Flood Induced Economic Damage Assessment from Satellite Imagery using Vision Transformers

Md. Ashrif Rahman Arian
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
ashrif.arian@northsouth.edu

Md. Mehedi Hasan Shishir
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
mehedi.shishir@northsouth.edu

Sadman Islam Chowdhury Samin
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
sadman.samin12@northsouth.edu

Shahnewaz Siddique
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
shahnewaz.siddique@northsouth.edu

***Abstract*—Every year floods cause substantial threats to lives, livelihoods, agriculture and infrastructure. Rapid assessment of economic damage caused by floods is necessary for disaster management, resource allocation and policy making. In this study, we propose a novel method for calculating flood induced economic damage using before and after flood satellite imagery. By leveraging Vision Transformer techniques, we perform semantic segmentation to identify land cover changes after the disaster. By measuring the area loss per class and assigning economic value to each class, we provide a method to estimate the monetary damage due to the disaster.We used Segformer B3 for segmentation which is a model of the Vision Transformer and achieved much higher pixel accuracy (0.98) and mIoU (0.53) compared to state of the art segmentation models UNet and DeeplabV3+. Moreover, Segformer B3 demonstrated considerably higher computational efficiency compared to the two other models experimented. Our approach offers an innovative and automated solution for post disaster flood damage assessment.**

***Index Terms*—Satellite images, Flood damage, Computer vision, Semantic segmentation, Vision transformer, Segformer.**

I. Introduction

Floods are one of the most devastating natural disasters causing more than 40 billion dollars of damage every year [1]. The timely assessment of flood damage is necessary for disaster management, search and rescue, delivering aid to the affected people, resource allocation and policy making. But the traditional ground survey method for damage determination is time consuming and also infeasible for inaccessible areas. As a result, it is necessary to assess flood damage automatically and accurately.

The advancement in satellite imagery and deep learning techniques paves the way to determine economic damage of flood automatically. Satellite images like Sentinel and Landsat data are publicly available which can provide us ”Before flood” and ”After flood” images. Deep learning-based computer vision techniques can help us to work on the satellite images and calculate the total flooded area. The lost area measurement can be used to get the approximate economic damage.

This study proposes a novel approach to estimate the economic damage of flood using computer vision. Segmentation is a very common task of computer vision. We use segmentation techniques to calculate economic damage by comparing the change of after flood images with before flood images. For segmentation, UNet and DeeplabV3+ are the most common techniques to use but here we used Vision Transformers as it offers advantages in contextual understanding. Transformers have become a necessity in NLP tasks as they enable models to capture long range dependencies and context in text. But recently transformers have also become prominent in Computer Vision field after Vision Transformer was introduced in the paper, ”An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale” [2].

In this paper, we discuss how vision transformers can be used for segmentation and utilizing the segmentation result how we can measure the change due to a natural disaster. By assigning the value for each lost class we showed the process of assessing the economic damage that occured due to the disaster. In the following sections, we discuss related works, outlined our methodology, presented experimental results and highlight the potential of determining flood damage from satellite imagery using Computer vision, more specifically vision transformer.

II. Literature Review

Till now, satellite images have not been used directly to assess the economic damage of floods. However, researchers have utilized satellite images in land area classification, semantic segmentation and crop mapping tasks. We found several research papers in which Vision Transformer (ViT) was used to work on satellite images.

In [3], the authors experimented the efficiency of ViT in classifying satellite images and they got 98.5% accuracy, which surpassed the accuracy of CNN, VGG16 and VGG19. Similarly, in [4], authors classified satellite images in two classes 'Damaged in Hurricane' and 'Not Damaged' using CCT model, a variant of ViT where convolutions are also incorporated and it achieved 98.79% accuracy. These papers demonstrate that Vision Transformer works well with satellite imagery.

In [5] the authors surveyed the viT in the semantic segmentation task. They found out that it reached an accuracy of 88.55% in ImageNet dataset and 94.55% on CIFAR-100 dataset which are higher than any convolution based method. It also proves that ViT works better in large dataset. So, it should be a good approach to use Vision Transformer in Transfer Learning. In [6], MeViT, a Medium-Resolution Vision Transformer, was proposed for semantic segmentation on Landsat imagery in Thailand. It outperformed models like HRViT and SegFormer with precision (92.22%), recall (94.69%). Here, Mixed Scale Convolutional Feedforward Network (MixCFN) is used which incorporates multiple depth-wise convolution paths to extract multi scale local information. The MixCFN in MeViT enhanced efficiency and boundary detection.

In [7], the Extended Vision Transformer (ExViT) achieved state-of-the-art results in land-use and land-cover (LULC) classification using multimodal data such as hyperspectral (HS), LiDAR, and SAR. This approach processes multimodal RS image patches with parallel branches of position shared ViTs extended with separable convolution modules. On benchmark datasets, it achieved an overall accuracy (OA) of 96.7% and a mean intersection-over-union (mIoU) of 83.6% and thus it demonstrated its potential for agricultural applications. In [8], ConvNext model is used with ViT. The ViT-ConvNeXt method for land cover classification achieved 99.38% outperforming CNN, ResNet, and VGG16.

In [9], multi-modal Temporal-Spatial Vision Transformers (TSViT) for crop mapping outperformed single-modal TSViT in Mean Accuracy (MA), Overall Accuracy (OA), and mean intersection-over-union (mIoU). The Synchronized Class Token Fusion (SCTF) method achieved the best performance, with a 12% improvement in MA and 11% in mIoU, emphasizing the benefits of multi-modal fusion for crop mapping.

So, the success of ViT in semantic segmentation, land cover classification and crop mapping tasks inspired us to use it in semantic segmentation on satellite imagery and to exploit it to assess the economic damage of floods.

III. Methodology

A. Overview

Our study proposes a transformer based framework for estimating economic damage caused by floods using before and after satellite images. The core idea is to segment land cover into predefined classes using a semantic segmentation model, measure area changes in these classes in the post flood images and compute economic losses by assigning per square kilometer monetary values to each class. Figure 1 depicts the pipeline of our system which includes data preprocessing, augmentation, training for segmentation with specialized loss function and damage quantification based on pixel wise area loss.

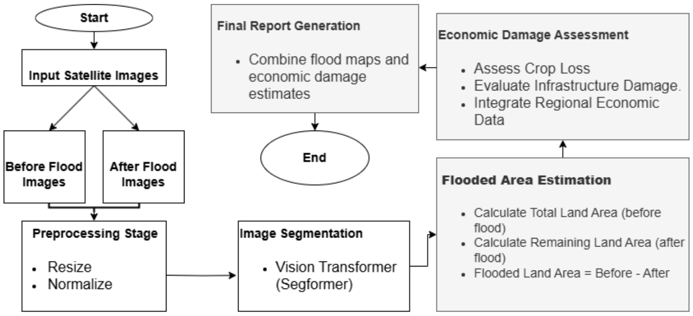

Start
Input Satellite Images
Before Flood Images
After Flood Images
Preprocessing Stage
Resize
Normalize
Image Segmentation
Vision Transformer (Segformer)
Flooded Area Estimation
Calculate Total Land Area (before flood)
Calculate Remaining Land Area (after flood)
Flooded Land Area = Before - After
Economic Damage Assessment
Assess Crop Loss
Evaluate Infrastructure Damage.
Integrate Regional Economic Data
Final Report Generation
Combine flood maps and economic damage estimates
End

Fig. 1: System Design of the Proposed Method

B. Dataset Preparation and Augmentation

For our task, we combined two publicly available segmentation dataset to create a unified collection of labeled satellite images. The datasets are DeepGlobe Land Cover Classification Dataset [10] and Semantic Segmentation of Aerial Imagery Dataset [11]. Both of these datasets are available in Kaggle. DeepGlobe dataset has 7 classes and the other dataset has 6 classes. In our combined dataset there are 12 classes as we kept all the unique classes from each of the datasets and combined the common class. The 12 classes of the combined dataset are: Urban land, Agriculture land, Range land, Forest land, Barren land, Building, Land, Road, Vegetation, Water, Unlabeled and Unknown. Each image was resized to 256×256 pixels to standardize input dimensions. From a total of 875 images we used 700 for training and 175 for validation. Train set was built with 642 DeepGlobe dataset images and 58 Semantic Segmentation with Aerial Imagery dataset images.

To improve generalization, data augmentation was performed using the Albumentations library, including random flips, contrast adjustments, elastic transformations and normalization. Augmentations were applied during training to ensure that the model was exposed to diverse spatial and spectral conditions without altering the semantic structure.

C. Model Architecture

The segmentation model we used is Segformer B3 [12] which is a hierarchical vision transformer architecture pre-trained on ADE20K. This model uses a multi scale encoder without relying on positional encoding which makes it robust to varying spatial input sizes and helps to balance accuracy and efficiency. As showed in figure 2, Segformer B3 consists of two main parts which are hierarchical transformer encoder and lightweight all MLP decoder. It can extract coarse and fine features better because of its hierarchical encoder. all MLP decoder is used to directly fuse multilevel features and predict the semantic segmentation mask. We finetuned the pretrained model on our custom dataset.

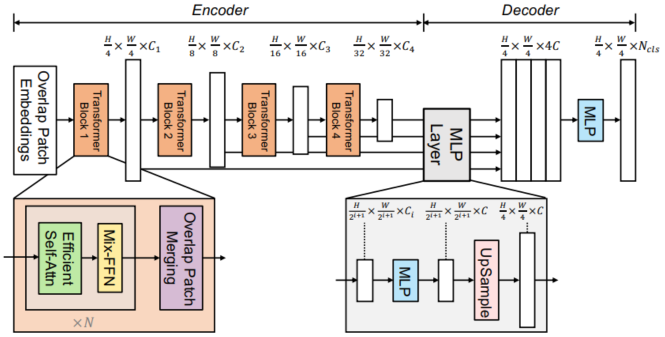

Encoder
Decoder
$\frac{H}{4}\times\frac{W}{4}\times C_1$
$\frac{H}{8}\times\frac{W}{8}\times C_2$
$\frac{H}{16}\times\frac{W}{16}\times C_3$
$\frac{H}{32}\times\frac{W}{32}\times C_4$
$\frac{H}{4}\times\frac{W}{4}\times 4C$
$\frac{H}{4}\times\frac{W}{4}\times N_{cls}$
Overlap Patch Embeddings
Transformer Block 1
Transformer Block 2
Transformer Block 3
Transformer Block 4
MLP Layer
MLP
Efficient Self-Attn
Mix-FFN
Overlap Patch Merging
×N
$\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}\times C_i$
$\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}\times C$
$\frac{H}{4}\times\frac{W}{4}\times C$
MLP
UpSample

Fig. 2: Architecture of Segformer B3 with hierarchical encoder and MLP decoder

D. Training Configuration

In the training process a composite loss function has been used. The loss function is built with Dice loss and Focal loss. It allows the model to handle class imbalance and challenging foreground background distinctions. The Dice Loss is defined as:

$$\mathcal{L}_{\text{dice}} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + \epsilon} \tag{1}$$

where p_i denotes the predicted probability for pixel i, g_i denotes the ground truth label (either 0 or 1) for pixel i, N is the total number of pixels, and ϵ is a small constant added to prevent division by zero.The Focal Loss is given by:

$$\mathcal{L}_{\text{focal}} = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{2}$$

where $p_t = p$ if the ground truth label $y = 1$, and $p_t = 1 - p$ if $y = 0$. Here, p is the predicted probability for the positive class, α_t is a weighting factor that balances the importance of positive/negative examples, and γ is the focusing parameter that reduces the relative loss for well-classified examples. The total loss used for training is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{dice}} + \mathcal{L}_{\text{focal}} \tag{3}$$

Training optimization was performed using the AdamW optimizer with and initial learning rate of 2e-4. The learning rate was dynamically adjusted using a ReduceLROnPlateau scheduler based on validation loss. Training was carried out over 30 epochs with a patience of 5 for early stopping to prevent overfitting. The batch size was set to 4 and training was conducted on a Kaggle hosted NVIDIA T4 GPU using Pytorch and HuggingFace's Transformers library.

E. Evaluation Metrics

The segmentation performance was evaluated using multiple standard metrics: pixel accuracy, mean Intersection over Union (mIoU), and F1 score. The metrics are defined as follows:

$$\text{Pixel Accuracy} = \frac{\sum_{i=1}^{K} TP_i}{\sum_{i=1}^{K} (TP_i + FP_i + FN_i)} \tag{4}$$

where TP_i, FP_i, and FN_i are the number of true positives, false positives, and false negatives for class i respectively and K is the total number of classes.

$$\text{mIoU} = \frac{1}{K} \sum_{i=1}^{K} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{5}$$

$$\text{F1 Score}_i = \frac{2 \cdot TP_i}{2 \cdot TP_i + FP_i + FN_i} \tag{6}$$

$$\text{Mean F1 Score} = \frac{1}{K} \sum_{i=1}^{K} \text{F1 Score}_i \tag{7}$$

These metrics were computed per epoch for the validation set and were used to determine the model checkpoint that performed the best.

F. Economic Loss Estimation

In the final step, economic damage is calculated. The number of pixels for each segmented class in both before and after flood images is counted. From the per pixel coverage area measurement we can get the total area of each segmented class. The difference in area per class represents the land lost due to flooding. The following formulas has been used for the calculation:

$$A_{\text{pixel}} = \frac{r \times r}{10000} m^2 \tag{8}$$

Where A_{pixel} represents coverage per pixel and r is pixel resolution in cm.

$$A_{\text{class}} = N \times A_{\text{pixel}} \tag{9}$$

Where A_{class} is the total area of one class and N is the number of total pixel of that class.

$$\Delta A_{\text{class}} = A_{class}^{pre} - A_{class}^{post} \tag{10}$$

Where ΔA_{class} is change in Area of a class due to flood, A_{class}^{pre} is Area of the class in before flood image and A_{class}^{post} is Area of the class in after flood image.

Now for economic damage calculation, we need to assign economic value to every class. For each class, we define a unit economic value, $\Delta Value_{class}$ (\$/m^2) reflecting replacement or opportunity cost. Agriculture values can be derived from crop yield (kg/m²) and farm-gate prices net of input costs; road values from standard reconstruction costs per m²; building values from replacement cost per m² for the prevailing building typologies; forest/vegetation values from restoration or stumpage estimates. Classes such as water or barren land were assigned $\Delta Value_{class} = 0$. So, the Economic Damage for a class, $Damage_{class}$ is calculated by the following equation,

$$Damage_{class} = \Delta A_{class} \times Value_{class} \tag{11}$$

Now we can calculate the total economic damage due to the flood by adding up the damages for all the classes. So, if there are total n classes then the total economic damage,

$$Damage_{Total} = \sum_{class=1}^{n} Damage_{class} \tag{12}$$

Where $Damage_{Total}$ represents the total damage in one image considering all classes.

By using these equations, pixel-wise semantic changes can be interpreted as economic damage due to flood, which will be effective for rapid post-disaster assessment.

IV. Results

We evaluated segmentation performance using the metrics and compared Segformer B3 with two state of the art models for segmentation, UNet and DeeplabV3+. The performance with each of the models are given in the following table:

TABLE I: Comparison of Segmentation Models' Performance on Combined Dataset

Model	Pixel Acc.	mIoU	F1 Score	Time/Epoch
Segformer B3	0.98	0.53	0.38	˜8 min
DeepLabV3+	0.85	0.50	0.44	˜17 min
UNet	0.75	0.31	0.31	˜20 min

From the Table I, we can see that Segformer B3 achieved the highest pixel accuracy and mIoU, while DeepLabV3+ showed a slightly higher F1 score. Segformer also showed efficiency in computing time as it trained each epoch faster than other models.

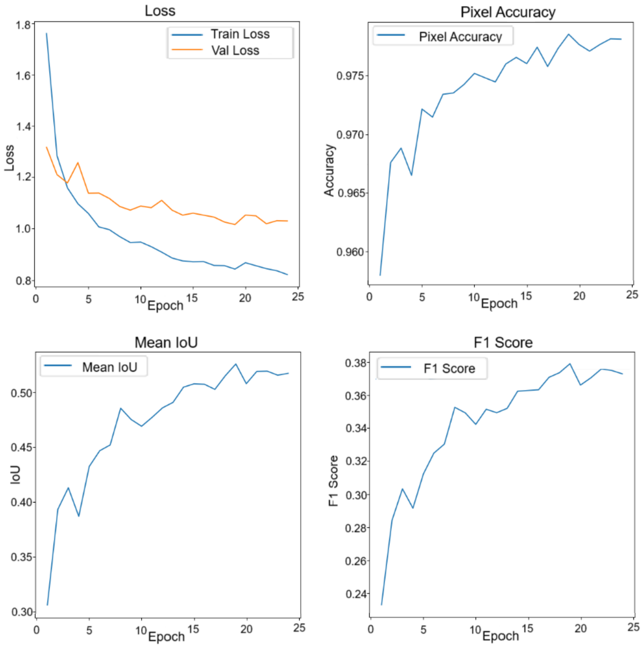


Fig. 3: Plots showing validation metrics (Training and Validation Loss, Pixel Accuracy, mIoU and F1 Score) over epochs. These plots represent the trend of the model's learning process. It can be observed that after the 20th epoch, most of the metrics started to mature and on the 25th epoch early stopping triggered.

As showed in Figure 3, during training with Segformer B3, we observed reduction of both training and validation loss and the metrics values kept increasing which indicates stable convergence and effective generalization of the model.

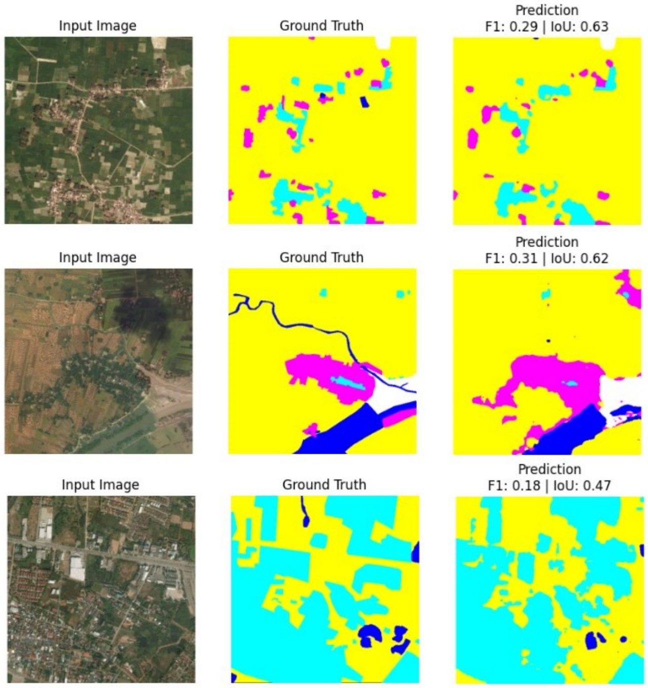


Fig. 4: Segmentation on the validation set with Segformer B3. Here in the first column there is satellite image, second column consists of ground truth mask for the satellite image and lastly the third column contains our predicted segmentation image. In the third column we can also see the F1 score and mIoU of the prediction.

From Figure 4 we can see that Segformer B3 segmented quite well compared to the ground truth masks. But still there are scopes of improvements, particularly in edges and narrowed classes.

Our system takes input of before and after flood images with pixel resolution and monetary value per unit for each class. As output it gives us the segmented images, lost area in each class, monetary lost value for each class and total monetary loss over all the classes. But for now, we tested our system with an arbitrary economic value for each class, as we could not find enough data of disaster damage and economic value of different types of land. In Figure 5, we can see after taking input images, the system generated segmented images. Lost area and economic damage has also been calculated.

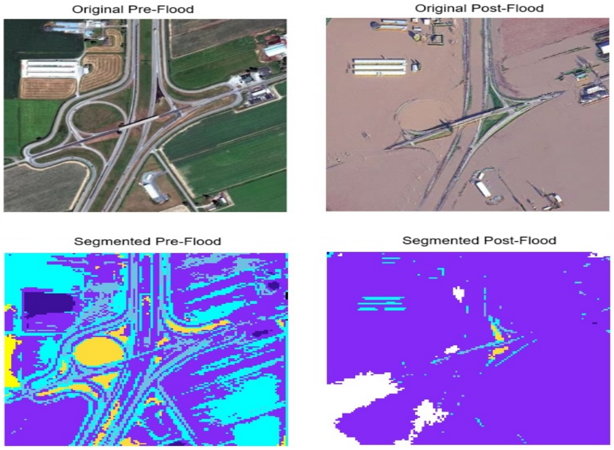
V. Discussion

A. Segmentation Performance Analysis

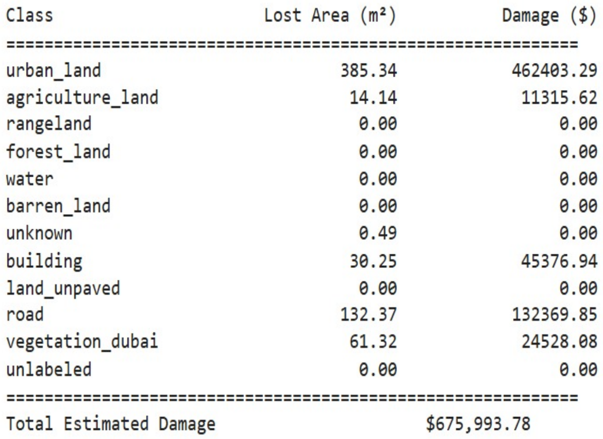
Our used model, Segformer's high pixel accuracy and mIoU suggests effective segmentation of complex land types. It also completed training faster. Overall, its better generalization capability indicates that this model is suitable for complex land cover classification and related task.

B. Economic Impact Results

Experiment on sample regions showed adequate estimation of monetary loss. The proposed approach is adaptable for vari-



(a)



```
Class                          Lost Area (m²)          Damage ($)
=============================================================
urban_land                              385.34           462403.29
agriculture_land                         14.14            11315.62
rangeland                                 0.00                0.00
forest_land                               0.00                0.00
water                                     0.00                0.00
barren_land                               0.00                0.00
unknown                                   0.49                0.00
building                                 30.25            45376.94
land_unpaved                              0.00                0.00
road                                    132.37           132369.85
vegetation_dubai                         61.32            24528.08
unlabeled                                 0.00                0.00
=============================================================
Total Estimated Damage                          $675,993.78
```

(b)

Fig. 5: (a) Original pre-flood and post-flood satellite images of a flood-affected area that were given as input along with the corresponding segmented outputs generated by our proposed system. (b) Based on the segmented images, the pixel coverage area of the satellite image and assigned arbitrary economic values for each class, our system computes the lost area and estimated economic damage for each class, as well as the total economic damage.

ous disaster types and different regions by adjusting economic coefficients.

C. Limitation and Future Work

We tested the economic loss value calculation part with assigned arbitrary economic value of each class as there is not enough publicly available economic value data for different types of land. Moreover, our proposed system cannot determine crop type or seasonal variance and assumes fixed economic value per m^2 according to the input. Besides, our training dataset is relatively small in size. In future, we will add crop type classification module using temporal data to calculate more accurate economic damage. Moreover, we will extend our research to estimate economic damage from other disasters like wildfire, drought etc.

VI. Conclusion

We presented a deep learning based framework for assessing flood induced economic damage from satellite images. By applying vision transformer based semantic segmentation model, we achieved high pixel accuracy and mIoU in segmenting satellite image classes. Exploiting segmented before and after flood images we measured the lost area. Assigning economic value to the measured lost area gives us a practical solution for rapid post disaster analysis. Future work will address crop type specific damage calculation and will extend the system for various types of disaster like drought, wildfire etc.

References

[1] OECD, *Financial Management of Flood Risk*. Paris: OECD Publishing, 2016. [Online]. Available: https://doi.org/10.1787/9789264257689-en

[2] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *ArXiv*, 2020, arXiv:2010.11929. [Online]. Available: https://arxiv.org/abs/2010.11929

[3] A. A. Adegun, S. Viriri, and J. Raymond-Tapamo, "Satellite images analysis and classification using deep learning-based vision transformer model," in *2023 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2023, pp. 1275–1279.

[4] M. F. Islam, S. Zabeen, M. M. Rahman, M. H. Khan, F. N. Khan, N. Z. Nahim, T. Anwar, and M. Kaykobad, "Identifying hurricane damage using explainable compact transformer with convolutional embedding," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*, 2022, pp. 833–838.

[5] H. Thisanke, C. Deshan, K. Chamith, S. Seneviratne, R. Vidanaarachchi, and D. Herath, "Semantic segmentation using vision transformers: A survey," *arXiv preprint arXiv:2305.03273*, 2023. [Online]. Available: https://arxiv.org/abs/2305.03273

[6] T. Panboonyuen, C. Charoenphon, and C. Satirapod, "Mevit: A medium-resolution vision transformer for semantic segmentation on landsat satellite imagery for agriculture in thailand," *Remote Sensing*, vol. 15, no. 5124, 2023.

[7] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, "Extended vision transformer (exvit) for land use and land cover classification: A multimodal deep learning framework," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, 2023, pp. 1–12.

[8] M. Gadiparthi, M. Kulkarni, R. R. Al-Fatlawy, and V. Malathy, "Land cover classification in high-resolution satellite images using vision transformer and convnext approach," in *2024 International Conference on Data Science and Network Security (ICDSNS)*, 2024, pp. 1–5.

[9] T. Follath, D. Mickisch, J. Hemmerling, S. Erasmi, M. Schwieder, and B. Demir, "Multi-modal vision transformers for crop mapping from satellite image time series," in *IGARSS 2024 - IEEE International Geoscience and Remote Sensing Symposium*, 2024.

[10] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[11] H. in the Loop, "Semantic segmentation of aerial imagery," https://www.kaggle.com/datasets/humansintheloop/semantic-segmentation-of-aerial-imagery, 2020, accessed: 2025-04-17.

[12] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," 2021. [Online]. Available: https://arxiv.org/abs/2105.15203