Cataract Classification Based on Fundus Images Using Convolutional Neural Network

Article *in* JOIV International Journal on Informatics Visualization · March 2022

DOI: 10.30630/joiv.6.1.856

Rita Magdalena
Telkom University

Sofia Saidah
Telkom University

Cataract Classification Based on Fundus Images Using Convolutional Neural Network

Richard Bina Jadi Simanjuntak[a], Yunendah Fu'adah [b,*], Rita Magdalena[b], Sofia Saidah[c], Abel Bima Wiratama[d], Ibnu Da'wan Salim Ubaidah[e]

[a, b, c, d, e] *School of Electrical Engineering, Telkom University, Jl. Telekomunikasi No.1 Terusan Buah Batu-Bojongsoang, Kec.Dayeuhkolot, Bandung, 40257, Indonesia*

Corresponding author: richardsimanjuntak@student.telkomuniversity.ac.id

***Abstract*— A cataract is a disease that attacks the eye's lens and makes it difficult to see. Cataracts can occur due to hydration of the lens (addition of fluid) or denaturation of proteins in the lens. Cataracts that are not treated properly can lead to blindness. Therefore, early detection needs to be done to provide appropriate treatment according to the level of cataracts experienced. In this study, a comparison of cataract classification based on fundus images was carried out using the Convolutional Neural Network. The dataset used comes from primary datasets with a total of 1600 datasets. The architecture used is GoogleNet, MobileNet, and ResNet. Comparison of the three architectures using Adam Optimizer with a Learning Rate 0.001. The best results were obtained from the GoogLeNet architecture with 90% accuracy, followed by the MobiLeNet architecture at 87% and the ResNet architecture at 85%. Researchers also make comparisons with previous research. Most of the previous studies only used two to three class categories. In this study, the system was improved by increasing the class category into four categories: Normal, Immature, Mature, and Hypermature. In addition, the accuracy obtained is also quite good compared to previous studies using manual feature extraction. This research is expected to help medical personnel in detecting cataracts better. In the future, researchers want to improve the accuracy of the cataract detection system to detect and classify cataracts more accurately.**

***Keywords*— Cataract; Convolutional Neural Network (CNN); GoogLeNet; MobileNet; ResNet.**

I. INTRODUCTION

A cataract is a disease that attacks the eye's lens and makes it difficult for sufferers to see. Cataracts can occur due to the lens's hydration (fluid increase) or denaturation of proteins in the lens. In general, cataract is a disease that attacks the elderly, but congenital abnormalities and eye diseases can also cause cataracts. Some eye diseases that trigger cataracts are glaucoma, ablation, uveitis, retinitis pigmentosa, and other intraocular disorders [1]. Based on the stage, senile cataract consists of 6 stages: incipient cataract, intumescent cataract, immature cataract, mature cataract, hypermature cataract, and Morgagni cataract.

Around 3.38% of the world's population or 253 million people have a visual impairment, of which 36 million people are blind, and 217 million people have moderate to severe visual impairment. The five countries with the most visually impaired populations are China, India, Pakistan, Indonesia, and the United States [2]. The most common causes of visual disturbances were uncorrected refractive errors at 48.99%, followed by cataracts at 25.81%, and Age-Related Macular Degeneration at 4.1%. The most common cause of blindness is cataracts with a percentage of 34.47%, followed by uncorrected refraction of 20.26% and glaucoma at 8.30% [2]. Based on the latest national data sourced from the Rapid Assessment of Avoidable Blindness in 2014-2016, the Data and Information Center of the Ministry of Health of the Republic of Indonesia stated that the leading cause of blindness and visual impairment in the population aged over 50 years in Indonesia is cataracts that are not operated on with a proportion of 77, 7%. In men, the leading causes of blindness are 71.7%, and women are 81.0% [3].

The impact given by cataracts can affect the productivity and mobility of sufferers, resulting in a decrease in people's quality of life [4]. Cataracts can be anticipated by early detection when the eye begins to experience disturbances. Currently, there are several methods used by ophthalmologists to diagnose cataracts in the form of visual acuity tests, slit-lamp tests, retina exams, and applanation tonometry. This is also not sufficient for the early detection of cataracts due to the duration of time required for detection and the limited stage of cataracts that can be detected. Therefore, a cataract detection system based on image processing has been developed that can perform early detection of cataracts in a quick but accurate manner as a tool for detecting cataracts.

Several studies based on fundus image processing have been developed for cataract detection. In 2018, Hadeer R. M. Tawfik, Rania A. K. Birry, and Amani A. Saad conducted research related to the early detection of cataracts using Combined 2D Log Gabor/Discrete Wavelet Transform with ANN and SVM. This study uses a dataset from the University of Aiwa eye rounds' Atlas of cataracts and Duane's Clinical Ophthalmology with three grade levels, namely normal, early-stage and advanced stage. The accuracy obtained when using SVM is 96.8%, while ANN is 92.3% [5].

In 2018, Riski Wahyu Hutabri et al. conducted a study entitled “Design of a Cataract Detection System using Principal Component Analysis (PCA) and K-nearest neighbor (K-NN) Methods” using 74 data with three categories: mature (20 images), immature (34 photos), and normal (20 images) with the highest accuracy of 67.57% [6].

In 2019, Yunendah Nur Fu'adah et al. utilized the optimization of the Gray Level Co-Occurrence Matrix (GLCM) method to extract information from the input in the form of eye images and classify eye images with K-Nearest Neighbor (K-NN) into three classes, namely normal, immature, and mature. System testing is carried out at the classification stage using K-NN by analyzing the influence of the Euclidean, Minkowski, Chebyshev, and City Block distance calculation methods. The effect of Minkowski and Euclidean distance produces the best accuracy is 93.33% [4].

In 2019, Ref. [7] conducted research titled “Mobile Application Based Cataract Detection System ."In this study, we compared the results of cataract detection with the KNN, SVM, and Naïve Bayes methods. In this study, cataract detection was limited to 2 classes, namely normal and cataract. The KNN classification method has the highest accuracy rate of 83.07%, followed by the SVM classification method of 75.2% and the Naïve Bayes classification method of 76.64%

In 2020, a cataract classification study using fundus images was also carried out by Mas Andam Syarifah et al utilizing the optimized Convolutional Neural Network and the Lookahead optimizer. This study uses the AlexNet architecture and classifies it into normal and cataract. With the highest accuracy of 97.50% [8].

In 2021, Indra Weni, et al. conducting research on cataract detection based on image features. Cataract detection was performed using Convolutional Neural Networks. This study utilizes the GoogleNet architecture and divides it into two classes, namely normal and cataract, with the highest accuracy of 88% [9].

Several previous studies using traditional techniques to classify glaucoma have shown satisfactory results. Traditional techniques for classifying systems usually involve feature extraction and classification [10]. However, the CNN network automatically extracts the relevant features and classifies them into different classes. The advantage of the CNN-based method is that it does not need to do feature extraction and classification explicitly. The author will compare several CNN architectures in this study, namely GoogLeNet, ResNet, and MobileNet. The fundus image of the eye will be classified into four classes, namely normal, immature, mature, and hypermature.

II. The Materials and Method

The cataract classification system uses the Convolutional Network method with the GoogLeNet model.
This system divides fundus images into four classes: normal, immature, mature, and hypermature. The data used comes from primary data collected from several hospitals in *.jpg format.

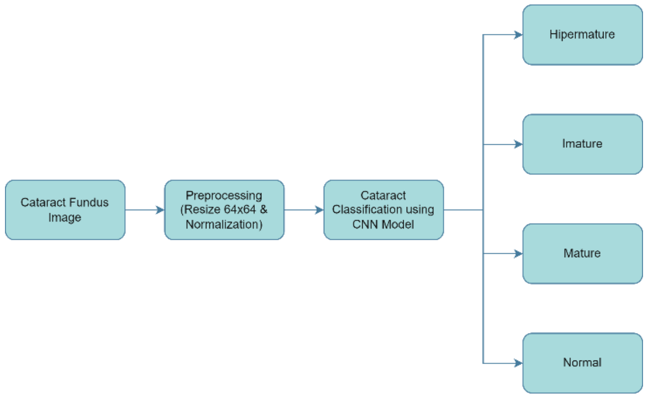

Fig 1. Cataract Classification Block Diagram

In general, the system block diagram is designed to get the results of cataract image classification. The system input used comes from the primary data of fundus images in images of cataracts. After that, pre-processing is done by resizing the image. Then model training is carried out by utilizing training data to the system in order to obtain maximum classification results. Classification is the output of model training in object recognition according to the specified class.

A. Dataset

This study utilizes primary datasets collected from several related hospitals in North Sumatra. The number of datasets used is 399 images with four classes: 181 hypermature images, 73 normal images, 74 immature images, and 72 mature images. An augmentation process is carried out To balance the size of each image. The process is random flip, rotation, zoom, and shift.

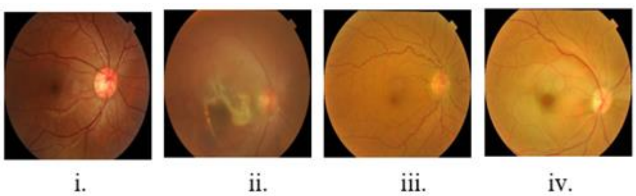

Fig 2. i. Hypermature; ii. Immature; iii. Mature; iv. Normal

B. Convolutional Neural Network (CNN)

Convolutional Neural Network is an artificial neural network that is used to perform image recognition and processing [11]. CNN imitates the way nerve cells communicate with interconnected neurons. The CNN concept is similar to MLP, but each neuron is represented in two dimensions in CNN, while in MLP, each neuron is one-dimensional [11].

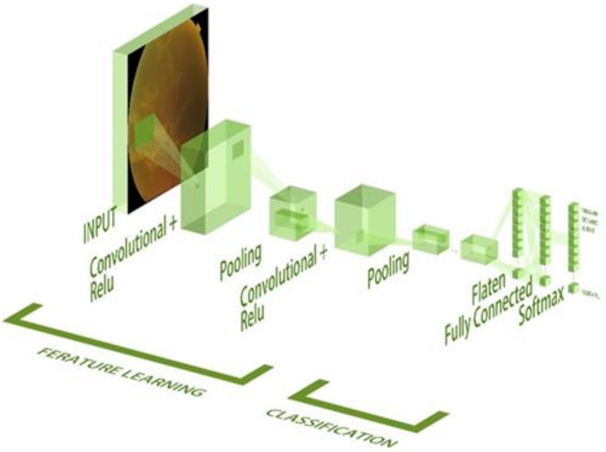

Fig 3. Convolutional Neural Network Architecture

In its application, CNN consists of two layers: the feature extraction layer, which consists of neurons connected to the local region. The convolutional layer is the first type of layer, and the pooling layer is the second. An activation function that alternates for each layer type is applied at each layer.

1. *Convolutional Layer:* is the core building block of CNN [12]. The Convolution Layer applies a convolution operation to the previous layer's output. The main process that underpins a CNN is this layer. Convolution is a mathematical term that refers to the process of applying a function to the output of another function repeatedly.

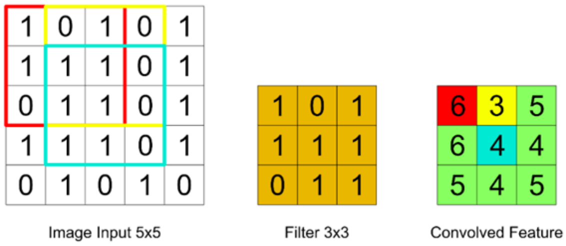

Fig 4. Convolutional Layer

2. *Pooling Layer:* This layer uses a function with a Feature Map as input and processes it with various statistical operations based on the nearest pixel value [13]. The Pooling layer is usually inserted regularly after several convolution layers in the CNN model. Pooling layers added between successive convolution layers in the CNN model architecture can reduce the size of the output volume on the Feature Map over time, lowering the number of parameters and calculations in the network and decreasing overfitting.

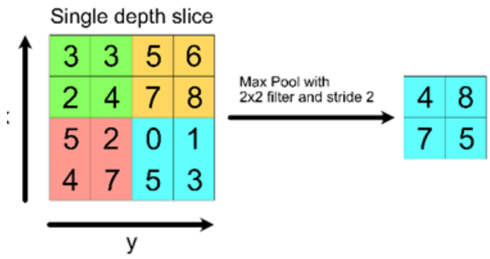

Fig 5. Max-Pooling Layer

3. *Rectified Linear Units (ReLU) Activation* is an activation function that can swiftly analyze vast amounts of input between the convolutional and pooling layers [14].

$$f\ (x) = \max(0, x) \qquad (1)$$

This function performs thresholding with a zero value of the pixel value in the input image. All negative values from the convolution process make the negative value equal to zero.

4. *Flatten:* The Feature Learning process has an output in the form of a multidimensional array, while the input to the fully connected layer must be a vector. Flatten serves to reshape multidimensional arrays into vectors [15]. Flatten is necessary to use this value as input to the Fully connected layer.

5. *Fully Connected Layer:* This layer is commonly employed in MLP applications, and its goal is to execute modifications on the data's dimensions so that it may be classified linearly. Before entering a fully connected layer, each neuron in the convolution layer must be turned into one-dimensional data [11]. A fully connected layer removes the spatial information from the data and is not reversible. Only at the end of the network, the fully connected layer can be implemented.

6. *Softmax Activation:* is a general logistic function with a vector output probability $p \in R^n$ with an input vector $x \in R^n$ with a softmax function at the end of the architecture.

$$p = \begin{pmatrix} p1 \\ \vdots \\ p2 \end{pmatrix} \text{ where } p_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x\,j}} \qquad (2)$$

The softmax activation function is used to get the final classification result. The activation function usually produces a value interpreted as an abnormal probability [13].

$$y_{ijk} = \frac{e^{x^{ijk}}}{\sum_{t=1}^{D} e^{x^{ijk}}} \qquad (3)$$

Based on equation 3, it can be seen that y_{ijk} is a vector containing values 0 and 1 while x is a vector generated by the fully connected layer.

C. GoogLeNet

GoogLeNet is a model and architecture based on a modified CNN, with the main builder series being the Inception Module commonly called the Inception Network. GoogLeNet has a total of 144 layers. The input layer on GoogLeNet is an image measuring 224 x 224 x 3 [16]. The advantage of GoogLeNet is that it has inception modules that consist of small convolutions and are designed to reduce the number of parameters without reducing network performance.

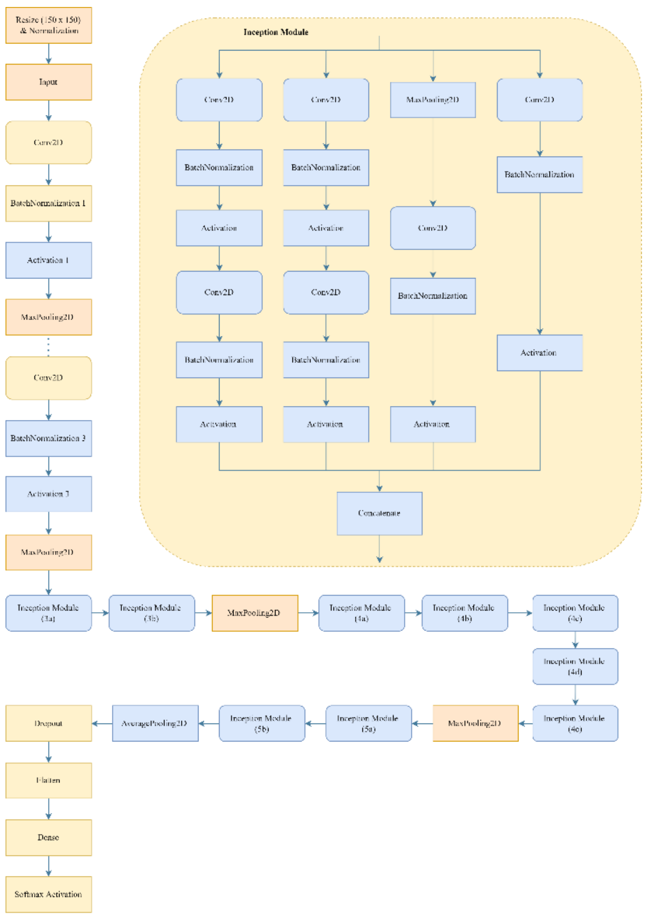

Fig 6. GoogLeNet Architecture

D. MobileNet

MobileNet is one of the CNN architectures that can overcome the need for excessive computing resources. MobileNet is designed to maximize accuracy results with limited resources. Therefore MobileNet has specifications, latency, and low energy consumption, so it is very suitable for mobile device applications [17]. MobileNet is based on Depthwise Separable Convolution, a substitute for standard convolution with factorization that divides convolution into two separate layers, namely depthwise convolution and pointwise convolution. By applying a single convolutional filter per input channel, the first layer or depthwise convolution filters is built [18]. The second layer or pointwise convolution is responsible for The second layer, also known as pointwise convolution, is in charge of creating new features by computing linear combinations of input channels using 1x1 convolution. MobileNet architecture consists of depthwise separable convolution blocks arranged repeatedly with one fully connected layer followed by a softmax layer.

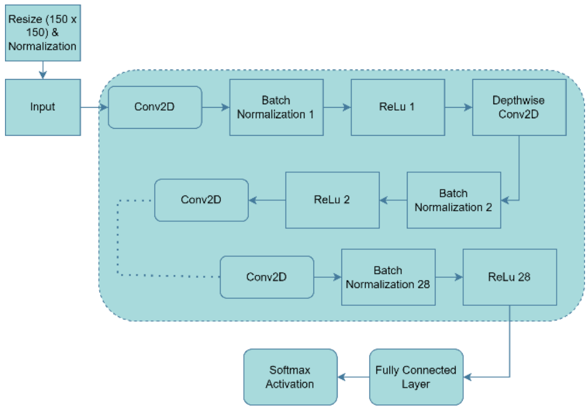

Fig 7. MobileNet Architecture

E. ResNet

ResNet Architecture is a 50-layer Residual Network architecture. As shown by the name, this network uses residual learning. Instead of attempting to learn certain features, the network in Residual Learning learns some residual. Residuals can be thought of as a reduced feature learned from the input of a layer. ResNet uses a shortcut connection to directly connect the input from the nth layer to some next layer $(n + x)$. It has been shown that training this type of network is easier than training deep convolutional neural networks, which can handle simple problems with decreasing accuracy [19].

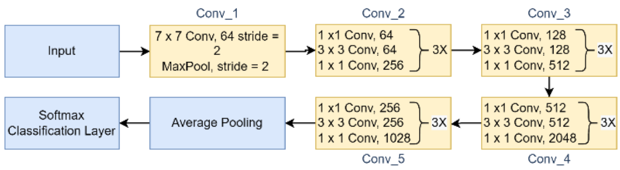

Fig 8. ResNet Architecture

III. System Performance

This study uses four parameters for measuring system performance: accuracy, recall, precision, and f1-score. The measurement of the system's performance is shown in 3, 4, 5, and 6 [20].

$$Accuracy=TP+TNTP+TN+FP+FN \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - Score = 2 \times \frac{recall \times precision}{recall + precision} \quad (7)$$

TTP (Total True Positive) is the amount of data for which the prediction model is positive. The actual data is positive, so it can be concluded that it has been classified correctly. TFP (Total False Positive) is the amount of negative data, and the prediction model is positive. TFN (Total False Negative) is the amount of actual positive data, and the prediction model is negative. TTN (Total True Negative) is the negative amount of actual data, and the prediction model is negative. The conclusion is that the classification is correct.

IV. RESULT AND DISCUSSION

The cataract fundus image dataset consists of 1600 immature, mature, normal, and hypermature fundus images. The dataset used for training consists of 1280 data, and for test consists of 320 data. The validation consists of twenty percent of the training data, which is 256 data. Training data used consists of 1024 data. This study compares several CNN models, such as GoogLeNet, MobileNet, and ResNet. The Optimizer used consistently utilizes the Adam Optimizer.

Table 1. The comparison of different CNN architectures on Train, Val, Test

Model	Train	Val	Test
GoogLeNet	0.92	0.91	0.90
MobiLeNet	0.95	0.86	0.87
ResNet	0.87	0.91	0.85

Based on the table above, we can see that the best accuracy when using a learning rate of 0.001 with the Adam optimizer is obtained from the GoogLeNet architectural model with a training accuracy of 0.92. This model has an accuracy of training data and data validation of 0.91 and 0.90, respectively. Table 1 shows the approach using the GoogLeNet architectural model has the best accuracy of 0.90. This is because GoogLeNet has the best efficiency compared to ResNet and MobileNet. The efficiency of GoogLeNet is affected by the complexity of the GoogLeNet model, which has branched networks.

Table 2. The comparison of different CNN architectures on Precision, Recall, and F1-Score

Model	Class	Precision	Recall	F1-Score
GoogLeNet	Hipermature	0.97	0.94	0.96
	Immature	0.83	0.87	0.85
	Mature	0.86	0.84	0.85
	Normal	0.99	1.00	0.99
MobiLeNet	Hipermature	0.86	0.89	0.87
	Immature	0.76	0.89	0.82
	Mature	0.86	0.73	0.79
	Normal	0.98	1.00	0.93
ResNet	Hipermature	0.94	0.93	0.93
	Immature	0.84	0.91	0.87
	Mature	0.87	0.83	0.85
	Normal	1.00	0.98	0.99

Table 2 shows the performance results of recall, precision, Furthermore, f1-score each model.

In the application of manual extraction, previous studies have produced satisfactory accuracy. However, the use of CNN can still be improved again. This study developed cataract classification using the CNN method utilizing the GoogLeNet architecture.

Table 3. Summary of Cataract Classification Techniques Used in This Study

Authors	Classifier	Architectures/ Features	Class	Accuracy
Hadeer R. M. Tawfik et al.	SVM	2D Log Gabor/DWT	Normal, early-stage, and advanced stage	96.8%
	ANN			92.3%
Riski Wahyu Hutabri et al.	KNN	PCA	Normal, immature, and mature	67.57%
		GLCM		
Yunendah Nur Fu'adah et al.	KNN	(Euclidean, Minkowski, Chebyshev, and City Block distance)	Normal, immature, and mature	93.33%
Vaibhav Agarwal et al.	KNN	EMOBPSO-GLS and EMOBPSO	Normal and cataract	83.07%
	SVM			75.2%
	Naïve Bayes			76.64%
Mas Andam Syarifah et al.	CNN	Alexnet	Normal and cataract	97.50%
Indra Weni et al.	CNN	GoogLeNet	Normal and cataract	88%
This Study	CNN	GoogLeNet	Normal, Immature, mature and Hypermature	90%

Table 3 briefly describes past techniques developed to classify cataracts automatically. Traditional techniques for classifying systems usually involve feature extraction and classification. However, the CNN network automatically extracts the relevant features and classifies them into different classes. The advantage of the CNN-based method is that there is no need to perform explicit feature extraction and classification.

In this study, a fundus image-based cataract classification system was developed using the GoogLeNet architecture. The improvement was made to the class category into four classes, namely Normal, Immature, Mature, and Hypermature. The proposed approach's advantages are that it can pre-screen fundus images to help medical staff classify cataracts and primary datasets that are owned have various amounts so that the system can classify cataracts into four classes, namely Normal, Immature, Mature, and Hypermature.

V. CONCLUSION

This paper discusses the classification of cataracts based on fundus images by comparing several CNN architectures, namely GoogLeNet, MobiLeNet, and ResNet. Comparison of the three architectures consistently uses the Adam Optimizer with a Learning Rate of 0.001. The best and most stable results were obtained from the GoogLeNet architecture with an accuracy of 0.90. In addition, the study also compared with previous studies by developing detectable cataract classes into four classes, namely Normal, Immature, Mature, and Hypermature, with fairly good accuracy of 90%. This research is expected to help medical staff carry out early detection of cataracts to carry out preventive actions appropriately. In the future, researchers want to improve the accuracy of the cataract detection system to be used to detect and classify cataracts more accurately.

ACKNOWLEDGMENT

The author would like to thank the hospital profusely for the fundus image dataset that can support this research.

References

[1] Ilyas S, *Penglihatan Turun Perlahan Tanpa Mata Merah*, 3rd ed. Jakarta: Balai Penerbit FKUI, 2007.

[2] P. Ackland, S. Resnikoff, and R. Bourne, "World blindness and visual impairment: despite many successes, the problem is growing," *Community eye health*, vol. 30, no. 100, pp. 71–73, 2017, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29483748

[3] R. Kemenkes, "Infodatin Situasi Gangguan Penglihatan," *Kementrian Kesehatan RI Pusat Data dan Informasi*, p. 11, 2018, [Online]. Available: https://pusdatin.kemkes.go.id/download.php?file=download/pusdatin/infodatin/infodatin-Gangguan-penglihatan-2018.pdf

[4] Y. N. Fuadah, R. Magdalena, S. Palondongan, and N. Kumalasari, "Optimasi K-Nearest Neighbor Untuk Sistem Klasifikasi Kondisi Katarak," *TEKTRIKA - Jurnal Penelitian dan Pengembangan Telekomunikasi, Kendali, Komputer, Elektrik, dan Elektronika*, vol. 4, no. 1, p. 16, 2019, doi: 10.25124/tektrika.v4i1.1832.

[5] R. A. K. B. A. A. S. Hadeer R. M. Tawfik, "Early Recognition and Grading of Cataract Using a Combined Log Gabor/Discrete Wavelet Transform with ANN and SVM," *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering*, vol. 12, pp. 1038–1043, 2018.

[6] R. W. Hutabri, R. Magdalena, and R. Y. N. Fu'adah, "Perancangan Sistem Deteksi Katarak Menggunakan Metode Principal Cmponent Analysis (PCA) dan K-Nearest Neighbor (K-NN)," *Seminar Nasional Inovasi dan Aplikasi Teknologi*, pp. 321–327, 2018.

[7] V. Agarwal, V. Gupta, V. Vashisht, K. Sharma, and N. Sharma, *Mobile Application Based Cataract Detection System*. 2019. doi: 10.1109/ICOEI.2019.8862774.

[8] M. A. Syarifah, A. Bustamam, and P. P. Tampubolon, "Cataract classification based on fundus image using an optimized convolution neural network with lookahead optimizer," *AIP Conference Proceedings*, vol. 2296, no. 1, p. 020034, Nov. 2020, doi: 10.1063/5.0030744.

[9] I. Weni, P. E. P. Utomo, B. F. Hutabarat, and M. Alfalah, "Detection of Cataract Based on Image Features Using Convolutional Neural Networks," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 1, p. 75, 2021, doi: 10.22146/ijccs.61882.

[10] S. Maheshwari, V. Kanhangad, and R. B. Pachori, "CNN-based approach for glaucoma diagnosis using transfer learning and LBP-based data augmentation," Feb. 2020, [Online]. Available: http://arxiv.org/abs/2002.08013

[11] W. S. Eka Putra, "Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101," *Jurnal Teknik ITS*, vol. 5, no. 1, 2016, doi: 10.12962/j23373539.v5i1.15696.

[12] J. Cao *et al.*, "DO-Conv: Depthwise Over-parameterized Convolutional Layer," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.12030

[13] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional Neural Networks for MATLAB," in *Proceedings of the 23rd ACM International Conference on Multimedia*, 2015, pp. 689–692. doi: 10.1145/2733373.2807412.

[14] R. Sulistyo Budi, R. Patmasari, and S. Saidah, "KLASIFIKASI CUACA MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN) WEATHER CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK (CNN) METHOD."

[15] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.

[16] C. Szegedy *et al.*, "Going Deeper with Convolutions," Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.4842

[17] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.04861

[18] Q. Wen, Z. Luo, R. Chen, Y. Yang, and G. Li, "Deep learning approaches on defect detection in high resolution aerial images of insulators," *Sensors (Switzerland)*, vol. 21, no. 4, pp. 1–26, Feb. 2021, doi: 10.3390/s21041033.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-December, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[20] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, "CataractNet: An automated cataract detection system using deep learning for fundus images," *IEEE Access*, vol. 9, pp. 128799–128808, 2021, doi: 10.1109/ACCESS.2021.3112938.