RESEARCH ARTICLE **OPEN ACCESS**

Manuscript received July 27, 2023; revised August 20, 2023; accepted September 21, 2023; date of publication October 30, 2023
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeemi.v5i4.314

How to cite: Anita Desiani, Sigit Priyanta, Indri Ramayanti, Bambang Suprihatin, Muhammat rio Halim, Ira Rayani, DIte Geovani, "Multi-Stage CNN: U-Net and Xcep-Dense of Glaucoma Detection in Retinal Images, vol. 5, no. 4, pp. 211–222, October 2023.

Multi-Stage CNN: U-Net and Xcep-Dense of Glaucoma Detection in Retinal Images

Anita Desiani[1], Sigit Priyanta[2], Indri Ramayanti[3], Bambang Suprihatin[1], Muhammat rio Halim[1], Ira Rayani[1], and Dite Geovani[1]

[1] Mathematics Department, Mathematics and Science Faculty, Universitas Sriwijaya, Inderalaya, 30862, Indonesia
[2] Departement of Computer Science and Electronics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta, 55281, Indonesia
[3] Departement of Parasitology, Faculty of Medicine, Universitas Muhammadiyah, Palembang, 30166, Indonesia

Corresponding author: Anita Desiani (e-mail: anita_desiani@unsri.ac.id).

The study was funded by Direktorat Riset, Teknologi dan Pengabdian Masyarakat, Direktorat Jenderal Pendidikan Tinggi, Riset dan Teknologi in accordance with the implementation contract of the state university research program operational assistance program No. 164/E5/PG.02.00.PL/2023.

ABSTRACT Glaucoma is a chronic neurological disease in the retina that causes vision loss. Glaucoma can be detected from abnormalities that occur in the optic disc and optic cup on the retina. To get the features, a segmentation process is needed. Segmentation can improve the performance of the classification. This study combines segmentation and classification with CNN architecture to detect glaucoma. At the segmentation stage, U-Net CNN architecture is applied. U-Net has encoder and decoder sections to get output in the form of images containing only the features needed. At the classification stage, the study proposes the Xcep-Dense Net. Xcep-Dense is a CNN architecture that combines Exception and Dense. The Xcep-Dense seeks the advantages possessed by Xception and Dense architecture and overcomes the weaknesses of each architecture. In the segmentation, The results of U-Net architecture are above 90% for the accuracy, recall, precision, and F1-score but Cohen's kappa is above 85%. The results show that U-Net is excellent for optic disc and optic cup segmentation. At the classification, the accuracy and precision are above 85%, the recall and F1-score are above 80%, and Cohen's kappa is 77%. These results show that the Xcep-Dense architecture is robust enough to classify glaucoma which consists of three classes advanced glaucoma, early glaucoma, and normal. Based on the results, it shows that the proposed method is feasible for detecting glaucoma. The results of this study are expected in the future to be developed into an automatic machine for early detection of glaucoma.

INDEX TERMS Glaucoma, Segmentation, Classification, U-Net, Xcep-Dense

I. INTRODUCTION

The optic disc is a circular spot on the retina formed by the axons of the retinal ganglion cells. The axons of these cells are responsible for sending signals from the eye's photoreceptors to the optic nerve so that the eye can see. The optic cup is the bright area in the center of the optic disc. The optic disc and optic cup are measurements used in the diagnosis of glaucoma. usually, the optic disc and optic cup can be measured horizontally or vertically on the patient [1]. Glaucoma is a chronic neurological disease in the human eye where the nerves that connect the eye to the brain are damaged gradually causing vision loss to blindness [2]. Glaucoma detection is done by direct observation of the optic disc and optic cup by an ophthalmologist through retinal images taken from the fundus camera. Abnormalities in the optic disc and optic cup can be helped by segmentation. Segmentation is needed to extract important features from images and remove parts that are not needed at the classification stage [7]. Segmentation is an important stage before carrying out a classification Segmentation of retinal images is necessary before classifying glaucoma. Segmentation is carried out to extract the features of the optic disc and optic cup from the background which are needed when carrying out the classification stage of glaucoma. Segmentation helps detect more significant glaucoma disorders on retinal image. There are 2 segmentation techniques, namely manual segmentation and automated segmentation. Manual segmentation requires expertise, is very tedious and time consuming, and the results are highly subjective[3], [4]. Automated segmentation has many benefits, including increased accuracy, time savings, cost savings, and robustness. Currently, many automated segmentations have been developed using deep learning. Deep learning has powerful capabilities for integrating very large data sets, learning complex relationships, and incorporating existing knowledge in data [5].

One of the deep learning methods developed for image segmentation and classification is the convolution neural network (CNN). CNN has been developed with various architectures for both image segmentation and classification. CNN architecture that is widely used for segmentation is the U-Net architecture [8]. The U-Net architecture consists of two parts, namely the encoder, and the decoder. The encoder section functions to extract features from the image while the decoder functions to reconstruct image features [18]. Several studies that have used the U-Net architecture have been carried out. Fu *et al.* [9] performed blood vessel segmentation using the U-Net architecture with performance results of accuracy above 90% and recall below 75% but did not measure precision and F1-score. Venkatesh *et al.* [10] segmented skin cancer using the U-Net architecture with performance results of accuracy above 90% but did not measure recall, precision, and F1-score. Saood and Hatem [11] performed lung segmentation using the U-Net architecture with performance results of accuracy, recall, and precision above 90% but did not measure the F1-score. Research conducted by Fu *et al.* [9], Venkatesh *et al.* [10], and Saood and Hatem [11] only performed segmentation to separate the required features and did not carry out classification. Segmentation only provides information about the boundaries and regions of the object being observed. Classification assigns labels to images or regions with a holistic understanding of those images. Segmentation only provides information for experts but classification provides information labels that can be understood by the public. segmentation is carried out before classification to be able to increase the validity and accuracy of the classification.

One of the automatic systems for detecting glaucoma is the classification of retinal images. Classification is the process of determining a category or label for an object that has been defined previously based on a particular model [12]. A method used for classification is the Convolutional Neural Network (CNN) because it can handle input data of type $m\ x\ n$ such as retinal image data [13]. The CNN method has been developed for several architectures, one of which is the Extreme Inception (Xception) architecture. Xception is a CNN architecture that improves efficiency in computing processes. Xception has depthwise separable convolution and residual connections in its architecture so that this architecture has small parameters and is computationally efficient [12], [14]. Several studies that have classified glaucoma using the Xception architecture on retinal images including Juneja *et al.* [15] obtained accuracy, recall, and precision values above 90%, but did not measure the F1-score and Cohen's kappa values. Diaz-Pinto *et al.* [16] obtained results for accuracy, recall, precision, and F1-score above 85%, but did not measure the Cohen's kappa. Juneja *et al.* [17] obtained accuracy, recall, and precision values above 90%, but did not measure the F1-score and Cohen's kappa. The study conducted by Juneja *et al.* [15], Diaz-Pinto *et al.* [16], and Juneja *et al.* [17] directly classify the original image without doing segmentation on the retinal image.

Xception helps in reducing the number of parameters and computations. The small number of parameters can avoid overfitting on Xception but having too few parameters can lead to underfitting, where the model fails to capture even the basic patterns in the training data, especially for image data. The use of residual blocks in Xception utilizes skip connections to help the network learn better representations, too many skip connections or a very deep architecture can increase the risk of overfitting, especially when the dataset is small or lacks diversity. The inclusion of skip connections might lead to the oversight or loss of several significant features and information from the preceding layers. [18]. The skip connection feature which causes a lot of information in the previous layer to be missed or lost (vanishing gradient) can be overcome by modifying the Xception architecture. Another CNN architecture is the Densely Connected Convolutional Network (DenseNet). DenseNet is a CNN architecture that uses dense connections and can overcome the problem of loss of gradients in deep networks [19]. DenseNet consists of several dense blocks and works by combining each layer without involving the skip connection feature. Each feature map is used as input for the next layer [20]. Layer merging causes parameters to be larger. Xception places its emphasis on utilizing depthwise separable convolutions to achieve effective feature extraction. While it does enable some feature reuse across several layers, it doesn't prioritize dense connectivity to the extent that DenseNet does. DenseNet, on the other hand, is recognized for its dense connectivity approach, involving the concatenation and transfer of feature maps from prior layers to successive layers. This dense reuse of features plays a key role in enhancing gradient flow, counteracting the vanishing gradient issue, and fostering the network's ability to acquire concise and distinctive representations. Several studies have used DenseNet in classification, including Wu *et al.* [21] obtained results for accuracy and F1-scores above 90%, but did not measure recall and precision. Liao *et al.* [22] obtained accuracy, recall, and precision values above 80%, but did not measure the F1-score. Hasan *et al.* [23] obtained accuracy, recall, and F1-score values above 90%, but did not measure precision. Research conducted by Wu *et al.* [21], Liao *et al.* [22], and Hasan *et al.* [23] only focused on classification, not segmentation.

This study proposes two stages in detecting glaucoma. In the first stage, segmentation is performed on the retinal image. Segmentation is carried out to separate the features of the optic disc and optic cup which are needed when classifying glaucoma. The segmentation stage is carried out using the U-Net architecture. In the second stage, glaucoma was classified. Classification is carried out based on the segmentation results of the optic disc and optic cup features of the retinal image.

The new architecture proposed in this study to be used at the classification stage is the Xcep-Dense architecture. Xcep-Dense is a combination of Xception and DenseNet architectures. The combination of the Xception and DenseNet architectures is done by adding a dense block at the end of Xception. The dense block is used to replace the residual connection in Xception. The addition of dense blocks is done to combine the output of previous blocks and future blocks so that they can store previous and current spatial information for a long time. has a lower number of parameters and complexity. The success rate of the architecture proposed in this study was measured by calculating performance evaluations such as accuracy, recall, precision, F1-score, and Cohen's kappa. The results obtained from this study accommodate the needs of experts in the medical world and the public. The results at the segmentation stage provide input to medical experts to be able to observe the optic disc and optic cup in the retina. The results at the classification stage provide information not only to medical experts but also to the public such as patients regarding the possibility of glaucoma in the eye. The results of this study can be further developed into automatic applications for detecting glaucoma in the medical world.

II. MATERIALS AND METHODS

The stages of this research method can be seen in FIGURE 1.

A. DATA DESCRIPTION

The data used in this study used secondary data, namely datasets collected from Harvard Dataverse which were obtained and uploaded by Ungsoo Kim from Kim's Eye Hospital. This dataset consists of 1,532 images consisting of 3 labels and is divided into 788 Normal Control images, 289 Early Glaucoma images, and 467 Advanced Glaucoma images. The dataset can be accessed online through the following website pages of the dataset on link https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/1YRRAC. An example of an enlarged retinal image display on the Harvard Dataverse dataset can be seen in FIGURE 2.

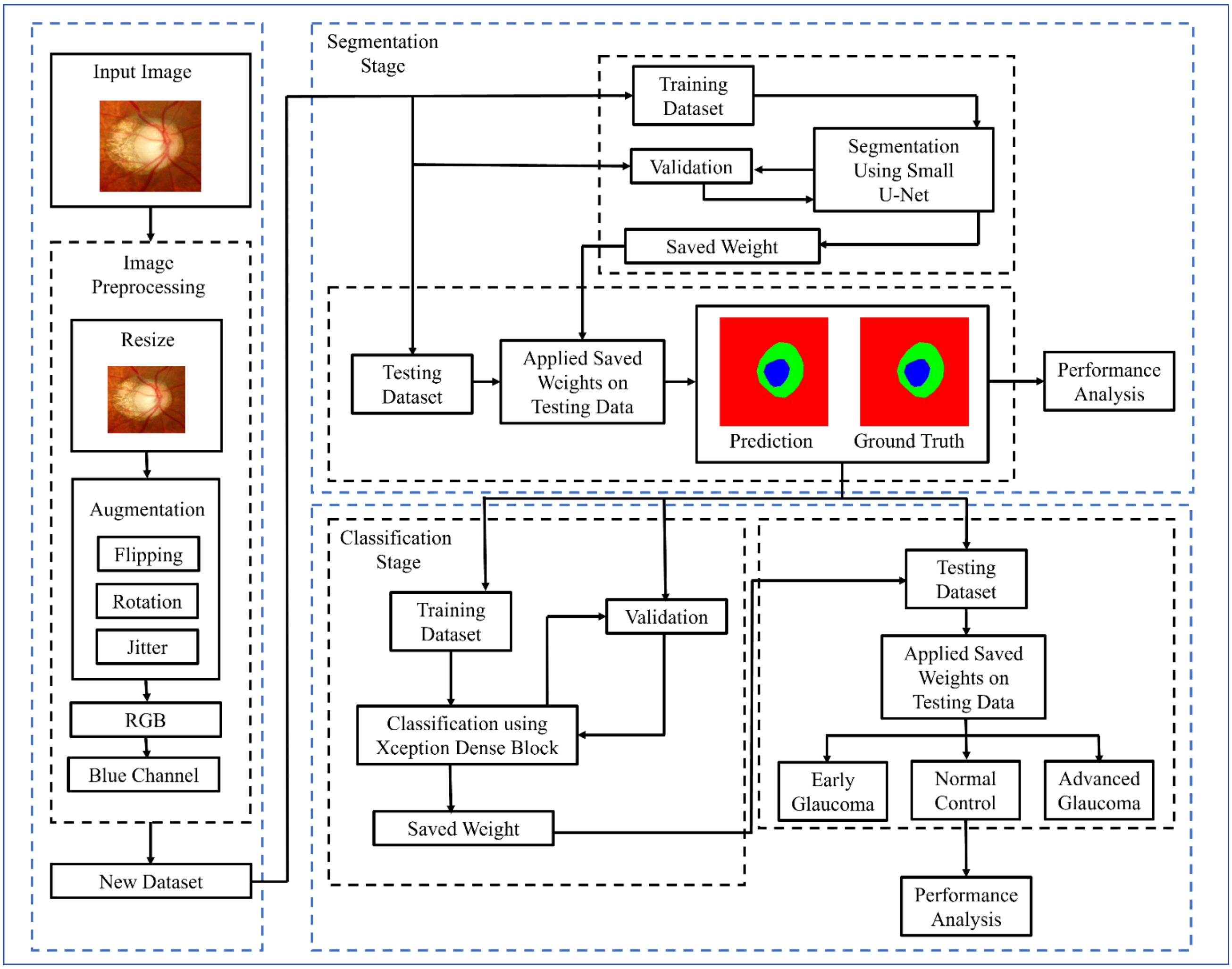


FIGURE 1. The flow of research stages

In FIGURE 2, you can see the appearance of the optic disc and optic cup in the retinal image. The optic disc is the bright part of the retinal image which is circled in green while the optic cup is the bright part inside the optic cup which is circled in blue.

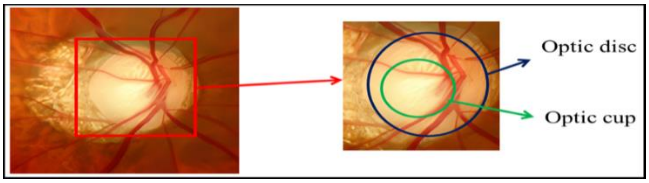


FIGURE 2. Views of the Optic Disc and Optic Cup on the Retina Image in Detail

B. PREPROCESSING

Pre-processing is carried out to improve the quality of the data for the better. The stages of research data pre-processing are as follows:

1) RESIZE

Image resizing aims to resize images adaptively for a more optimal display so that the data used is not too large and adapted to the model to be used. The resizing performed on all images conforms to the model used and, in this study, all images are converted to a size of 224×224 pixels.

2) DATA AUGMENTATION

The process of augmentation is to increase data to increase model capabilities and overcome the problem of limited data at the time of research [24]. In this study, the augmentations used are rotation image, image flipping, and color jitter. Augmentation with the rotation technique, namely random rotation with an angle of $[0^{\circ}, 20^{\circ}]$, the restriction of the rotation angle in this augmentation is intended so that other features that are not needed such as the background are not taken away so that the results are not too different from the original image. An example of an image resulting from augmentation rotation can be seen in FIGURE 3.

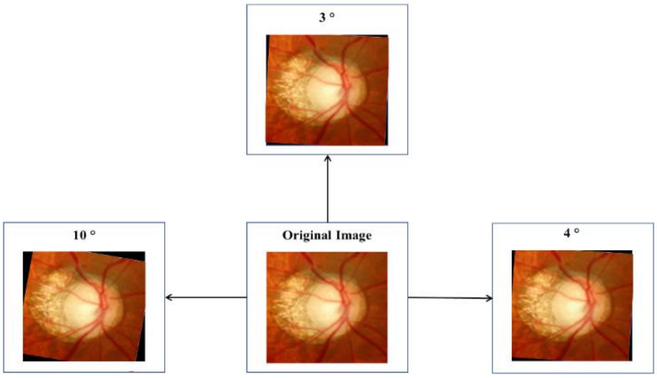


FIGURE 3. Example of an Augmented Rotation Image Result

In FIGURE 3, the original image is rotated randomly with angles of $3^{\circ}, 4^{\circ}$, and 10°respectively. The flipping augmentation used is a random flip vertically and horizontally which aims to increase data without reducing the features in it, an example of an image augmentation flipping results can be seen in FIGURE 4. In FIGURE 4, the original image is augmented by flipping vertically and flipping horizontally, so a new image is obtained. Furthermore, augmentation with color jitter, color jitter functions to increase the amount of data by changing the brightness, contrast, saturation, and color levels of the image randomly.

3) RGB CONVERSION

The image received from the input data is still in the form of BGR (Blue Green Red), then the image is converted from BGR (Blue Green Red) to RGB (Red Green Blue) which aims to make channel acquisition more accurate and the image easier to model.

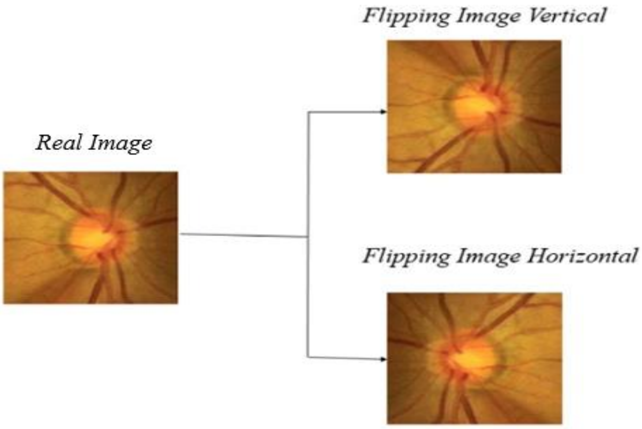


FIGURE 4. Example of Flipping Augmentation Result Image

Color jitter works by using a number matrix, namely the pixel values on a computer with each pixel combined into RGB to produce a variety of colors so that the contrast, saturation, and brightness of the image increase. An example of color jitter augmentation can be seen in FIGURE 5.

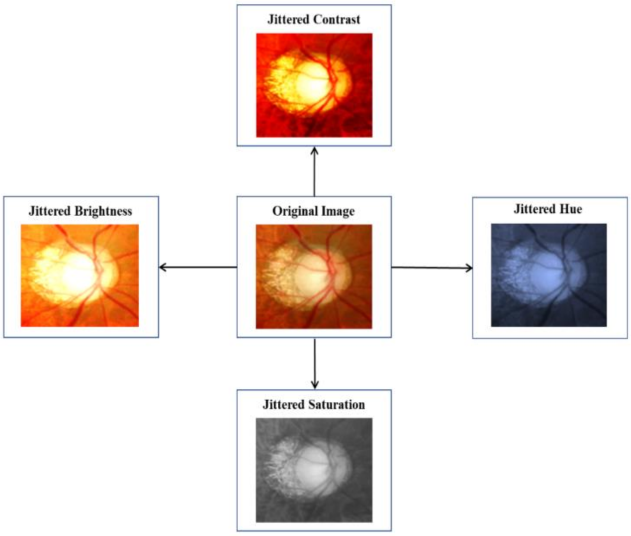


FIGURE 5. Example of Image Result of Color Jitter Augmentation

4) SPLIT CHANNEL

After converting the image from BGR to RBG, each channel is taken from the RBG image which is divided into red, green, and blue channels. The comparison results for each channel can be seen in FIGURE 6. Based on FIGURE 6, the red channel has a display that is too bright so that the optical disc and optical cup parts will be difficult to detect because they will include a background, the green channel has a display of the optical disc and optical cup parts to be taken which are quite clear but other features are not needed such as blood vessels, and the blue channel has the appearance of the optic disc and optic cup features which will be taken very clearly compared to other parts and other features are also not too flashy. The channel that will be taken for this research is the blue channel because the blue channel has a very clear display of optical disc and optical cup features than the other channels.

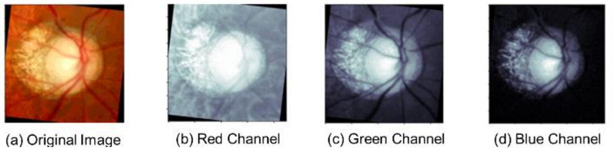

(a) Original Image (b) Red Channel (c) Green Channel (d) Blue Channel

FIGURE 6. Comparison of Display of Image Channels (a) Original image (b) Red Channel (c) Green Channel and (d) Blue Channel

C. SEGMENTATION WITH U-NET ARCHITECTURE

After the image preprocessing stage segmentation is carried out on the retinal image. Segmentation is carried out to separate the features that will be used at the classification stage, namely the optical disc and the optical cup. The segmentation stage uses the U-Net architecture. The U-Net architecture consists of two parts, namely the decoder, and encoder. The encoder part functions to extract useful features from the input image and the decoder is used to reconstruct the features to get the final segmentation results. The encoder has a convolution layer process, ReLu activation function, batch normalization, and max pooling. The U-Net architecture used at the segmentation stage can be seen in FIGURE 7. The convolution layer aims to learn the feature representation of the input. This layer consists of a set of convolutional kernels for extracting local features from input [25]. The process of calculating the convolution operation on the convolutional layer uses Equation (1).

$$v_{i,j} = \left(\sum_{u=0}^{n-1} \sum_{v=0}^{n-1} a_{u+i,v+j} \times k_{u+1,v+1} \right) + b_q \qquad (1)$$

for $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,n$, where $v_{i,j}$ is the convolution matrix entry in the baris i-th row, j-th column, $a_{u+i,v+j}$ is the input matrix entry $u+i$-row, $v+j$-th column, $k_{u+1,v+1}$ is the kernel matrix entry $u+1$-th row, $v+1$-th column, and b_q is the bias for the q-th kernel. Then the process of calculating the ReLu activation function is carried out from the results of the convolution layer. Rectified Linear Unit (ReLU) is one of the activation functions used in CNN where if the input of the activation function is negative then the output changes to zero. Meanwhile, if the input of the activation function is positive, then the output is the value of the input itself. Mathematically, ReLU can be defined in Equation (2).

$$t_{i,j} = r(v_{i,j}) = \max(0, x) = \begin{cases} v_{i,j} & jika\ v_{i,j} \geq 0 \\ 0 & jika\ v_{i,j} < 0 \end{cases} \qquad (2)$$

where, $t_{i,j}$ is the output result of the ReLu activation function and $v_{i,j}$ is the input pixel value from the result of the convolution layer operation.

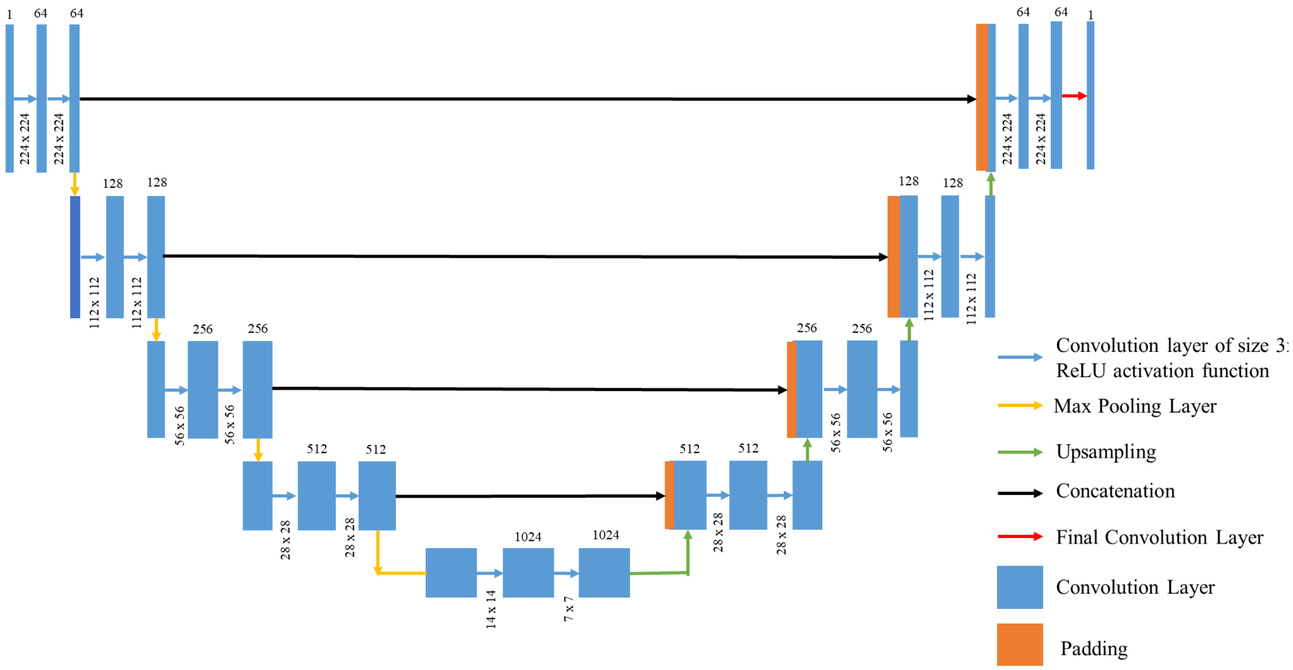


FIGURE 7. U-Net Architecture for Optic Disc and Optic Cup Segmentation

The results of the ReLU activation function are normalized using batch normalization. Batch Normalization is a normalization process for each layer in the network that is applied before or after the activation function [26]. Batch Normalization results are calculated by first calculating the average (μ_j) and variance $\left(\sigma_j^2\right)$, then normalizing them. The process of calculating the average (μ_j), variance $\left(\sigma_j^2\right)$, and normalization is carried out using Equations (3), (4), and (5).

$$\mu_j = \frac{1}{m}\sum_{i=1}^{m} t_{i,j} \quad (3)$$

$$\sigma_j^2 = \frac{1}{m}\sum_{i=1}^{m}\left(t_{i,j} - \mu_j\right)^2 \quad (4)$$

$$g = \hat{t}_{ij} = \frac{t_{ij} - \mu_j}{\sqrt{\sigma_j^2 + \varepsilon}} \quad (5)$$

where, μ_j is the average value of each mini-batch, σ_j^2 is the variance value for each mini-batch, j s the number of mini-batches, m is the amount of data in a mini-batch, $\hat{t}_{ij}$ the result of normalizing input values in the i-th row and the j-th column, t_{ij} is the input matrix entry resulting from the operation of the ReLU activation function in i-th row and j-th column, and ε is the smallest constant value. Then the dimension reduction is carried out on the feature map resulting from batch normalization using max pooling. In the decoder section, the convolution layer operation is performed, the ReLu activation function, and the same batch normalization as in the encoder section. Then, the feature map dimensions are increased by using up-sampling on the decoder section. The results of operations on the encoder and decoder are combined using concatenate. Then perform the calculation operation of the SoftMax activation function using Equation (6).

$$s = s(g)_j = \frac{e^{g_j}}{\sum_{k=1}^{K} e^{g_k}} \quad (6)$$

for $k = 1, \ldots, K$ where K is the number of classes. s is the output result of the softmax activation function and g is the input result of batch normalization.

The final stage is to calculate the loss function using categorical loss entropy. Categorical cross-entropy is a loss function that has more than 2 object classes or multi-class. Categorical cross-entropy is calculated using Equation (7) [27].

$$L(y, s) = -\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij} \cdot \log s_{ij} \quad (7)$$

where m is the predicted result matrix row, n is the predicted matrix column, s_{ij} is the output matrix entry of the softmax activation function operation result in the i-th row of the j-th column, y_{ij} is the actual result matrix entry in the i-th row of the j-th column, and L are the results of categorical cross-entropy.

D. CLASSIFICATION WITH XCEP-DENSE ARCHITECTURE

After the segmentation stage, a combination of Xception and Dense Block architectures is carried out. The combination of the Xception and Dense Block architectures is done by adding a dense block at the end of the Xception model after the flattening process to overcome the vanishing gradient problem caused by the use of skip connections. The combination of the Xception and Dense Block architectures forms a new architecture, namely Xcep-Dense which can be seen in FIGURE 8.

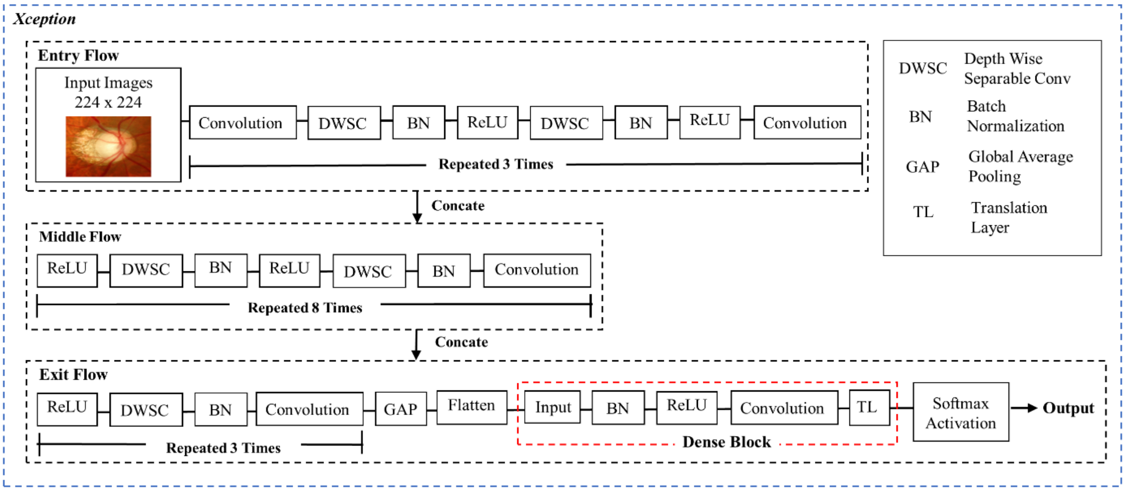


FIGURE 8. Xcep-Dense Architecture

Based on FIGURE 8, the basic architecture used is Xception where the Xception architecture consists of 3 main blocks, namely entry flow, middle flow, and exit flow. Each block in Xception uses a depthwise separable convolution in its architecture which has a residual connection in it. The Xception architecture is composed of several convolution layers, batch normalization, and ReLU activation. In the basic Xception architecture, image input is first performed on the entry flow block which is repeated 3 times according to the Xception model which has many convolution layers, then it is continued into the middle flow block which is repeated 8 times and finally enters the exit flow block which is done 3 times with the addition of dense block. In the dense block, there are input sections, batch normalization, ReLU, convolution, and transition layers where this process is carried out to overcome the vanishing gradient problem which is carried out as many as d-layers according to the model and then ends with class classification using the softmax activation function.

III. RESULT AND DISCUSSION

A. PREPROCESSING

In the initial preprocessing stage, the image is resized to a size of 224 × 224 pixels. Then, augmentation is performed using image rotating, image flipping, and color jitter techniques. The rotating image technique produced 5,902 new images, the flipping image technique produced 5,932 new images, and the color jitter technique produced 5,918 new images so a total of 19,032 new data were obtained. The data consists of three classes, namely 6,370 images of the advanced glaucoma class, 6,358 images of the early glaucoma class, and 6,304 images of the normal control class. Next, a conversion from BGR to RGB is performed to avoid errors when selecting channels, and the blue channel is retrieved from the image because the blue channel has optical disc and optical cup features that are brighter and clearer than other channels. The results of the preprocessing stages carried out can be seen in FIGURE 9.

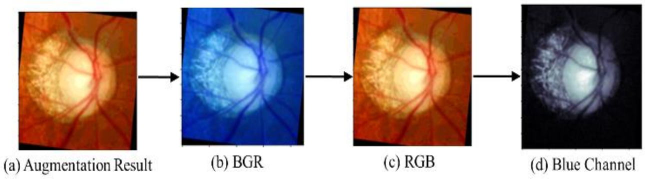


FIGURE 9. Image Preprocessing Results in Stages (a) Augmentation (b) BGR (c) RGB and (d) Blue Channel

B. SEGMENTATION WITH U-NET ARCHITECTURE

Segmentation is carried out to separate the optical disc and optical cup features that will be used at the classification stage. The segmentation process is divided into two, namely training and testing. In the training process, measurement of accuracy and loss is carried out on the training data and data validation. Accuracy is used to measure the success of the segmentation stage in extracting the desired features from the image while loss is measured to see the level of error in the segmentation stage in recognizing features from the image. Graph of accuracy and loss values in the segmentation stage training process can be seen in FIGURE 10.

In FIGURE 11(a) it can be seen that the recall in the training process is close to 100%. The recall on training data and validation data continues to increase and begins to stabilize in the 5th epoch for training data and the 30th epoch for data validation. In FIGURE 11(b) it can be seen that the precision value in the training process for training data and validation data continues to increase above 95% and begins to stabilize in the 10th epoch. At the segmentation training stage, F1-score and Cohen's kappa measurements were also carried out. The F1-score is the average of the weighted recall and precision values. Cohen's kappa indicates a measure of the degree of agreement between the predicted features produced and the actual features on a nominal scale. The graph of the F1-score and Cohen's kappa in the segmentation stage of the training process can be seen in FIGURE 12.

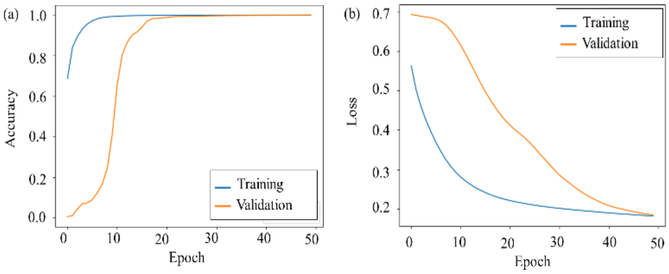


FIGURE 10. Graph of (a) Accuracy and (b) Loss in the Segmentation Training Process

In FIGURE 10(a), it can be seen that the accuracy value in the training process is above 90%. Accuracy values in training data and validation data continued to increase and began to stabilize from the 20th epoch. In FIGURE 10(b) it can be seen that the loss in the training process for training data and data validation continues to decrease towards a loss below 25%. In addition to accuracy and loss, at the segmentation training stage recall and precision measurements are also carried out. The measured recall indicates the success of the segmentation stage in recognizing each feature of the image correctly, while the precision is measured to see the success rate of the segmentation stage in recognizing each feature correctly compared to the features predicted correctly. The graph of recall and precision in the segmentation stage training process can be seen in FIGURE 11.

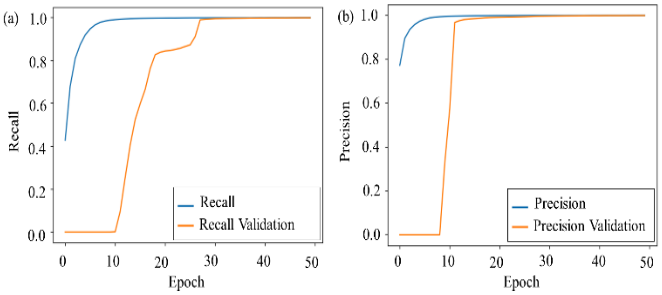


FIGURE 11. Graph of (a) Recall and (b) Precision in the Segmentation Training Process

In FIGURE 12(a) it can be seen that the F1-score in the training process is close to 100%. The F1-score for training data and validation data continued to increase and began to stabilize in the 5th epoch for training data and the 30th epoch for data validation. In FIGURE 12(b) it can be seen that Cohen's kappa in the training process for data training and data validation continues to increase above 85%. In the testing process of the segmentation stage, a comparison is made between the predicted results of the segmentation stage and the ground truth. A comparison between the results of segmentation and ground truth at the segmentation stage can be seen in Table 1.

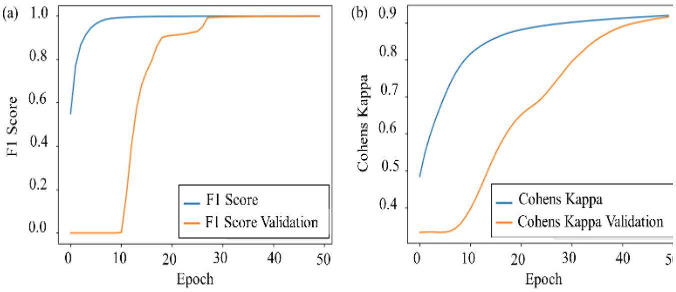


FIGURE 12. Graph of (a) F1-Score and (b) Cohen's Kappa in the Segmentation Training Process

TABLE 1
Comparison between segmentation results and ground truth

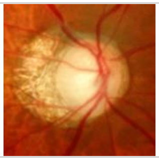
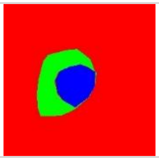
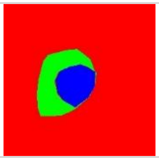
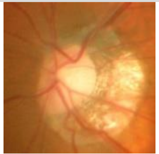
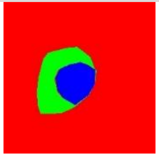
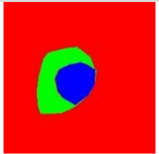
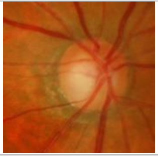
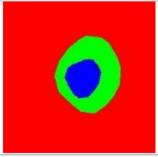
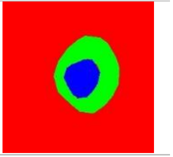
No	Original Image	Prediction Results	*Ground Truth*
1			
2			
3			

TABLE 1, It can see a comparison of the results of segmentation and ground truth at the segmentation stage. At the segmentation stage, the resulting prediction results have a similar appearance to ground truth. This shows that the segmentation stage was successful in recognizing the features of the optic disc and optic cup from the retinal image according to ground truth.

C. CLASSIFICATION WITH XCEP-DENSE ARCHITECTURE

The classification stage consists of two processes, namely training, and testing. Before entering the training stage, the data was divided into two parts, namely 80%, namely 15,226 data as training data, and as much as 20%, namely 3,806 data as testing data. In the data training process, the data used will be trained using the Xcep-Dense model with each parameter used including the number of epochs of 200 and a batch size of 32. The labels used are 3 labels, namely advanced glaucoma, early glaucoma, and normal control. Then, a random split was performed on the training data into two parts, namely 90%, namely 13,703 data as training data, and as much as 10%, namely 1,523 data as validation data. Furthermore, the results of the split data are carried out by a training process for each layer which is carried out until the 200th epoch. In the training process, the accuracy value of the training data and data validation is measured to see the ability of the proposed classification model. In addition, the loss value is also measured to see the level of error between the predicted class and the actual class. Graph of accuracy and loss in the classification training process can be seen in FIGURE 13. In FIGURE 13(a) it can be seen that the accuracy in the training process is above 85%. The accuracy of the training data and data validation continues to increase and begins to stabilize from the 50th epoch. In FIGURE 13(b) it can be seen that the loss value in the training process for training data and data validation continues to decrease below 20%.

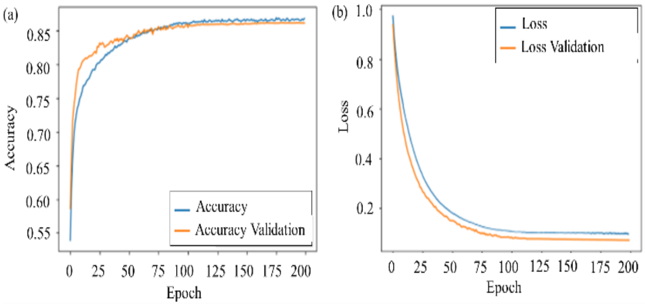


FIGURE 13. Graph of (a) Accuracy and (b) Loss in the Classification Training Process

In addition to the accuracy and loss, at the classification training stage recall and precision measurements are also carried out. The recall indicates the success of the classification stage in correctly predicting the pixels of each label. Precision is a measure of how well the architecture correctly predicts each label compared to the correctly predicted result. The graph of recall and precision in the segmentation stage training process can be seen in FIGURE 14.

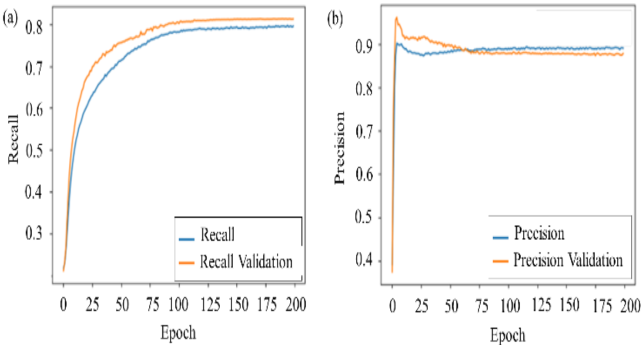


FIGURE 14. Graph of (a) Recall and (b) Precision in the Classification Training Process

In FIGURE 14(a) it can be seen that the recall value on the training data and data validation in the training process is very good. This can be seen from the results that continue to increase in each epoch. In the 100th epoch, the recall value has begun to stabilize and has not decreased. The recall value in the training process is 80%. In FIGURE 14(b) it can be seen that the precision value of the training data and data validation in the training process is 88%. The precision value begins to stabilize at the 60th epoch. During the training process, the F1-score and Cohen's kappa were also measured. The F1-score is the average of the weighted recall and precision values. Cohen's kappa shows a measure of the degree of agreement between the predicted results generated with the actual label on a nominal scale with two or more classes. The graph of the F1-score and Cohen's kappa in the training process can be seen in FIGURE 15. In FIGURE 15(a) it can be seen that the F1-score on the training data and data validation training process continues to increase in each epoch. The F1-score at the 100th epoch has begun to stabilize and has not decreased. In FIGURE 15(b) it can be seen that Cohen's kappa in the training data and data validation in the training process is towards 80%. However, the graph of Cohen's kappa is not overfitting and continues to increase in each epoch and begins to stabilize at the 100th epoch.

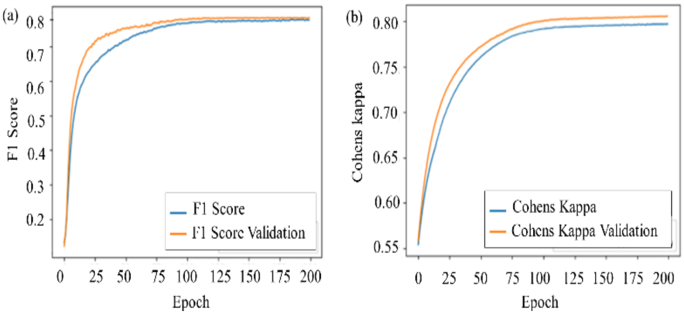


FIGURE 15. Graph of (a) F1-Score and (b) Cohen's Kappa in the Classification Training Process

In the training process, the accuracy of the forecast results of a model used is measured, and the smaller the error results obtained, the better the model used by calculating the RMSE result. The RMSE shows that the difference between the predicted results and the actual results has a small error or error. The graph of the RMSE in the training process can be seen in FIGURE 16.

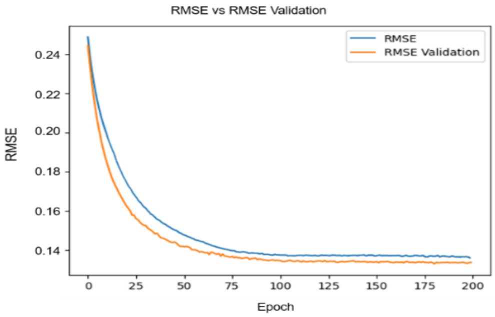


FIGURE 16. Graph of RMSE in the Classification Training Process

In FIGURE 16 it can be seen that the RMSE on the training data and data validation training process is very good. The RMSE obtained is close to 0 and continues to decrease at each epoch. In the 100th epoch, the RMSE has begun to stabilize and no longer increases. The RMSE indicates that the image quality is very good. At the testing stage, predictions are made based on the results of the best weight that has been obtained at the training stage. The testing data obtained from the split data is 3,806 data. The data is used to see the results of the model's performance in classifying images. At this stage, the results of model performance will be obtained in the form of accuracy, recall, precision, F1-score, and Cohen's kappa. A comparison of the performance results for each label is shown in FIGURE 17.

Based on FIGURE 17, each label is marked with orange advanced glaucoma, yellow early glaucoma, and green normal eyes. In the graph, it can be seen that the results of the green early glaucoma label have smaller results compared to other labels. This occurs because the initial data before augmentation on the early glaucoma label is less than the other label data. The highest accuracy, recall, F1-score, and Cohen's Kappa are owned by the advanced glaucoma label, but the precision on the advanced glaucoma label is still below the precision on the normal control label. The precision on the normal control label has the highest value compared to the precision on the other labels. The results show that Xcep-Dense is excellent for classifying glaucoma from the data provided.

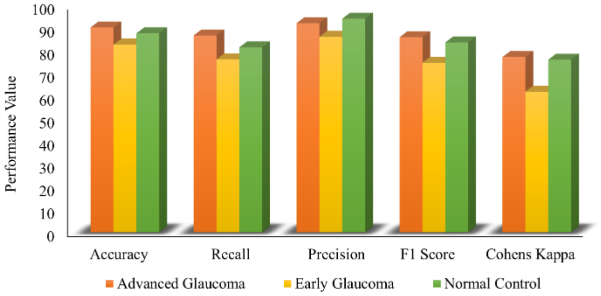


FIGURE 17. Comparison of results Classification performance on each label

At the testing stage, the AUC for each class was also measured. The AUC obtained for each class can be seen in the ROC graph in FIGURE 18. Based on FIGURE 18, it can be seen that the higher the false positive rate, the higher the true positive rate for each label. The ROC graph shows the performance of the proposed classification model at all classification thresholds. On the ROC chart, there are AUC values for each label. The AUC indicates the model's ability to group each label. The lowest AUC is owned by the normal control label and the highest is obtained by the advanced glaucoma label.

D. ANALYSIS AND DISCUSSION

In this study, two stages were carried out in detecting glaucoma which consisted of segmentation and classification stages. At the segmentation stage, the performance results on data testing obtained accuracy, recall, precision, and an F1-score of 98% while Cohen's kappa was 88%. the result of Cohen's Kappa indicates that there is a very high degree of agreement between actual observations and predictive observations than expected. F1-score above 98% indicates that the U-Net architecture works very well in optical disc and optical cup segmentation. The F1-score results also show that the model obtained has a good balance between precision and recall, or in other words a very low number of false positives and false positives.

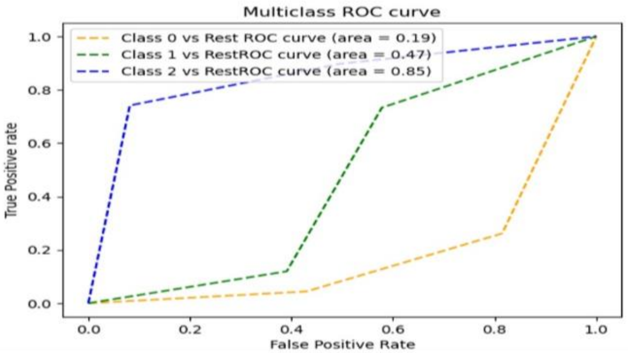


FIGURE 18. **Multiclass ROC Graph**

At the classification stage, the performance results of Xcep-Dense on data testing obtained an accuracy (Acc) of 87.58%, recall (Rec) of 81.43%, precision (Prec) of 90.73%, F1-score (F1) of 81.37%, and Cohen's Kappa (Kappa) of 77%. The obtained F1-score shows a good balance between precision and recall, with a high average harmonization. Cohen's Kappa of 77% indicates a good degree of agreement between the two observers. Overall, the performance results show that the model produced by Xcep-Dense has good performance in the classification of glaucoma disorders with high accuracy, precision, and recall on the labels of advanced glaucoma, early glaucoma, and normal control based on segmented images. However, the F1-score and Cohen's Kappa obtained are still below the results of the F1-score and Cohen's Kappa from the segmentation stage. These results show that Xcep-Dense can classify glaucoma on the labels of advanced glaucoma, early glaucoma, and normal controls based on segmented images. To determine the success of the Xcep-Dense Net in the classification of glaucoma disorders, the performance results of the Xcep-Dense Net are compared with existing studies. The comparison of the classification results in the study with the classification results in other studies can be seen in Table 2.

Based on Table 2, the classification results in this study obtained the highest accuracy, precision, and F1-score compared to other studies. The study by Ahn *et al.* [28] has a higher accuracy than the studies of Velpula [29] and Esengonul and Cunha [30], however, the research of Ahn *et al*. [28] did not measure recall, precision, and F1-score. The highest recall was obtained by Esengonul and Cunha's [30] study, but this study did not measure the F1-score. Esengonul and Cunha's [30] study has the highest recall, but the accuracy and precision are lower than the accuracy and precision obtained by the proposed method. The results of the recall, F1-score, and Cohen's Kappa in the classification still need to be below 85% so the classification results need to be improved to be more valid and accurate. The Xcep-Dense architecture can be further developed to obtain better classification performance results in the detection of glaucoma disorders. The results obtained in this study indicate that the proposed model can be developed as an automatic system for the early detection of glaucoma so that treatment planning can be carried out better. The involvement of segmentation in this study is to improve the performance of the classification.

TABLE 2
Comparison of performance evaluation results with other research

Methods	Acc.	Rec.	Prec.	F1	Kappa
	(%)	(%)	(%)	(%)	(%)
InceptionV3 [28]	84,5	-	-	-	-
Fusion [29]	75	71	72	63	-
Mobile Use [30]	72	87	90	-	-
ResNet50 [29]	83	53	90	55	-
Xcep-Dense	87.58	81.43	90.73	81.37	77

IV. CONCLUSION

This study proposes two stages to classify glaucoma which consists of segmentation and classification stages. Segmentation is performed using the U-Net architecture to separate the features of the optic disc and optic cup from other features on the retinal image. The classification is carried out based on image segmentation results to classify glaucoma into three classes, namely advanced glaucoma, early glaucoma, and normal control. At the segmentation stage, the performance results for accuracy, recall, precision, and F1-score were above 90%, while Cohen's kappa was above 85%. These results conclude that the model produced by the U-Net architecture has excellent and robust capabilities in segmenting the optic disc and optic cup in retinal images. The performance results in the classification obtained accuracy and precision above 85%, while the recall, F1-score, and Cohen's Kappa are still below 85%, but the classification results using Xcep-Dense can still be considered very good because the average performance value is quite high above 75%. These results can be concluded that the proposed method is good in the classification of glaucoma based on retinal segmented images. The results of this study are models that can be used directly in detecting glaucoma disorders so the future work of the study can be developed to build intelligent automatic applications that can assist the medical world in identifying glaucoma disorders.

REFERENCES

[1] B. B. Naik and R. Mariappan, “Classification of Eye Diseases Using Optic Cup Segmentation and Optic Disc Ratio,” *IOSR J. Comput. Eng.*, vol. 18, no. 05, pp. 87–94, 2016, doi: 10.9790/0661-1805038794.

[2] M. U. Akram, A. Tariq, S. Khalid, M. Y. Javed, S. Abbas, and U. U. Yasin, “Glaucoma Detection Using Novel Optic Disc Localization, Hybrid Feature Set and Classification Techniques,” *Australas. Phys. Eng. Sci. Med.*, vol. 38, no. 4, pp. 643–655, 2015, doi: 10.1007/s13246-015-0377-y.

[3] M. Han *et al.*, “Automatic Segmentation of Human Placenta Images With U-Net,” *IEEE Access*, vol. 7, pp. 180083–180092, 2019, doi: 10.1109/ACCESS.2019.2958133.

[4] A. Desiani, Erwin, B. Suprihatin, F. Efriliyanti, M. Arhami, and E. Setyaningsih, “VG-DropDNet A Robust Architecture for Blood Vessels Segmentation on Retinal Image,” *IEEE Access*, vol. 10, no.

August, pp. 1–1, 2022, doi 10.1109/access.2022.3202890.

[5] A. Septiarini, D. M. Khairina, A. H. Kridalaksana, and H. Hamdani, “Automatic Glaucoma Detection Method Applying A Statistical Approach to Fundus Images,” *Healthc. Inform. Res.*, vol. 24, no. 1, pp. 53–60, 2018, doi: 10.4258/hir.2018.24.1.53.

[6] A. Desiani, Erwin, B. Suprihatin, Ermatita, F. R. Husein, and Y. Wahyudi, “A Novelty Patching of Circular Random and Ordered Techniques on Retinal Image to Improve CNN U-Net Performance,” *Eng. Lett.*, vol. 30, no. 4, pp. 1217–1229, 2022.

[7] A. Desiani, B. Suprihatin, S. Yahdin, A. I. Putri, and F. R. Husein, “Bi-path Architecture of CNN Segmentation and Classification Method for Cervical Cancer Disorders Based on Pap - smear Images,” *IAENG Int. J. Comput. Sci.*, vol. 48, no. 3, 2021.

[8] V. Sathananthavathi and G. Indumathi, “Encoder Enhanced Atrous (EEA) Unet architecture for Retinal Blood vessel segmentation,” *Cogn. Syst. Res.*, vol. 67, pp. 84–95, 2021, doi: 10.1016/j.cogsys.2021.01.003.

[9] H. Fu, Y. Xu, D. W. K. Wong, and J. L. Ocular, “Retinal Vessel Segmentation Via Deep Learning Network And Fully-Connected Conditional Random Fields,” in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 698–701.

[10] G. M. Venkatesh, Y. G. Naresh, S. Little, and N. E. O’Connor, *A deep residual architecture for skin lesion segmentation*, vol. 11041 LNCS. Springer International Publishing, 2018.

[11] A. Saood and I. Hatem, “COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet,” *BMC Med. Imaging*, vol. 21, no. 1, pp. 1–10, 2021, doi: 10.1186/s12880-020-00529-5.

[12] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, “An Improved Method to Construct basic Probability Assignment based on The Confusion Matrix for Classification Problem,” *Inf. Sci. (Ny).*, vol. 340–341, pp. 250–261, 2016, doi: 10.1016/j.ins.2016.01.033.

[13] I. R. I. Haque and J. Neubert, “Deep learning approaches to biomedical image segmentation,” *Informatics Med. Unlocked*, vol. 18, p. 100297, 2020, doi: 10.1016/j.imu.2020.100297.

[14] I. Kandel and M. Castelli, “Transfer Learning with Convolutional Neural Networks for Diabetic Retinopathy Image Classification. A Review,” *Appl. Sci.*, vol. 10, no. 6, pp. 1–24, 2020, doi: 10.3390/app10062021.

[15] M. Juneja, S. Thakur, A. Uniyal, A. Wani, N. Thakur, and P. Jindal, “Deep Learning-Based Classification Network for Glaucoma in Retinal Images,” *Comput. Electr. Eng.*, vol. 101, no. April, p. 108009, 2022, doi: 10.1016/j.compeleceng.2022.108009.

[16] A. Diaz-Pinto, S. Morales, V. Naranjo, T. Köhler, J. M. Mossi, and A. Navea, “CNNs for Automatic Glaucoma Assessment using Fundus Images: An extensive validation,” *Biomed. Eng. Online*, vol. 18, no. 1, pp. 1–19, 2019, doi 10.1186/s12938-019-0649-y.

[17] M. Juneja, N. Thakur, S. Thakur, A. Uniyal, A. Wani, and P. Jindal, “GC-NET for Classification of Glaucoma in The Retinal Fundus Image,” *Mach. Vis. Appl.*, vol. 31, no. 5, pp. 1–18, 2020, doi: 10.1007/s00138-020-01091-4.

[18] K. Wu, S. Zhang, and Z. Xie, “Monocular Depth Prediction with Residual DenseASPP Network,” *IEEE Access*, vol. 8, no. 1, pp. 129899–129910, 2020, doi: 10.1109/ACCESS.2020.3006704.

[19] J. Zhang, C. Wu, X. Yu, and X. Lei, “A Novel DenseNet Generative Adversarial Network for Heterogenous Low-Light Image Enhancement,” *Front. Neurorobot.*, vol. 15, no. June, pp. 1–10, 2021, doi: 10.3389/fnbot.2021.700011.

[20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[21] J. Wu, W. Hu, Y. Wen, W. Tu, and X. Liu, “Skin Lesion Classification Using Densely Connected Convolutional Networks with Attention Residual Learning,” *Sensors (Switzerland)*, vol. 20, no. 24, pp. 1–15, 2020, doi: 10.3390/s20247080.

[22] T. Liao *et al.*, “Classification of Asymmetry in Mammography via The DenseNet Convolutional Neural Network,” *Eur. J. Radiol. Open*, vol. 11, no. July, p. 100502, 2023, doi: 10.1016/j.ejro.2023.100502.

[23] N. Hasan, Y. Bao, A. Shawon, and Y. Huang, “DenseNet Convolutional Neural Networks Application for Predicting COVID-19 Using CT Image,” *SN Comput. Sci.*, vol. 2, no. 5, pp. 1–11, 2021, doi: 10.1007/s42979-021-00782-7.

[24] C. Bhardwaj, S. Jain, and M. Sood, “Diabetic retinopathy severity grading employing quadrant-based Inception-V3 convolution neural network architecture,” *Int. J. Imaging Syst. Technol.*, vol. 31, no. 2, pp. 592–608, 2021, doi: 10.1002/ima.22510.

[25] T. Guo, J. Dong, and H. Li, “Simple convolutional neural network on image classification,” in *IEEE International Conference on Big Data Analysis*, 2017, pp. 721–724.

[26] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” *Journal. Pract.*, vol. 10, no. 6, pp. 730–743, 2016, doi: 10.1080/17512786.2015.1058180.

[27] 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, doi: 10.1109/ACCESS.2019.2962617.

[28] J. M. Ahn, S. Kim, K. S. Ahn, S. H. Cho, K. B. Lee, and U. S. Kim, “A Deep Learning Model for The Detection of Both Advanced and Early Glaucoma using Fundus Photography,” *PLoS One*, vol. 14, no. 1, pp. 1–8, 2019, doi: 10.1371/journal.pone.0207982.

[29] V. K. Velpula and L. D. Sharma, “Multi-Stage Glaucoma Classification using Pre-Trained Convolutional Neural Networks and Voting-Based Classifier Fusion,” *Front. Physiol.*, vol. 14, no. June, pp. 1–17, 2023, doi: 10.3389/fphys.2023.1175881.

[30] M. Esengonul and A. Cunha, “Glaucoma Detection using Convolutional Neural Network Mobile Use,” *Computer Science.*, vol. 219, pp. 1153–1160,2023.

BIOGRAPHY

Anita Desiani. In 2020, She is currently working on a project for her Doctoral Program at Mathematics and Natural Science Faculty, Universitas Sriwijaya. She received mathematics bachelor from Universitas Sriwijaya in 2000, a magister degree in Computer Science from Universitas Gadjah Mada in 2003, and a doctoral degree in Mathematics and science department in 2022. In 2004, she joined as a lecturer at Mathematics Department at Universitas Sriwijaya until now. Her current research interests include the field of data mining, image processing, pattern recognition and computer vision, and artificial intelligence.

Sigit Priyanta obtained his Doctorate in 2016 from Doctoral Program in Computer Science, Universitas Gadjah Mada, Yogyakarta, Indonesia. He is a lecturer in the Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia. His research interests are in geographic information systems, location-based services, and mobile information systems.

Indri Ramayanti was born in Palembang, Indonesia, in 1983. She received his Bachelor of Biology from Universitas Sriwijaya, Indonesia, in 2005, and an M.Sc degree in Tropical Medicine from Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia, in 2008. In 2009, she joined Universitas Muhammadiyah Palembang, as a lecturer, she now works in the Faculty of Medicine at Universitas Muhammadiyah Palembang. Then, in 2022, she received his environmental science doctorate at Sriwijaya University Postgraduate Program. Her current research interests include parasitology, entomology, and medical informatics.

Bambang Suprihatin was born in Salatiga, Indonesia, in 1971. He received his Bachelor of Mathematics from Universitas Sriwijaya, Indonesia, in 1994, and an M.Sc. degree in Mathematics from the Bandung Institute of Technology (ITB), Bandung, Indonesia, in 2002. In 1994, he joined Universitas Sriwijaya, as a Lecturer. He was Associate Professor in 2011. In 2012, he received his Doctorate in Mathematics, Universitas Gadjah Mada (UGM) in 2016. His current research interests are statistics and modeling.

Muhammat Rio Halim was born in Palembang, 2000, Indonesia He is currently working on a project for his undergraduate degree at Mathematics Department, science and Nature Faculty, Universitas Sriwijaya. In 2021, he joined the Laboratory of Computation Mathematics and Natural Science Faculty, Universitas Sriwijaya as Assistant Lecturer. His current research interests include the field of data structure, image processing, pattern recognition and computer vision, data mining, and artificial intelligence.

Ira Rayani was born in Pagaralam, Mei 2001. She currently working on a project for his undergraduate degree at Mathematics Department, science and Nature Faculty, Universitas Sriwijaya. In 2023, She joined the Laboratory of Computation Mathematics and Natural Science Faculty, Universitas Sriwijaya as an Assistant Lecturer. Her current research includes the field of image processing, pattern recognition and computer vision, data mining, and artificial intelligence.

Dite Geovani was born in Palembang, Januari 2002. She currently working on a project for his undergraduate degree at Mathematics Department, science and Nature Faculty, Universitas Sriwijaya. In 2021, She joined the Laboratory of Computation Mathematics and Natural Science Faculty, Universitas Sriwijaya as an Assistant Lecturer. Her current research includes the field of image processing, pattern recognition and computer vision, data mining, and artificial intelligence.