Comparative Analysis of Fine-Tuned Foundation Models for Land Cover Classification using Sentinel-2 Imagery, Study Area: Sumatra and Kalimantan, Indonesia

Wahyu Wiratama , Min Kass Chong, Yi Lin Lim, and Chun Jian Ho

ST Engineering Geo-Insights, Singapore, chunjian.ho@stengg.com

Keywords: Foundation Model, Deep Learning, CNN, Land Cover Classification, Satellite Imagery, Sentinel-2.

Abstract

Land cover classification plays a pivotal role in understanding and managing Earth's resources, influencing decisions in agriculture, forestry, urban planning, and environmental conservation. This study evaluates the performance of fine-tuning the Prithvi Foundation Model, Clay Foundation Model, and U-Net++ with Sentinel-2 imagery for land cover classification in the regions of Kalimantan and Sumatra, Indonesia. Using a dataset of Sentinel-2 image tiles labelled with 12 land cover classes, the models were trained and assessed using Intersection over Union (IoU) metrics. Results demonstrate the superior performance of the Clay Foundation Model, achieving a mean IoU of 0.4819. This research paper highlights the potential of Vision Transformer-based foundation models in distinguishing complex land cover categories and suggests directions for future research.

1. Introduction

Land cover classification is a critical tool for sustainable resource management, providing insights into agriculture, forestry, and urban planning. High-resolution satellite imagery has transformed land cover mapping by offering unprecedented detail for mapping land cover features. Sentinel-2, a mission by the European Space Agency (ESA), is particularly notable for its capabilities in this field. Consisting of two twin satellites, Sentinel-2A and Sentinel-2B, it provides multispectral imagery with a revisit time of 5 days. Sentinel-2 captures data in 13 spectral bands, including visible, near-infrared, and shortwave infrared wavelengths, with spatial resolutions ranging from 10 to 60 meters. Its frequent coverage and high-resolution data make it an invaluable resource for monitoring dynamic land cover changes. However, effectively leveraging this data requires robust machine learning models capable of handling the complexity and diversity of land cover types.

Deep learning has revolutionized image classification, with architectures like U-Net++ demonstrating strong performance in semantic segmentation. Additionally, the rise of foundation models has further revolutionized the field of remote sensing and computer vision. Foundation models, like Prithvi and Clay Foundation Models, are large-scale pre-trained architectures designed to be general-purpose and highly transferable across tasks. By being trained on vast datasets in a self-supervised manner, these models can extract rich feature representations that can be fine-tuned for specific tasks, such as land cover classification. Their ability to capture global context and complex relationships within data makes them particularly suited for handling the spectral and spatial intricacies of satellite imagery. This study aims to compare the effectiveness of the Prithvi and Clay Foundation Models, alongside a non- foundation model, the U-Net++ architecture with a ResNeXt backbone, in classifying land cover types across Sumatra and Kalimantan. These regions are characterised by diverse and dynamic land cover changes, providing an ideal testbed for assessing the capabilities of different deep learning models in remote sensing applications.

2. Related Work

The task of land cover classification has seen substantial advancements with the adoption of deep learning models and the availability of high-resolution satellite imagery. Early methods relied on traditional machine learning algorithms, but recent developments have shifted towards more sophisticated approaches, such as convolutional neural networks (CNNs) and Vision Transformer (ViT)-based architectures.

CNNs have long been a cornerstone of land cover classification due to their ability to extract hierarchical spatial features from imagery. Their widespread adoption stems from their ability to capture finegrained details in high-resolution data. Enhanced architectures like U-Net++ have further advanced the field, achieving state-of-the-art results in complex segmentation tasks. For instance, Zhou et al. (2018) demonstrated that U-Net++ significantly improved segmentation accuracy in medical imaging tasks by effectively distinguishing intricate anatomical structures. This ability to handle complex segmentation challenges highlights its potential for other domains, such as land cover classification, where precise boundary delineation and feature differentiation are similarly critical.

However, despite their success, CNN-based approaches are inherently constrained by their reliance on convolutional filters with fixed receptive fields. This limitation restricts their ability to capture long-range dependencies between distant regions of an image which can hinder performance in tasks where spatial context and relationships between distant features are crucial. For example, distinguishing fragmented plantations from forests or understanding urban sprawl.

The emergence of Vision Transformer (ViT)-based architectures offers a compelling solution to these challenges by leveraging self-attention mechanisms to model global contextual relationships. Unlike CNNs, which focus on local feature extraction, ViTs analyze all patches or pixels within an image simultaneously, facilitating a more comprehensive understanding of spatial dependencies. Dosovitskiy (2021) demonstrated that ViT models, when pre-trained on large-scale datasets, can outperform state-of-the-art CNN-based architectures in image classification tasks due to their ability to encode long-range dependencies and global context. Furthermore, their study

showed that scaling ViTs in terms of both dataset size and model complexity, consistently improved performance. Building on these findings, this study explores the potential of foundation models that utilize ViT architectures, investigating how large-scale pre-training can enhance land cover classification and improve generalization across remote sensing data.

3. Data Acquisition

3.1 Study Area

Sumatra and Kalimantan are among Indonesia's largest islands, featuring diverse land cover types such as tropical rainforests, peatlands, plantations, and urban regions. These areas are experiencing significant changes due to activities such as deforestation and agricultural expansion, making them ideal for evaluating land cover classification models.

3.2 Data Collection

The dataset comprises 140 Sentinel-2 image tiles, each measuring 1024x1024 pixels at 10-meter resolution. The images, acquired between 2023 and 2024, were visually inspected to ensure minimal cloud cover and proximity to peatland areas, as identified by Global Forest Watch. Each image was meticulously labelled with 12 distinct land cover classes, namely Background, Barren, Built-up, Canals, Cropland, Forest, Grassland and Shrubland, Oil Palm Plantation, Roads, Waterbodies, General Plantations and Cloud.

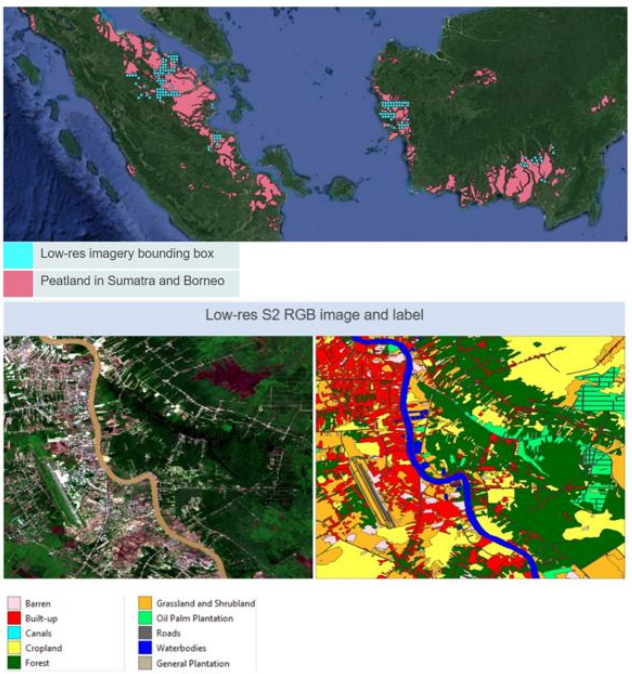

Figure 1. Location of 140 Sentinel 2 images over Sumatra and Kalimantan (Top), and Sentinel-2 imagery and its label (Bottom)

3.3 Data Annotation

The data labelling process was guided by a structured annotation guideline that provided precise definitions and characteristics for each land cover classes. This guideline ensured consistency and accuracy in the classification process by standardizing the identification of spatial boundaries and distinctive features for each class. Once the annotations were completed, the annotated images underwent a rigorous quality control process to identify and rectify any misclassifications or drawing inaccuracies. The annotations were then rasterized to align with the spatial resolution of the Sentinel-2 imagery and formatted appropriately for model training.

4. Methodology

4.1 Data Preprocessing

To facilitate model training and evaluation, the 1024x1024-pixel Sentinel-2 image tiles were divided into smaller 224x224-pixel tiles. This tiling process ensured that each smaller image retained its original 10-meter resolution, which is critical for capturing fine-grained land cover details. This resulted in a total of 2240 image tiles, of which 1792 were allocated to the training set, and 448 were reserved for the validation set.

Additionally, the background class was excluded from the dataset prior to model training. This decision was made because the background class does not provide meaningful information for the land cover classification task and could introduce noise into the model, potentially hindering its ability to learn relevant features. Removing the background class helps the model focus on distinguishing between actual land cover types, thereby improving classification performance.

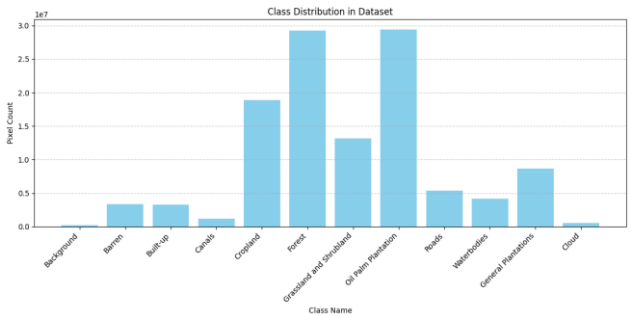

Figure 2. Class distribution in the training set, highlighting the variation in pixel counts across different land cover classes.

4.2 Model Selection

The selection of models for land cover classification using Sentinel-2 imagery was guided by the need to evaluate the performance of different deep learning architecture, including both foundation models and traditional convolutional neural network (CNN) frameworks. Three models were chosen: the Prithvi Foundation Model, the Clay Foundation Model, and the U-Net++ model with a ResNeXt backbone.

4.2.1 Clay Foundation Model: The Clay Foundation Model (Clay Foundation, 2024) is built on a ViT architecture, which has been pre-trained using a self-supervised Masked Autoencoder (MAE) technique. In this pre-training approach, random portions of the input images are masked, and the model learns to reconstruct the missing parts (He et al., 2022), as illustrated in Figure 3. This encourages the model to understand the global structure and underlying patterns of the image, rather than relying solely on localized pixel-level information. Such pre-training is particularly advantageous for multispectral data, where inter-band relationships and spatial patterns are integral to accurate land cover classification. Notably, the Clay model is pre-trained on a diverse dataset, totaling 33.8 TB and comprising approximately 70 million image chips from various satellite sensors, including Sentinel-2, Landsat, Sentinel-1 SAR, LINZ, NAIP, and MODIS, with 18 million chips sourced from Sentinel-

2. This diversity allows the Clay Foundation Model to be well-suited for handling multispectral and multimodal remote sensing data.

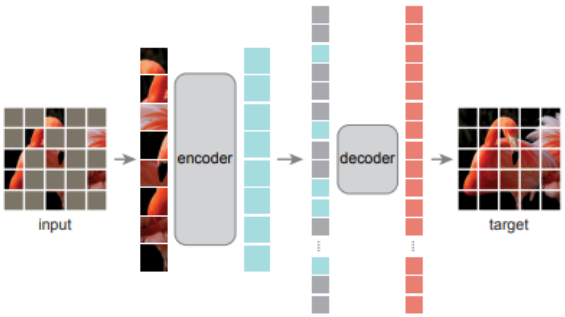

Figure 3. MAE Architecture. Adapted from "Masked Autoencoders Are Scalable Vision Learners" by He et al., 2022, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)

4.2.2 Prithvi Foundation Model: The Prithvi Foundation Model (Jakubik et al., 2023) shares architectural similarities with the Clay Foundation model, as both are ViT-based architectures employing self-supervised learning with a MAE pre-training strategy to capture global contextual information. A key distinction lies in their training data. The Prithvi Foundation Model was pre-trained on about 1TB of multi-temporal Sentinel-2 imagery, giving it a unique advantage in understanding the spectral and spatial characteristics of the dataset used in this study.

4.2.3 U-Net++: The U-Net++ model is a convolutional neural network (CNN)-based architecture designed specifically for semantic segmentation tasks, as illustrated in Figure 4. It extends the original U-Net (Ronneberger, Fischer, & Brox, 2015) architecture by introducing nested skip connections and deep supervision, which bridge the encoder and decoder pathways. These enhancements facilitate multiscale contextual information flow and improve gradient propagation during training, resulting in superior feature representation and improved segmentation performance across various datasets and tasks.

In this study, the U-Net++ model incorporates a ResNeXt backbone (Xie et al., 2017) pre-trained on ImageNet (Iakubovskii, 2019), as shown in Figure 5. ResNeXt enhances representation learning through aggregated residual transformations and parallel pathways within its residual blocks. This approach enables the model to efficiently learn diverse features, while its modular design allows for easy scalability and adaptability to various tasks and datasets by adjusting cardinality and depth.

To further improve the model's capability, a squeeze-and-excitation (SE) block is included in the ResNeXt backbone, forming the SE-ResNeXt architecture. This block improves spatial encoding by recalibrating features across channels, resulting in richer and more precise feature representations. Together, these architectural improvements enable U-Net++ to excel at extracting fine-grained spatial details, which are crucial for achieving high performance in semantic segmentation tasks.

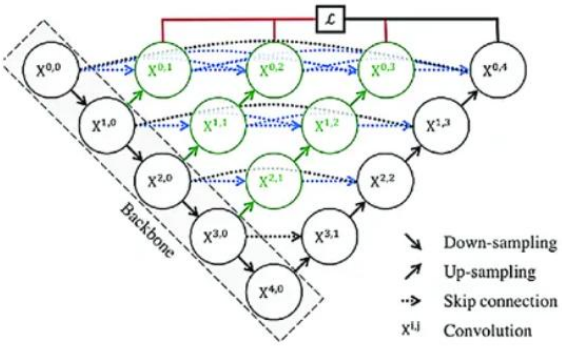

Figure 4. Schematic representation of the Unet++ architecture. Adapted from "UNet++: A Nested U-Net Architecture for Medical Image Segmentation" by Zhou et al., 2018, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support.

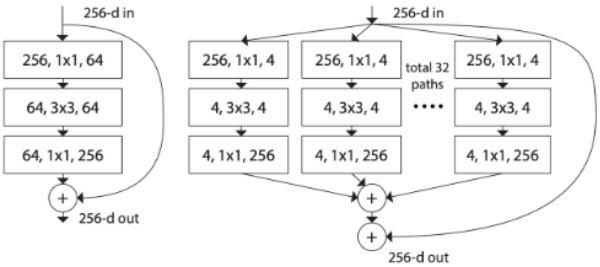

Figure 5. Comparison of ResNet (left) architecture and ResNeXt (right) architecture. Adapted from "Aggregated residual transformations for deep neural networks." by Xie et al., 2017, EEE Conference on Computer Vision and Pattern Recognition (CVPR).

4.3 Transfer Learning Strategy

The U-Net++ model utilized pre-trained weights from ImageNet, which is based on natural imagery rather than satellite data. Given this domain difference, we opted to train the entire U-Net++ model, allowing both the encoder and decoder pathways to be updated during fine-tuning. This strategy ensures that the model can fully adapt to the unique spectral and spatial characteristics of Sentinel-2 multispectral data, which differ significantly from the features present in natural images.

On the other hand, the foundation models were initialized with pre-trained weights derived from large-scale satellite imagery datasets. To preserve the rich feature representations learned during this domain-specific pre-training, we froze the backbone layers during fine-tuning, allowing only the segmentation head to be updated. This approach reduces computational complexity, mitigates the risk of overfitting, and leverages the models' inherent understanding of satellite imagery, which aligns closely with the Sentinel-2 data used in this study.

To ensure a consistent training framework across all models, we used the same labeled Sentinel-2 dataset prepared in Section 4.1 for the experiments mentioned above. This uniform setup enables a fair and unbiased comparison of model performance.

4.3.1 Spectral Bands Utilized

Different models were fine-tuned using specific combinations of Sentinel-2 spectral bands, tailored to their architectural design and pre-training strategies. The spectral bands used for each model are summarized in Table 1.

Model	Spectral Bands Used
Clay Foundation Model	B02, B03, B04, B05, B06, B07, B08, B8A, B11, B12
Prithvi	B02, B03, B04, B8A, B11, B12
U-Net++	B01, B02, B03, B04, B05, B06, B07, B08, B8A, B09, B11, B12

Table 1. Spectral bands used for training each model. Refer to Table 4 in Appendix for corresponding band function.

4.3.2 Hyperparameters and Optimization

The **cross-entropy (CE)** loss function was used to optimize model performance in our multi-class land cover classification task. It measures the dissimilarity between the predicted probability distribution and the actual class labels, effectively quantifying the penalty for incorrect predictions. Figure 2 illustrates the class distribution in the training set, highlighting the presence of class imbalance, where certain classes are overrepresented while others are underrepresented. This imbalance can bias the model towards predicting the majority classes, leading to suboptimal performance on minority classes. To mitigate this issue, CE loss penalizes misclassifications more severely for minority classes, as incorrect predictions for these classes typically result in higher loss values. This mechanism encourages the model to learn more robust and discriminative features for underrepresented classes, thereby improving classification performance across all categories.

The models were optimized using the **AdamW** optimizer, an improved variant of the Adam optimizer that decouples weight decay from the gradient update process. While traditional Adam incorporates L2 regularization as part of the update rule, AdamW applies weight decay directly to the model's parameters (Loshchilov and Hutter, 2017), providing more effective regularization. This decoupling has been shown to improve generalization, especially in deep learning models with complex architectures. For all models, the initial learning rates is set to 0.0001. However, it is not kept constant throughout the training process due to the learning rate scheduler, which followed the default configuration provided in the respective configuration files.

The **trainable parameters**, which are updated during backpropagation, varies across the models due to difference in architectural and representational capacity. Table 2 summarizes both trainable and total parameters for each model used in this study.

Model	Untrainable Parameters	Trainable Parameters	Total Parameters
clay_mae_base	92.1M	10.0M	102.1M
Prithvi-EO-1.0-100M	86.4M	13.6M	100.0M
U-Net++	0.0M	51.0M	51.0M

Table 2. Comparison of trainable and total parameters each model. The foundation models have a frozen backbone, resulting in a lower number of trainable parameters compared to its total parameters, while U-Net++ allows all parameters to be trainable.

4.4 Model Evaluation

The performance of the models was evaluated using the Intersection over Union (IoU) metric, a widely adopted measure in semantic segmentation tasks (Rezatofighi et al., 2019). IoU quantifies the overlap between the predicted and ground truth regions for each land cover class, providing a robust assessment of the model's classification accuracy.

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

where A = set of pixels belonging to the ground truth region
B = set of pixels in the predicted region

The numerator, $|A \cap B|$, quantifies the intersection, representing the correctly classified pixels, while the denominator, $|A \cup B|$, captures the total number of unique pixels in either the ground truth or predicted region.

For this study, the mean Intersection over Union (mIoU) was used to evaluate the overall model performance across the 11 land cover classes. mIoU is calculated as the arithmetic mean of the IoU values for all classes, providing a single performance metric that reflects the models' ability to generalize across diverse land cover types. This metric is particularly advantageous for evaluating models trained on imbalanced datasets, as it considers class-specific performance, ensuring that smaller or less frequent land cover types are not overshadowed by dominant classes.

5. Results and Discussion

This section presents the quantitative and qualitative evaluation of the models – Prithvi Foundation Model, Clay Foundation Model and U-Net++.

The performance of each model was assessed using the IoU metric introduced in Section 4.4. Table 3 presents the class-wise IoU scores, along with the mIoU as an aggregate measure of overall performance.

Classes	Clay	U-Net++	Prithvi
Barren	**0.2895**	0.2462	0.2009
Built-up	**0.4568**	0.2851	0.2618
Canals	**0.4661**	0.0782	0.0512
Cropland	0.3913	**0.3957**	0.3068
Forest	0.4955	**0.6502**	0.6317
Grassland and Shrubland	0.2550	0.3123	**0.3438**
Oil Palm Plantation	0.5129	**0.5500**	0.5482
Roads	**0.2497**	0.1839	0.1615
Waterbodies	0.5755	**0.6847**	0.6628
General Plantations	**0.7789**	0.6404	0.5931

Clouds	**0.8296**	0.4536	0.2840
Mean	**0.4819**	0.4073	0.3678

Table 3. IoU scores for each land cover class across models. The highest IoU score for each class is in bold.

The Clay Foundation Model achieves the highest mIoU score of 0.4819. Notably it showed strong segmentation capability in classes such as ‘Built-up’ (IoU: 0.4568), ‘Canals’ (IoU: 0.4661), ‘General Plantations’ (IoU: 0.7789), and ‘Clouds’ (IoU: 0.8296). This suggests that the model effectively captures global context and complex spatial relationships, likely due to its extensive pre-training on diverse remote sensing datasets.

The U-Net++ model, despite not being a foundation model, demonstrated competitive performance in ‘Cropland’ (IoU: 0.3957), ‘Forest’ (IoU: 0.6502), ‘Oil Palm Plantation’ (IoU: 0.5500), and ‘Waterbodies’ (IoU: 0.6847). These results indicate that the CNN-based model is highly effective in segmenting large, homogeneous land cover classes with distinct spectral characteristics. However, its performance dropped significantly for Canals (IoU: 0.0782) and Built-up areas (IoU: 0.2851), which may be attributed to its limited ability to capture global dependencies compared to the transformer-based models.

The Prithvi Foundation Model, while expected to perform well due to its ViT-based architecture, underperformed relative to the Clay Foundation Model across most categories. Its mIoU of 0.3678 was the lowest among the three models. However, it showed relatively high IoU scores for ‘Forest’ (IoU: 0.6317), ‘Grassland and Shrubland’ (IoU: 0.3438) and ‘Waterbodies’ (IoU: 0.6628). The lower performance of Prithvi may be attributed to several key factors. Firstly, its pre-training dataset size (~1TB of Sentinel-2 imagery) is significantly smaller compared to Clay’s 70 million+ image chips sourced from a diverse range of satellite sensors. This limited dataset likely restricts Prithvi's exposure to varied spatial and spectral patterns. Secondly, the limited number of spectral bands utilized during fine-tuning—only 6 bands (B02, B03, B04, B8A, B11, B12)—may have constrained the model’s ability to fully capture the spectral diversity necessary for distinguishing complex land cover classes. In contrast, the Clay Foundation Model leveraged 10 spectral bands, and U-Net++ utilized spectral bands, providing these models with richer spectral information to enhance feature discrimination. This limitation in spectral inputs may have affected Prithvi’s performance, particularly in heterogeneous environments where subtle spectral variations are critical for accurate classification.

To further analyse model performance, Figure 6 provides a visual comparison of the segmentation outputs across each land cover types. We observe a distinct difference in the models’ ability to accurately delineate land cover boundaries. The Clay Foundation Model produces more coherent and spatially consistent segmentations, closely resembling the ground truth. The annotated black rectangle shows how the boundaries between different classes are well-defined, with minimal noise or fragmented patches.

In contrast, both U-Net++ model and Prithvi Foundation Model exhibit a tendency toward over-segmentation, resulting in fragmented predictions. This effect is particularly noticeable in the regions representing built-up areas. Despite this, U-Net++ excels at capturing subtle boundary variations in waterbodies, as shown in the annotated blue box, which aligns with its higher IoU score in that category.

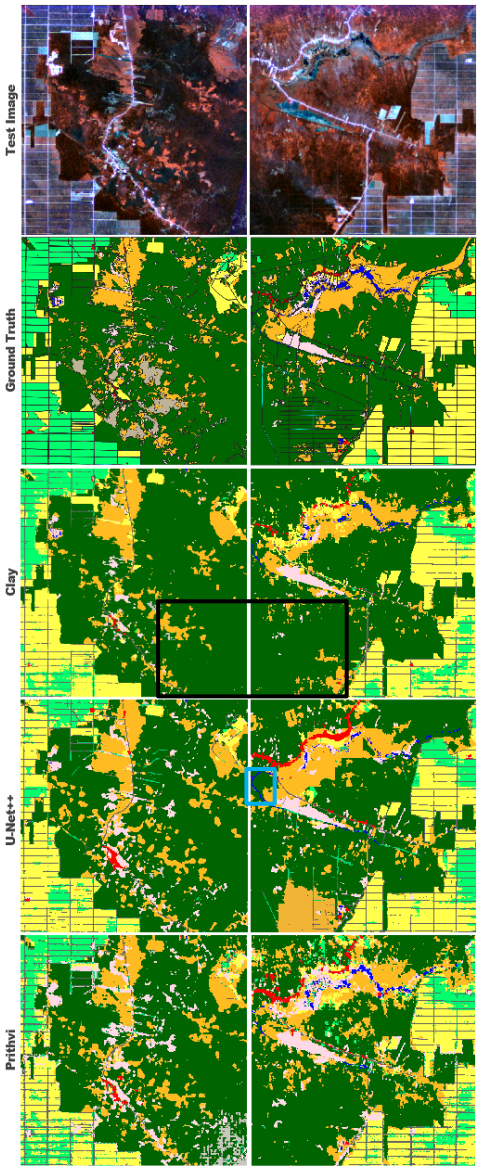

Figure 6. Qualitative comparison of prediction results. Ground truth is provided for reference. Legend in Appendix.

6. Conclusion

This study presents a comparative analysis of three deep learning models—the Prithvi Foundation Model, the Clay Foundation Model, and U-Net++—for multi-class land cover classification using Sentinel-2 imagery across Sumatra and Kalimantan, Indonesia. The evaluation, based on the Intersection over Union (IoU) metric, highlights key differences in model performance driven by architectural design, pre-training strategies, spectral band utilization, and computational efficiency.

The Clay Foundation Model offers a robust framework for analysing multispectral satellite data, showing particular promise in handling complex and heterogeneous land cover categories such as Built-up areas, Canals, General Plantations, and Clouds. This performance can be attributed to its extensive pre-training on a large, diverse, multi-sensor dataset, enabling robust feature

extraction and effective generalization across diverse land cover types.

The U-Net++ model delivered competitive performance, particularly excelling in classes like Forest and Waterbodies. A key advantage of U-Net++ is its computational efficiency. As shown in Table 2, U-Net++ has a smaller number of parameters (51 million) compared to the foundation models, with both the Clay and Prithvi models having 100 million total parameters each. This reduced parameter count results in lower computational load, faster inference times, and reduced memory requirements, making U-Net++ a practical choice for real-time applications or environments with limited computational resources. Despite its simpler model size, U-Net++ demonstrated strong performance across multiple land cover classes, highlighting its effectiveness without the need for extensive pre-training.

In conclusion, this study underscores the potential of foundation models, particularly those trained on diverse, multi-sensor datasets, for advancing land cover classification tasks. However, it also highlights that conventional models like U-Net++ remain highly competitive due to their computational efficiency and robust performance, making them suitable for real-world applications where resource constraints are a key consideration. The findings open avenues for future research to optimize model design, pre-training strategies, and data utilization to achieve more accurate, scalable, and efficient land cover mapping solutions.

For future work, it is recommended to explore the integration of additional data sources, such as Sentinel-1 SAR, DSEO, or NEUSAR, to complement Sentinel-2 imagery and enhance the models' ability to capture diverse land cover characteristics. Additionally, applying these models to different geographic regions will help validate their generalizability and adaptability across varied environmental conditions. Another promising direction is the adoption of next-generation models, particularly the newly released Prithvi 2.0 (Szwarcman et al., 2024), which features more parameters and potentially improved representation capabilities. Investigating Prithvi 2.0’s performance could address some of the limitations observed in this study, offering new insights into the role of model capacity in land cover classification.

References

Clay Foundation. (n.d.). “Clay Foundation model v1.0” *GitHub Repository*, 2024. https://github.com/Clay-foundation/model

Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16000-16009).

Iakubovskii, Pavel. "Segmentation Models Pytorch." *GitHub Repository*, 2019. https://github.com/qubvel/segmentation_models.pytorch

Jakubik, J., Roy, S., Phillips, C. E., Fraccaro, P., Godwin, D., Zadrozny, B., ... & Ramachandran, R. (2023). Foundation models for generalist geospatial artificial intelligence. *CoRR*.

Loshchilov, I., & Hutter, F. (2017). Fixing weight decay regularization in adam. *arXiv preprint arXiv:1711.05101*, *5*.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 658-666).

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer International Publishing.

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1492-1500).

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4* (pp. 3-11). Springer International Publishing.

Szwarcman, D., Roy, S., Fraccaro, P., Gíslason, Þ. E., Blumenstiel, B., Ghosal, R., ... & Moreno, J. B. (2024). Prithvi-EO-2.0: A Versatile Multi-Temporal Foundation Model for Earth Observation Applications. *arXiv preprint arXiv:2412.02732*.

Appendix

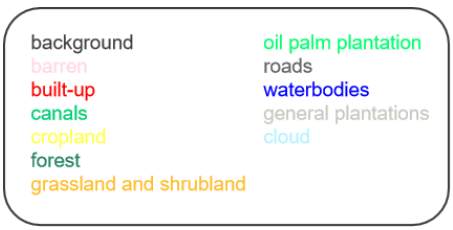

Figure 7. Legend for land cover classification classes. Corresponds to the class representations used in Figure 6.

Band Number	Function
B01	Coastal Aerosol
B02	Blue
B03	Green
B04	Red
B05	Red Edge 1
B06	Red Edge 2
B07	Red Edge 3
B08	NIR
B8A	Narrow NIR
B09	Water Vapour
B11	SWIR 1
B12	SWIR 2

Table 4. Band number and its function.