Classification of eye diseases in fundus images using convolutional neural network (CNN) method with efficientnet architecture

Zhafeni Arif*), R. Yunendah Nur Fu'adah, Syamsul Rizal, Divo Ilhamdi
Teknik Telekomunikasi, Fakultas Teknik Elektro, Telkom University, Bandung, Indonesia

Article Info

ABSTRACT

This research is designed to classify eye disease conditions into three classes, namely normal, cataract, and glaucoma using a system. The system will use Convolutional Neural Network (CNN) with EfficientNet architecture. The EfficientNet that will be used is EfficientNet-B0. The dataset in this paper is obtained from Kaggle totaling 300 images. From this data, augmentation is carried out so that 3.600 images are obtained consisting of "normal" (1200 images), "cataract" (1200 images), and "glaucoma" (1200 images). This data is processed into 4 different datasets, namely the original dataset, augmentation dataset, augmentation dataset that has been preprocessed grayscale, and augmentation dataset that has been preprocessed thresholding. The best results are obtained using the Adam optimizer, learning rate 0.00001, and batch size 32, and iterations of 20 epochs. The best dataset is the augmentation dataset that has been preprocessed grayscale obtained quite good results with 79.22% accuracy, 80.3% precision value, 79.22% recall value, 78.87% F1-Score.

Corresponding Author:

Arif, Z.,
Telkom University, Bandung, Indonesia
Email: zhafeniarif@student.telkomuniversity.ac.id

Introduction

The eye is one of the most important vital organs owned by humans. Humans can obtain 80% of their information from sight alone (G. P. Kumasela, 2013). So the eye organ needs to be taken care of as well as possible. There are several eye diseases that can threaten humans. Eye health has a significant impact on the quality of human life, such as physical activity, mental and social well-being. In 2015 an estimated 36 million people worldwide were living with blindness. Two hundred and seventeen million people have moderate to severe visual impairment, and 188 million people have mild visual impairment (Bourne R et al, 2020). The World Health Organization (WHO) revealed that the main factors leading to the increase in the number of people with visual impairment are an aging population, lifestyle changes, and limited access to eye care in low- and middle-income countries (WHO, 2020). Vision loss has many causes that require comprehensive prevention, treatment and care. Cataracts, and glaucoma account for the majority of global visual impairment (M. J. Burton et al., 2020).

Detection of eye abnormalities can indeed be done directly with the naked eye. But this can only detect abnormalities on the outside of the eye. For other eye disorders, a more in-depth examination is needed using a Fundus Camera or Ophthalmoscope that produces a fundus image. Detection and classification of eye diseases with a fundus camera is done by medical examination, which is directly observed by a doctor.

However, this method takes a long time (Vania Annisa Queentinela, Y. T. ,2021). An accurate funduscopic examination takes 30 minutes or longer. In a state of practice overwhelmed by increasing time pressure, doctors are unlikely to spend a lot of time conducting a thorough funduscopic examination (Devin D. Mackay, et al, 2015). Therefore, to process a large number of fundus images quickly and accurately, deep learning-based digital image processing is required.

In 2021, research was conducted (Geza Jeremia Bu'ulölö, A. J. ,2021) Identification of Cataract Eye Disease Using Convolutional Neural Network which is divided into 2 classes, namely normal and cataract. Using data from kaggle as many as 220 fundus images. With the accuracy values obtained, namely accuracy of 91.41% on the RMSProp Optimizer, 92.93% on the Adam Optimizer, 81.56% on the SGD Optimizer, and 68.65% on the AdaDelta Optimizer. In 2021, research was conducted (Fani Nurona Cahya, N. H. ,2021) which classified eye diseases using a convolutional neural network (CNN) using the AlexNet method into 4 types of normal, cataract, glaucoma, and retinal disease. The dataset used consists of 610 data. The accuracy results of the study were obtained 98.37%. Another study in 2022 which has researched (Syafiq Hilmi Abdullah, R. M. ,2022) the classification of diabetic retinopathy based on fundus image processing and deep learning. There are 5 classes namely no DR, mild NPDR, moderate NDPR, severe NDPR and proliferate DR. The dataset used in this study contains 3662 images and the argumentation result dataset contains 5100 images. This research uses CNN method with EfficientNet-B0 architecture. Accuracy of 88.863%, precision value of 89.2%, recall value of 89%, and F1-Score of 88.8% were obtained. In 2022, research was conducted (Diki Hananta Firdaus, B. I ,2022) on the classification of cataract disease in the eye using the web-based CNN. The dataset used was 512 images into two classes, namely normal eyes and cataract eyes which resulted in the highest accuracy of 99.74% at epoch 25. In 2022, research was conducted (Rarasmaya Indraswari, W. H. ,2022) to detect eye disease in fundus images using Convolutional Neural Network (CNN) using MobileNetV2 architecture with a dataset of fundus images consisting of 601 images divided into 2 normal and abnormal classes (cataract, glaucoma, and retinal disease). get an accuracy value of 72%, precision of 72%, recall of 72%, and F1-Score of 72%. In 2023 researchers detected (Ericco Andreas, 2023) cataract eye disease classification using CNN with inception V3 architecture with a dataset of 400 fundoscopy images. From this dataset there are 2 classes, namely cataract and normal. The best classification results by augmenting the training data so that 100% accuracy is obtained.

Previous researchers mostly detected one type of eye disease. However, research on eye diseases with various types of eye conditions using the CNN method is still small. One of them is research that uses the AlexNet architecture. However, these researchers did not include tests on the testing data. Therefore, the author wants to conduct a more complete research to perfect the previous researchers. Because AlexNet is old from 2015, therefore the author wants to use the more recent EfficientNet architecture, it is expected to get more optimal results because this architecture has higher accuracy than other CNN architectures, and EfficientNet is more efficient to help improve performance.

Method

Dataset
At this stage, a dataset is taken which is the input of this system that has been designed. The fundus image dataset was obtained from Kaggle (https://bit.ly/3xuAo42). Researchers chose this dataset because it had only been developed a few times so researchers wanted to try to develop it with the latest architecture. In addition, researchers want to help detect eye diseases accurately.

The datasets used were 300 original datasets and 3,600 augmented datasets in jpg format. The original dataset is grouped into 3 classes, namely "normal" (100 images), "cataract" (100 images) and "glaucoma" (100 images). As for the augmentation dataset, there are 3 classes with class divisions as "normal" (1200 images) "cataract" (1200 images) and "glaucoma" (1200 images). In this study, grayscale preprocessing and Thresholding used the augmented dataset.

Preprocessing
Preprocessing is an important step in processing data. Preprocessing is an image processing stage that aims to improve the quality of the image obtained. There are four preprocessing stages in this research, namely resize, normalization, grayscale and Thresholding. (1) Resize is a process of changing the large size of several images to become the same size to facilitate the image classification detection process (Zen, 2019). (2) Normalization is a process of changing the scale of image pixel values to have the same range of values. Pixel value is a numerical value that has a level of brightness or brightness of a pixel. The greater the pixel value, the brighter the image (Darnisa Azzahra Nasution, 2019). (3) Grayscale as an image with gray, black and white colors. grayscale is a collection of monochromatic shades, from pure white in the brightest corner to pure black in the

opposite corner (Jasman Pardede, 2017). (4) Thresholding is one of the image segmentation methods where the process is based on the difference in the degree of grayness of the image (Abdul Khair Tarigan, 2016).

Deep Learning

Deep Learning is one of the Machine Learning fields based on artificial neural networks (JST). In deep learning there are many hidden layers that are modeled in such a way as to provide accurate output. The performance of deep learning teaches computers to process data like the human brain. Deep Learning makes it possible to recognize and classify text, images, moving images, and audio (LM Azizah, 2018).

Convolutional Neural Network

One of the artificial neural networks, which is very common for analyzing visual images, is CNN (William, 2022). *Convolutional Neural Network* (CNN) is one of the artificial neural networks developed from Multilayer Perceptron (MLP). The number of dimensions in CNN is more than MLP. CNN is an architecture that can be trained and consists of several stages. In CNN, input arrays ranging from two dimensions to more. CNN functions to analyze visual images, detect and recognize objects, which are high-dimensional vectors involving many parameters in characterizing the network (Ibnu Dawan Ubaidullah, Y. N. ,2022). Figure 2.4 is an image of the convolutional Neural Network architecture.
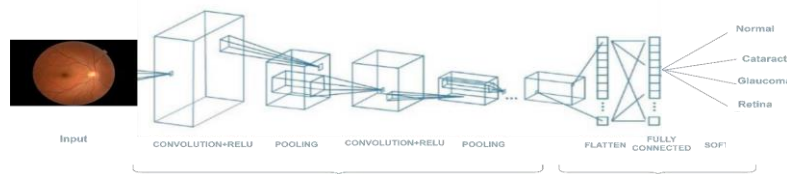

Figure 1 Archithecture CNN

The first stage in the CNN architecture is the convolution layer. The convolution layer stage is performed using a kernel of a certain size. The calculation of the number of kernels used is based on the number of features generated. The second stage performs the activation function, generally using the ReLU (Rectified Linear Unit) activation function. The third stage is the pooling layer, this process is repeated until a sufficient feature map is obtained to proceed to the fully connected neural network. Fully connected network is the output class. The high accuracy obtained by CNN occurs due to the complex feature extraction from image convolutions and the fusion of each image using data that is updated for each specific condition. The better the fusion, the higher the accuracy. (1) Convolution Layer. Convolution layer is a stage in the CNN architecture. At this stage, the convolution operation is performed from the previous convolution operation. Convolutional layer consists of neurons arranged in such a way that a filter with length and height (pixels) is formed (Ibnu Dawan Ubaidullah, Y. N. ,2022). (2) Aktivasi Rectified Linear Units (ReLU). Aktivitas Rectified Linear Units (ReLU) is an operation that aims to introduce nonlinearity and improve the representation of the model. With ReLU, all negative pixel values become zero. This helps to decompose the feature map closer to the corresponding image (Ibnu Dawan Ubaidullah, Y. N. ,2022). (3) Pooling Layer. Pooling is a process of reducing the matrix size by performing a pooling operation. The pooling process divides the input image into sections of a certain size. In each part, only 1 pixel will be taken that represents a certain static value, depending on the type of pooling used. If max pooling is used, the highest value in the window is taken. If average pooling is used, then the average value of the window is taken, and so on. The pooling process makes the image size smaller so that the computation required is lower (Ibnu Dawan Ubaidullah, Y. N. ,2022). (4) Fully Connected Layer. Fully Connected Layer is a layer where all activity neurons from the previous layer are connected to neurons in the next layer. This layer is basically used in MLP which aims to transform the dimensions so that it can be classified linearly (Ibnu Dawan Ubaidullah, Y. N. ,2022).

EfficientNet

There are 8 models in EfficientNet, namely B0-B7. EfficientNet has 7 blocks, where each block has various sub-blocks. The number of sub-blocks will increase when going from EfficientNet-B0 to EfficientNet-B7. EfficientNet uses depthwise and pointwise convolutions, splitting the original convolution into two stages to reduce computational cost with optimal accuracy.

In general, the EfficientNet architecture obtains high accuracy results and better performance compared to other CNN architectures, reduces parameter size and sorts FlOPS by magnitude (Wahyuni Rizky Perdani, R. M. ,2022). In this study using the EfficientNet-B0 model. Figure 2.5 shows the EfficientNet-B0 architecture.

Figure 2 < Archithecture *EfficientNet-B0*>

System Performance

For performance measurement parameters, the following confusion matrix parameters will be used: 1) Accuracy. Accuracy is a parameter used to determine the feasibility of the system in data classification. Accuracy is calculated by comparing the correctly classified data with the total number of classified data (3.1). 2) Precision. Precision is a parameter to determine the positive prediction data that has been classified correctly from all positive predicted data. Mathematically, the precision value can be calculated using equation (3.2). 3) Precision. Precision is a parameter to determine the positive prediction data that has been classified correctly from all positive predicted data. Mathematically, the precision value can be calculated using equation (3.2) 4) Recall. Recall is a parameter used to determine positive predicted data that has been correctly classified from all actual positive data. Mathematically, the recall value can be calculated using equation (3.3). 5) F-1 score. F1-Score is a parameter used to make low precision values and high recall values or vice versa become balanced. The F1-Score equation can be seen in equation (3.4). 6) Loss Function. Loss is a parameter to determine the error rate of the system in data classification.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \qquad (3.1)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \qquad (3.2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \qquad (3.3)$$

$$F1 - Score = \frac{2 \ x \ Recall \ x \ Presisi}{Recall \ + \ Presisi} \times 100\% \qquad (3.4)$$

Results and Discussions

In this study, scenario testing was carried out with several different CNN parameters and different preprocessing techniques. The first scenario is carried out by comparing several Optimizer parameters to get the best Optimizer. Optimizer parameters tested are Adam, NAdam, SGD, and RMSprop. Each Optimizer parameter will be tested with the original dataset, augmented dataset, augmented dataset using grayscale preprocessing, and augmented dataset using Thresholding preprocessing with a learning rate parameter of 0.0001 and a batch size of 32.

The second scenario is carried out by comparing the learning rate parameter based on the best Optimizer parameter that has been produced in the first scenario and batch size 32. The learning rate parameter values used include 0.01, 0.001, 0.0001, 0.00001 and 0.000001. Each Optimizer parameter will be tested with the original data set, the augmented data set, the augmented data set using grayscale pre-processing, and the augmented data set using Thresholding pre-processing.

The third scenario is carried out by comparing the batch size parameters based on the best optimizing parameters that have been produced in the first scenario and the best learning rate from the second scenario. The batch size parameter values used include 32, 64, and 128. Each Optimizer parameter will be tested with the original dataset, augmented dataset, augmented dataset using grayscale preprocessing, and augmented dataset using Thresholding preprocessing.

All scenarios will be compared with the results from all original, augmented, and pre-processed datasets based on the best Optimizer parameters, batch size, and learning rate that have been generated in the first, second, and third scenarios. Thus, the EfficientNetB0 architecture CNN model with the best parameters is obtained. Table 4.1 shows the best results from each dataset used.

Table 1 Best Model Hyperparamter

Hyperparameter	Original Dataset	Augmentation Dataset	Grayscale Dataset	Thresholding Dataset
Optimizer	Adam	Nadam	Adam	Nadam
Learning rate	0.01	0.0001	0.00001	0.00001
Batch Size	32	32	32	32
Akurasi	73.33	78.44	79.22	71.84
Loss	1.12	1.11	0.72	0.92

The table 1 shows the best parameters of each dataset, namely the best optimizer, the best learning rate, and the best batch size. The best results were obtained using the grayscale dataset using the Adam optimizer, learning rate 0.0001, batch size 32. Table 4.2 shows the best model simulation results using the original dataset.

Table 1 Simulation Results of Best Model for Grayscale Dataset

Akurasi	**79.22**
Presisi	80.3
Recall	79.22
F1-Score	78.87

Commented [MOU1]: Bagian pembahasan perlu ditambahkan hasil penelitian terdahulu ataupun referensi yang relevan dengan hasil penelitian yang ditemukan.

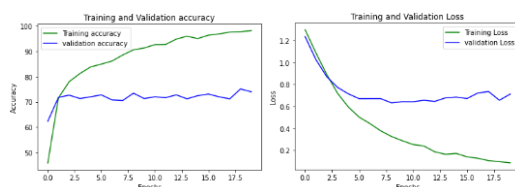**Commented [MOU2]:** Sebutkan tabel berapa (Tabel 1)

Figure 3 Training, validation, and loss graphs best model using grayscale dataset

Table 2 shows the results of accuracy, precision, recall, and f1-score obtained using the best model in this study. The accuracy value is 72.22%, precision value is 80.3%, recall value is 79.22%, and f1-score is 78.87%. Figure 1 shows the training, validation, and loss graphs for the best model using the grayscale dataset. The graph shows the occurrence of overfitting. In the training graph, it can be seen that it is very stable, although the validation graph has ups and downs, but the accuracy is quite high. The loss graph itself is very good because the loss value is already below 1.
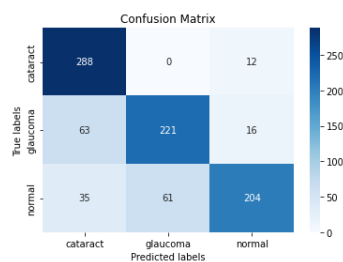

Figure 4 *Confusion Matrix*

Confusion matrix testing the best model using grayscale dataset, As shown in the figure, the system is able to classify cataract as many as 288 images, Glaucoma as many as 221 images, and normal as many as 204 images.

This study has improved previous research that classifies eye diseases using convolutional neural network (CNN) using the AlexNet method including, by performing augmentation, balancing the amount of data for each class and including testing results that were not previously done by previous researchers. The validation results obtained by previous researchers were 60% while in this study higher validation results were obtained, namely 74%. For the training results obtained by previous researchers amounted to 98.37, not much different from this study of 98.16%.

Conclusions

This research has succeeded in designing a system to classify eye disease into three classes, namely normal, cataract, and glaucoma using a fundus image dataset with CNN-based EfficientNet-B0 architecture. The image is processed first through resizing, normalization, grayscale, and thresholding preprocessing processes. Based on the test scenarios that have been carried out, the parameters that affect system performance are the Optimizer, the learning rate value, and the batch size value. Dataset variations were tested by optimizing Adam, Nadam, SGD, and RMSprop, then testing the best learning rate using learning rate values of 0.01, 0.001, 0.0001, 0.00001 and 0.000001, and testing the best batch size using batch size values of 32, 64, and 128.

Based on the test results and analysis of various test scenarios, the designed system is able to classify fundus images based on their class using the original dataset, Augmentation dataset, Augmentation dataset with grayscale preprocessing, and Augmentation dataset with Thresholding preprocessing randomly with fairly good accuracy. . The best performance results were obtained in augmented datasets with grayscale preprocessing using the Adam Optimizer, learning rate 0.00001, batch size 32, and iterations of 20 epochs. The system can detect and classify three classes of eye disease in the augmented dataset with grayscale

preprocessing with a performance accuracy value of 79.22%, precision value of 80.3%, recall value of 79.22%, F1-Score 78.87%.

References

G. P. Kumasela, J. S. (2013). Hubungan Waktu Penggunaan Laptop Dengan Keluhan Penglihatan Pada Mahasiswa Fakultas Kedokteran Universitas Sam Ratulangi. *Jurnal e-Biomedik (eBM), Volume 1, Nomor 1*, 291-299.

Bourne R, Steinmetz J, Flaxman S, et al., Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. Lancet Glob Health. 2020. Accessed via the IAPB Vision Atlas (https://www.iapb.org/learn/vision-atlas)

WHO. (2023, March 29). Blindness and vision impairment. Retrieved from https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

M. J. Burton et al., "The Lancet Global Health Commission on Global Eye Health: vision beyond 2020," Lancet Glob. Heal., vol. 9, no. 4, pp. e489–e551, 2021, doi: 10.1016/S2214-109X(20)30488-5.

Vania Annisa Queentinela, Y. T. (2021). KlasifikasiI Penyakit Diabetic Retinopathy Pada Citra Fundus Berbasis Deep Learning. *Applied Business and Engineering*, 1007-1018.

Devin D. Mackay, et al (2015). The Demise Of Direct Ophthalmoscopy. *American Academy of Neurology*, 150-157.

Geza Jeremia Bu'ulölö, A. J. (2021). Identification of Cataract Eye Disease Using Convolutional Neural Network. *Jurnal Teknik Informatika vol. 16 no. 4* , 375-382.

Fani Nurona Cahya, N. H. (2021). Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Nerwork . SISTEMASI Volume 10, Nomor 3, 618-626.

Syafiq Hilmi Abdullah, R. M. (2022). Klasifikasi Diabetic Retinopathy Berbasis Pengolahan Citra. JESCE (Journal of Electrical and System Control Engineering), 5 (2), 84-9

Diki Hananta Firdaus, B. I. (2022). Klasifikasi Penyakit Katarak Pada Mata Menggunakan Metode CNN Berbasis Web. Jurnal Kecerdasan Buatan dan Teknologi Informasi (JKBTI), 18-26.

Rarasmaya Indraswari, W. H. (2022). Deteksi Penyakit Mata Pada Citra Fundus Menggunakan CNN . Techno.COM, Vol. 21, No. 2, 378-389.

Ericco Andreas, W. W. (2023). Klasifikasi Penyakit Mata Katarak Menggunakan Convolutional Neural Network Dengan Arsitektur Inception V3. MDP Student Conference, 107-113.

Zen, M. (2019). Perbandingan Metode Dimensi Fraktal Dan Jaringan Syaraf Tiruan Backpropagation Dalam Sistem Identifikasi Sidik Jari Pada Citra Digital. JITEKH, Vol 7, No 2, 42-50.

Darnisa Azzahra Nasution, H. H. (2019). Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN. CESS (Journal of Computer Engineering System and Science) Vol. 4 No. 1, 78-82.

Jasman Pardede, D. B. (2017). Implementasi "Principal Component Analysis -Scale Invariant Feature Transform" pada Content Based Image Retrieval. Jurnal Teknik Informatika dan Sistem Informasi, Volume 3, Nomor 3, 565-574.

Abdul Khair Tarigan, S. D. (2016). Aplikasi Pembelajaran Citra Dengan Menggunakan Metode Computer Assisted Instruction (CAI). Jurnal Riset Komputer (JURIKOM), Vol. 3 No. 4, 1-4.

LM Azizah, S. U. (2018). Deteksi Kecacatan Permukaan Buah Manggis Menggunakan Metode Deep Learning dengan Konvolusi Multilayer. SEMESTA TEKNIKA Vol. 21, No. 2, 230-236

William, C. L. (2022). Klasifikasi Penyakit Mata Menggunakan CNN. Jurnal Ilmu Komputer dan Sistem Informasi Vol. 10 No. 1, 1-4.

Ibnu Dawan Ubaidullah, Y. N. (2022). Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma. ELKOMIKA Vol. 10 No. 3, 728 - 741.

Wahyuni Rizky Perdani, R. M. (2022). Deep Learning untuk Klasifikasi Glaukoma dengan menggunakan Arsitektur EfficientNet . Jurnal Teknik Energi Elektrik Vol. 10 No. 2, 322 - 333.

Commented [MOU3]: Tambahkan referensi minimal 20.