the results showcasing the performance of the Prithvi model and the U-Net model at detecting forest disturbances for two dataset sizes detailed in Table 1. It further details the performance of these models when the two SWIR bands of Sentinel-2 are replaced with SAR backscatter and InSAR coherence bands. Finally, ta single run was sequentially trained on burn scars detection followed by forest disturbance detection.

Model	mAcc (%)	mIoU (%)	Forest		Disturbed Forest	
			Avg.IoU (%)	Avg.Acc (%)	Avg.IoU (%)	Avg.Acc (%)
U-Net 15k alerts coherence	69.287	**60.053**	92.237	96.080	**27.877**	42.493
Prithvi 15k alerts	67.067	59.357	92.667	96.883	26.047	37.257
U-Net 15k alerts backscatter	**69.823**	58.430	91.053	94.697	25.807	**44.947**
U-Net 15k alerts	64.830	59.370	93.596	98.250	25.136	31.410
Prithvi 15k alerts backscatter	64.520	58.473	93.100	97.737	23.843	31.303
Prithvi 15k alerts coherence	61.977	57.017	93.283	98.307	20.753	25.647

Table 3: Performance evaluation results on unseen test data. Avg.: Average. Acc: Accuracy. mAcc: mean Accuracy mIoU: mean Intersection over Union.

Model	Disturbed Forest	
	Avg.IoU (%)	Avg.Acc (%)
U-Net 15k alerts coherence	**27.877**	42.493
Prithvi 15k alerts swir	26.047	37.257
U-Net 15k alerts backscatter	25.807	**44.947**
U-Net 15k alerts swir	25.136	31.410
Prithvi 15k alerts backscatter	23.843	31.303
Prithvi 15k alerts coherence	20.753	25.647

Table 4: Performance evaluation results on unseen test data. Avg.: Average. Acc: Accuracy. mAcc: mean Accuracy mIoU: mean Intersection over Union.

4.2 Visual Analysis

Figures 7 - 16 present inference conducted on various test sites, utilizing some of the models found in Table 4. Each of these input RGB, backscatter, and coherence composites are globally normalized images utilized for model training. Ground truth refers to RADD alert labels in this case.

Prithvi Model Visualization Images 7 and 8 are of two distinct test locations with inference applied from the Prithvi model, the first of which, tile T49MDU contains dense RADD alerts, while the second tile T49MKE contains a larger number of dispersed, small alerts.

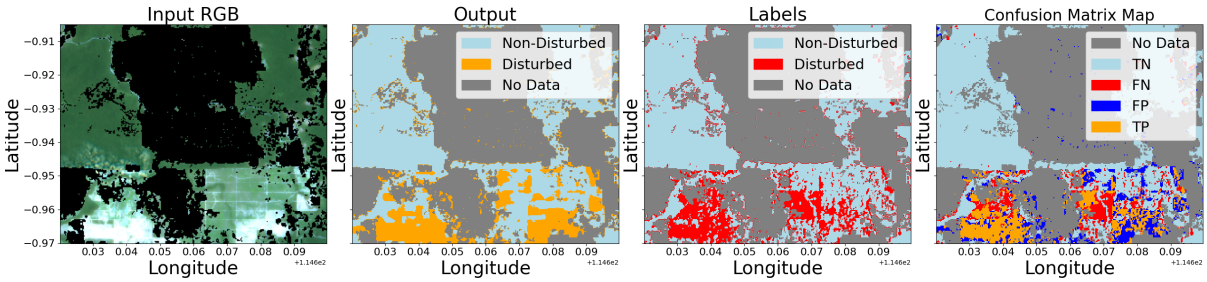


Figure 7: Prithvi 15k model inference on test data tile T49MDU, with input RGB, inference output, RADD labels, and confusion matrix map.

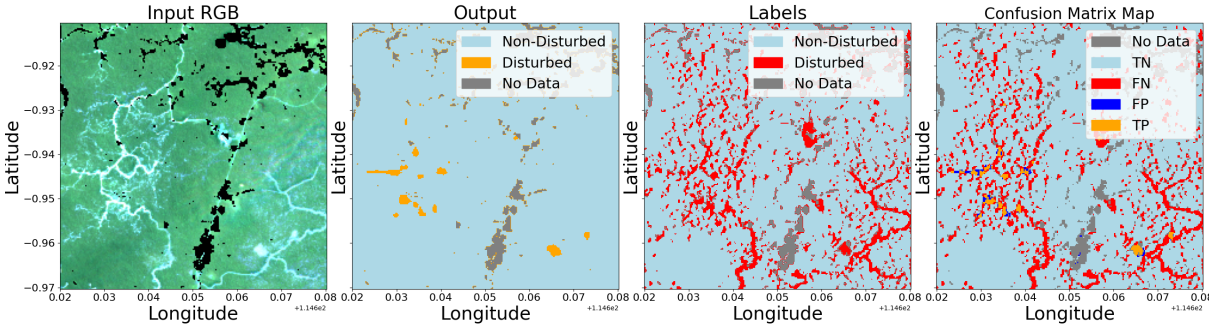


Figure 8: Prithvi 15k model inference on test data tile T49MKE, with input RGB, inference output, RADD labels, and confusion matrix map.

Prithvi Model, burn scars fine-tuning visualization

Images 9 and 10 are of two distinct test locations with inference applied from the Prithvi model with burn scars fine-tuning, the first of which, tile T49MDU contains dense RADD alerts, while the second tile T49MKE contains a larger number of dispersed, small alerts.

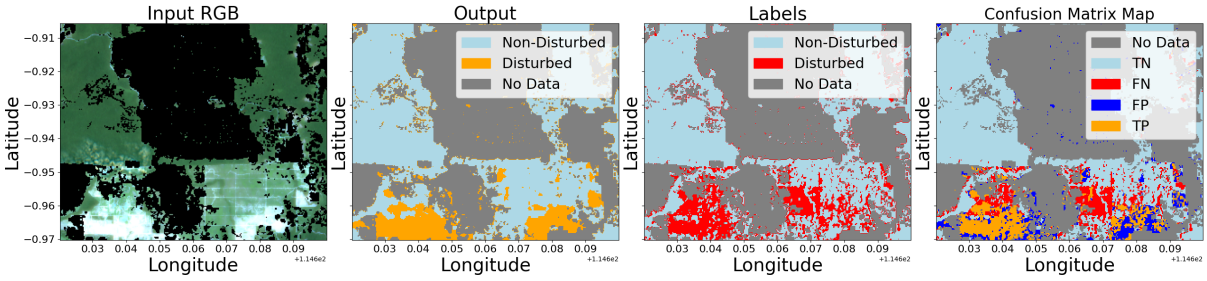


Figure 9: Prithvi burn scars 15k model inference on test data tile T49MDU, with input RGB, inference output, RADD labels, and confusion matrix map.

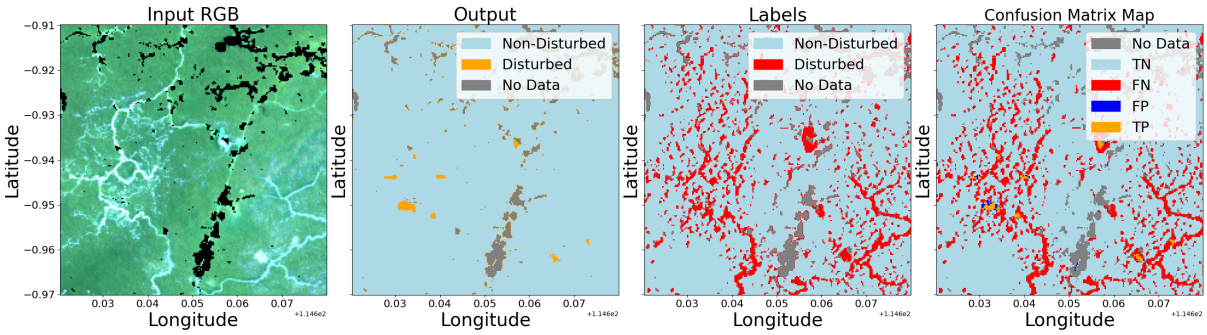


Figure 10: Prithvi burn scars 15k model inference on test data tile T49MKE, with input RGB, inference output, RADD labels, and confusion matrix map.

U-Net Visualization

Images 11 and 12 are of two distinct test locations with inference applied from the U-Net model, the first of which, tile T49MDU contains dense RADD alerts, while the second tile T49MKE contains a larger number of dispersed, small alerts.

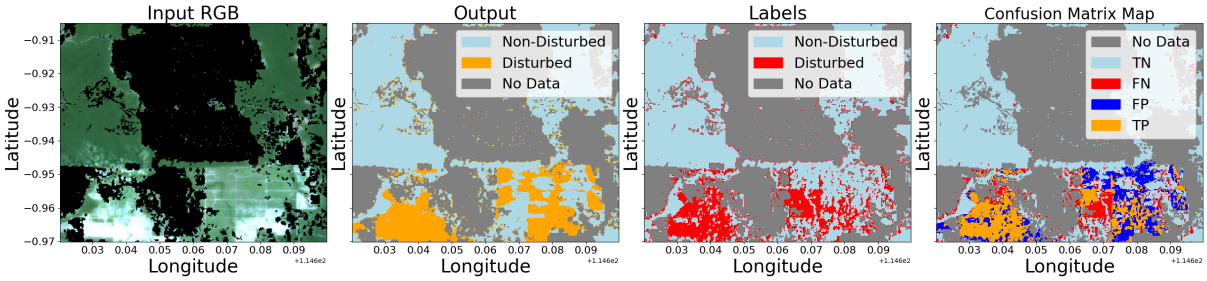


Figure 11: U-Net 15k model inference on test data tile T49MDU, with input RGB, inference output, RADD labels, and confusion matrix map.

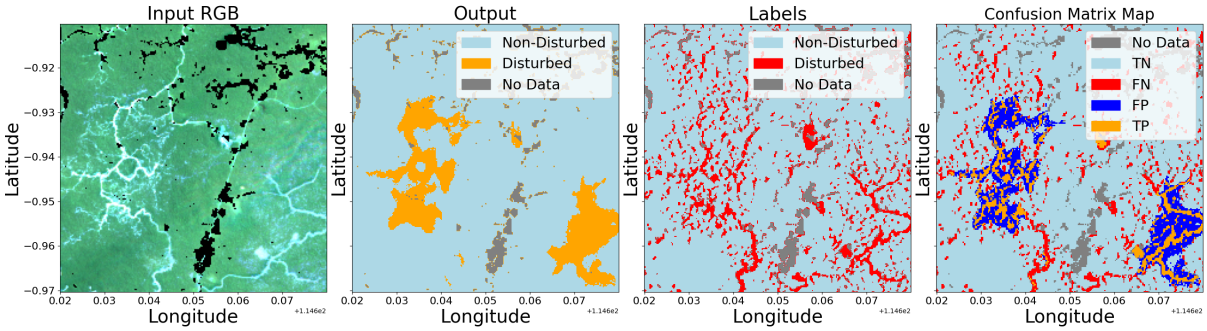


Figure 12: U-Net 15k model inference on test data tile T49MKE, with input RGB, inference output, RADD labels, and confusion matrix map.

Prithvi & U-Net Backscatter visualization

Images 13 and 14 are both of the same test location with inference applied from both the Prithvi and U-Net models, both utilizing SAR backscatter data in the place of SWIR. The scene from tile T49MET seen here, contains some regions of dense RADD alerts.

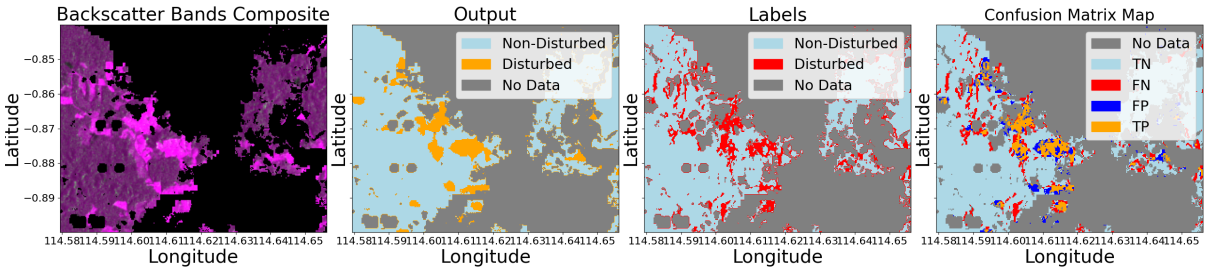


Figure 13: Prithvi backscatter 15k model inference on test data tile T49MET, with backscatter VV and VH, inference output, RADD labels, and confusion matrix map.

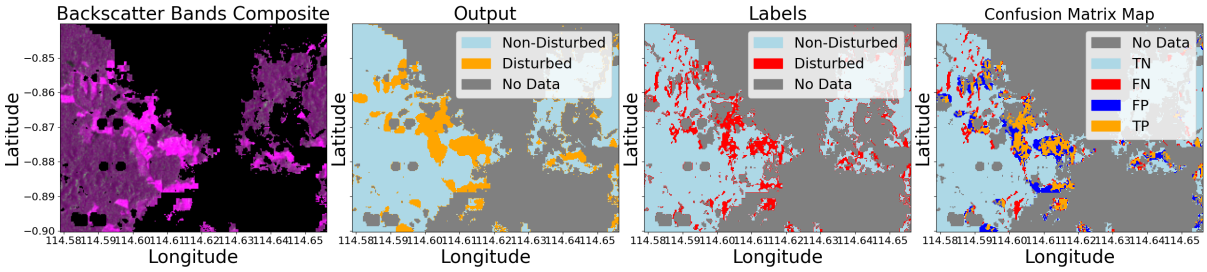


Figure 14: U-Net backscatter 15k model inference on test data tile T49MET, with backscatter VV and VH bands, inference output, RADD labels, and confusion matrix map.

Prithvi & U-Net Coherence visualization

Images 15 and 16 are both of the same test location with inference applied from both the Prithvi and U-Net models, both utilizing InSAR coherence data in the place of SWIR. The scene from tile T49MET seen here, contains some regions of dense RADD alerts.

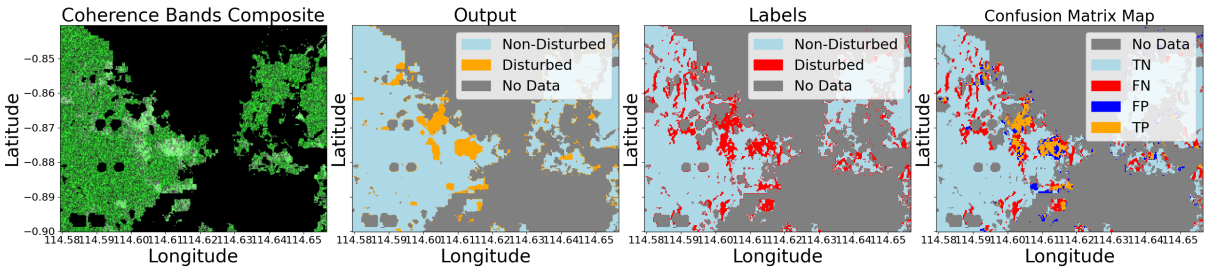


Figure 15: Prithvi coherence 15k model inference on test data tile T49MDU, with coherence VV and VH bands, inference output, RADD labels, and confusion matrix map.

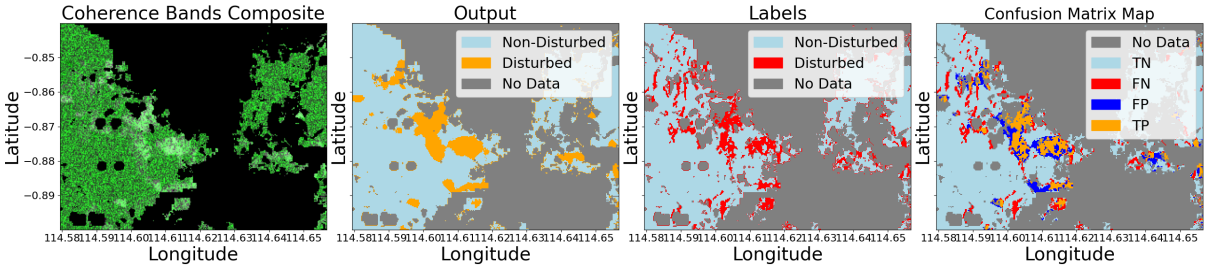


Figure 16: U-Net coherence 15k model inference on test data tile T49MET, with coherence VV and VH bands, inference output, RADD labels, and confusion matrix map.

5 Discussion

Summary
The objective of this study was to evaluate the Prithvi-100m model's efficacy in fine-tuning for forest disturbance detection over Borneo using Sentinel-2 data and RADD alerts and to examine the model's ability to generalize to new, unseen data types during fine-tuning. The findings reveal that the Prithvi model, even with sparse data comprising only 5-8% forest disturbance coverage across 30 to 58 images, displayed commendable performance. Notably, it outperformed a Sentinel-2 based U-Net model, particularly under the constraints of limited and imbalanced data. This underscores the Prithvi model's potential robustness and applicability in real-world environmental monitoring.
Furthermore, this study aimed to gauge the Prithvi model's generalization to different data types. Specifically, the substitution of SWIR bands with SAR backscatter and InSAR coherence data. Results indicated the model's difficulty in adapting to this new data. The Prithvi model experienced a reduction in average IoU of 2% and 6% for backscatter and coherence data respectively. Notably, the Prithvi model showed distinct difficulty in interpreting and fitting to InSAR coherence data, indicating the distinct challenges that this data type comprises when compared to SWIR data. Conversely, the U-Net model utilizing InSAR coherence showcased an enhanced ability to determine disturbed forest, achieving the highest average IoU of 27.8%. Similarly, the U-Net model with backscatter data exhibited increased performance vs the SWIR-based model, achieving the highest average accuracy with 44.9%, an increase of 13% over the SWIR model. Both of these outcomes not only emphasize the potential of SAR and InSAR data in forest disturbance detection but also highlight the need for tailored approaches in model training and data integration to harness the full capabilities of these advanced data sources in environmental monitoring tasks.

5.1 Prithvi Model Performance

In this research, the change in data volume between the two datasets found in Table 1 did not significantly impact the Prithvi model's performance, aligning with findings by Jakubik et al. (2023), where substantial data reduction—up to 75% for wildfire scar detection and 87.5% for flood mapping did not notably impact the resulting disturbed forest IoU. This research utilized a notably sparse dataset with only 30 to 58 images, where the disturbed forest class accounted for 5-8% of the data at 15,000 alerts, featuring extensive no-data regions. Despite these limitations, the Prithvi model demonstrated commendable performance achieving an Avg. IoU of 26% and Avg. accuracy of 37% for 30 images, illustrating its capability to learn from limited data and underscoring its potential robustness in various environmental contexts. These results are reflected in visual outputs Figures 7 and 8.
The datasets in research by Jakubik et al. (2023) particularly for burn scar detection, consisted of 804 images with a class distribution of 11% burn scar to 88% non-burned areas , a stark contrast to this study's dataset sparsity. Similarly, in research conducted by Li et al. (2023) using the Sen1Floods11 dataset which comprises of 446 tiles, also contrasts with the current study's data sparsity. The effectiveness of the model in this context is noteworthy, managing to extract relevant features from minimal data, a characteristic that is crucial for practical applications where extensive, balanced datasets may not always be available.

The difference in data modality, in particular, the disparity between the label acquisition process (utilizing 10m resolution SAR data in the RADD alert system, resampled to 30m) and the analysis data (30m resolution HLS data from Sentinel-2, encompassing RGB and additional spectral bands), significantly influenced the Prithvi model's performance. This modality difference potentially resulted in a label-data correlation mismatch, with Sentinel-2 data potentially not reflecting the disturbances that SAR data detects. This discrepancy underscores a critical area for methodological improvement in model training and data alignment, and is potentially the main contributor to the relatively low IoU and Accuracy values observed across Table 4.

Transfer learning, a powerful tool in machine learning, exhibits limitations when sequentially fine-tuning models on distinct tasks, as evidenced by the Prithvi model's performance dip when transitioning from burn scar detection to forest disturbance detection. The burn scar dataset utilized by Jakubik et al. (2023), with its large singular and centered burn scars, starkly contrasts with the target dataset's characteristics, which feature numerous smaller and dispersed disturbances typical of wet tropical rainforests. This significant disparity in data characteristics likely hampers the model's ability to adapt the learned features from the burn scar context to the more complex and varied forest disturbance patterns.

The performance degradation observed in the Prithvi model, particularly in detecting low-density scattered events such as observed in Figure 10, underscore the challenges of applying features learned from one domain (dry areas prone to burns) to another (wet tropical rainforests). Figure 9 reflects the model's efficacy in areas resembling the burn scar dataset's characteristics suggesting that while transfer learning has occurred, it is not wholly beneficial for the current application. The distinct vegetation types and disturbance patterns between the datasets potentially lead to a negative transfer, where knowledge from the burn scar domain may interfere with or obscure relevant features for forest disturbance detection.

These challenges observed in transitioning from burn scar to forest disturbance detection align with broader findings in machine-learning research regarding "concept forgetting" during sequential fine-tuning. As highlighted by Mukhoti et al. (n.d.), which indicates that once fine-tuned for a particular task such as burn scar detection, a model may exhibit reduced capability in subsequent tasks such as forest disturbance detection, compared to if it were directly fine-tuned on the latter from its original pre-trained state. The prior fine-tuning on burn scars is observed to impair the model's ability to effectively learn and adapt to the new context of forest disturbances in tropical rainforests, reflecting a reduction in Avg. IoU and Avg. accuracy by 3% and 7% respectively compared to the 15,000 alert Prithvi model.

5.2 U-Net Model Performance Comparison

The observed performance metrics in Table 4 reveal that the Prithvi model outperforms the U-Net model for both the 10,000 and 15,000 RADD alert datasets, highlighting a nuanced difference in their capabilities. While the Avg. IoU for disturbed forest only shows a marginal 1% difference between the models, a more significant disparity of 5-6% in accuracy suggests that Prithvi is more effective in correctly identifying disturbed forest areas, despite the close Avg. IoU values. This difference could indicate that while U-Net is proficient in segmenting the general area of disturbed forests, as reflected in IoU, it may be less adept at precisely classifying individual pixels within those areas.

U-Net shows strong performance at similar tasks when compared to the Prithvi model. There are numerous examples of U-Net's strong performance in semantic segmentation of forested regions and disturbances, achieving high Avg. IoU values (Pyo et al., 2022; Wagner et al., 2019). Li et al. (2023) found that when tested on flood detection and compared, a U-Net model obtained an Avg. IoU of 70.5%, 6% less than that of the Prithvi model. The differences in performance metrics between U-Net and Prithvi observed here suggest that while U-Net is robust, there are scenarios where Prithvi's advanced features may offer an edge, which is particularly crucial for precise environmental monitoring and disturbance detection. Figures 11 and 12 show the U-Net model performance over two scenes. An over-estimation of central disturbances, in particular, is observed in Figure 12, when compared to 8, in which we see an underestimation of these similar central disturbances.

Given the Prithvi model's advanced pre-training, it was anticipated to significantly outperform U-Net, particularly in smaller sample sets where effective learning from limited data is crucial. However, the relatively narrow performance gap may point to the influence of data disparity issues. The Prithvi model, despite its pre-training advantages, is potentially constrained by the same data limitations affecting U-Net, particularly the discordance between Sentinel-2 analysis data and Sentinel-1 derived labels. This mismatch might be obscuring the full extent of Prithvi's capabilities. Despite this, the modest performance lead over U-Net underscores the value of utilizing a pre-trained foundation model, which demonstrates improvements in detecting forest disturbances compared to U-Net.

5.3 Model Adaption to Unseen Data

SWIR bands are particularly effective in distinguishing variations in moisture content, which are crucial for identifying changes in forest cover and disturbances (Holzman et al., 2021). SAR backscatter provides insights based on the interaction of radar signals with surface features, offering details on surface roughness and geometric structure (Musthafa et al., 2021). InSAR coherence, tracks the change in SAR information over time, providing a measure of temporal stability or change (Kim & van Zyl, 2000). Both the Prithvi and U-Net models underwent a significant alteration in their input data, where the two SWIR bands used in Sentinel-2 HLS data were replaced with the VV and VH polarizations from Sentinel-1 SAR backscatter and InSAR coherence data. This change represents a fundamental shift in the type of information the models are processing.

Upon this new data type introduction, the Prithvi model demonstrated a notable decline in its ability to detect disturbances, attributed largely to the model's significant pre-training. This resulted in some of the lowest average Intersection over Union (Avg.IoU) values for the disturbed forest class observed in Table 4. Specifically, the capacity reduction ranges between 2-5% in Avg.IoU, with InSAR coherence data showing a further decrease of 3% from backscatter data. This notable drop underscores the model's difficulty in adapting to data with vastly different data distributions. Figures 13 and 15 represent inference of the Prithvi backscatter and coherence models respectively, with both showing a significant under-determination of labeled regions.

In contrast, the U-Net model's performance was positively influenced by the data alteration. The introduction of backscatter data led to a modest improvement in Avg. IoU of 1%, but a pronounced increase in Avg. accuracy of 13% for the disturbed forest class, leading to the highest resulting Avg. Accuracy of the models tested, with 44.9%. This

result is in line with the expectation that SAR backscatter data will align strongly with SAR-based labels. InSAR coherence data yielded a notable enhancement in performance over the U-Net model with SWIR, with an increase of 3% in Avg.IoU and a significant 11% rise in Avg.Accuracy. These outcomes suggest that the transition from SWIR to InSAR coherence data is beneficial for forest disturbance detection within the U-Net model, highlighting the potential advantages of leveraging radar-based data in environmental monitoring tasks. Example inference of these results are visualized in Figures 14 and 16 for backscatter and coherence respectively.

While SAR data has been shown to perform well at interpreting land use classes, Sentinel-1 coherence measures have been shown to outperform backscatter intensity in detecting clear-cut areas (Akbari & Solberg, 2022). Research conducted by (Jacob et al., 2020) demonstrated that when compared to backscatter intensities, coherence measures resulted in higher accuracies in various land classification algorithms. Similar research has been conducted showcasing the strength of InSAR coherence in classifying a variety of land use classes with an accuracy of 90%, further underscoring the potential of InSAR coherence for land use classification (Engdahl & Hyyppa, 2003). The results of this research, with the U-Net coherence-based model achieving the highest Avg. IoU of the models studied, further add to this growing consensus on the strength and potential of InSAR coherence for forest disturbance mapping.

6 Recommendations

Potential gains in performance may be made by investing in foundation models that are pre-trained in different regions of the world. The Prithvi model utilized here is pre-trained on contiguous data from the U.S. which is oftentimes markedly different from those found in the tropics. Therefore, foundation models that are pre-trained on data from predominantly tropical regions may benefit the application of forest disturbance detection (Jakubik et al., 2023).

Furthermore, the intersection of InSAR data and foundation models presents a novel avenue in forest disturbance detection. While applications of InSAR data and foundation models are well-established, their combined use remains less explored, particularly in contrast to RGB-based geospatial foundation models. Research like Han et al. (2024) has explored the potential of multi-sensor geospatial foundation models, albeit with a focus on SAR backscatter data, yielding promising results on a multitude of tasks. The research conducted here underscores the potential use of InSAR-based foundation models, suggesting that their combination could offer significant enhancements in forest disturbance detection and environmental monitoring, marking an important direction for future research in this vital domain.

7 Conclusion

The Prithvi model, integrating vision transformer and masked autoencoder architectures, represents a novel approach in the study of forest disturbance detection, an area where foundation models have been under-explored. The need for advanced data-driven solutions are underscored by the critical challenges facing tropical rainforests in Borneo and other regions, where accurate detection and analysis of disturbances are essential for conservation efforts and ecosystem management. This research suggests the expansion

of the application of foundation models like Prithvi in forest disturbance detection, addressing a significant gap in current environmental monitoring practices.
To address this, this study developed a dataset based on RADD alerts and Sentinel-2 HLS data, targeting forest disturbance detection over Borneo. The analysis compared the Prithvi model against a U-Net model and examined the effects of integrating SAR and InSAR data into the Prithvi model during fine-tuning by replacing the two Sentinel-2 SWIR bands. The findings indicate that foundation models like Prithvi show substantial potential in forest disturbance detection, delivering commendable results even with small and sparsely labeled datasets. Specifically, Prithvi achieved the second and third highest Avg. IoU values of 26.1% and 26% for dataset sizes of 58 and 30 images, respectively. Moreover, when InSAR data was applied with the U-Net model, this model demonstrated exceptional efficacy, achieving the highest Avg. IoU value of 27.8%, on 30 images. Conversely, the Prithvi model exhibited poor performance when introduced to SAR and InSAR data, highlighting the critical role of the pre-training stage for foundation models. These outcomes affirm the effectiveness of the Prithvi model in detecting forest disturbances and illustrate the potential of InSAR data in this domain.
This research underscores the role that foundation models may play in the future of forest disturbance detection. Their demonstrated potential is a promising avenue for advancing the accuracy and efficiency of environmental monitoring. Directions for future research include exploring the development of pre-trained foundation models specifically tailored to tropical regions, which could enhance their effectiveness and adaptability to the unique challenges of these environments. Additionally, the promising results with InSAR data suggest that foundation models designed to leverage this type of data could represent a significant step forward in detecting and analyzing forest disturbances, offering valuable insights for conservation strategies and the sustainable management of forested landscapes.

8 Additional Information

Software & Hardware Used
The Prithvi-100m model is built on OpenMM Lab architecture. OpenMM Lab provides libraries, wrappers, and toolchains built on top of the Pytorch deep learning library, and is hosted in Python (OpenMMLab, 2023). Training was conducted on a single 4090 NVIDIA GPU.

Time Schedule
Figure 17 details the time schedule for this internship. A significant amount of time is given to the first research question, as much of the data acquisition, pre-processing and model analysis will be conducted in this period. A significant portion of time is also given to writing up the report, given the depth of knowledge required for the study subject, length of report and complicated results that may appear.
A midterm meeting will be conducted in early to mid-January, with a midterm presentation to be given at VITO near the end of January. Final submission will occur by the 20th of April, with an oral presentation and reflection paper to follow.

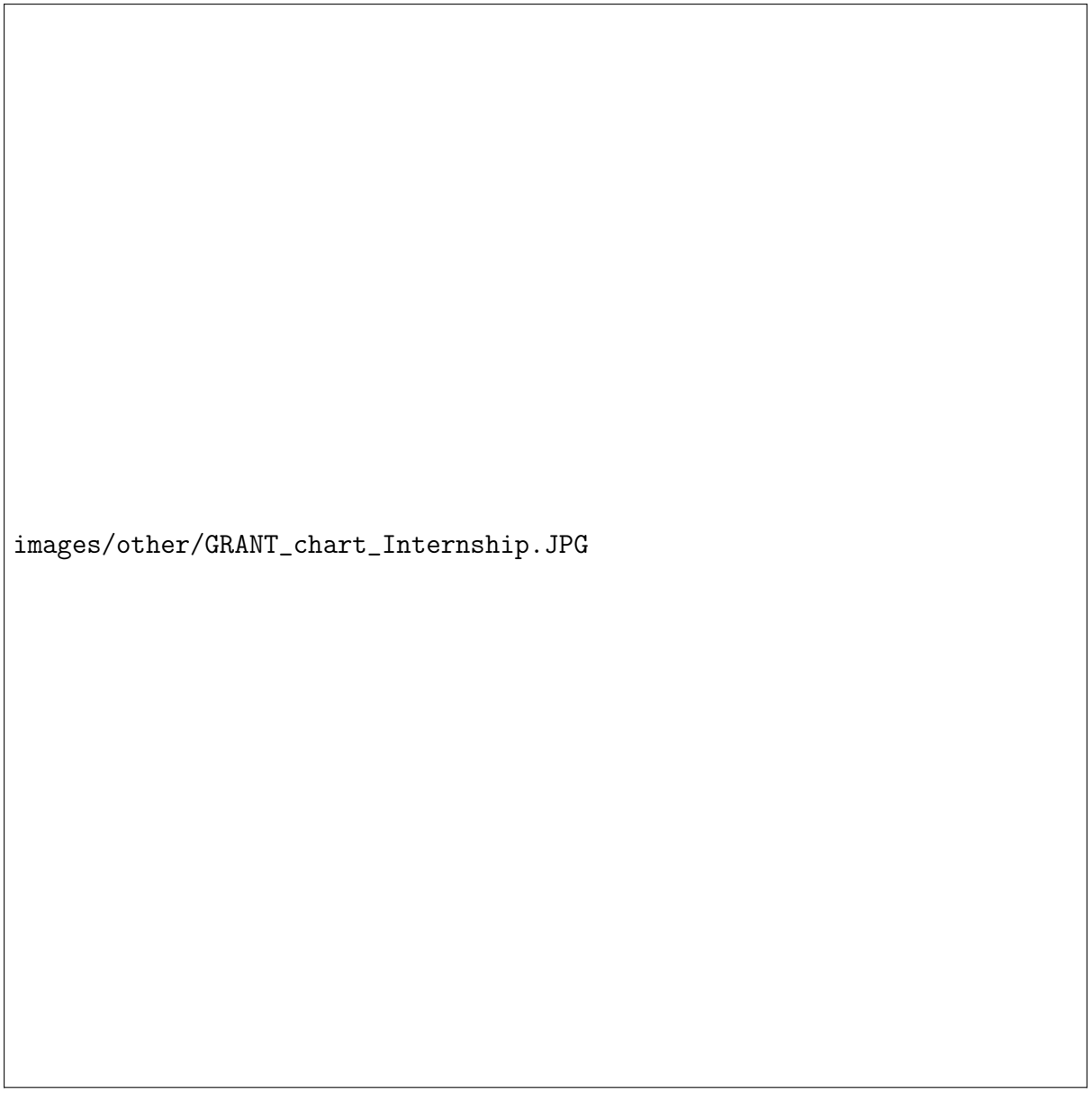


Figure 17: Grant chart, detailing internship steps with task length in orange, important dates in red and vacation days in blue.

Feasibility

The size of the model decoder, as well as data requirements, may add some feasibility issues to this project. If further refinement of the model fine-tuning parameters are needed, this would require multiple re-runs. On the burn scars sample dataset supplied, training time averaged 1-1.5 days on a single 3070 GPU. With this in mind, the parameters from the burn scars example will initially be heavily relied upon ("Prithvi-100M burn scar", 2023).

A limitation appears also from the data labeling utilizing the RADD alert detections. This detection system details an exact date for detections. However, there is little literature that investigates the actual length of a forest disturbance event. Thus, when connecting a RADD alert detection to an HLS S30 image by date, there is no definitive date range that defines when a disturbance occurs. With data requirements in the center of the data labeling step, an initial attempt will be made to correlate the detection dates

to Sentinel-2 images as closely as possible, with increasing difference between detection date and imagery date, until sufficient data is obtained for model training. This may lead to inaccuracies in the dataset, where alerts are not visually present in the Sentinel-2 images.

Data Co-registration is a significant step in multi-modal data fusion and carries with it risks of inherited errors propagating throughout both the fine-tuning and analysis steps. Here, a visual quality assessment will be conducted on images, to ensure that no significant geolocation errors are occurring. The persistence of these errors could cause significant feasibility issues for the third research question. Area-based image matching processes may be explored to minimize these effects if time permits, however, the data fusion step may incur significant challenges to this research question.

University Supervisor	Name: Johannes Reiche
	Address: Droevendaalsesteeg 3, Wageningen, Gelderland
	Email: Johannes.reiche@wur.nl
Internship Supervisor	Name: Xenia Ivashkovych
	Function: Remote Sensing Data Scientist
	Address: Boertang 280 (TAP), Mol, Belgium
	Email: xenia.ivashkovych@vito.be

Table 5: Information on Supervisors

References

Akbari, V., & Solberg, S. (2022). Clear-cut detection and mapping using sentinel-1 backscatter coefficient and short-term interferometric coherence time series. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1–5. https://doi.org/10.1109/lgrs.2020.3039875

Angelsen, A. (2016). REDD as result-based aid: General lessons and bilateral agreements of norway. *Review of Development Economics*, *21*(2), 237–264. https://doi.org/10.1111/rode.12271

Ballère, M., Bouvet, A., Mermoz, S., Le Toan, T., Koleck, T., Bedeau, C., André, M., Forestier, E., Frison, P.-L., & Lardeux, C. (2021). Sar data for tropical forest disturbance alerts in french guiana: Benefit over optical imagery. *Remote Sensing of Environment*, *252*, 112159. https://doi.org/10.1016/j.rse.2020.112159

Bank, D., Koenigstein, N., & Giryes, R. (2020). Autoencoders. https://arxiv.org/pdf/2003.05991.pdf

Başkent, E. Z., Keleş, S., Kadıoğulları, A. İ., & Bingöl, Ö. (2010). Quantifying the effects of forest management strategies on the production of forest values: Timber, carbon, oxygen, water, and soil. *Environmental Modeling &amp Assessment*, *16*(2), 145–152. https://doi.org/10.1007/s10666-010-9238-y

Bouvet, A., Mermoz, S., Ballère, M., Koleck, T., & Toan, T. L. (2018). Use of the SAR shadowing effect for deforestation detection with sentinel-1 time series. *Remote Sensing*, *10*(8), 1250. https://doi.org/10.3390/rs10081250

Boyce, P. C., Wong, S., Ting, A., Low, S., Low, S., Ng, K., & Ooi, I. (2010). The araceae of borneo–the genera. *Aroideana*, *33*, 3–74.

Bunting, P., Lucas, R., Rosenqvist, A., Rebelo, L.-M., Hilarides, L., Thomas, N., Hardy, A., Itoh, T., Shimada, M., & Finlayson, M. (2018). The global mangrove watch - a new 2010 baseline of mangrove extent. *Remote Sensing*, *10*, 1669. https://doi.org/10.3390/rs10101669

Cha, K., Seo, J., & Lee, T. (2023). A billion-scale foundation model for remote sensing images. https://arxiv.org/pdf/2304.05215.pdf

Chung, M. G., Frank, K. A., Pokhrel, Y., Dietz, T., & Liu, J. (2021). Natural infrastructure in sustaining global urban freshwater ecosystem services. *Nature Sustainability*, *4*(12), 1068–1075. https://doi.org/10.1038/s41893-021-00786-4

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., & Justice, C. (2018). The harmonized landsat and sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, *219*, 145–161. https://doi.org/10.1016/j.rse.2018.09.002

Cole, L. E. S., Bhagwat, S. A., & Willis, K. J. (2014). Recovery and resilience of tropical forests after disturbance. *Nature Communications*, *5*(1). https://doi.org/10.1038/ncomms4906

Corbera, E., & Schroeder, H. (2011). Governing and implementing REDD. *Environmental Science &amp Policy*, *14*(2), 89–99. https://doi.org/10.1016/j.envsci.2010.11.002

Cramer, W., Bondeau, A., Schaphoff, S., Lucht, W., Smith, B., & Sitch, S. (2004). Tropical forests and the global carbon cycle: Impacts of atmospheric carbon dioxide, climate change and rate of deforestation (Y. Malhi & O. L. Phillips, Eds.). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1443), 331–343. https://doi.org/10.1098/rstb.2003.1428

Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. https://arxiv.org/pdf/2010.11929.pdf

Durieux, A. M., Calef, M. T., Arko, S., Chartrand, R., Kontgis, C., Keisler, R., & Warren, M. S. (2019, September). Monitoring forest disturbance using change detection on synthetic aperture radar imagery. In M. E. Zelinski, T. M. Taha, J. Howe, A. A. Awwal, & K. M. Iftekharuddin (Eds.), *Applications of machine learning*. SPIE. https://doi.org/10.1117/12.2528945

Engdahl, M., & Hyyppa, J. (2003). Land-cover classification using multitemporal ERS-1/2 insar data. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(7), 1620–1628. https://doi.org/10.1109/tgrs.2003.813271

Ferraz, A., Saatchi, S., Xu, L., Hagen, S., Chave, J., Yu, Y., Meyer, V., Garcia, M., Silva, C., Roswintiart, O., Samboko, A., Sist, P., Walker, S., Pearson, T. R. H., Wijaya, A., Sullivan, F. B., Rutishauser, E., Hoekman, D., & Ganguly, S. (2018). Carbon storage potential in degraded forests of kalimantan, indonesia. *Environmental Research Letters*, *13*(9), 095001. https://doi.org/10.1088/1748-9326/aad782

Foumelis, M., Blasco, J. M. D., Desnos, Y.-L., Engdahl, M., Fernandez, D., Veci, L., Lu, J., & Wong, C. (2018). Esa snap - stamps integrated processing for sentinel-1 persistent scatterer interferometry. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. https://doi.org/10.1109/igarss.2018.8519545

Geudtner, D., Torres, R., Snoeij, P., Davidson, M., & Rommen, B. (2014). Sentinel-1 system capabilities and applications. *2014 IEEE Geoscience and Remote Sensing Symposium*. https://doi.org/10.1109/igarss.2014.6946711

Goldstein, J. E. (2015). Knowing the subterranean: Land grabbing, oil palm, and divergent expertise in indonesia's peat soil. *Environment and Planning A: Economy and Space*, *48*(4), 754–770. https://doi.org/10.1177/0308518x15599787

Han, B., Zhang, S., Shi, X., & Reichstein, M. (2024). Bridging remote sensors with multisensor geospatial foundation models. http://arxiv.org/pdf/2404.01260

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Hansen, M. C., Krylov, A., Tyukavina, A., Potapov, P. V., Turubanova, S., Zutta, B., Ifo, S., Margono, B., Stolle, F., & Moore, R. (2016). Humid tropical forest disturbance alerts using landsat data. *Environmental Research Letters*, *11*(3), 034008. https://doi.org/10.1088/1748-9326/11/3/034008

Hayasaka, H., Noguchi, I., Putra, E. I., Yulianti, N., & Vadrevu, K. (2014). Peat-fire-related air pollution in central kalimantan, indonesia. *Environmental Pollution*, *195*, 257–266. https://doi.org/10.1016/j.envpol.2014.06.031

He, K., Chen, X., Xie, S., et al. (2021). Masked autoencoders are scalable vision learners. https://arxiv.org/pdf/2111.06377.pdf

Holzman, M., Rivas, R., & Bayala, M. (2021). Relationship between tir and nir-swir as indicator of vegetation water availability. *Remote Sensing*, *13*(17), 3371. https://doi.org/10.3390/rs13173371

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. https://arxiv.org/pdf/1502.03167.pdf

Izsak, P., Berchansky, M., & Levy, O. (2021). How to train bert with an academic budget. http://arxiv.org/pdf/2104.07705

Jacob, A. W., Vicente-Guijalba, F., Lopez-Martinez, C., Lopez-Sanchez, J. M., Litzinger, M., Kristen, H., Mestre-Quereda, A., Ziolkowski, D., Lavalle, M., Notarnicola, C., Suresh, G., Antropov, O., Ge, S., Praks, J., Ban, Y., Pottier, E., Franquet, J. J. M., Duro, J., & Engdahl, M. E. (2020). Sentinel-1 InSAR coherence for land cover mapping: A comparison of multiple feature-based classifiers. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 535–552. https://doi.org/10.1109/jstars.2019.2958847

Jakubik, J., & Chu. (2023). Hls foundation. https://doi.org/10.57967/hf/0952

Jakubik, J., Roy, S., Phillips, C. E., et al. (2023). Foundation models for generalist geospatial artificial intelligence. https://arxiv.org/pdf/2310.18660.pdf

Keydel, W. (1992). Basic principles of sar. *In AGARD*.

Keyes. (2023). Assessing sentinel-1 coherence measures for tropical forest disturbance mapping. *MSc Thesis*, -, -.

Kim, Y., & van Zyl, J. (2000). Overview of polarimetric interferometry. *3*, 231–236 vol.3. https://doi.org/10.1109/AERO.2000.879850

Li, W., Lee, H., Wang, S., et al. (2023). Assessment of a new geoai foundation model for flood inundation mapping. https://arxiv.org/pdf/2309.14500.pdf

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. http://arxiv.org/pdf/1708.02002

Lindenmayer, D., & McCarthy, M. A. (2002). Congruence between natural and human forest disturbance: A case study from australian montane ash forests. *Forest Ecology and Management*, *155*(1-3), 319–335. https://doi.org/10.1016/s0378-1127(01)00569-2

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. https://arxiv.org/pdf/1711.05101.pdf

Lp daac 2023 [Accessed: 2023-11-15]. (2023).

Masek, J. G. (2023). *Harmonized landsat sentinel-2 (hls) product user guide* [Available online at https://hls.gsfc.nasa.gov/documents]. Version 1.5. NASA/GSFC. Greenbelt, MD.

McNeill, D. (2015). Norway and REDD in indonesia: The art of not governing? *Forum for Development Studies*, *42*(1), 113–132. https://doi.org/10.1080/08039410.2014.997791

Michelucci, U. (2022). An introduction to autoencoders. https://arxiv.org/pdf/2201.03898.pdf

Misiukas, J. M., Carter, S., & Herold, M. (2021). Tropical forest monitoring: Challenges and recent progress in research. *Remote Sensing*, *13*(12), 2252. https://doi.org/10.3390/rs13122252

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., & Gascon, C. (2011). Global biodiversity conservation: The critical role of hotspots. In *Biodiversity hotspots* (pp. 3–22). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20992-5_1

Morris, R. J. (2010). Anthropogenic impacts on tropical forest biodiversity: A network structure and ecosystem functioning perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1558), 3709–3718. https://doi.org/10.1098/rstb.2010.0273

Mukhoti, J., Gal, Y., Torr, P., & Dokania, P. K. (n.d.). Fine-tuning can cripple foundation models; preserving features may be the solution. https://openreview.net/pdf?id=VQ7Q6qdp0P

Musthafa, M., Singh, G., & Nela, B. R. (2021). Time-series analysis of c-band and l-band sar backscatter in detecting forest disturbance and regrowth dynamics. *2021 IEEE International India Geoscience and Remote Sensing Symposium (InGARSS)*. https://doi.org/10.1109/ingarss51564.2021.9792018

Nasa earthdata search [Accessed: 2023-11-15]. (2023).

Olier, I., Orhobor, O., Vanschoren, J., & King, R. (2018). Transformative machine learning.

OpenMMLab. (2023). Openmmlab computer vision foundation [Accessed: 2023-10-31].

Oubara, A., Wu, F., Amamra, A., & Yang, G. (2022). Survey on remote sensing data augmentation: Advances, challenges, and future perspectives. In *Advances in computing systems and applications* (pp. 95–104). Springer International Publishing. https://doi.org/10.1007/978-3-031-12097-8_9

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. https://arxiv.org/pdf/1712.04621.pdf

Pillay, R., Venter, M., Aragon-Osejo, J., González-del-Pliego, P., Hansen, A. J., Watson, J. E., & Venter, O. (2021). Tropical forests are home to over half of the world's vertebrate species. *Frontiers in Ecology and the Environment*, *20*(1), 10–15. https://doi.org/10.1002/fee.2420

Prithvi-100m [Accessed: 2023-10-01]. (2023). https://huggingface.co/ibm-nasa-geospatial/Prithvi-100M

Prithvi-100M burn scar. (2023, August). https://doi.org/10.57967/hf/0953

Pulella, A., Santos, R. A., Sica, F., Posovszky, P., & Rizzoli, P. (2020). Multi-temporal sentinel-1 backscatter and coherence for rainforest mapping. *Remote Sensing*, *12*(5), 847. https://doi.org/10.3390/rs12050847

Pyo, J., Han, K.-j., Cho, Y., Kim, D., & Jin, D. (2022). Generalization of u-net semantic segmentation for forest change detection in south korea using airborne imagery. *Forests*, *13*(12), 2170. https://doi.org/10.3390/f13122170

Qin, Y., Hu, S., Lin, Y., Chen, W., Ding, N., Cui, G., Zeng, Z., Huang, Y., Xiao, C., Han, C., Fung, Y. R., Su, Y., Wang, H., Qian, C., Tian, R., Zhu, K., Liang, S., Shen, X., Xu, B., … Sun, M. (2023). Tool learning with foundation models. *ArXiv*, *abs/2304.08354*. https://api.semanticscholar.org/CorpusID:258179336

Qiu, S., Zhu, Z., & He, B. (2019). Fmask 4.0: Improved cloud and cloud shadow detection in landsats 4–8 and sentinel-2 imagery. *Remote Sensing of Environment*, *231*, 111205. https://doi.org/10.1016/j.rse.2019.05.024

Rahman, A. A. A., Majid, N. A., Ahli, N. A., Latip, A. S. A., & Taib, A. M. (2023). The capability of SNAP software application to identify landslide using InSAR technique. *Physics and Chemistry of the Earth, Parts A/B/C*, *131*, 103427. https://doi.org/10.1016/j.pce.2023.103427

Randhir, T. O., & Erol, A. (2013). Emerging threats to forests: Resilience and strategies at system scale. *American Journal of Plant Sciences*, *04*(03), 739–748. https://doi.org/10.4236/ajps.2013.43a093

Reiche, J., Mullissa, A., Slagter, B., Gou, Y., Tsendbazar, N.-E., Odongo-Braun, C., Vollrath, A., Weisse, M. J., Stolle, F., Pickens, A., Donchyts, G., Clinton, N., Gorelick, N., & Herold, M. (2021). Forest disturbance alerts for the congo basin

using sentinel-1. *Environmental Research Letters*, *16*(2), 024005. https://doi.org/10.1088/1748-9326/abd0a8

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. https://arxiv.org/pdf/1505.04597.pdf

Roy, D. P., Li, J., Zhang, H. K., Yan, L., Huang, H., & Li, Z. (2017). Examination of sentinel-2a multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, *199*, 25–38. https://doi.org/10.1016/j.rse.2017.06.019

Roy, D., Zhang, H., Ju, J., Gomez-Dans, J., Lewis, P., Schaaf, C., Sun, Q., Li, J., Huang, H., & Kovalskyy, V. (2016). A general method to normalize landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, *176*, 255–271. https://doi.org/10.1016/j.rse.2016.01.023

Ruan, B.-K., Shuai, H.-H., & Cheng, W.-H. (2022). Vision transformers: State of the art and research challenges. https://arxiv.org/pdf/2207.03041.pdf

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/https://doi.org/10.1016/j.neunet.2014.09.003

Shafique, A., Cao, G., Khan, Z., Asad, M., & Aslam, M. (2022). Deep learning-based change detection in remote sensing images: A review. *Remote Sensing*, *14*(4), 871. https://doi.org/10.3390/rs14040871

Singh, A., Kushwaha, S. K. P., & Kumar, S. (2020). Backscatter and coherence analysis using space borne c-band data for forest characterization. *14*, 39–48.

Sobien, D., Higgins, E., Krometis, J., Kauffman, J., & Freeman, L. (2022). Improving deep learning for maritime remote sensing through data augmentation and latent space. *Machine Learning and Knowledge Extraction*, *4*(3), 665–687. https://doi.org/10.3390/make4030031

Sunderlin, W. D., Angelsen, A., Belcher, B., Burgers, P., Nasi, R., Santoso, L., & Wunder, S. (2005). Livelihoods, forests, and conservation in developing countries: An overview. *World Development*, *33*(9), 1383–1402. https://doi.org/10.1016/j.worlddev.2004.10.004

Tarasiou, M. (2021). DeepSatData: Building large scale datasets of satellite images for training machine learning models. https://doi.org/10.36227/techrxiv.16558482.v1

Tuli, S., Dasgupta, I., Grant, E., et al. (2021). Are convolutional neural networks or transformers more like human vision? https://arxiv.org/pdf/2105.07197.pdf

Turubanova, S., Potapov, P. V., Tyukavina, A., & Hansen, M. C. (2018). Ongoing primary forest loss in brazil, democratic republic of the congo, and indonesia. *Environmental Research Letters*, *13*(7), 074028. https://doi.org/10.1088/1748-9326/aacd1c

Uppuluri, A., & Jost, R. (n.d.). An application of the SAR image processing toolkit: InSAR. *2006 IEEE Conference on Radar*. https://doi.org/10.1109/radar.2006.1631826

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. https://arxiv.org/pdf/1706.03762.pdf

VITO. (2023). Vision on technology for a better world | vito [Accessed: 2023-10-31].

VITORemoteSensing. (2023). Homepage | vito remote sensing [Accessed: 2023-10-31].

Wagner, F. H., Sanchez, A., Tarabalka, Y., Lotte, R. G., Ferreira, M. P., Aidar, M. P. M., Gloor, E., Phillips, O. L., & Aragão, L. E. O. C. (2019). Using the u-net convolutional network to map forest types and disturbance in the atlantic

rainforest with very high resolution images (N. Pettorelli & N. Clerici, Eds.). *Remote Sensing in Ecology and Conservation*, *5*(4), 360–375. https://doi.org/10.1002/rse2.111

Wang, M., Yang, B., Hu, F., & Zang, X. (2014). On-orbit geometric calibration model and its applications for high-resolution optical satellite imagery. *Remote Sensing*, *6*(5), 4391–4408. https://doi.org/10.3390/rs6054391

Wang, X., Chen, G., Qian, G., Gao, P., Wei, X.-Y., Wang, Y., Tian, Y., & Gao, W. (2023). Large-scale multi-modal pre-trained models: A comprehensive survey. *Machine Intelligence Research*, *20*(4), 447–482. https://doi.org/10.1007/s11633-022-1410-8

Woodhouse, I. H. (2017, July). *Introduction to microwave remote sensing*. CRC Press. https://doi.org/10.1201/9781315272573

Wu, Y., & He, K. (2018). Group normalization. https://arxiv.org/pdf/1803.08494.pdf

Yang, S., Xiao, W., Zhang, M., et al. (2022). Image data augmentation for deep learning: A survey. https://arxiv.org/pdf/2204.08610.pdf

Zhang, T., Zeng, T., & Zhang, X. (2023). Synthetic aperture radar (SAR) meets deep learning. *Remote Sensing*, *15*(2), 303. https://doi.org/10.3390/rs15020303

Zhang, T., Gao, P., Dong, H., Zhuang, Y., Wang, G., Zhang, W., & Chen, H. (2022). Consecutive pre-training: A knowledge transfer learning strategy with relevant unlabeled data for remote sensing domain. *Remote Sensing*, *14*(22), 5675. https://doi.org/10.3390/rs14225675

Zhou, C., Li, Q., Li, C., et al. (2023). A comprehensive survey on pretrained foundation models: A history from bert to chatgpt. https://arxiv.org/pdf/2302.09419.pdf

Zupanc, A. (2023). Improving cloud detection with machine learning | by anze zupanc | sentinel hub blog | medium [Accessed: 2023-10-30].