



## Optimization Contrast Enhancement and Noise Reduction for Semantic Segmentation of Oil Palm Aerial Imagery

Maura Widyaningsih<sup>1,2</sup>      Tri Kuntoro Priyambodo<sup>3\*</sup>      Moh Edi Wibowo<sup>3</sup>      Muhammad  
 Kamal<sup>4</sup>

<sup>1</sup> *Doctoral Program, Department of Computer Science and Electronics,  
 Faculty of Mathematics and Computer Sciences, Gadjah Mada University, Yogyakarta, Indonesia*

<sup>2</sup> *Department of Informatica Technic, STMIK Palangkaraya Palangka Raya, Indonesia*

<sup>3</sup> *Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences,  
 Gadjah Mada University, Yogyakarta, Indonesia*

<sup>4</sup> *Department of Geographic Information Science, Faculty of Geography,  
 Gadjah Mada University, Yogyakarta, Indonesia*

\* Corresponding author's Email: [mastri@ugm.ac.id](mailto:mastri@ugm.ac.id)

---

**Abstract:** Contrast improvement and noise reduction are needed to improve aerial images prone to interference during data collection. Contrast limited adaptive histogram equalization (CLAHE) is a histogram equalization (HE) development method that is commonly used for contrast improvement. Median blur (MB) is one of the methods used for noise reduction. Combining these two techniques helps optimize the preprocessing process before semantic segmentation analysis is carried out in image maps. The results of testing experiments with U-Net VGG-16 and VGG-19 on image maps show a detailed representation of the predicted pixel class. Comparing accuracy with state-of-the-art methods shows that contrast enhancement and noise reduction are better than the previous method. The highest average result for combining CLAHE+MB with U-Net VGG-16 was 76.5 and VGG-19 was 73.8, and the highest accuracy for image sample testing was 87.94 with U-net VGG-16.

**Keywords:** Noise reduction, Aerial imagery, Semantic segmentation.

---

### 1. Introduction

Quality data supply supports the process of semantic segmentation in producing map images. Enhancement aims to improve image quality during pre-processing, such as segmentation tasks, feature extraction, and image analysis [1,2], and make it more informative [3]. The purpose of optimization during pre-processing is to maintain the image's edges, structure, and detail [4, 5]. A good map image can represent the results of segmenting different objects in the vegetation area. Its success is also influenced by the improvements made during pre-processing.

The cause of aerial image interference can be influenced by the results of shooting with different quality fishing gear, for example, drones with

unmanned aerial vehicle (UAV) facilities that are not the same. This impact allows for overlapping and loss of quality, such as defects (noise), contrast disturbances, sharpness, and blur [6] in the image. Other causes include poor user experience and inappropriate data collection times [7], along with the effects of light, weather, and viewing angles. So, this needs to be considered for pre-processing to improve image quality.

So far, the challenge of the pre-processing problem is when analyzing the separation of objects that must still be consistent with the results of the object segmentation representation. The issue of object separation during class prediction is that the images have considerable intra-class variability due to large-scale object/scene variations, affecting low-class separation results.

The histogram equalization (HE) technique has so far been used to overcome previous challenges in aerial imagery problems, but with this technique, it is still necessary to increase image artifacts [1]. HE is the simplest method to adjust or enhance color contrast globally [8] and update the histogram pixel intensity distribution. CLAHE is also an approach to reduce noise and increase the contrast value in images [9], resulting in artificial boundaries [10]. CLAHE also can consider feature brightness [11] and reduce noise artifacts [12] that can occur in aerial image acquisition. The noise reduction approach using median blur (MB) is the most effective method for removing image noise, such as salt-and-pepper interference [4].

Color image segmentation is a color-space-dependent procedure; choosing a suitable color space also improves the performance of segmentation results and feature compatibility, in addition to applying contrast and noise reduction. Determination of color space also aims to reduce the complexity of the segmentation process. The proposed method uses YUV color space conversion to emphasize intensity in segmentation correction.

The proposed approach to improve image quality uses contrast enhancement and noise reduction. The proposed method for optimizing the improvement of contrast quality uses histogram equalization (HE) and contrast-limited adaptive histogram equalization (CLAHE) combined with median blur (MB) noise reduction. Combining the proposed image improvement methods with their respective capabilities, along with distribution to the YUV color space, aims to optimize the characteristics of the pixels when predictive learning is carried out on the proposed network to produce a good map image.

The segmentation method is based on pixel classification using the U-Net VGG-16 and VGG-19 architecture on Keras, which will process the enhanced image. Semantic segmentation is the process of generating map images using deep learning (DL) methods across a convolution network (FCN). FCN developed a CNN framework that is applied to represent images on a map using a segmentation process [13, 14]. CNN can build a good model by modifying the model's hyperparameters and defining the network structure [15, 16]. The modified CNN is like the general U-Net framework form in the Encoder-Decoder layer. This segmentation network is used to annotate pixels in the pixel-labeling process so that the final labeling result is an image map.

The proposed method to improve image quality combines HE+MB and CLAHE+MB using YUV color space processing to improve segmentation

results. The results of the correction help improve the intensity of pixels in aerial images; thus helping pixels predict their class in the segmentation process. The proposed enhancement method will be compared with the results of segmentation accuracy with other state-of-the-art methods, such as the contrast enhancement primer method with HE and CLAHE [17], noise reduction with MB [18], and contrast enhancement with dynamic clip-limit histogram equation boundary window size (DCWHE) [1], and adaptive gamma correction with weighted histogram distribution (AGCWHHD) [19]. The segmentation process used in this experiment is with VGG-16 and VGG-19 on Keras.

Besides comparing the results of the accuracy of image segmentation testing, comparing the representation results with ground-truth images is also carried out. Ground-truth images are processed from conventional annotations with the help of humans to annotate pixels with labeling using tools. The purpose of comparing the segmentation representation results between the annotation processes carried out by humans and those carried out by machine learning as a recommendation material for providing map images that meet the needs of further analysis in a large area.

The conventional way humans do the pixel labeling process has several weaknesses that can occur; namely, the annotation process must be thorough and takes quite a long time. In addition, the annotation work process cannot be done automatically; the work is carried out through a long process stage with tools to improve image quality, annotations, and segmentation. The tools used for labeling are also licensed and require funding. Another possibility is that annotation errors occur when marking out pixels. So in this study, the authors offer a segmentation process using the U-net shape in producing map images.

## 2. Image quality enhancement method

Aerial images are susceptible to noise, contrast, blur, overlap, and image size differences. Pre-processing supports good data representation results to produce image maps in the segmentation process.

Pre-processing helps improve and reduce dire image predictions to produce good image maps in the semantic segmentation process. The noise reduction approach allows aerial imagery to be corrected for noise caused by shooting, such as image blur. Contrast enhancement is also an approach to help with the distribution of the image's color intensity, to be precise. One effective aerial imagery enhancement

solution introduces enhanced local detail lighting [19,20].

## 2.1 Contrast correction method

### 2.1.1. Histogram equalization (HE)

HE functions to evenly distribute the most frequent pixel values (which have the most significant number in the histogram apart from spreading the intensity, the function density also increases the contrast in the gradient domain and then equalizes a single value to adjust the brightness [19].

Good histogram equalization if the image histogram is limited to a particular area. HE changes the intensity value of the image such that the distribution is uniform by changing the intensity level of a pixel with a new intensity level with the transformation function. HE will perform poorly in places with a considerable variation in intensity and where the histogram covers a large extent; there are light and dark pixels.

HE starts by processing each pixel channel to correct for contrast, using a histogram or probability density function (PDF) distribution first calculated as [11]:

$$h(i) = n(i) \text{ and } p(i) = \frac{h(i)}{N} \quad (1)$$

Where  $i$  = pixel intensity index,  $p(i)$  = probability of each intensity index,  $n(i)$  is the number of pixels in the image with intensity  $i$ , and  $N$  is the image size expressed by height and width. Then calculate the value of the cumulative density function (CDF) distribution.

$$C(i) = \sum_{j=0}^i p(j) \quad (2)$$

The final step of the HE process is followed by the improvement of the pixel intensity of the image, where  $I_E$  is the enhanced intensity, and  $L$  is the color channel range of intensity between 0 and 255 expressed in each pixel.

$$I_E = (L - 1) \times C(i) \quad (3)$$

### 2.1.2. Contrast limited adaptive histogram equalization (CLAHE)

Contrast and reduce noise by creating multiple image histograms and using them to redistribute the image brightness. Two parameters must be set clip limit (Clip Limit/ CL) and block size (BS). CL measures image quality by setting a threshold for contrast changes. CL is also a truncated histogram

that depends on two-factor histogram normalization and the size of the neighbor's region. The BS regulates the number of titles in the rows and columns [12].

The CLAHE method does not eliminate artifacts but at least reduces artifact interference [17]. The block size and clip boundary are parameters that control image quality and are defined. If we specify and set parameters incorrectly, CLAHE results will not be as good as using HE.

CLAHE process is explained that the original image is divided into sub-images of size  $M \times N$ , followed by calculating the histogram of each sub-image, clip limit the histogram limit of each image. The clip limit of a histogram can be defined as follows:

$$\beta = \frac{M}{N} \left( 1 + \frac{\alpha}{100} (S_{max} - 1) \right) \quad (4)$$

Variable  $M$  represents the area size,  $N$  represents the intensity value (0-255) aims the clip factor represents the addition of the histogram boundary by a value between 0 and 100. Histograms above the clip limit value were considered excess pixels distributed over the area. Below the clip boundary, the histogram has evenly distributed an illustration of the distribution of leftover pixels [17].

## 2.2 Noise reduction method

### 2.2.1. Median Blur (MB)

The median filter method uses the value generated by the filter sequence. The effectiveness of noise reduction depends on the filter mask's size and shape, while the algorithm's complexity depends on the median value [18]. The median filter is a nonlinear function used to remove salt or pepper noise from the image and is usually applied to pre-processing before the data is segmented. This algorithm provides a smooth flow of ideas, can effectively maintain edges while reducing impulsive noise [21], and can preserve image detail and improve retouching performance [7].

This method focuses on the median value of the total number of surrounding pixel values that affect the central pixel. This technique works by filling the value of each center pixel (noisy value) with the median value of a sequence of neighbor pixels. The median selection process begins by first sorting the neighbor's pixel values and then selecting the middle value. The output of the median filter is:



Figure. 1 Illustration of calculating the value in an image with a kernel size of 3x3

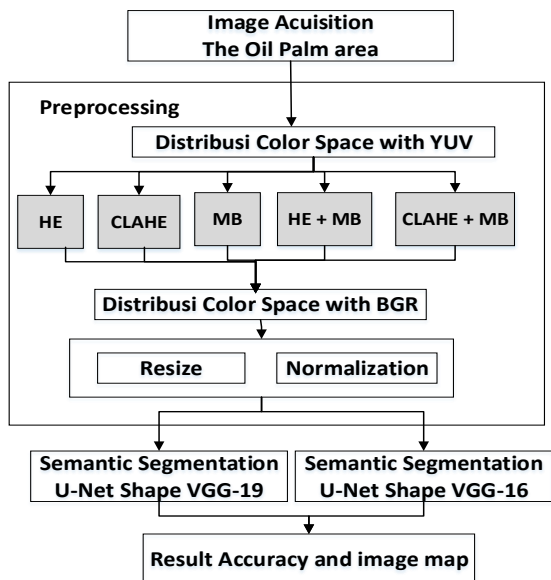


Figure. 2 Process Enhancement and semantic segmentation with U-Net Shape

$$g(x, y) = med\{f(x - i, y - i), i, j \in W\} \quad (5)$$

Where  $f(x, y)$  is the original image, and  $g(x, y)$  is the image representation of the median filter output. The values of  $i$  and  $j$  represent the kernel filter sizes.  $W$  is a two-dimensional mask with a mask size of  $n \times n$  (where  $n$  is usually odd), such as  $3 \times 3$ ,  $5 \times 5$ , and so on, and the shape of the mask is linear, square, circular, and cross [7]. The replacement of noisy values in the median blur technique is as follows:

Fig. 1 with neighbors  $3 \times 3$ , sorted, and then taking the median value of the pixel sequence, 86 will replace the value 61, and the process continues to shift throughout the image to replace noisy pixels.

### 3. Research method

The data sample for this study is an aerial image of the oil palm vegetation areas taken from the Malaysian oil palm plantation (MOPD) dataset ([https://github.com/dongrunmin/oil\\_palm\\_data](https://github.com/dongrunmin/oil_palm_data)).

The process begins with pre-processing, color space conversion, image quality improvement methods, size determination, and image normalization. The process is continued to pre-processing with the semantic segmentation method. The final result is the calculation of the class pixel

classification accuracy in each image and the representation of the segmentation process. The whole process can be seen in Fig. 2.

### 3.1 Preprocessing

Color selection is one of the tasks supporting the improvement of ideal color results in segmentation tasks. A color space is a numerical replica representing color details as distinct color channels (components) in a three or four-dimensional polar or Cartesian system. The color space used in the experiment uses BGR to YUV conversion because it is more suitable for recovering cloud-free images. After all, the luminance can change independently without affecting the chromatic information [22].

YUV models human perception of color more closely than the standard red-green-blue (RGB) model used in computer graphics hardware. The YUV color space is  $Y$ , which is the component of luma, or brightness, ranging from 0 to 100% in most applications;  $UV$  is chromium (an element of blue and red lighting). The color space application function provides information on sharpening the visualization results of the image. The conversion from the RGB to YUV color space is linear [23], and the standard formula for converting the RGB to YUV color space is :

$$\begin{aligned} Y &= 0,299R + 0,587G + 0,114B \\ U &= -0,1687R - 0,3313G + 0,5B + 128 \\ V &= 0,5R - 0,4187G 0,813B + 128 \end{aligned} \quad (6)$$

This experiment converts the input image from the blue green red (BGR) color space to YUV. The task is continued in the process of improving image quality using the HE contrast improvement method as in Eq. (1-3) and CLAHE with Eq. (4), MB with Eq. (5), and combining HE+MB, and CLAHE+MB.

Implementing the HE and CLAHE methods uses the functions available in open CV. The aerial image used as input data still requires quality improvement before being forwarded to the segmentation using the previously discussed method. The noise reduction procedure with a median blur set filter of kernel size ( $3 \times 3$ ) will help in more detail on the affirmation of pixel values. In contrast, the CLAHE function sets two parameters with a clip limit value of 0.52, which is quite suitable for color images with large image resolutions, and  $4 \times 4$  tile sizes.

The pre-processing results are continued, return the color to BGR mode, then set all sizes equal to  $512 \times 512$  and normalize the image. Image normalization

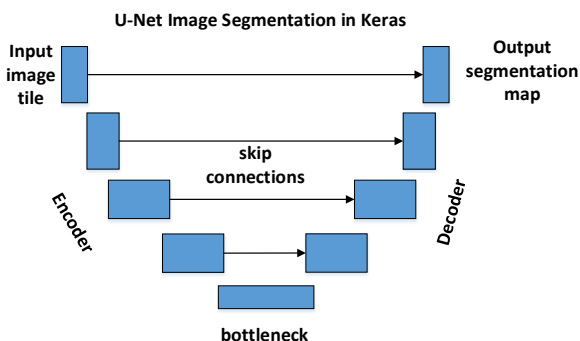


Figure. 3 U-Net shape of image segmentation to VGG-16 and VGG-19 architecture

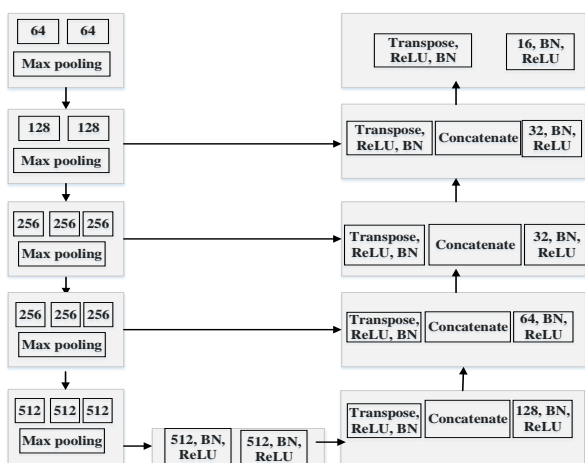


Figure. 4 The process structure VGG-16

process by dividing each intensity value by 255 to provide a smaller pixel intensity distribution value, not reducing feature information and simplifying the segmentation process.

### 3.2 Segmentation processing with U-Net shape

Image segmentation is also an important stage of operations describing vegetation objects, from the process also segmentation also affects the results of classification accuracy. The implementation of image segmentation is to separate vegetation types, in this case, the challenge is because of irregular shapes and various patterns and colors [24].

The result of segmentation is assessed quantitatively based on the image map which is visually interpreted as a reference. Semantic segmentation using U-Net Shape architecture is suitable for handling multiclass problems, so the determination is based on the pixel class (not the entire image/segment). In other words, semantic segmentation is tasked with classifying object classes for each pixel in an image, which means assigning a label to each pixel.

The purpose of image segmentation is to partition an image into homogeneous and connected regions

without using additional knowledge about the objects in the picture. Regional homogeneity in color image segmentation involves natural colors and sometimes color textures [25].

The semantic segmentation method was developed by increasing the sampling time, involving full convolution (FC) in the network to recover every input detail, and balancing network work processes and invariants [26]. The architecture of the U-Net form consists of two parts: the encoder and decoder layers. U-Net VGG-16 is similar to the encoder and decoder layer, which modifies the inference layer on the CNN with a de-convolutional layer using the backbone [27]. VGG-19 is similar to VGG-16 and differs only in each convolutional encoder layer.

The U-Net form architecture consists of paths defined to capture the evolving context symmetrically to allow for precise localization. The network path with convolution operations performs back-and-forth operations on the sampling feature map alternately and in sampling. Convolution (left layer) is a process to get a smaller output size. In comparison, de-convolution (suitable layer) is derived from up-sampling or transposing of the convolution function to get a larger output size, a fractional convolution step to produce an image map. Between the two layers, there is a bottleneck layer, a convolution layer to bridge the two network processes of the encoder and decoder.

In the encoder, convolutional layers and a continuous down-sampling layer extract in-depth features with large receptive fields. The decoder samples the extracted inner features from the input resolution for pixel-level semantic prediction. The high-resolution features of the different scales of the encoder are combined with bypass connections to reduce the loss of spatial information caused by down-sampling.

The task of the decoder is to project the discriminatory features semantically (lower resolution) studied by the encoder into the pixel space (higher resolution) to obtain a dense classification [26]. The general framework of the U-Net is illustrated in Fig. 3.

U-Net shapes VGG-16 and VGG-19 use the Keras standard for their implementation. In this experiment, we set the adjusted parameter functions, namely the activation function, decoder line layer arrangement, and optimization function. The layer sizes specified in the decoder are 128, 64, 32, 32, and 16, which are manageable measures of the number of parameters in accelerating the segmentation process. The activation function uses ReLU to use convolution and softmax to the final operation in

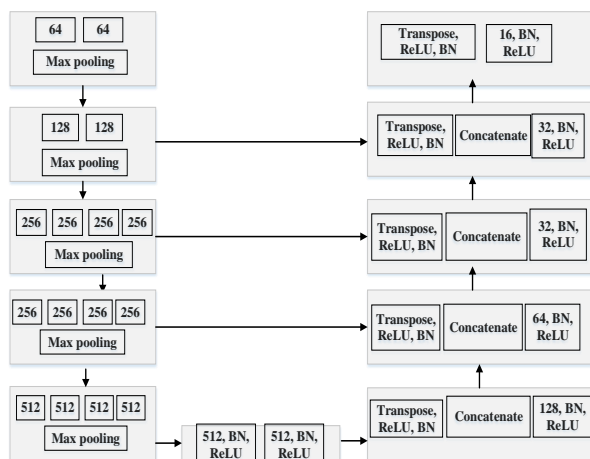


Figure. 5 The process structure VGG-19

network, which supports the categorical cross-entropy loss function, which changes the scale of the model output so that it has the right properties [26, 28]. This function is also very effective for segmentation tasks [29]. Up-sampling function with Transpose function, normalization to the convolution function use batch normalization (BN).

The decoder layer is also set to the size of the convolution layer with standard parameters that are manageable on the network. Softmax activation function to represent more than two class pixels so that the color is distributed according to the pixel class visualized by the image map.

Fig. 4 and Fig. 5 are the process structure on the VGG-16 and VGG-19 architectures that are used to perform segmentation testing on the pixel class. The VGG-16 architecture of Keras with a total of 21,877,363 parameters and VGG-19 consists of 27,187,059 parameters. The structure of VGG-16 and VGG-19 consists of a convolution layer on the encoder layer, a bottleneck with the center layer, and deconvolution with the decoder layer. Concatenate function used the merger process with layers convolution of encoder with decoder.

### 3.3 Evaluation performance

Performance measurement determines the model's ability to train pixels to recognize their classes. The overall accuracy considers all categories that enter metrically from the class pixel prediction component.

$$OA = \frac{\sum_{k=1}^N TP_k}{\sum_{k=1}^N TP_k + FP_k + FN_k + TN_k} \quad (7)$$

Where  $TP_k$ ,  $FP_k$ ,  $TN_k$ , and  $FN_k$  indicate true positives, false positives, true negatives, and false negatives, respectively, for object indication as class  $k$  [30], the accuracy result is multiplied by 100 [31].

## 4. Discussion and result

Pre-processing was carried out on six samples of aerial images, in pre-processing image quality using contrast correction using the HE function, CLAHE, noise correction with MB, and merging HE with MB, and CLAHE with MB. Overall, as shown in Fig. 4, the result of processing the contrast equalization and noise reduction process in Image1, Image2, Image3, Image4, Image5, and Image6.

Parameters are used and applied in noise reduction and contrast improvement, taking into account the size of the input image and the picture's condition. The HE parameter is aligned globally in the embodiment using the default value in the open CV function, which has three color channels, namely RGB.

When implementing CLAHE, two parameters are used: clip limit with a value of 0,52, which is a parameter to set the threshold or contrast. The second parameter is the 4x4 tile grid size to set the number of kernels in rows and columns. The kernel size applied to MB using is 3x3; this smallest kernel is defined to support stricter filtering on pixel-based classification.

A comparison of histogram results for each image shows that the original image processed with the MB process slightly changes in intensity. The visualization results show that the picture looks sharper than the original. Contrast processing with HE and CLAHE, the histogram results of both processes show a more even distribution of color towards the light intensity. The HE histogram results show a more even contrast than CLAHE, and with CLAHE, there are still some non-contrast intensities. The image visualization between the HE and CLAHE results shows the emphasis of dark blue on plant areas with old vegetation, light blue on areas of young, and yellow and white color on road or soil areas (objects which are not vegetation).

The histogram results show the difference in better results after combining the HE process with MB and CLAHE with MB. The histogram results show that both approaches experience better color distribution and reduced image noise. Image visualization shows a change in color intensity in a particular pixel area so that this color variation strengthens the detail of the pixel intensity value, which is processed into a semantic segmentation network. The incorporation of retouching techniques on the image makes it sharper; the contrast is increased, and noise is noticeably reduced.

Fig. 6 is one of the Image2 samples with a dense vegetation area that was preprocessed; the difference

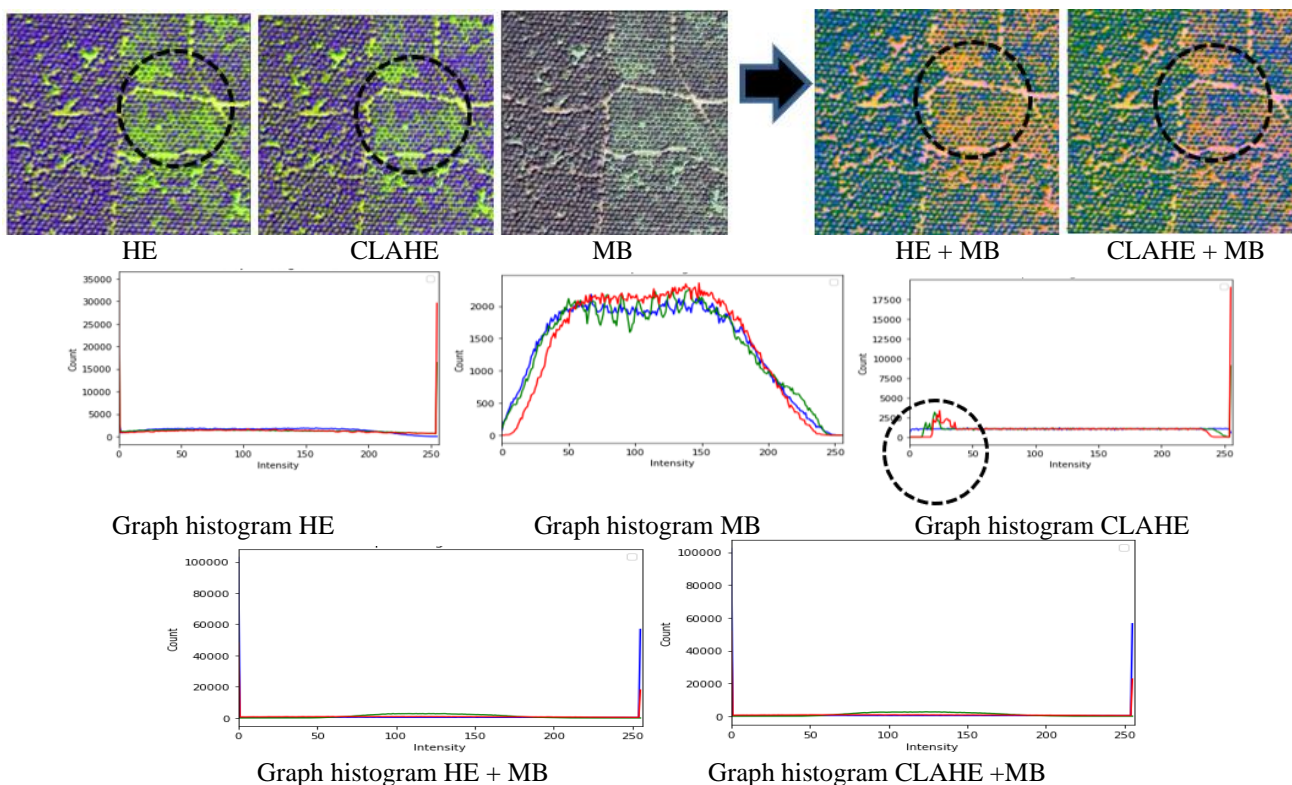


Figure. 6 Representation sample Image2 for contrast result with HE , CLAHE and noise reduction with MB; Graph histogram results of distribution pixel-intensity from the image representation

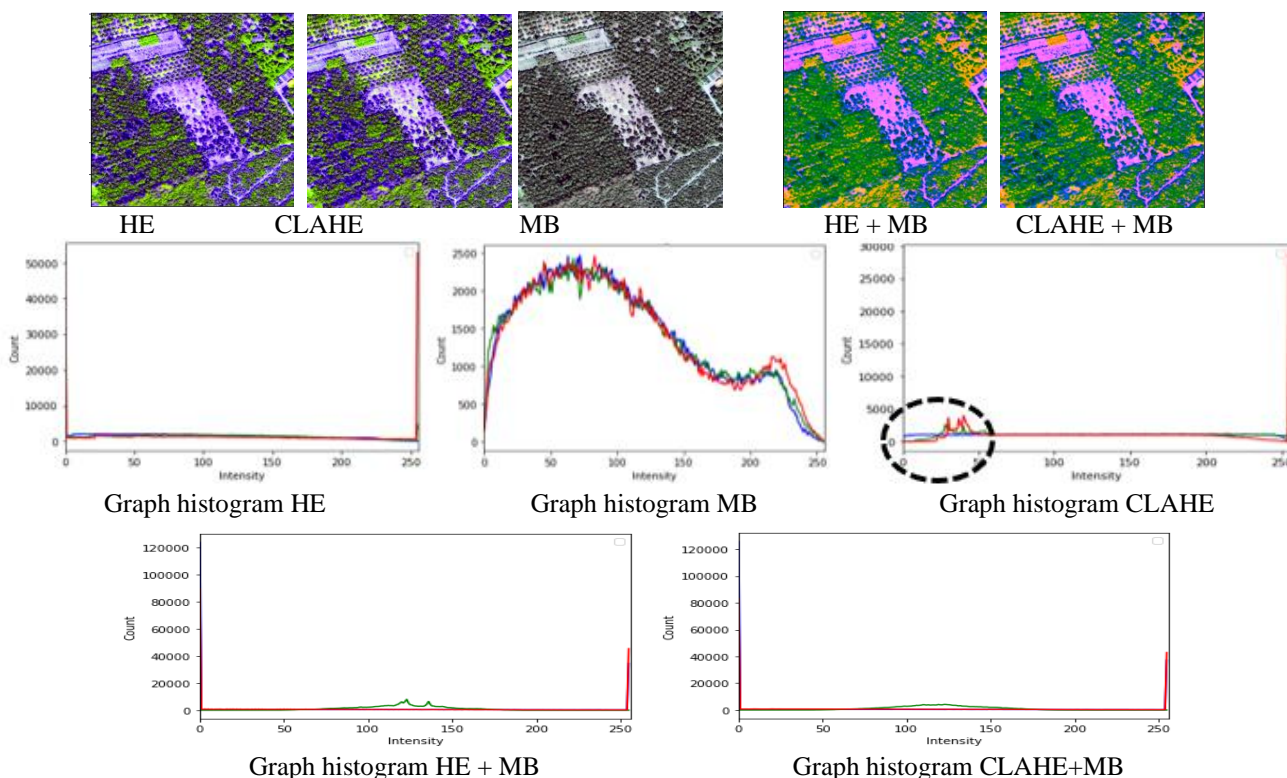


Figure. 7 Representation sample Image4 for contrast result with HE, CLAHE and noise reduction with MB; Graph histogram results of distribution pixel-intensity from the image representation

in contrast results is visible after the HE and CLAHE processes, as an example of a partial image representation marked with a circle. The HE

histogram graph shows that the green color is more distributed to the contrast, while with CLAHE, the

Table 1. Training semantic segmentation VGG-16 with six aerial image (Image1, Image2, Image3, Image4, Image5, Image6 ) use accuracy value in percent

Metode	Data Sample six image					
	1	2	3	4	5	6
Original	50,13	39,97	42,01	26,87	15,33	43,01
HE	47,43	58,75	46,28	38,97	48,73	37,5
CLAHE [17]	40,78	47,7	39,7	30,57	23,41	40,25
MB [18]	39,84	47,34	41,85	28,79	19,57	42,09
DCWHE [1]	43,8	43,55	41,36	31,23	22,83	37,96
AGCW HD [19]	31,96	39,57	36,93	37,73	42,51	40,07
Proposed HE+MB	70,11	64,69	65,91	67,63	<b>74,18</b>	58,8
Proposed CLAHE+MB	70,85	74,05	78,6	77,12	<b>87,94</b>	70,24

Table 2. Training semantic segmentation VGG-19 with six aerial image (Image1, Image2, Image3, Image4, Image5, Image6 ) use accuracy value in percent

Method	Data Sample six image					
	1	2	3	4	5	6
Original	35,16	42,04	38,51	32,5	33,21	22,71
HE	44,9	54,06	46,35	48,53	51,36	20,61
CLAHE [17]	42,59	49,97	43,27	33,18	35,04	19,49
MB [18]	40,88	47,32	38,25	33,1	33,43	21,59
DCWHE [1]	41,86	47,24	34,59	30,62	35,03	20,79
AGCW HD [19]	34,45	30,53	32,82	33,59	34,66	40,52
Proposed HE+MB	66,74	67,07	62,86	69,09	<b>72,26</b>	65,4
Proposed CLAHE+MB	72,34	68,45	70,18	72,31	75,02	<b>84,47</b>

red color is more many contrasted. CLAHE indicates that some prominent intensities and distributions differ from the HE results. However, combining the noise reduction process with MB gives good results from both the HE and CLAHE functions; the image representation shows the pixel color distribution providing additional intensity variations. It can be seen that the blue color CLAHE+MB histogram is more highly distributed than the red HE+MB color; in this case, the blue color confirms the vegetation area.

Fig. 7 is an example of an area not dense with vegetation objects; several other objects exist. The results of the HE improvement compared to CLAHE show a different distribution of pixel intensity, as seen in the histogram graph. The pixel color representation also looks different in the area marked

in the image. The results of combining HE+MB and CLAHE+MB show a different histogram in the even distribution of red and green colors, where the combination of the two colors emphasizes the non-vegetated object area, namely roads or soil. From the image representation, CLAHE+MB retains the blue color element to describe the vegetation object area, as seen from the circle marked.

After improving the image quality, the segmentation process continued using the U-Net VGG-19 and VGG-16 forms. The parameters set in U-Net use the standard set in Keras; parameters given to the decoder function are layer size and the output activation and transpose process as up-sampling.

Evaluation of network performance is based on the segmentation process's accuracy function for six aerial image samples. U-Net shape network training is carried out with a loss cross entropy function with 30 epochs and activation operations with Softmax for multiclass identification on pixels. The results show an increase when the enhancement approach is applied to the image compared, as shown in Table 1 and Table 2.

The information shown in Table 1 and Table 2 calculates the prediction accuracy of the pixel class with the proposed HE, MB, CLAHE, DCWHE, AGCWHD, HE+MB, and CLAHE+MB processes. The results of the accuracy of the pixel testing in the segmentation process with 6 sample images show that the combination of HE+MB and CLAHE+MB functions is better than HE and CLAHE without combining with MB. Likewise, comparisons with state-of-the-art methods such as DCWHE and AGCWHD show better results than we propose.

Accuracy with the VGG-16 process showed that the highest accuracy was in combining CLAHE + MB in the Image5 sample with an accuracy test result of 87.94, while in combining HE + MB with an accuracy of 74.18, as shown in Table 1. Testing accuracy with VGG-19, the highest result was the combination of CLAHE+MB in the Image6 sample, which was 84.47, while with HE+MB in the Image5 example, the accuracy was 72.26, as in Table 2.

Overall it shows that the highest accuracy when testing the segmentation results of 6 images is the combination of the CLAHE+MB approach compared to other methods. The application of YUV color space conversion also supports the sharpening characteristics of the pixels themselves. This approach helps improve image quality so that pixels can optimize that noise in the resulting image.

Fig. 8 shows the results of the highest accuracy in the Image5 sample of the six samples tested for pixel prediction in the image segmentation process with



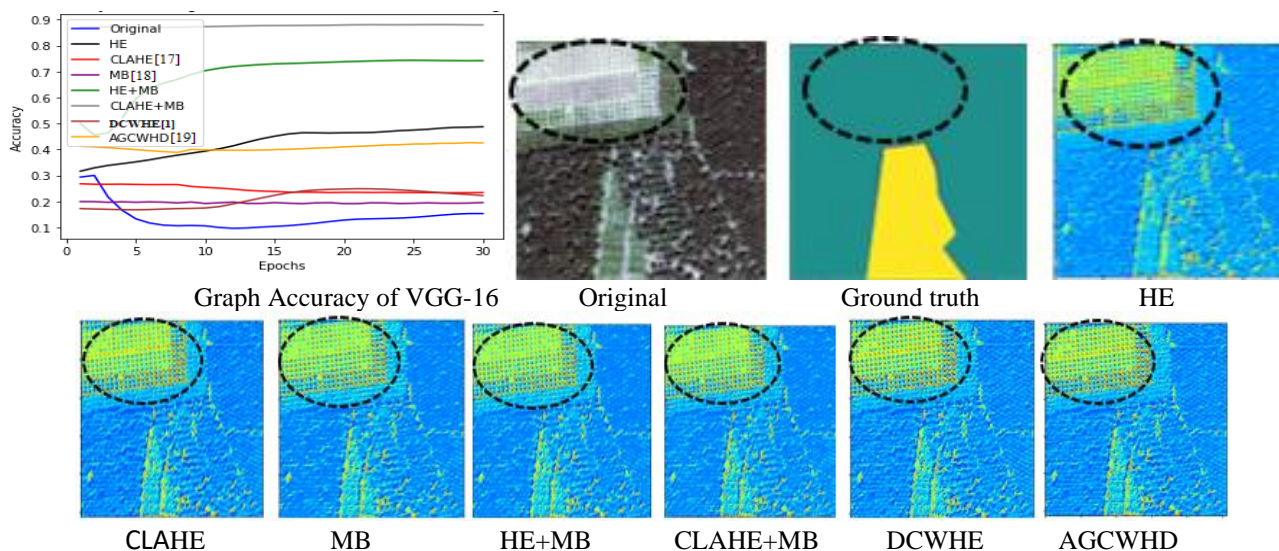


Figure. 8 The highest accuracy of results the pixel training using VGG-16 and representation result prediction segmentation of Image5

VGG-16 compared to other methods using epoch 30. The accuracy graph shows that the highest accuracy is using quality improvement by incorporating CLAHE+MB. The second sequence is the combination of HE+MB, as shown in Table 1 for accuracy value. The test graph shows that the processing of the CLAHE+MB image repair results when testing shows an increase in accuracy, which from the start maintains the stability of the accuracy results, while the HE+MB test experience the process of increasing the accuracy. The representation of the segmentation process shows results that are almost difficult to distinguish from several comparisons, but using U-Net segmentation shows the level of detail of pixels in its class, compared to the ground truth process that is carried out conventionally, which causes pixels to work not optimally which causes pixels to be recognized as other objects. The difference in representation results between the ground truth and the consequences of segmentation processing with U-Net shows that the different segmentation object is more visible and transparent than the ground truth. The map image shows apparent differences in the area of soil and vegetation.

Segmentation from the CLAHE+MB and HE+MB merging process shows the density of older vegetation objects marked in dark blue, young vegetation in light blue, and soil areas marked in yellow and orange. The difference is not too striking with other methods. Still, there are slight differences with other methods, namely the density of vegetation intensity, old vegetation is somewhat blurred closer to young color, and some areas of land are recognized

as vegetation. In contrast, from the ground-truth, the land object is detected as a vegetation class object, as shown in the image marked with a circle.

The success of the pixels trained in the U-net network to carry out the segmentation process can produce image maps that represent objects with precision and detail. Fig. 9 shows pixel training in the segmentation process on the Image6 sample, which shows the highest accuracy with VGG-19. The displayed accuracy graph shows the same thing in the previous approach: the combined CLAHE+MB and HE+MB are still superior compared to other methods. It can see from the ground-truth representation all objects are identified with soil, while segmentation processing with U-Net shows a clear representation between land and vegetation areas. The two proposed methods show better representation results because some vegetation objects retain their class sufficiently, so they are not recognized as soil objects. In other scenarios, some intensities experience blurred conditions so that other class objects still realize some things; for example, vegetation is identified as soil or vice versa.

Table 3 shows the average accuracy results from testing the data of 6 sample images against a comparison of image quality improvement methods, namely HE, CLAHE, MB, DCWHE, AGCWHD, and the proposed combination of HE+MB and CLAHE+MB. The processing results show that VGG-16 performs better on six image samples in thesegmentation testing process compared to VGG-19. Improving the image quality of the combined

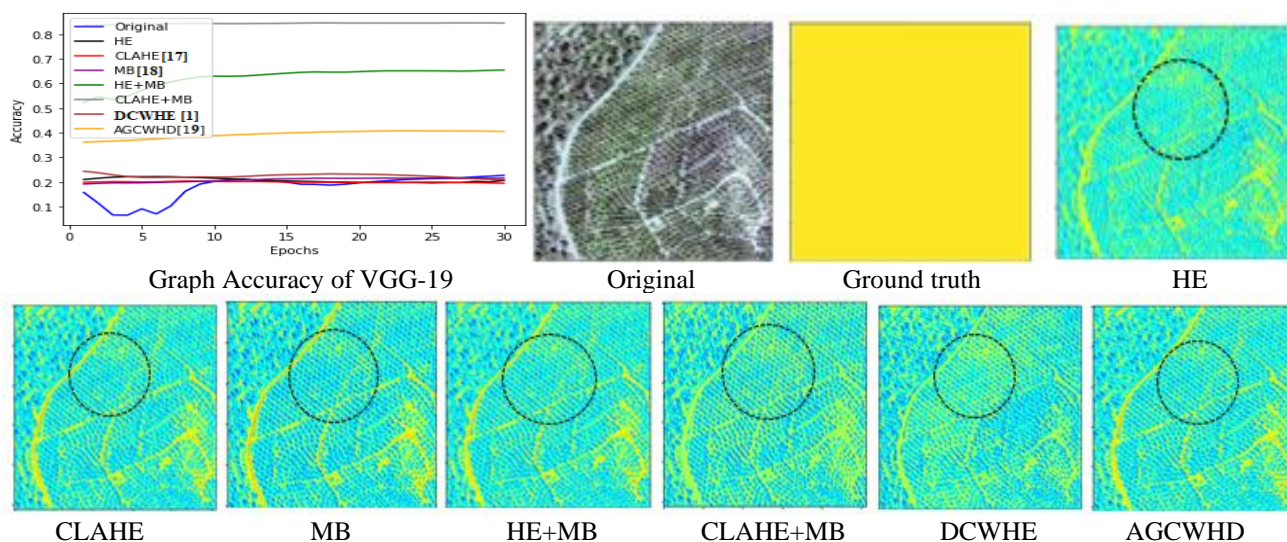


Figure. 9 The highest accuracy of results the pixel training using VGG-19 and representation result prediction segmentation of Image6

Table 3. The accuracy average testing six of the sample images with U-Net shape

Method	Segmentation Method	
	VGG-16	VGG-19
Original	36,2	34,0
HE	46,3	44,3
CLAHE [17]	37,1	37,3
MB [18]	36,6	35,8
DCWHE [1]	36,8	35,0
AGCWHD [19]	38,1	34,4
Proposed (HE+MB)	66,9	67,2
Proposed (CLAHE+MB)	<b>76,5</b>	<b>73,8</b>

processing between CLAHE+MB gave the highest accuracy results in the process with VGG-16, namely 76.5, while with VGG-16, it was 73.8.

The experimental results of applying the YUV color space conversion help increase the color intensity of pixels before image problems are corrected. The role of YUV color space conversion before contrast enhancement and noise reduction can strengthen pixel characteristics, thereby reducing the complexity of pixel classes to predict.

An approach that combines contrast optimization and noise reduction can be detailed pixel intensity so that when area segmentation is performed, the pixel can predict its class quite well. Improving image quality can help prepare data to support the quality of pixel intensity before the segmentation process is carried out. The proposed approach combines contrast enhancement and noise reduction like CLAHE+MB and HE+MB, giving feature pixel enhancement. The improved pixel characteristics help the pixel learning process in segmentation and

can produce good-quality map images to separate objects in the vegetation area.

Testing with the appropriate parameters in determining the Clip-limit and Window-size in the CLAHE process determines the results of the contrast quality improvement process. Although the HE and CLAHE methods are commonly used and developed for image improvement processes, implementation with aerial image data has yet to yield good results. Likewise, when image repair noise reduction with MB, the accuracy results also do not give good results. Compared with other improvement methods, such as DCWHE and AGCWHD also do not provide good accuracy compared to the two proposed approaches.

The segmentation process with U-Net is used to perform labeling through the machine learning of pixels, and the process works automatically, including the image quality improvement process. Map image representation of the results of the segmentation test on U-Net as machine learning has better results than ground-truth images. Pixels successfully trained in the U-net network to carry out the segmentation process can produce image maps representing objects with sufficient precision and detail.

Ground-truth image representation itself does not give to provide maximum results from the segmentation process carried out by humans, and it can be seen that several classes are recognized as other classes. Labeling done conventionally by humans still has the risk of identifying pixels as another class, so the map image results may not necessarily display the appropriate segmentation.

The segmentation process performed by machine learning helps make pixel labeling work better. Comparing the results of the machine learning

process allows pixels to make predictions about their class to generate map images that can distinguish objects from one another, such as differences in soil and old and young vegetation over a large area.

## 5. Conclusion

Overall, the proposed process of improving image quality by combining HE+MB and CLAHE+MB can improve segmentation results for aerial images, with the highest average accuracy in VGG-16. The highest Accuracy of the VGG-16 process with combining CLAHE + MB in the Image5 sample is 87.94. The average accuracy of the six sample data in the semantic segmentation process for pixels is 76.5 on VGG-16 and 73.8 on VGG-19. The settings still often experience weight changes which usually affect the accuracy of the results between the two VGG-16 and VGG-19 networks. However, the two still provide better and superior results in combining image enhancement compared to other image repair methods.

The characteristics of the aerial image features are strengthened by the distribution of the YUV color space and the suggested approach for optimizing contrast improvement coupled with noise reduction. The distribution of color space is one of the supports in providing quality to the intensity of image pixels. Consideration of good results is also affected by using CLAHE-compliant parameters such as the Clip-limit and Window-size parameters and the kernel size for MB.

The role of machine learning work on pixel annotations can provide immediate results both in the pre-processing process and the segmentation that works on pixels to produce map images. Pixel annotation with U-net as machine learning for segmentation helps reduce labeling processing time in providing better image map data because it works automatically. The weaknesses of the annotation process carried out by humans are to minimize the possibility of errors that occur during annotations, which lead to unfavorable segmentation results.

U-Net is the recommended network for segmenting and training pixels to class prediction. This network needs further improvement and optimization research to provide even better segmentation results. One of these others problems of aerial imagery is also influenced by the differences in the characteristics of each image at the time of data collection. Therefore, image repair is always carried out for the availability of good data on aerial imagery before the segmentation process is carried out. Further development can improve the U-net shape architectural model to provide better pixel machine

learning results of enhanced image quality, thus producing good-quality map images.

## Conflicts of interest

The authors declare there is no conflict of interest in this work.

## Author contributions

The contributions of each author are described as follows: "Conceptualization and methodology, Priyambodo, Wibowo, Widyaningsih; implementation the of methodology, Widyaningsih; data validation, Kamal and Widyaningsih ; formal analysis, Priyambodo, Wibowo, and Kamal; investigation, Priyambodo, Wibowo, Kamal and Widyaningsih; writing—original draft preparation, Widyaningsih; writing—review and editing, Priyambodo, Wibowo and Kamal; supervision, Priyambodo, Wibowo and Kamal; funding acquisition, Widyaningsih".

## Acknowledgments

The authors would like to thank STMIK Palangkaraya (for collect data) and the Department of Computer Science and Electronics for supporting publication funds and using laboratories for experiments and testing.

## References

- [1] M. B. Mortatha, S. S. Thabit, H. R. A. Ameer, and R. R. Nuiiaa, "Dynamic Clip Limit Window Size Histogram Equalization for Poor Information Images", *International Journal of Intelligent Engineering & System*, Vol. 15, No. 5, pp.57-70, 2022, doi: 10.22266/ijies2022.1031.06.
- [2] G. Tsagkatakis, A. Aidini, K. Fotiadou, M. Giannopoulos, A. Pentari, and P. Tsakalides, "Survey of deep learning approaches for remote sensing observation enhancement", *Sensors*, Switzerland, Vol. 19, No. 18, pp. 1-38, 2019.
- [3] A. Shah, J. I. Bangash, A.W. Khan, I. Ahmed, A. Khan, A. Khan, and A. Khan, "Comparative analysis of median filter and its variants for removal of impulse noise from grayscale images", *Journal of King Saud University - Computer and Information Sciences*, Vol. 34, No. 3, pp. 505-519, 2022.
- [4] U. Erkan, S. Enginoğlu, D. N. H. Thanh, and L. M. Hieu, "Adaptive frequency median filter for the salt and pepper denoising problem", *The Institution of Engineering and Technology*

- Image Processing*, Vol. 14, No. 7, pp. 1240-1247, 2020.
- [5] P. Srestasathiern and P. Rakwatin, "Oil Palm Tree Detection with High Resolution Multi-Spectral Satellite Imagery", *Remote Sensing*, Vol. 6, No. 10, pp. 9749-9774, 2014.
- [6] A. Mitra, S. Roy, S. Roy, and S. K. Setua, "Enhancement and restoration of non-uniform illuminated fundus image of retina obtained through thin layer of cataract", *Computer Methods and Programs in Biomedicine*, Vol. 156, pp. 169-178, 2018.
- [7] B. Desai, S. Jha, and U. Kushwaha, "Image Filtering - Techniques, Algorithm and Applications", *GIS Science Journal*, Vol. 7, No. 11, pp. 970-975, 2020.
- [8] M. J. Alwazzan, M. A. Ismael, and A. N. Ahmed, "A Hybrid Algorithm to Enhance Colour Retinal Fundus Images Using a Wiener Filter and CLAHE", *Journal of Digital Imaging*, Vol. 34, No. 3, pp. 750-759, 2021.
- [9] P. Dai, H. Sheng, J. Zhang, L. Li, J. Wu, and M. Fan, "Retinal fundus image enhancement using the normalized convolution and noise removing", *International Journal of Biomedical Imaging*, pp. 1-12, 2016.
- [10] Z. Zou and Z. Shi, "Random Access Memories: A New Paradigm for Target Detection in High Resolution Aerial Remote Sensing Images", *IEEE Transactions on Image Processing*, Vol. 27, No. 3, pp. 1100-1111, 2018.
- [11] A. M. Pour, H. Seyedarabi, S. H. A. Jahromi, and A. Javadzadeh, "Automatic Detection and Monitoring of Diabetic Retinopathy Using Efficient Convolutional Neural Networks and Contrast Limited Adaptive Histogram Equalization", *IEEE Access*, Vol. 8, pp. 136668-136673, 2020.
- [12] G. Ulutas and B. Ustubioglu, "Underwater image enhancement using contrast limited adaptive histogram equalization and layered difference representation", *Multimedia Tools and Applications*, Vol. 80, No. 10, pp. 15067-15091, 2021.
- [13] M. Lan, Y. Zhang, L. Zhang, and B. Du, "Global Context based Automatic Road Segmentation via Dilated Convolutional Neural Network", *Information Sciences*, Vol. 535, pp. 156-171, 2020.
- [14] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 8, pp. 1798-1828, 2013.
- [15] Chrisantonius, T. K. Priyambodo, F. H. Raswa, and J. C. Wang, "Partial Fingerprint on Combined Evaluation using Deep Learning and Feature Descriptor", In: *Proc. of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1611-1614, 2021.
- [16] K. Hidjah, M. Wibowo, A. Harjoko, and R. R. Shantiningsih, "Periapical Radiograph Texture Features for Osteoporosis Detection using Deep Convolutional Neural Network", *International Journal of Advanced Computer Science and Applications*, Vol. 13, No. 1, pp. 223-232, 2022.
- [17] F. Hana and I. D. Maulida, "Analysis of contrast limited adaptive histogram equalization (CLAHE) parameters on finger knuckle print identification", *Journal of Physics: Conference Series*, Vol. 1764, No. 1, 2021.
- [18] A. Pati, S. K. Sagarar, and P. R. Rajvanshi, "A New Improved Statistical Algorithm for Image Noise reduction By", *International Journal of Research*, Vol. 5, No. 12, pp. 2672-2679, 2018.
- [19] M. Veluchamy and B. Subramani, "Image contrast and color enhancement using adaptive gamma correction and histogram equalization", *International Journal for Light and Electron Optics - Optik*, Vol. 183, pp. 329-337, 2019.
- [20] X. Fu, J. Wang, D. Zeng, Y. Huang, and X. Ding, "Remote Sensing Image Enhancement Using Regularized-Histogram Equalization and DCT", *IEEE Geoscience and Remote Sensing Letters*, Vol. 12, No. 11, pp. 2301-2305, 2015.
- [21] R. Sabah, R. Ngadiran, and D. A. Hammood, "Image denoising using wavelet and median filter based on raspberry Pi", *Jurnal Informatika*, Vol. 15, No. 2, pp. 91-102, 2021.
- [22] X. Wen, Z. Pan, Y. Hu and J. Liu, "Generative Adversarial Learning in YUV Color Space for Thin Cloud Removal on Satellite Imagery", *Remote Sensing*, Vol. 13, No. 1079, pp. 3-22, 2021.
- [23] C. E. Prema, S. S. Vinsley, and S. Suresh, "Multi-Feature Analysis of Smoke in YUV Color Space for Early Forest Fire Detection", *Fire Technology*, Vol. 52, No. 5, pp. 1319-1342, 2016.
- [24] S. M. U. Sinaga, and M. Kamal, "Image segmentation for vegetation types extraction using WorldView-2: a case study in parts of Dieng Plateau, Central Java", In: *Proc Sixth Geoinformation Science Symposium*, p. 2, 2019.
- [25] S. Tammina, "Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images", *International Journal of*

- Scientific and Research Publications (IJSRP)*, Vol. 9, No. 10, p. 9420, 2019.
- [26] S. Avalos, and J. Ortiz, *Convolutional Neural Networks Architecture: A Tutorial I*, Queens University, pp. 159-168, 2019.
- [27] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 12, pp. 2481–2495, 2017.
- [28] Y. Ho and S. Wookey, “The Real-World-Weight Cross-Entropy Loss Function : Modeling the Costs of Mislabeling”, *IEEE Access*, Vol. 8, pp. 4806-4813, 2020.
- [29] D. J. Doggalli and S. B. S. Kumar, “The Efficacy of U-Net in Segmenting Liver Tumors from Abdominal CT Images”, *International of Intelligent Engineering and Systems*, Vol. 15, No. 5, pp. 151-161, 2022, doi: 10.22266/ijies2022.1031.14.
- [30] K. M. Ting, “Confusion Matrix”, *Encyclopedia of Machine Learning and Data Mining*, pp. 260-260, 2017.
- [31] M. S. Madni and C. Vijaya, “Hand Gesture Recognition using Auto Encoder with Bi-direction Long Short Term Memory”, *International of Intelligent Engineering & System*, Vol. 14, No. 6, pp. 168 – 176, 2021, doi: 10.22266/ijies2021.1231.16.