Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 2, April 2025, pp. 1391~1406
ISSN: 2302-9285, DOI: 10.11591/eei.v14i2.8454

Transformers for aerial images semantic segmentation of natural disaster-impacted areas in natural disaster assessment

Deny Wiria Nugraha[1,3], Amil Ahmad Ilham[2], Andani Achmad[1], Ardiaty Arief[1]
[1]Department of Electrical Engineering, Faculty of Engineering, Hasanuddin University, Gowa, Indonesia
[2]Department of Informatics, Faculty of Engineering, Hasanuddin University, Gowa, Indonesia
[3]Department of Information Technology, Faculty of Engineering, Tadulako University, Palu, Indonesia

Article Info

Article history:

Received Mar 9, 2024
Revised Oct 26, 2024
Accepted Nov 19, 2024

Keywords:

Aerial image
Natural disaster
Natural disaster assessment
Semantic segmentation
Transformer

ABSTRACT

Aerial image segmentation of natural disaster-impacted areas and detailed and automatic natural disaster assessment are the main focus of this study. Detecting and recognizing objects on aerial images of areas impacted by natural disasters and assessing natural disaster-impacted areas are still difficult problems. To solve these problems, this study utilizes four of the latest transformer-based semantic segmentation network models, bidirectional encoder representation from image transformers (BEiT), dense prediction transformer (DPT), OneFormer, and SegFormer, and proposes a detailed and automatic natural disaster assessment of the segmented image. The SegFormer model achieved the first-best result, and the OneFormer model achieved the second-best result. The SegFormer model outperformed OneFormer by 1.58% higher for the mean accuracy value and 4.28% for the mean intersection over union (mIoU) value. All receiver operating characteristics (ROC) curves have mean area under curve (AUC) values above 0.9, which means that the SegFormer model performs well in generating semantic segmentation images. The fuzzy c-means (FCM) clustering algorithm performed well and could automatically cluster the natural disaster assessments into four categories. This study has produced semantic segmentation of aerial images of areas impacted by natural disasters and natural disaster assessments, which can be used in natural disaster management systems.

Corresponding Author:

Andani Achmad
Department of Electrical Engineering, Faculty of Engineering, Hasanuddin University
St. Poros Malino Km. 6, Bontomarannu, Gowa 92171, South Sulawesi, Indonesia
Email: andani@unhas.ac.id

1. INTRODUCTION

Catastrophic events known as natural disasters arise from natural occurrences, including floods, landslides, earthquakes, volcanic eruptions, tsunamis, hurricanes, and droughts. Natural calamities are occurring more frequently. Given the rising frequency of disasters, all parties must comprehend and apply disaster management strategies. Accurately assessing the areas impacted by natural disasters is crucial in the natural disaster management system to enable quick, effective, and efficient emergency responses. Unmanned aerial vehicles (UAVs) or drone technology are used to acquire aerial images of natural disaster damage and impacted areas. Analyzing aerial images of natural disaster-impacted areas presents challenges in segmenting important objects impacted by natural disasters, particularly those with irregular shapes, varying sizes, and small sizes. Additionally, conducting detailed and automated assessments of disaster-impacted areas based on segmented images is complex. Until now, researchers have used several methods to

try to produce accurate image segmentation in the field of natural disasters. The state-of-the-art methods used today can be broadly classified into two categories: models that rely on traditional handcrafted features created by researchers [1], [2], and models based on deep learning. Research using deep learning is divided into two models: convolutional neural network (CNN) and transformer. Studies using CNN-based segmentation models have been conducted by researchers [3]-[13] and studies using transformer-based models have been conducted by researchers [14]-[27].

The studies that have used segmentation in solving natural disaster problems with transformers. Researcher [14] detected building damage using DamFormer, research [15] detected building damage using swin transformer encoder and UPerNet semantic segmentation head, and research [16] detected building damage using SDAFormer. The studies [14]-[16] used the xBD dataset. Researcher [17] classified building damage based on satellite images (xBD and LEVIR-CD datasets) after natural disasters using a transformer-based network called DAHiTrA. Researcher [18] automatically detected damaged buildings using improved Swin-Unet on the Gaofen-2/Jilin-1 dataset, satellite images dataset (xBD), and the AIST building change detection (ABCD) dataset. Researcher [19] segmented post-earthquake solid buildings using the swin transformer from remote sensing images with complex backgrounds. The studies [14]-[17] have the advantage of detecting four types of building damage: no damage, minor damage, major damage, and destroyed. However, these studies suffer from the drawback of relying on a comparison between pre-disaster and post-disaster image pairs to assess building damage. This results in a lengthier process for generating segmentation images. This is because each time a natural disaster takes place at a specific location, change detection is performed on both pre-disaster and post-disaster images to identify and detect building objects impacted by the disaster. Researcher [15] also has shortcomings that result in low model performance values. Researcher [18] has the advantage of testing to detect building damage in several remote sensing image datasets but has the disadvantage of producing very low model performance values.

The studies addressing the problem of flood natural disasters have been conducted by research [20] which detects and segments flooded areas from aerial images at the disaster site using FloodTransformer on the SWOC flood segmentation dataset, research [21] detects flooded areas using synthetic aperture radar (SAR) images on the S1GFloods dataset with differential attention metric-based network (DAM-Net), and studies [22], [23] detect flood-impacted objects using several models, one of which is SegFormer on the FloodNet dataset. The studies addressing the problem of landslide natural disasters have been conducted by [24] which detects landslides on the Bijie and Iburi small datasets using vision transformer (ViT), research [25] determines the exact extent of landslide areas using separable channel attention network (SCANet) on loess plateau landslide image dataset, and studies [26], [27] identifies coseismic landslides from coseismic landslide dataset using SegFormer semantic segmentation network. Researcher [20] has the advantage of using multiple open-source datasets but has the disadvantage of using very different image resolutions. Researcher [22] has the advantage of comprehensively testing several encoder-decoder and two-path-based architectures. The studies [22], [23] evaluated their performance on aerial image datasets but have the disadvantage of producing model performance values that are still lacking. Researcher [24] has the advantage of using two datasets with different landslide characteristics, and research [25] has the advantage of conducting extensive experiments on landslide datasets using several mainstream semantic segmentation networks compared with transformer architecture, but studies [24], [25] have the disadvantage of using image datasets with small image size and a small number of images for training, validation, and testing.

Based on a search of related previous studies, it is concluded that some studies use satellite-derived image datasets, namely studies [1], [4], [14]-[19], [24]. However, these datasets have drawbacks, such as being captured from a higher altitude and being susceptible to cloud and smoke interference. Additionally, they lack comprehensive information about disaster areas and objects impacted by natural disasters, which hinders their usefulness for detailed assessment of natural disasters. In contrast, this study uses a dataset of low-altitude, UAV-derived aerial images, which has many advantages over satellite images. The dataset used in this study is high-resolution aerial images of natural disaster-impacted areas to generate segmentation images that can be used automatically for detailed natural disaster assessment. Previous studies have also produced segmented images that only display a small number of object classes, namely studies [2]-[4], [14]-[18], [20], [21], [24]-[27], which have the disadvantage of not being able to detect and identify other important objects impacted by natural disasters that appear in post-disaster images, making them incomplete for use in a detailed and automated natural disaster assessment process. In contrast, this study includes nine object classes that represent various objects impacted by natural disasters, like buildings and roads impacted by floods, irregularly shaped objects like trees and grass, and small objects like vehicles, so that they can be used in the natural disaster assessment process in more detail and automatically. This cannot be done if only a small number of object classes are displayed. To produce a detailed and automatic assessment of areas impacted by natural disasters, it is also important to calculate the number of objects in each object class, calculate the area

of objects in each object class, and calculate the percentage impacted by natural disasters (PIND) so that natural disaster assessments can be calculated and clustered automatically into several categories.

Deep learning-based semantic segmentation has advanced quickly and is commonly applied in satellite image recognition for remote sensing. However, its utilization in recognizing aerial images from UAVs for addressing natural disaster issues, particularly with transformers for segmenting multiple objects, is limited. This study has the advantage of using post-disaster aerial images so that it can produce segmentation images more quickly without the need to compare with pre-disaster images. All of the previous research described above has the disadvantage of only segmenting the image and visually displaying the segmented image without conducting a detailed and automatic natural disaster assessment on the segmented image and still requiring human intervention in performing manual calculations to obtain detailed information about the impact of the disaster. In contrast, this study performs a detailed and automatic assessment of natural disaster-impacted areas (natural disaster assessment) on the segmented image. The main challenge is accurately performing aerial image semantic segmentation of natural disaster-impacted areas and performing detailed and automatic natural disaster assessments with several categories. The main contribution in this study is twofold: first, applying transformers for aerial image semantic segmentation of natural disaster-impacted areas that have complexity and diversity of objects visible in aerial images with four latest models of semantic segmentation networks, namely: bidirectional encoder representation from image transformers (BEiT), dense prediction transformer (DPT), OneFormer, and SegFormer, to segment various objects impacted by natural disasters, irregularly shaped and sized objects, and small objects in post-disaster aerial images; second, perform automatic natural disaster assessment on segmented images with categories: areas not impacted by natural disasters, areas lightly impacted by natural disasters, areas moderately impacted by natural disasters, and areas heavily impacted by natural disasters, using k-means and fuzzy c-means (FCM) clustering algorithms.

This paper is structured into multiple sections. Section 1 discusses the introduction and related research relevant to this study. Section 2 discusses the proposed framework or method, segmentation, transformer model built, datasets used, data augmentation and transfer learning, natural disaster assessment consisting of k-means clustering algorithm, FCM clustering algorithm, calculation of PIND, and calculation of object area in aerial images of areas impacted by natural disasters, and implementation details. Section 3 displays quantitative and qualitative results produced by the transformer model for semantic segmentation, displays quantitative and qualitative comparisons of the transformer model test results for semantic segmentation of aerial images of natural disaster-impacted areas with previous studies, and displays the results of detailed and automatic natural disaster assessments on segmented images. Section 4 summarizes the results of this study and explains further research.

2. METHOD

This section presents the proposed framework or method for semantic segmentation of aerial images of areas impacted by natural disasters using transformers for natural disaster assessment. An overview of the process is shown in Figure 1. The proposed method comprises eight stages, i.e.,: pre-processing stage; pre-training stage for transfer learning; training, validation, and testing stage; semantic segmentation model evaluation (performance evaluation) stage; segmentation result display stage (display of segmented objects in the image and object class label information); calculation stage of the number of objects and object area in each object class and PIND; natural disaster assessment stage using k-means clustering algorithm and FCM clustering algorithm; and natural disaster assessment result display stage. Each stage will be discussed in detail in this section.

2.1. Image segmentation

In the field of computer vision, one of the most crucial subjects is image segmentation. It has various applications, including scene understanding, analysis of medical images, perception for robots, video surveillance, augmented reality, image compression, and many other applications. Several researchers have developed various methods for image segmentation. Lately, there has been a considerable effort to build image segmentation methods utilizing deep learning models, driven by the achievements of these models in several computer vision tasks [28]. Image segmentation is the task of classifying pixels according to semantic labels or separating them into distinct objects. Semantic segmentation is the process of labeling individual pixels in an image based on a predefined set of categories, such as buildings, roads, and trees.

2.2. Transformer

Transformer-based models have achieved promising results in many computer vision and natural language processing (NLP) tasks [29]. The success of transformers in NLP has generated attention not just in various fields of computer vision but also in the aerial image semantic segmentation of natural disaster-impacted areas, where this study explores several transformer-based models. This study uses BEiT [30], DPT [31], OneFormer [32], and SegFormer [33] as transformer models. BEiT is a self-supervised vision

representation model inspired by bidirectional encoder representations from transformers (BERT) [34] in NLP. BEiT is designed for pre-training ViT using a masked image modeling (MIM) task. BEiT uses a pre-training task to help the model recognize images and identify objects in the images. Each image has two views in the pre-training phase: image patches and visual tokens. The model is pre-trained to recover the original visual tokens based on the corrupted image. MIM aims to recover masked image patches based on the encoding vector. The pre-training task aims to predict the original image's visual tokens based on the corrupted image's encoding vectors. After pre-training, the model parameters can be fine-tuned on downstream tasks, such as image classification and semantic segmentation, by appending task layers to the pre-trained model.

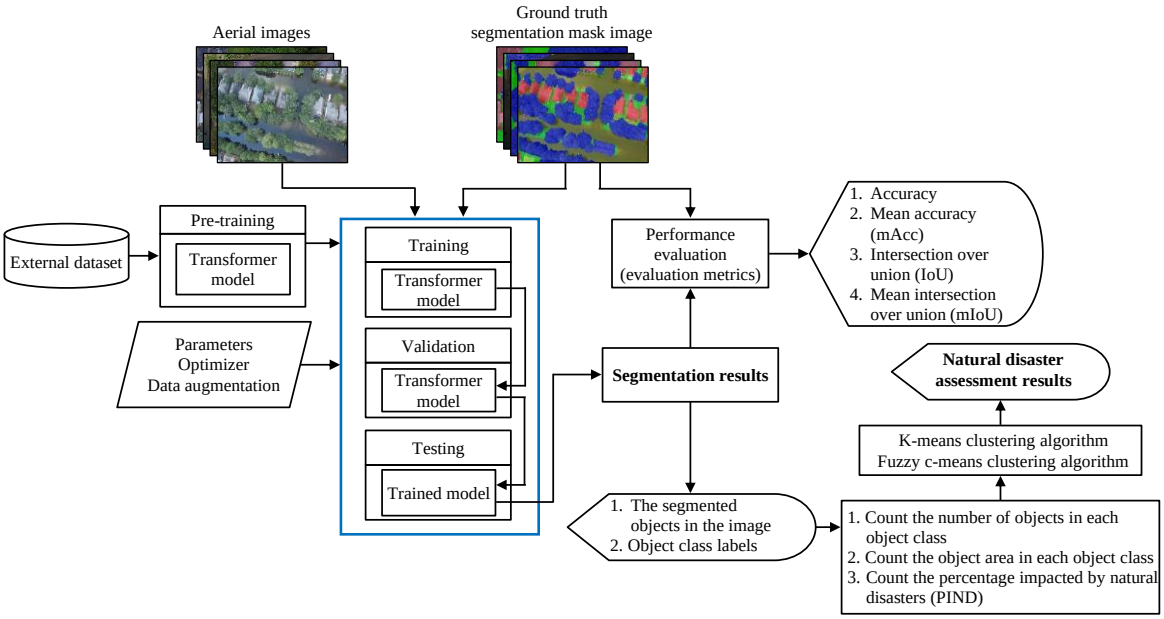

Figure 1. Overview of the proposed overall method

DPT is a neural network model architecture that combines image segmentation techniques with transformer architecture. DPT uses the ViT as the backbone for dense prediction tasks such as semantic image segmentation and monocular depth estimation. DPT is a dense prediction architecture based on an encoder-decoder design that utilizes the transformer as the basic computational building block of the encoder. DPT processes images for semantic segmentation by converting the input image into tokens through transformer stages. Thereafter, the tokens are recombined into image representations at several different resolutions. This process is followed by progressive merging to generate predictions at full resolution using a convolutional decoder.

OneFormer is a transformer designed to overcome the problems of using transformers as models for image segmentation. OneFormer uses a single transformer to perform image segmentation. Thus, the name "OneFormer" is derived from "One Transformer." OneFormer is a multi-task universal image segmentation framework that utilizes transformers. It requires training only once using a single universal architecture, model, and dataset. OneFormer utilizes a task-conditioned joint training approach, which uniformly selects several ground truth domains (semantic, instance, or panoptic) by utilizing all labels from panoptic annotations to train its multi-task model. OneFormer utilizes a pair of inputs: a sample image and a task input. OneFormer utilizes a backbone and pixel decoder to extract multiscale features from the input image.

SegFormer is a semantic segmentation model architecture used to identify and segment objects in images. SegFormer is a streamlined and effective semantic segmentation system that combines the transformer with a lightweight multilayer perceptron (MLP) decoder. SegFormer possesses two distinct characteristics: SegFormer comprises a novel transformer encoder with a hierarchical structure that produces multiscale features. The transformer encoder is positional encoding free and hierarchical. The model utilizes a hierarchical transformer encoder to extract high-resolution coarse and low-resolution fine features from the input image. Additionally, SegFormer simplifies the decoding process by avoiding using sophisticated decoders. The MLP decoder aggregates input from multiple layers, integrating local and global attention to provide a robust representation. The MLP decoder, which is lightweight, utilizes multi-level features to provide the final semantic segmentation mask.

In this study, BEiT and DPT use the ViT [35] architecture, OneFormer uses the swin transformer-large (Swin-L) [36] architecture, and SegFormer uses the mix transformer-B0 (MiT-B0) [33] architecture as their respective backbones. The parameters for BEiT, DPT, and SegFormer, 0.00006 and 0.01, were used for the learning rate and weight decay, respectively, and the optimizer used was Adam. Crop size and batch size for BEiT and DPT are set at 256×256 and 2, and for SegFormer are set at 512×512 and 4. The parameters for OneFormer are the learning rate is 0.0001, weight decay is 0.05, the optimizer used is Adam, crop size is set at 1024×1024, and batch size is set at 2. Cross entropy is a loss function in BEiT, DPT, and SegFormer. Contrastive query is a loss function in OneFormer.

2.3. Datasets

This study uses FloodNet [6] as a dataset in semantic segmentation of natural disaster-impacted area images, a dataset of post-disaster aerial images derived from UAVs, especially in flooded areas. FloodNet images are labeled at the pixel level, which makes them useful for semantic segmentation tasks. FloodNet has 2,343 images with a total of 55,739 objects annotated. FloodNet images and annotations have an average resolution of 3000×4000 pixels, with nine classes including building non-flooded, road non-flooded, building flooded, road flooded, vehicle, tree, grass, water, and pool. The FloodNet dataset is divided into 1686 images for training, 422 for validation, and 235 for testing.

This study uses an external dataset, namely ADE20K [37], for the pre-training dataset. ADE20K is a densely annotated dataset, which includes various annotations of scenes, objects, parts of objects, and, in some cases, even parts of parts. There are 25,000 images of complex daily scenes containing various objects in their natural spatial context.

2.4. Data augmentation and transfer learning

Data augmentation in aerial image semantic segmentation of natural disaster-impacted areas has several primary objectives that can enhance the semantic segmentation model performance. The following are some of the objectives of data augmentation in the semantic segmentation of aerial images of areas impacted by natural disasters: i) increase data diversity: by performing data augmentation, the variety and diversity of images can be increased. This helps the model to handle better scale differences, rotation, and perspective changes that may occur in aerial images of natural disaster-impacted areas; ii) reduce overfitting: data augmentation helps reduce overfitting by providing more variety in the training data. Models trained on diverse data are more likely to handle well-test data or situations in the field that are not encountered during training; iii) increase robustness: data augmentation can help make models more resilient to environmental variations that may occur in aerial images during or after natural disasters. This helps the model perform well under varying conditions; iv) improve segmentation accuracy: data augmentation can help improve segmentation accuracy by creating more variation in image detail and texture. This helps the model to identify better and map important objects or features in the aerial images of natural disaster-impacted areas; and v) minimize the limitations of limited data: sometimes, limited data availability can be a problem in semantic segmentation tasks. Data augmentation can help expand the available dataset and optimize the use of existing data. Designing appropriate data augmentation can improve the model's ability to deal with variations that may occur in aerial images of natural disaster-impacted areas, resulting in a more reliable and responsive semantic segmentation model.

Transfer learning aims to utilize the knowledge gained from models trained on related tasks in other environments or domains. In this context, some of the objectives of transfer learning involve: i) improved model performance: transfer learning can enhance the performance of the model on aerial image semantic segmentation of natural disaster-impacted areas. By using a model trained on a large and diverse dataset, the model can extract more general and in-depth features, which may be difficult to obtain from a limited dataset; ii) reduction of training data requirements: by using transfer learning, the model can utilize existing knowledge without the need to train it from the beginning. This is very important in natural disaster situations where training data may be limited or difficult to obtain. Transfer learning can help models adapt to relatively small datasets; iii) adapt to environmental variability: areas impacted by natural disasters often have unique and complex environmental conditions. Transfer learning can help models adapt more quickly to changing environmental conditions and characteristics that may not be present in the original training dataset; iv) resource utilization optimization: training a model from the beginning requires significant computational resources. Transfer learning can help optimize the use of resources by leveraging existing knowledge, thereby reducing training time and costs; v) handling contextual changes: natural disasters can result in major changes in the conditions of the impacted area. Transfer learning can help models adjust quickly to contextual changes, such as landform changes, infrastructure damage, or land cover changes; and vi) model generalization: transfer learning can help create more general models that can be used for different types of natural disasters. The obtained model can recognize patterns and features relevant to semantic segmentation without relying on the specific details of one particular disaster type. As such, transfer learning can be a

powerful tool in improving model performance on aerial image semantic segmentation of natural disaster-impacted areas and help better understand the emergency situation.

In this study, BEiT and DPT use flip and resize for data augmentation. OneFormer uses random scale, and SegFormer uses random scale, random brightness contrast, vertical flip, resize, horizontal flip, and Gaussian noise. This study uses the ADE20K pre-training dataset for transfer learning, which helps improve the semantic segmentation model's performance in detecting and recognizing objects in aerial images of areas impacted by natural disasters. Transfer learning is applied to each semantic segmentation model in a way that the model is first trained on the pre-training dataset, transferred, and used to retrain on the FloodNet dataset for semantic segmentation of aerial images of areas impacted by natural disasters efficiently.

2.5. Natural disaster assessment

2.5.1. K-means clustering algorithm

The k-means clustering algorithm is a frequently employed method for clustering data, including in the context of clustering areas affected by natural disasters. It does so by analyzing the area of objects in aerial images of such areas. The K-means algorithm is widely recognized as a prominent unsupervised clustering technique utilized for the automated clustering of data into coherent clusters. The algorithm aims to cluster the data into k clusters by identifying the centroid or cluster center and thereafter clustering the data with the nearest points. To correctly cluster data points into k clusters, a cost function is required to minimize the distance metric (e.g., Euclidean distance and Manhattan distance) [38]. The k-means clustering algorithm is a simple approach to clustering a dataset into k clusters (C1, C2, C3,..., Ck), which are represented by their cluster centers (centroids) [39]. The k-means clustering algorithm tries to cluster existing data into several clusters, where the data in one cluster has the same characteristics/features as each other and has characteristics/features that are different from the data in other clusters. The k-means algorithm can be used for image segmentation. The k-means technique is based on clustering similar pixels and median allocation. Repeating the same process multiple times provides better object identification [40]. In the context of clustering areas impacted by natural disasters based on the area of objects in aerial images of areas impacted by natural disasters, the following are the steps of the k-means clustering algorithm:

Step 1: data representation. Use all aerial images of natural disaster-impacted areas and calculate the areas of natural disaster-impacted objects in the ground truth image to obtain information about the area of objects in each image.

Step 2: cluster center (centroid) initialization. Select the desired number of clusters (k), and randomly select k points from the data as initial cluster centers (centroids) (C_i); the equation is shown in (1).

$$C_i = (x_i, y_i) \tag{1}$$

Step 3: clustering the data. Calculate the distance between each data point and each centroid. Determine the cluster for each data point based on the closest distance. The Euclidean distance between a data point $P(x_p, y_p)$ and a cluster center $C_i(x_i, y_i)$ is shown in (2).

$$D_{i,p} = \sqrt{(x_p - x_i)^2 + (y_p - y_i)^2} \tag{2}$$

Step 4: centroid update. Calculate the mean of all data points in each cluster to get a new centroid. Move the centroid to the mean point. The new cluster center (centroid) formula is shown in (3).

$$C_j = \frac{1}{nj}\sum_{i=1}^{nj} X_i \tag{3}$$

where C_j is the j-th cluster center, nj is the number of data points in the j-th cluster, and X_i is the i-th data point in the j-th cluster.

Step 5: iteration. Repeat step 3 of clustering the data and step 4 of updating the cluster centers until convergence (until there is no significant change in the clustering) or reaching the specified maximum number of iterations. The objective function used for clustering is shown in (4).

$$J = \sum_{j=1}^{K}\sum_{i=1}^{nj} D(X_i, C_j)^2 = \sum_{j=1}^{K}\sum_{i=1}^{nj} \left\| X_i - C_j \right\|^2 \tag{4}$$

where J is an objective function that tries to minimize the sum of squared distances between each data point and its cluster center.

Step 6: final result. The algorithm stops when the cluster center point does not change significantly or the maximum number of iterations is reached. After convergence, clusters are formed, and the object area data on aerial images of natural disaster-impacted areas are clustered into four clusters of natural disaster-impacted areas.

2.5.2. Fuzzy c-means clustering algorithm

FCM clustering is a clustering technique commonly employed inside the framework of the hard k-means method. FCM employs a fuzzy clustering technique to allow data to be included in all established clusters, each with varying degrees or membership levels ranging from 0 to 1. The extent to which data is present within a cluster is contingent upon its level of membership. The FCM algorithm is one of the most widely used clustering techniques, and it uses one of the Euclidean distance metrics as a similarity measurement. An important component of clustering algorithms is the measurement of similarity between data points [41]. The basic concept of FCM is to determine the cluster center, which will mark the mean location for each cluster. Among the many algorithms for image segmentation, the FCM algorithm is one of the most popular. The prominent advantage of the FCM algorithm is that the segmentation process is unsupervised, and the algorithm can be applied to images that have noise [42]. In the context of clustering areas impacted by natural disasters based on the area of objects in aerial images of areas impacted by natural disasters, the FCM algorithm can be adapted to determine clusters that reflect the level of damage or the level of impacted areas. The following are the steps to adapt FCM in that context:

Step 1: data representation. Use all aerial images of natural disaster-impacted areas and calculate the area of natural disaster-impacted objects in the ground truth image to obtain information about the area of objects in each image.

Step 2: initialization. Determine the desired number of clusters (c) to describe the level/category of natural disaster impacted. Initialize the membership matrix ($U=\{u_{ij}\}$) with a random value between 0 and 1, where u_{ij} is the membership level of the i-th data in cluster j. Ensure that each row of the U matrix sums to 1 ($\sum_j u_{ij} = 1$).

Step 3: cluster center calculation. Calculate the cluster center (v_j) using the (5).

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_i}{\sum_{i=1}^{n} u_{ij}^{m}} \tag{5}$$

where v_j is the j-th cluster center, m is a fuzzification parameter usually set between $1<m<\infty$, x_i is the i-th data, and n is the number of data.

Step 4: membership function calculation. Calculate the new membership matrix using the adjusted FCM in (6).

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|}\right)^{\frac{2}{m-1}}} \tag{6}$$

where u_{ij} is the membership level of data i to cluster j, $d_{ij}=\|x_i - v_j\|$ is the Euclidean distance between data i and cluster j center, $d_{ik}=\|x_i - v_k\|$ is the Euclidean distance between data i and cluster k center, c is the number of clusters, and m is the fuzzification parameter.

Step 5: iteration. Repeat steps 3 and 4 until the membership matrix converges, i.e., does not change significantly. When the iteration is complete, each data will have a membership level to each cluster, and the cluster centers will be updated. The result is a fuzzy cluster where each data has a membership level to each cluster. The FCM objective function is shown in (7).

$$J = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ij}^{m}\, d_{ij}^{2} = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ij}^{m} \left\|x_i - v_j\right\|^2 \tag{7}$$

Step 6: cluster determination. Each data will belong to the cluster with the highest membership level.

Step 7: result analysis. Analyze the cluster results to determine the clustering of natural disaster-impacted areas with a high degree of membership.

2.5.3. Percentage impacted by natural disasters

Calculating the percentage impacted by natural disasters is an important first step in disaster management and recovery. PIND is used to detect how much impact a natural disaster has on the impacted objects and areas. In post-disaster aerial images, PIND is needed to inform the percentage of objects and areas impacted by natural disasters. Calculating PIND can provide a clearer understanding of the scale and impact of natural disasters in an area and is used to analyze the percentage of areas vulnerable to natural disasters. The PIND value is used for damage assessment or assessment of areas impacted by natural disasters. The formula for calculating PIND is shown in (8)-(10):

$$PIND = \frac{Number\ of\ pixels\ of\ objects\ and\ areas\ impacted\ by\ natural\ disasters}{Total\ number\ of\ pixels} \times 100 \tag{8}$$

$$PIND = \frac{\sum_{i=1}^{n}(pixBF_i + pixRF_i)}{width\ of\ an\ image\ resolution \times length\ of\ an\ image\ resolution} \times 100 \quad (9)$$

or $$PIND = \sum_{i=1}^{n} \frac{pixBF_i + pixRF_i}{pixBNF_i + pixRNF_i + pixBF_i + pixRF_i + pixV_i + pixT_i + pixG_i + pixW_i + pixP_i + pixB_i} \times 100 \quad (10)$$

Where *pixBNF* is the number of pixels of the building non-flooded object, *pixRNF* is the number of pixels of the road non-flooded object, *pixBF* is the number of pixels of the building flooded object, *pixRF* is the number of pixels of the road-flooded object, *pixV* is the number of pixels of the vehicle object, *pixT* is the number of pixels of the tree object, *pixG* is the number of pixels of the grass object, *pixW* is the number of pixels of the water object, *pixP* is the number of pixels of the pool object, *pixB* is the number of background pixels, *n* is the total number of each object, and *i* symbolizes each object respectively.

2.5.4. Area of objects in aerial images of natural disaster-impacted areas

For the assessment of natural disasters, it is necessary to calculate the area of objects in aerial images of areas impacted by natural disasters, which has several important purposes, including: i) the area of objects impacted by natural disasters in images resulting from semantic segmentation of aerial images of areas impacted by natural disasters can provide a real description of the extent to which the area is impacted and the area obtained matches the area in the real world; ii) this information can be used to measure the level of damage and assist in emergency response planning and post-disaster recovery; iii) data on the area of objects impacted by natural disasters can be used to plan recovery and rehabilitation efforts; iv) a better understanding of how large the impacted area is can help in the efficient allocation of resources to restore the area, v) information on the area and type of objects impacted by natural disasters helps in prioritizing recovery, vi) information on the area impacted by natural disasters can help in evacuation planning and support evacuation route planning, and vii) information on the area of objects can be utilized to help the development of disaster mitigation policies and as reference data for further research in the field of natural disaster management. The formula for calculating the object area in each object class is shown in (11)-(13).

$$Actual\ width = 2 \times camera\ altitude \times \tan\left(\frac{FOV_h}{2}\right) \quad (11)$$

$$Actual\ height = 2 \times camera\ altitude \times \tan\left(\frac{FOV_v}{2}\right) \quad (12)$$

$$Segmentation\ area\ of\ each\ object\ class = \left(\frac{Total\ pixel\ of\ each\ object}{w \times l}\right) \times Actual\ width \times Actual\ height \quad (13)$$

where *FOV* is the field of view, FOV_h is the angle of view that the camera can capture horizontally, FOV_v is the angle of view that the camera can capture vertically, *w* is the width of an image resolution, and *l* is the length of an image resolution.

Using (11)-(13), the area of the objects in all ground truth segmentation mask images of the FloodNet dataset was calculated, resulting in the number of images, number of objects, and the size of the object area corresponding to each object class shown in Table 1.

Table 1. Number of images, number of objects, and area of objects that correspond with the object classes in FloodNet

Object class	Number of images	Number of objects	Area of objects (m^2)
Building non-flooded	880	3427	919080.46
Road non-flooded	1175	2155	1625586.62
Building flooded	245	3248	460791.95
Road flooded	264	495	742013.49
Vehicle	813	4535	49385.88
Tree	1882	19682	5078292.70
Grass	2161	19682	16008212.91
Water	1059	1374	3124001.20
Pool	531	1141	56425.02
Total		55739	28063790.23

Data normalization is then carried out after calculating the number of objects and the area of objects. Data normalization is the process of adjusting data values in a dataset so that they can have a uniform or normal scale. The data normalization function involves data transformation so that the values are within a certain range or have a more controlled distribution. For the normalization method, this study chose min-max normalization. Min-max normalization transforms values in a certain range into a range between 0 and 1. The min-max normalization formula is shown in (14):

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (14)$$

where X_{norm} is the normalized value, X is the original value to be normalized, X_{min} is the minimum value in the original range, and X_{max} is the maximum value in the original range.

Furthermore, using the normalized data, the k-means clustering algorithm and the FCM clustering algorithm are used to calculate and cluster the data of the area of objects impacted by natural disasters into four clusters of natural disaster-impacted areas, which are the four categories of natural disaster assessment, namely areas not impacted by natural disasters, areas lightly impacted by natural disasters, areas moderately impacted by natural disasters, and areas heavily impacted by natural disasters. The clustering of natural disaster-impacted areas reflects the level of damage or impact of an area. The results of the calculation and clustering of natural disaster assessments are used to automatically calculate and predict natural disaster assessments on images resulting from semantic segmentation. These results are very useful for conducting natural disaster assessments on all aerial images of natural disaster-impacted areas. The clustering results of the k-means algorithm are shown in Table 2 and the clustering results of the FCM algorithm are shown in Table 3.

Table 2. The centroid value of each cluster using the k-means clustering algorithm

Cluster number	Building flooded	Road flooded	Cluster name
0	0.00130667	0.0012569	Areas not impacted by natural disasters
1	0.09480807	0.75438608	Areas lightly impacted by natural disasters
2	0.12038973	0.30203463	Areas moderately impacted by natural disasters
3	0.25212445	0.17462374	Areas heavily impacted by natural disasters

Table 3. The centroid value of each cluster using the FCM clustering algorithm

Cluster number	Building flooded	Road flooded	Cluster name
0	0.000856336	0.00074293	Areas not impacted by natural disasters
1	0.091169744	0.772654604	Areas lightly impacted by natural disasters
2	0.131145276	0.278849165	Areas moderately impacted by natural disasters
3	0.255703145	0.173164929	Areas heavily impacted by natural disasters

2.6. Implementation details

This study builds the transformer model and implements the segmentation network using the deep learning framework PyTorch 1.13.1. All transformer models were trained for 50 epochs. This study uses personal computer hardware with an Intel Core i7 processor, 32 GB RAM, and NVIDIA GeForce RTX 3080 GPU (10 GB), with CUDA version 11.7 and Windows 10 Pro operating system.

To evaluate and compare the transformer model performance, this study displays evaluation metrics consisting of accuracy, mean accuracy (mAcc), intersection over union (IoU), and mean intersection over union (mIoU). Accuracy is a commonly employed evaluation metric for assessing the performance of models. Accuracy measures how well the model predicts correctly compared to all predictions made. The accuracy metric can be defined as the proportion of accurately predicted pixels compared to the total number of pixels. Mean accuracy (mAcc) is the ratio of the sum of the accuracy of each object class to the total number of object classes. IoU is a metric that measures the pixel-level similarity between the segmentation result image and the ground truth segmentation mask image. mIoU is the ratio of the number of IoU of each class to the total number of classes. The formulas for calculating accuracy, mean accuracy (mAcc), IoU, and mIoU are respectively shown in (15)-(18):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$mAcc = \frac{1}{k}\sum_{i=1}^{k} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (16)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (17)$$

$$mIoU = \frac{1}{k}\sum_{i=1}^{k} \frac{TP_i}{TP_i + FP_i + FN_i} \quad (18)$$

Where TP is true positive, TN is true negative, FP is false positive, FN is false negative, k is the total number of object classes, and i represents each object class respectively, namely building non-flooded, road non-flooded, building flooded, road flooded, vehicle, tree, grass, water, and pool.

3. RESULTS AND DISCUSSION

A comparison of the accuracy and performance (quantitative results) of all transformer models for semantic segmentation of aerial images of areas impacted by natural disasters is shown in Table 4, with the accuracy and IoU values for each object class (the values highlighted in bold are the best). The visualization results (qualitative results) of all transformer models are provided in Figure 2. The results presented in Table 4 demonstrate that the SegFormer model outperforms other models on the mAcc and mIoU evaluation metrics. The second best result is achieved by the OneFormer model, while the BEiT and DPT models have lower mIoU values. The IoU values for small shaped object classes, like vehicles and pools, produced by the BEiT and DPT models, are still very low.

Table 4. Comparison of transformer model test results with accuracy and IoU values (in %) for each object class

Object class	BEiT		DPT		OneFormer		SegFormer	
	Accuracy	IoU	Accuracy	IoU	Accuracy	IoU	Accuracy	IoU
Building non-flooded	96.83	31.58	97.39	39.60	99.35	80.47	98.64	65.74
Road non-flooded	97.10	55.13	97.94	65.44	99.41	75.45	98.78	76.99
Building flooded	97.94	29.86	97.49	26.59	93.90	72.19	98.77	58.29
Road flooded	96.06	31.86	97.28	38.49	93.47	65.80	97.01	50.33
Vehicle	99.81	4.37	99.65	10.33	99.59	43.14	99.87	49.60
Tree	90.82	56.25	93.25	68.33	91.98	71.69	96.35	81.85
Grass	79.12	66.33	85.70	76.96	88.28	61.39	91.40	85.28
Water	86.08	33.27	93.45	59.05	96.39	40.41	95.52	66.36
Pool	99.37	2.29	99.56	3.91	99.52	42.05	99.83	56.70
	mAcc=93.68	mIoU=34.55	mAcc=95.75	mIoU= 43.19	mAcc=95.77	mIoU= 61.40	mAcc=**97.35**	mIoU=**65.68**

The SegFormer model performed well, outperforming the second-best model, OneFormer, by 97.35% in the mAcc value and 65.68% in the mIoU value. The difference in mIoU of BEiT, DPT, and OneFormer models with SegFormer is 31.13% (65.68−34.55%) between SegFormer and BEiT, 22.49% (65.68−43.19%) between SegFormer and DPT, and 4.28% (65.68−61.40%) between SegFormer and OneFormer. The difference in mAcc between the first-best model (SegFormer) and the second-best model (OneFormer) is 1.58% (97.35−95.77%). The segmented image shows that buildings and roads impacted by floods (building flooded and road flooded) and small shaped objects (vehicles and pools) are not well segmented by the BEiT and DPT models. These results show that the BEiT and DPT models can still not learn feature mapping well and have limited ability to segment objects impacted by natural disasters and small shaped objects.

Figure 2 visually compares the segmentation results of the overall transformer model. Figure 2(a) shows aerial images of natural disasters-impacted areas, Figure 2(b) shows the ground truth segmentation mask images, and Figures 2(c)-(f) are the predicted images (segmented images) generated by all transformer models, namely BEiT, DPT, OneFormer, and SegFormer. The SegFormer model shows that it is capable of segmenting objects of buildings and roads impacted by natural disasters and various irregularly shaped and sized objects, which are important objects in natural disaster events. The SegFormer model is also able to segment vehicles and pools, which are the smallest objects. In this study, it is very important to distinguish objects impacted by natural disasters and objects that are not impacted by natural disasters. The SegFormer model successfully segmented and distinguished buildings that were flooded, buildings that were not flooded, roads that were flooded, roads that were not flooded, and other objects.

Figure 3 shows the receiver operating characteristics (ROC) curve and the area under curve (AUC) value for each example of the predicted images (segmentation results) in Figure 2, providing a visual representation of the performance of the SegFormer model. Figure 3(a) shows aerial images of natural disasters-impacted areas, Figure 3(b) shows the segmented images using SegFormer, and Figure 3(c) shows the ROC curve and AUC values, which have colors corresponding to their respective object classes, with the x-axis is the false positive rate (FPR) and the y-axis is the true positive rate (TPR). The AUC value is used to measure the SegFormer model performance in segmenting objects in aerial images of natural disaster-impacted areas according to their respective object classes. From the four ROC curves, it can be seen that all object class curves are above the discontinuous black baseline line or diagonal line crossing from point 0.0, which means that the SegFormer model performs well in producing semantic segmentation images for their respective object classes. All ROC curves have a mean AUC value above 0.9, indicating that each object's semantic segmentation results in aerial images of natural disaster-impacted areas are excellent, and the objects are segmented according to their respective object classes.

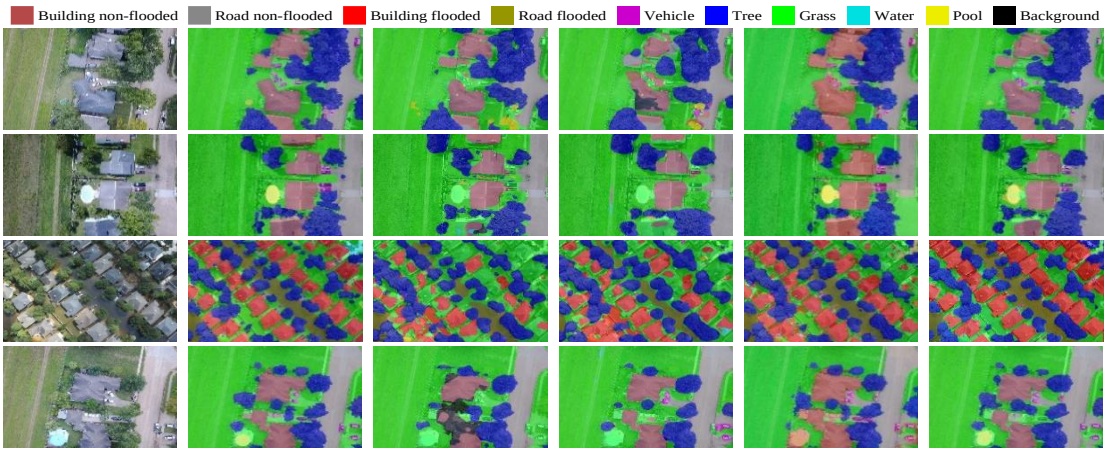

Figure 2. Visual comparison of transformer models; (a) aerial images of natural disasters-impacted areas, (b) ground truth, (c) BEIT, (d) DPT, (e) OneFormer, and (f) SegFormer

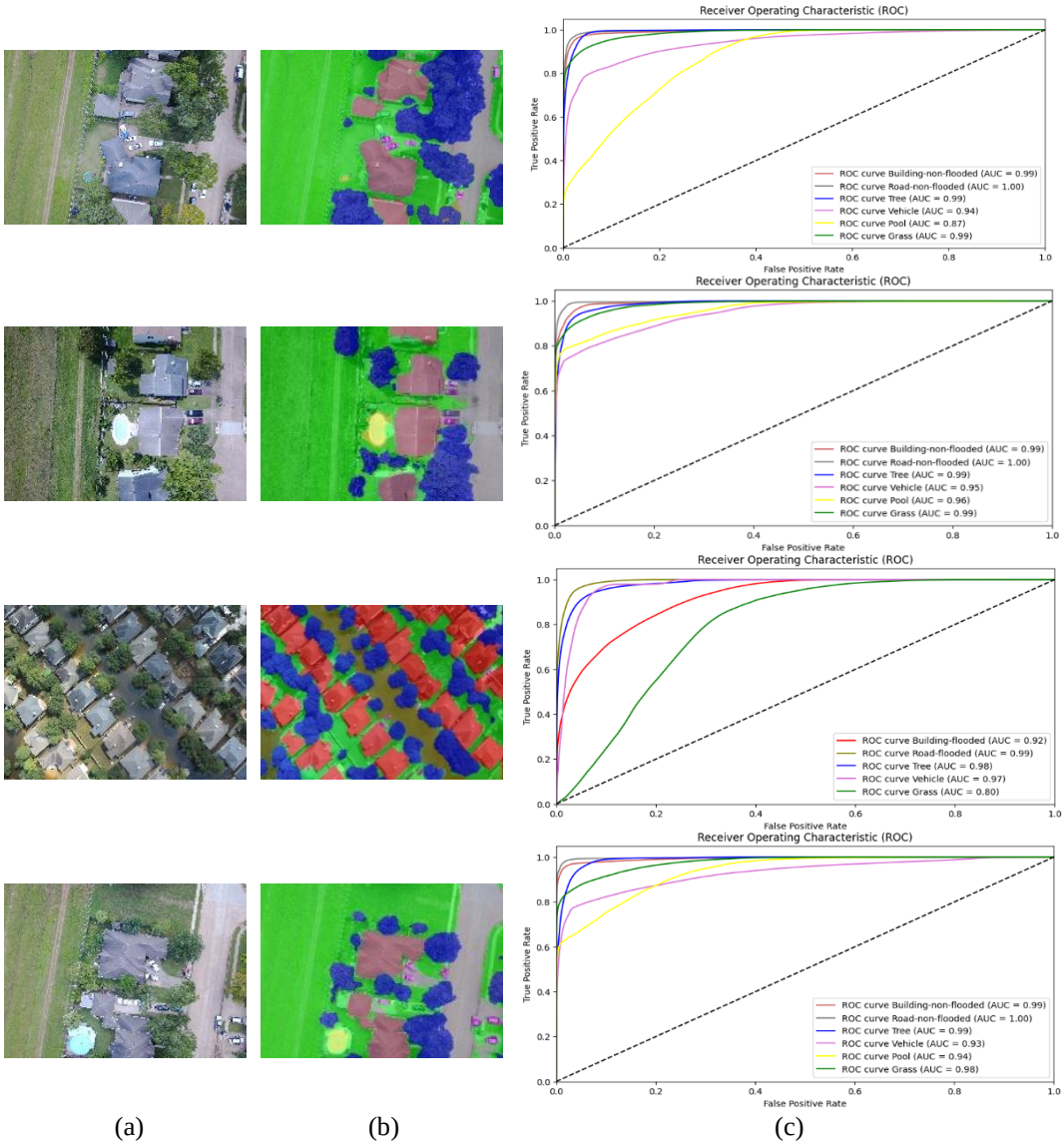

Figure 3. ROC curve and AUC values of segmentation results; (a) aerial images of natural disasters-impacted areas, (b) segmentation results using SegFormer, and (c) ROC curve and AUC values

The test results of the SegFormer model produced in this study outperformed the test outcomes produced by other studies in the literature review that also used the transformer model and the same dataset (FloodNet) on aerial image recognition in the context of natural disasters using semantic segmentation. Researcher [22] produced mAcc values of 89.50% and mIoU of 61.60%, and research [23] produced mAcc values of 88.50% and mIoU of 59.69%, for the SegFormer model, which are the highest mAcc and mIoU values in these studies. In comparison, this study produces higher mAcc and mIoU values of 97.35% and 65.68% for the same model, SegFormer. The results of this study also outperformed the test outcomes produced by other studies that used convolutional neural network-based models and used the same dataset for semantic segmentation. Researcher [5] used the U-Net, PSPNet, and DeepLabV3+ models, which resulted in the highest mIoU value of 56.22% in the DeepLabV3+ model. Researcher [7] used the PSPNet, DeepLabV3, and U-Net models, which resulted in the highest mIoU value of 56% in the PSPNet model. Researcher [8] used several encoder-decoder models, which resulted in the highest mIoU value of 50.44% in the UNet-MobileNetV2 model. Researcher [13] used the DeepLabV3 model with the EfficientNet-B4 backbone, which resulted in the best mIoU and mAcc values of 48.1% and 90%, respectively. The results in studies [5], [7], [8], [13] have lower mIoU performance values compared to the results of this study. The studies [6], [9]-[12] have weaknesses in the visual display of segmented images that are less accurate; important objects impacted by natural disasters, irregularly shaped and sized objects, and small objects in the segmented image samples are not segmented properly in these studies. This study produces a better visual appearance of semantic segmentation results of aerial images of areas impacted by natural disasters. A quantitative comparison of the advantages of the SegFormer model produced in this study compared to previous studies is shown in Table 5, the values highlighted in bold are the best. Figure 4 shows a qualitative comparison of the advantages of the SegFormer model produced in this study compared to previous studies. Figure 4(a) shows aerial images of natural disasters-impacted areas, Figure 4(b) shows the ground truth segmentation mask images, Figure 4(c) shows the segmented images generated by the deep learning models for semantic segmentation used in previous studies, and Figure 4(d) shows the segmented images generated by the SegFormer model in this study.

Table 5. Quantitative comparison of transformer model test results against previous studies using evaluation metrics (in %)

Model	Building -flooded	Building-non flooded	Road-flooded	Road-non-flooded	Water	Tree	Vehicle	Pool	Grass	Evaluation metrics
DeepLabV3+ [5]	48.00	69.00	48.00	75.00	72.00	76.00	15.00	18.00	85.00	mIoU=56.22
PSPNet (152) [7]				N/A						mIoU=56.00
UNet-MobileNetV2 [8]	43.50	59.30	21.20	61.20	73.30	64.90	15.10	32.70	82.80	mIoU=50.44
DeepLabV3-EfficientNet-B4 [13]				N/A						mAcc=90.00
				N/A						mIoU=48.10
SegFormer [22]				N/A						mAcc=89.50
	66.80	59.30	44.60	67.70	67.10	69.60	45.60	48.30	85.50	mIoU=61.61
SegFormer [23]				N/A						mAcc=88.50
	44.31	69.80	44.50	77.62	74.77	78.95	19.35	40.35	87.57	mIoU=59.69
SegFormer (this study)	98.77	98.64	97.01	98.78	95.52	96.35	99.87	99.83	91.40	mAcc=**97.35**
	58.29	65.74	50.33	76.99	66.36	81.85	49.60	56.70	85.28	mIoU=**65.68**

The segmentation results of the model with the best accuracy and performance, SegFormer, are then integrated with the k-means and FCM clustering algorithms to automatically assess areas impacted by natural disasters with four categories: areas not impacted by natural disasters, areas lightly impacted by natural disasters, areas moderately impacted by natural disasters, and areas heavily impacted by natural disasters. The results of semantic segmentation using the model with the best accuracy and performance, SegFormer, and the results of automatic natural disaster assessment using k-means and FCM clustering algorithms are shown in Figure 5. Figure 5(a) shows aerial images of natural disaster-impacted areas, Figure 5(b) shows the ground truth segmentation mask images and the natural disaster assessment results on the ground truth images, Figure 5(c) shows the segmented images generated by the SegFormer model, Figure 5(d) shows the natural disaster assessment results on the segmented images using the k-means clustering algorithm, and Figure 5(e) shows the natural disaster assessment results on the segmented images using the FCM clustering algorithm. The first row in Figure 5 shows an example of natural disaster assessment with the category of areas not impacted by natural disasters, the second row shows an example of natural disaster assessment with the category of areas lightly impacted by natural disasters, the third row shows an example of natural disaster assessment with the category of areas moderately impacted by natural disasters, and the fourth row shows an example of natural disaster assessment with the category of areas heavily impacted by natural disasters.

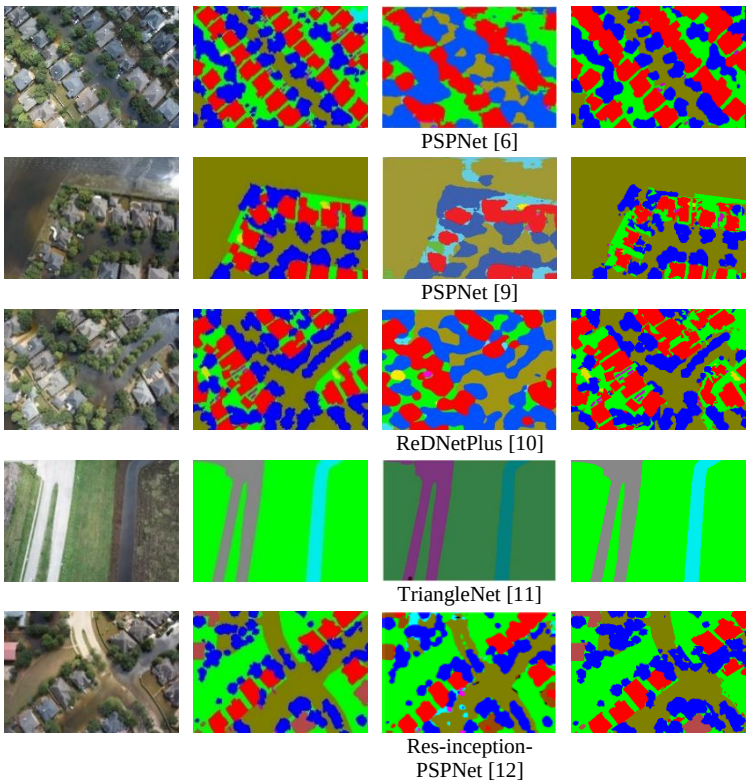

Figure 4. Visual (qualitative) comparison of transformer model test results against previous studies; (a) aerial images, (b) ground truth, (c) previous studies, and (d) SegFormer (this study)

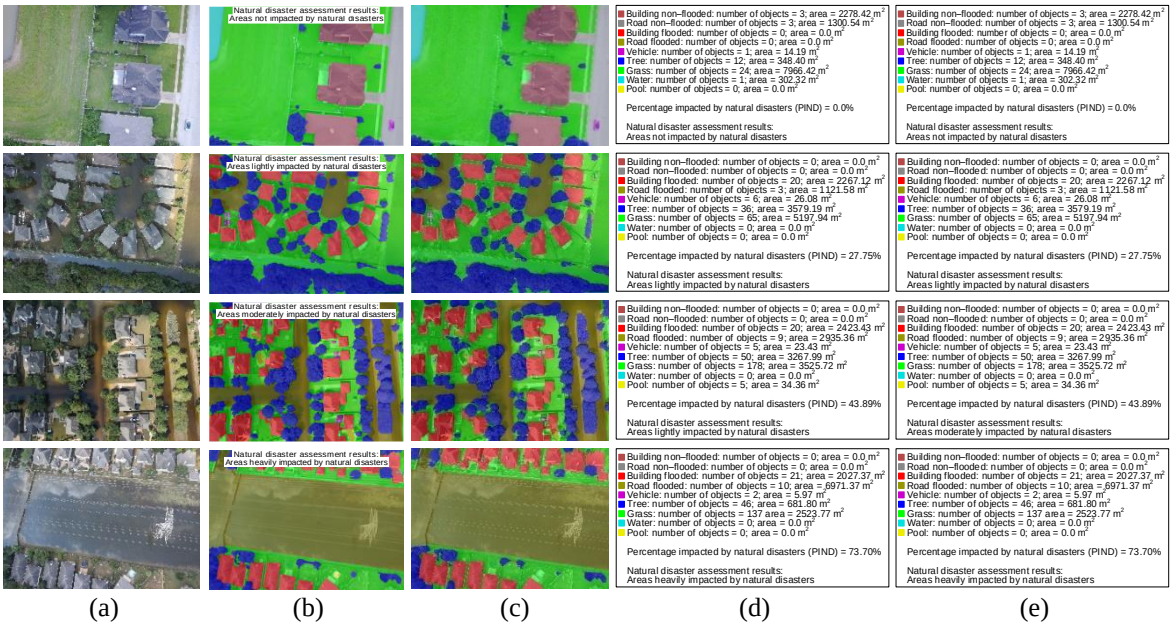

Figure 5. Semantic segmentation results using SegFormer and natural disaster assessment results using k-means and FCM clustering algorithms; (a) aerial images of natural disasters-impacted areas, (b) ground truth and the results of natural disaster assessment on ground truth, (c) segmentation results using SegFormer, (d) the results of natural disaster assessment on segmented images using the k-means clustering algorithm, and (e) the results of natural disaster assessment on segmented images using the FCM clustering algorithm

The SegFormer model and FCM clustering algorithm work well, as evidenced by the natural disaster assessment results in the ground truth images (Figure 5(b)) and the natural disaster assessment results in the segmented images (Figure 5(e)), showing the same assessment results and PIND values are in accordance with the respective natural disaster impacted area categories. Figure 5 shows the effective performance of the FCM clustering algorithm in categorizing the natural disaster assessment into four different clusters. This study succeeded in performing a detailed and automatic natural disaster assessment on the segmented images, which completely displays information on the number of each object class, displays information on the area of objects in each object class, displays information on the PIND, and displays information on the assessment of areas impacted by natural disasters, all of which have not been done by previous studies. This study accurately produces semantic segmentation of aerial images of areas impacted by natural disasters and natural disaster assessments, which can be used for emergency disaster management quickly, effectively, and efficiently in natural disaster management systems.

4. CONCLUSION

This study has succeeded in producing semantic segmentation of aerial images of areas impacted by natural disasters, and detailed and automatic assessment of natural disasters on the segmented images. The use of technology transformer model, aerial images of natural disaster-impacted areas derived from UAVs, semantic segmentation, counting the number of objects in each object class, counting the area of objects impacted by natural disasters, counting the percentage impacted by natural disasters, and clustering algorithms in natural disaster assessment can provide accurate and extensive information to support effective decisions and actions in the face of natural disasters. It can be a crucial component in effective disaster management efforts and post-disaster recovery. The SegFormer model achieved good segmentation results and outperformed other transformer models on semantic segmentation of natural disaster-impacted areas using post-disaster aerial images. The best natural disaster assessment results are shown by the FCM clustering algorithm, which can automatically cluster natural disaster assessments into four categories accurately on the segmented images. Further study will integrate the aerial image segmentation system of natural disaster-impacted areas, UAV video results, and web-based natural disaster assessment to build a natural disaster management system that can assist disaster emergency management more quickly, effectively, and efficiently.

ACKNOWLEDGEMENTS

This research was funded by the Directorate of Research, Technology, and Community Service, Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia in the Doctoral Dissertation Research (*Penelitian Disertasi Doktor-PDD*) program, with contract numbers 020/E5/PG.02.00.PL/2023 and 01676/UN4.22/PT.01.03/2023.

REFERENCES

[1] J. Tu, D. Li, W. Feng, Q. Han, and H. Sui, "Detecting damaged building regions based on semantic scene change from multi-temporal high-resolution remote sensing images," *ISPRS International Journal of Geo-Information*, vol. 6, no. 131, pp. 1–15, 2017, doi: 10.3390/ijgi6050131.

[2] N. S. Ibrahim, M. K. Osman, S. B. Mohamed, S. H. Y. S. Abdullah, and S. M. Sharun, "The application of UAV images in flood detection using image segmentation techniques," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 2, pp. 1219–1226, Aug. 2021, doi: 10.11591/ijeecs.v23.i2.pp1219-1226.

[3] H. S. Munawar, F. Ullah, S. Qayyum, S. I. Khan, and M. Mojtahedi, "UAVs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection," *Sustainability*, vol. 13, no. 14, pp. 1–22, Jul. 2021, doi: 10.3390/su13147547.

[4] A. Gupta, S. Watson, and H. Yin, "Deep learning-based aerial image segmentation with open data for disaster impact assessment," *Neurocomputing*, vol. 439, pp. 22–33, Jun. 2021, doi: 10.1016/j.neucom.2020.02.139.

[5] S. Khose, A. Tiwari, and A. Ghosh, "Semi-supervised classification and segmentation on high resolution aerial images," *arXiv preprint arXiv:2105.08655*, 2021.

[6] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. R. Murphy, "FloodNet: A high resolution aerial imagery dataset for post flood scene understanding," *IEEE Access*, vol. 9, pp. 89644–89654, Jun. 2021, doi: 10.1109/ACCESS.2021.3090981.

[7] D. Hernández, J. M. Cecilia, J.-C. Cano, and C. T. Calafate, "Flood detection using real-time image segmentation from unmanned aerial vehicles on edge-computing platform," *Remote Sensing (Basel)*, vol. 14, no. 1, pp. 1–20, Jan. 2022, doi: 10.3390/rs14010223.

[8] F. Safavi *et al.*, "Efficient semantic segmentation on edge devices," *arXiv*, Dec. 2022, doi: 10.48550/arXiv.2212.13691.

[9] F. Safavi, T. Chowdhury, and M. Rahnemoonfar, "Comparative study between real-time and non-real-time segmentation models on flooding events," in *2021 IEEE International Conference on Big Data (Big Data)*, Dec. 2021, pp. 4199–4207. doi: 10.1109/BigData52589.2021.9671314.

[10] T. Chowdhury and M. Rahnemoonfar, "Attention for damage assessment," in *CML 2021, Workshop on Tackling Climate Change with Machine Learning*, 2021.

[11] D. Zhang, R. Zheng, G. Luosang, and P. Yang, "TriangleNet: Edge prior augmented network for semantic segmentation through cross-task consistency," *International Journal of Intelligent Systems*, vol. 2023, pp. 1–16, 2023, doi: 10.1155/2023/1616055.

[12] S. D. Khan and S. Basalamah, "Multi-scale and context-aware framework for flood segmentation in post-disaster high resolution aerial images," *Remote Sensing (Basel)*, vol. 15, no. 8, pp. 1–16, 2023, doi: 10.3390/rs15082208.

[13] R. D. I. Puspitasari, F. Q. Annisa, and D. Ariyanto, "Flooded area segmentation on remote sensing image from unmanned aerial vehicles (UAV) using DeepLabV3 and EfficientNet-B4 model," in *2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, Oct. 2023, pp. 216–220 doi: 10.1109/IC3INA60834.2023.10285752.

[14] H. Chen, E. Nemni, S. Vallecorsa, X. Li, C. Wu, and L. Bromley, "Dual-Tasks Siamese Transformer Framework for Building Damage Assessment," *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, Kuala Lumpur, Malaysia, 2022, pp. 1600-1603, doi: 10.1109/IGARSS46834.2022.9883139.

[15] Z. Xia, Z. Li, Y. Bai, J. Yu, and B. Adriano, "Self-supervised learning for building damage assessment from large-scale xBD satellite imagery benchmark datasets," in *Database and Expert Systems Applications*, 2022, pp. 373–386, doi: 10.1007/978-3-031-12423-5_29.

[16] Y. Da, Z. Ji, and Y. Zhou, "Building damage assessment based on Siamese hierarchical Transformer framework," *Mathematics*, vol. 10, no. 11, pp. 1–23, May 2022, doi: 10.3390/math10111898.

[17] N. Kaur, C. C. Lee, A. Mostafavi, and A. Mahdavi-Amiri, "Large-scale building damage assessment using a novel hierarchical Transformer architecture on satellite images," *Computer-Aided Civil and Infrastructure Engineering*, pp. 1–20, Feb. 2023, doi: 10.1111/mice.12981.

[18] S. Xu, X. He, X. Cao, and J. Hu, "Damaged building detection with improved Swin-Unet model," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–10, Jul. 2022, doi: 10.1155/2022/2124949.

[19] L. Cui, X. Jing, Y. Wang, Y. Huan, Y. Xu, and Q. Zhang, "Improved Swin Transformer-based semantic segmentation of postearthquake dense buildings in urban areas using remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 369–385, Nov. 2022, doi: 10.1109/JSTARS.2022.3225150.

[20] A. Roy, S. S. Kulkarni, V. Soni, and A. Chittora, "Transformer-based flood scene segmentation for developing countries," in *arXiv preprint arXiv:2210.04218*, 2022.

[21] T. Saleh, X. Weng, S. Holail, C. Hao, and G.-S. Xia, "DAM-Net: Global flood detection from SAR imagery using differential attention metric-based Vision Transformers," *arXiv preprint arXiv:2306.00704*, 2023.

[22] F. Safavi and M. Rahnemoonfar, "Comparative Study of Real-Time Semantic Segmentation Networks in Aerial Images During Flooding Events," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 4-20, 2023, doi: 10.1109/JSTARS.2022.3219724.

[23] K. Gupta and P. Mishra, "Post-disaster segmentation using FloodNet," *Studies*, vol. 13, p. 8, 2021.

[24] Z. Yang, C. Xu, and L. Li, "Landslide detection based on ResU-Net with Transformer and CBAM embedded: Two examples with geologically different environments," *Remote Sensing (Basel)*, vol. 14, no. 12, pp. 1–19, Jun. 2022, doi: 10.3390/rs14122885.

[25] Z. Wang, T. Sun, K. Hu, Y. Zhang, X. Yu, and Y. Li, "A deep learning semantic segmentation method for landslide scene based on Transformer architecture," *Sustainability*, vol. 14, no. 23, pp. 1–22, Dec. 2022, doi: 10.3390/su142316311.

[26] X. Tang, Z. Tu, Y. Wang, M. Liu, D. Li, and X. Fan, "Automatic detection of coseismic landslides using a new Transformer method," *Remote Sensing (Basel)*, vol. 14, no. 12, pp. 1–19, May 2022, doi: 10.3390/rs14122884.

[27] X. Wang, C. Fang, X. Tang, L. Dai, X. Fan, and Q. Xu, "Research on emergency evaluation of landslides induced by the Luding Ms 6.8 earthquake," *Geomatics and Information Science of Wuhan University*, vol. 48, no. 1, pp. 25–35, Jan. 2023, doi: 10.13203/j.whugis20220586.

[28] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3523–3542, Feb. 2021, doi: 10.1109/TPAMI.2021.3059968.

[29] A. A. Aleissaee *et al*., "Transformers in remote sensing: A survey," *Remote Sensing (Basel)*, vol. 15, no. 7, pp. 1–31, Mar. 2023, doi: 10.3390/rs15071860.

[30] H. Bao, L. Dong, S. Piao, and F. Wei, "BEiT: BERT pre-training of image Transformers," ArXiv, vol. abs/2106.08254, pp. 1–18, 2021, doi: 10.48550/arXiv.2106.08254.

[31] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision Transformers for dense prediction," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 12159–12168, doi: 10.1109/ICCV48922.2021.01196.

[32] J. Jain, J. Li, M. Chiu, A. Hassani, N. Orlov, and H. Shi, "OneFormer: One Transformer to rule universal image segmentation," in *CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 2989–2998, doi: 10.48550/arXiv.2211.06220.

[33] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with Transformers," in *Advances in neural information processing systems*, vol. 34, no. 1, pp. 2077-12090, Dec. 2021.

[34] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional Transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Jun. 2019, pp. 4171–4186. doi: 10.18653/V1/N19-1423.

[35] A. Dosovitskiy *et al*., "An image is worth 16x16 words: Transformers for image recognition at scale," *in International Conference on Learning Representations (ICLR 2021)*, Aug. 2021, pp. 1–22, doi: 10.48550/arXiv.2010.11929.

[36] Z. Liu *et al*., "Swin Transformer: Hierarchical Vision Transformer using shifted windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 9992–10002. doi: 10.1109/ICCV48922.2021.00986.

[37] B. Zhou *et al*., "Semantic understanding of scenes through the ADE20K dataset," *International Journal of Computer Vision*, vol. 127, no. 3, pp. 302–321, Dec. 2019, doi: 10.1007/s11263-018-1140-0.

[38] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Student engagement level in an e-learning environment: Clustering using k-means," *American Journal of Distance Education*, vol. 34, no. 2, pp. 137–156, Mar. 2020, doi: 10.1080/08923647.2020.1696140.

[39] M. N. Reza, I. S. Na, S. W. Baek, and K.-H. Lee, "Rice yield estimation based on k-means clustering with graph-cut segmentation using low-altitude UAV images," *Biosystems engineering*, vol. 177, pp. 109–121, 2019, doi: 10.1016/j.biosystemseng.2018.09.014.

[40] K. Deeparani and P. Sudhakar, "Efficient image segmentation and implementation of k-means clustering," *Materials Today: Proceedings*, vol. 45, no. 9, pp. 8076–8079, 2021, doi: 10.1016/j.matpr.2021.01.154.
[41] J. Arora, K. Khatter, and M. Tushir, "Fuzzy c-means clustering strategies: A review of distance measures," in *Software Engineering*, 2019, pp. 153–162, doi: 10.1007/978-981-10-8848-3_15.
[42] J. Miao, X. Zhou, and T.-Z. Huang, "Local segmentation of images using an improved fuzzy c-means clustering algorithm based on self-adaptive dictionary learning," *Applied Soft Computing*, vol. 91, pp. 1–15, 2020, doi: 10.1016/j.asoc.2020.106200.

BIOGRAPHIES OF AUTHORS

Deny Wiria Nugraha is a doctoral student at the Department of Electrical Engineering, Faculty of Engineering, Hasanuddin University, Indonesia. He is currently a member of the Cloud Computing and Information System (CCIS) Laboratory. He is also a lecturer at the Department of Information Technology, Faculty of Engineering, Tadulako University, Indonesia. He received his Bachelor of Engineering degree in 2002 from Adhi Tama Institute of Technology Surabaya and his Master of Engineering degree in 2010 from Gadjah Mada University, Yogyakarta. His current research interests include artificial intelligence, computer vision, disaster informatics, algorithm optimization, and software engineering. He can be contacted at email: nugrahadw20d@student.unhas.ac.id or deny.wiria.nugraha@untad.ac.id.

Amil Ahmad Ilham is a Professor at the Department of Informatics, Faculty of Engineering, Hasanuddin University, Indonesia, and has been a lecturer since 1998. He received his bachelor's degree in Electrical Engineering from Hasanuddin University, Indonesia, in 1997 and his Master of Information Technology degree from The University of Newcastle, Australia, in 2003. Then, he received his doctoral degree from Kyushu University, Japan, in 2009. He is currently the Vice Dean for Academic and Student Affairs at the Faculty of Engineering, Hasanuddin University. His research interests include various fields in information systems, big data, cloud computing, and computer systems. Additionally, he also has interests in image processing, clustering and similarity analysis, sentiment analysis, machine learning, deep learning, and artificial intelligence. He can be contacted at email: amil@unhas.ac.id.

Andani Achmad is a Professor at the Department of Electrical Engineering, Faculty of Engineering, Hasanuddin University, Indonesia, and has been a lecturer since 1987. He received his bachelor's degree in 1986, master's degree in 2000, and doctoral degree in 2010 from Hasanuddin University in Indonesia. He is currently the Head of the Electrical Engineering Doctoral Study Program at the Faculty of Engineering, Hasanuddin University. His research interests are primarily in the field of information and communication technology. Additionally, he also has interests in artificial intelligence, image processing, machine learning, deep learning, and the internet of things. He can be contacted at email: andani@unhas.ac.id.

Ardiaty Arief is a Professor at the Department of Electrical Engineering, Faculty of Engineering, Hasanuddin University, Indonesia. She received her Bachelor's degree in electrical engineering as first-class honor from Hasanuddin University, Makassar, Indonesia, in 2001, her Master of Technology Management degree from the University of New South Wales, Sydney, Australia, in 2004, and her Doctor of Philosophy degree in power systems stability from the University of Queensland, Brisbane, Australia, in 2013. She is currently a senior researcher and head of the Power and Energy System Research Group at the Department of Electrical Engineering, Hasanuddin University, Indonesia. In addition, she has been involved in many research and projects such as: Renewable Energy Technologies in Eastern Indonesia Short Term Award by Australia Award Indonesia - Murdoch University (2019-2020) and USAID INDONESIA's Sustainable Energy Workshop: "Solar Photovoltaic (PV) and Diesel/Solar PV Hybrid System Training Workshop" (2016). Presently, she is involved as a research partner in international collaborative research with Australian institutions through the Reliable, Affordable, Clean Energy for 2030 (RACE for 2030) program. Her research interests include power system planning, power system stability, power system reliability, load modeling, computational intelligence, and its application in power systems engineering. She can be contacted at email: ardiaty@eng.unhas.ac.id.