OPEN ACCESS
ISSN 2356-5462
http://socj.telkomuniversity.ac.id/ijoict/

Intl. Journal on ICT
Vol. 9, No. 2, Dec 2023. pp. 66-74
doi: doi.org/10.21108/ijoict.v9i2.839

Convolutional Neural Network Implementation with AlexNet Architecture for 4 Classes Face Recognition

Dyah Anggun Sartika [1], Denny Hardiyanto [2*], Imam Junaedi [3], Sukamto [4]

[1,3,4]*Department of Computer Control Engineering,* [1,3,4]*Politeknik Negeri Madiun*
Serayu street No.84, Madiun, East Java, Indonesia

[2*]*Department of Electrical Engineering Education, Universitas PGRI Madiun*
Setia Budi street No.85, Madiun, East Java, Indonesia

[2*] denny.hardiyanto@unipma.ac.id

Abstract

In today's digital era, the process of facial recognition has a very big role. Face recognition has benefits for authentication and identification processes. The development of facial recognition research continues to be carried out with the aim of being able to get the right algorithm, more accurate, faster processing, to be able to recognize faces from various angles. In this study, a performance test was performed on the Convolutional Neural Network (CNN) algorithm with the AlexNet architecture, which is one of the deep learning algorithm developments for facial recognition. The research contribution is to provide the results of the AlexNet CNN architecture work in facial classification and recognition. AlexNet has 8 convolution layers so that it will not leave even the slightest feature of the object. The number of datasets used is 400 image data which is divided into 360 training image data and 40 test image data. The 400 data come from 4 classes of facial images that have been labeled with names and each classes have 100 images. The training process produces an accuracy of 100% and the testing process produces an accuracy of 95%.

Keywords: Convolutional Neural Network, Deep Learning, AlexNet, Face Recognition

I. INTRODUCTION

AT the moment, the process of facial recognition has a very big role. Face recognition has benefits for authentication and identification processes. In the authentication process, facial recognition is used by users to access devices/systems such as opening smartphones, accessing systems on laptops, and other benefits. Meanwhile, face identification is a combination of face detection and face recognition processes. Viola and Jones were the initial researchers who initiated the face detection algorithm and published their scientific work [1], [2]. There are many applications of this facial recognition process, including: 1) security systems (for example authentication when opening a smartphone screen), 2) security systems (for example to search for missing persons and for accessing control/control rooms), 3) recognition systems (for example on social media), 4) attendance system (for example in automatic attendance of employees or students at school by [3], [4], [5],[6]), 5) forensic systems (for example identification of corpses), and many other applications. Various face recognition problems include faces that are difficult to recognize when there is insufficient light, faces that are difficult to recognize when using accessories such as headscarves, and facial recognition angles. To overcome

Received on 10 Aug 2023. Revised on 23 Aug 2023. Accepted and Published on 14 Dec 2023.

these various facial recognition problems and to improve accuracy in facial recognition, the CNN algorithm is applied. The development of facial recognition research continues to date to be able to get the right algorithm, more accurate, faster processing, to be able to recognize faces deeper at various angles.

Convolutional Neural Networks or known as CNN is one of the Multi-Layer Neural Network learning algorithms based on Deep Learning. The CNN algorithm is widely used in various applications because it has high accuracy and is an efficient algorithm to use. The advantage of the CNN Algorithm is that it can imitate image recognition systems such as the human visual cortex. This model can reduce several independent parameters and can handle deformation problems in the input image. CNN is very useful in image classification and recognition. CNN has various types of constituent architectures such as AlexNet, Visual Geometry Group (VGG) 16, VGG19, Residual Network (ResNet)50, ResNet101, GoogleNet, Inception-V3, InceptionResNetV2, and Squeezenet. Each architecture has its advantages and disadvantages. In the case flower image classification research, the performance of the Resnet architecture produces higher accuracy than the AlexNet architecture. This is because Resnet froze the top 2 layers so that it gets more effective results [7]. In the case of the American Sign Language recognition classification, the LeNet architecture obtains higher accuracy than the AlexNet architecture [8]. The advantages of AlexNet architecture are AlexNet has a network that is quite complex so it is possible for AlexNet to be able to work properly and have high accuracy in recognizing faces through images.

In this research, the researcher wants to develop a facial recognition algorithm using CNN with AlexNet architecture. In addition, this study also aims to determine the performance of the CNN algorithm that uses the AlexNet architecture in the facial recognition process which is measured based on the parameters of accuracy and sensitivity (recall). The research contribution is to provide the results of the AlexNet CNN architecture work in facial classification and recognition.

II. LITERATURE REVIEW

The CNN algorithm that will be applied in this study has an AlexNet structure. AlexNet is one of the convolution structures on CNN which has 8 layers/layers consisting of 5 convolution layers and 3 fully connected layers, and allows for a more detailed classification process. This AlexNet has a network pre-training process, which is trained on more than one million images of the ImageNet database [9].

The convolution technique is a technique used for image quality improvement and softening. The convolution process is the sum of the multiplication of each kernel with each point in the input function. Image processing is a form of processing an image with numerical processing of an image. In this case, the process is performed on each pixel or point of an image. In this case, the researcher uses the CNN AlexNet algorithm, because AlexNet has a more detailed structure. AlexNet has won the competition from its competitor LeNet because AlexNet has a more detailed architecture and has a lower error rate than LeNet. Previous studies using Alexnet CNN as the main algorithm include Alexnet being used to detect pathological brains in MRI images. In this study resulted in an accuracy of 100% [10]. Then Alexnet CNN is also used for fundus image classification to classify the severity of Diabetic Retinopathy (DR). This study obtained an accuracy of 96.6% [11]. From previous studies, the Alexnet architecture produces high accuracy, so this becomes our reference for conducting research on the performance of the Alexnet CNN architecture on facial recognition. This research is expected to produce a fairly high accuracy.

Other research related to facial identification and recognition and using Deep Learning-based algorithms, among others, was carried out by [12] is to create a system called Retina Face Mask, a mask detection tool using the ResNet and MobileNet architectures as the main algorithms for deep learning. Where in each image will be trained to have a comparison of each of the results. The results showed that ResNet had a recall/sensitivity of 94% higher than MobileNet which only reached 89%. But the drawback of ResNet is that it requires large resources, especially for using a graphics card from a PC to conduct data training.

Other research by [13], [14] used the Raspberry Pi as the main processor for facial image detection. In this study using Raspberry Pi 3 B+ and Arduino [15],[16]. The system used to perform facial image recognition uses the Haar Cascade Classifier which is implemented in the OpenCV library. While the facial pattern recognition method uses the PCA (Principal Component Analysist) and LDA (Linear Discriminant Analysis) methods. The drawback of his research is that it only detects facial conditions and still cannot detect faces using masks.

Next research by [17] is to use CNN in facial recognition. In his research, the system is able to recognize faces with accuracy values of around 81%, 72%, 69% which are differentiated based on the shooting distance. Of course, this accuracy value can still be improved by utilizing other algorithms.

III. RESEARCH METHOD

This study implements the AlexNet-architectured CNN (Convolutional Neural Network) algorithm which works on the MATLAB programming platform version 2021a. This MATLAB version has been made easier by the Deep Learning Toolbox. In this program, it consists of a training program that will be used to train training image data and a testing program that will be used for test image data. Furthermore, the accuracy value is obtained by matching the training data with the test data. In addition, accuracy and sensitivity (recall) values are presented. These parameter values are used as an assessment of the performance of algorithm.

A. Dataset

The dataset used in this study is a facial image dataset of someone who has been recognized and taken directly by photographing it using a smartphone camera or laptop camera. As for the unknown group dataset, the data was obtained from the internet. Then the image data is collected in a database. In this study there were 4 classes, namely 3 classes of faces that were recognized and 1 class of faces that were not recognized. Each group of person's facial images has 100 image data so that 400 image data are obtained. Each class is divided into 90 image data for the training process and 10 image data for testing. The face image used is limited to only the frontal face image. In one of the classes, there is an image of a woman's face wearing an accessory in the form of a headscarf/veil. For the unknown face class, it is a class that consists of other facial images that are categorized in the unknown person folder. This facial image is obtained from images of public figures' faces obtained from the internet [18]. The image data samples used in this study are shown in Fig. 1. Meanwhile, Table I shows the number of image class name datasets used.

TABLE I
CLASS NAME AND DATASET

No	Class Name	Training Data	Test Data
1	Denny	90	10
2	Abil	90	10
3	Anggun	90	10
4	(*Unknown face*)	90	10

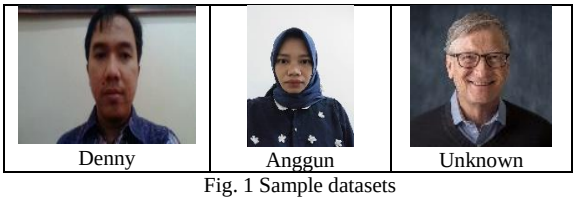

Denny Anggun Unknown

Fig. 1 Sample datasets

B. Method

The algorithm used includes a training algorithm using the Convolutional Neural Network algorithm with the AlexNet architecture in MATLAB to obtain training accuracy. The training algorithm is shown in Fig. 2. Then Fig. 3 shows the test algorithm for testing the performance of the system.

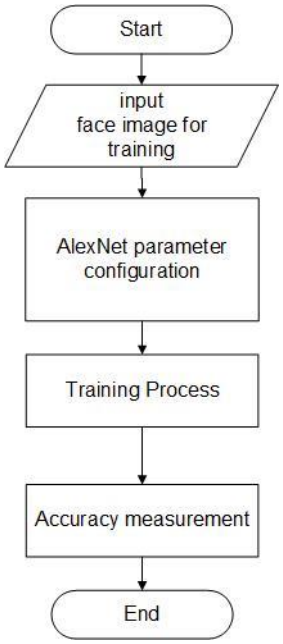


Fig. 2 Training Algorithm Using AlexNet

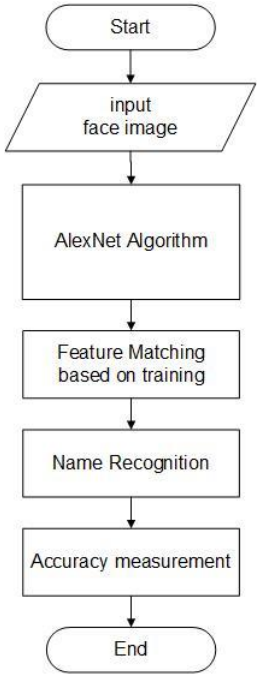


Fig. 3 Testing Algorithm Using AlexNet

Fig. 2 shows the system training algorithm using CNN AlexNet. Stages, starting with entering a dataset in the form of facial images. In this case using 3 recognized face image classes and 1 unrecognized class. Each face category has 100 image data samples. A total of 90 image data from each face category were used as training data. Next, set the CNN AlexNet parameters in the form of the number of epochs, learning rate, and mini batch/batch size. The AlexNet configuration parameters are shown in Table III in detail. The results of the training are in the form of facial recognition features and obtaining training accuracy.

Fig. 3 shows the stages of system testing. At this stage, the AlexNet algorithm parameters used for the test process are the same as the training parameters. The process that occurs at this stage is feature matching between the test data and the training data at the previous stage. The output of this system is in the form of name identification according to the facial image that is entered into the system. Then do the calculation of the accuracy value.

C. AlexNet's CNN theorem

AlexNet is the name given to the Convolutional Neural Network architecture designed by Alex Krizhevsky. AlexNet consists of 8 layers, namely 5 convolution layers and 3 fully connected layers. ReLu is activated at the end of each convolution layer. In the final layer there is softmax to determine the class distribution of detected objects. Sotfmax itself has more than 1000 distribution classes. Fig. 4 shows the architecture of AlexNet [19]. Fig. 5 shows the steps taken by AlexNet in processing image data [20].

D. Analysis Method

At this stage, the performance measurement of the system is carried out. The performance parameters measured are shown in equations (1) to (2). The values of TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) are obtained from the confusion matrix table shown in Table II [21].

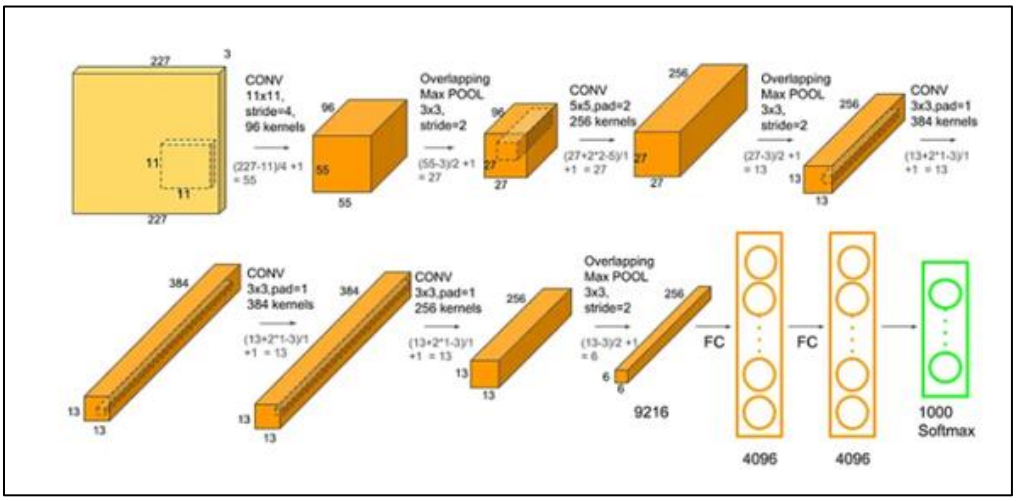


Fig. 4. AlexNet architecture

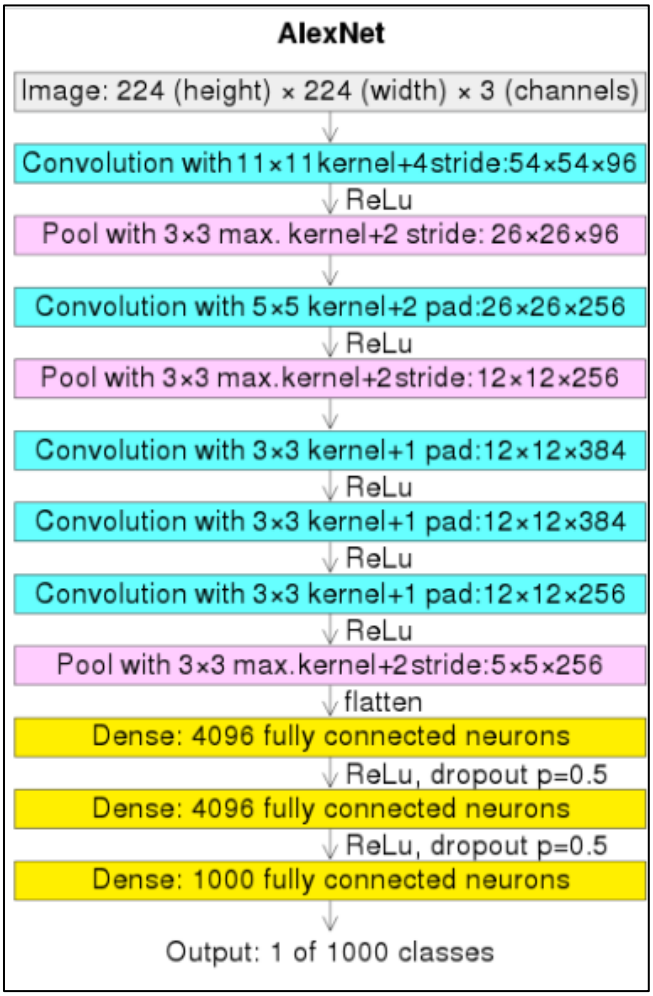
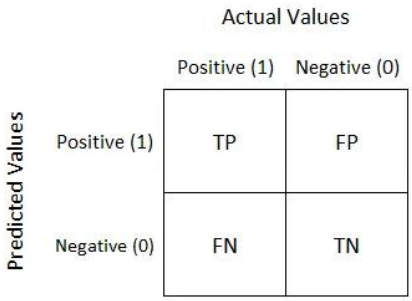


Fig. 5 AlexNet Work Stages

TABLE II
CONFUSION MATRIX



This Confusion Matrix determines the TP, TN, FP, and FN values from the training and testing processes. It can be seen in Figure 6 that TP shows the correct (positive) value/calculation occurs between the prediction and the actual condition. Meanwhile, TN has the right correlation (negative/opposite TP) between predictions and actual (actual) conditions. Then FN and FP are prediction errors that do not correspond to actual conditions. Based on the confusion matrix table, then the accuracy and sensitivity values are calculated as system performance value parameters shown in equations (1) and (2).

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} x\ 100\% \quad (1)$$

$$\boldsymbol{Sensitivity/recall} = \frac{TP}{TP + FN} x\ 100\% \quad (2)$$

IV. RESULTS AND DISCUSSION

The training process is carried out 4 times with different combinations of AlexNet parameters. The combined parameters include (Epoch, Learning Rate, Mini Batch) to get the highest accuracy value. The highest accuracy value was obtained with the following parameters.

TABLE III
ALEXNET PARAMETER COMBINATION DURING THE TRAINING PROCESS

Training Test ID	***Max Epoch***	***Learning Rate***	***Mini Batch***	***Time***	**Accuracy**
Train1	5	0.0001	8	55 seconds	95%
Train2	5	0.0001	10	55 seconds	100%
Train3	6	0.0001	10	1 minutes, 11 seconds	100%
Train4	6	0.0001	8	1 minutes, 28 seconds	95%

From Table III, the highest accuracy value and fast learning time were obtained in the second experiment, namely ID Train2. This ID has parameters Max epoch 5, Learning Rate 0.0001, Mini Batch 10, and has the fastest learning time of around 55 seconds. One of the most important hyperparameters is the batch size (Mini batch), which is the number of images used to train one pass forward and backward [22]. In this study, higher batch sizes result in higher accuracy than lower batch sizes.

The training process with ID Train2 is shown in Fig. 6. Then, the best training result network is stored and reused in the testing process. The testing process is carried out by calling the network resulting from previous training to identify the person's name from the test image. Fig. 7 shows the results of the network test in the test images. Fig. 7a shows correct identification, while Fig. 7b shows wrong identification. There is a name identification error as shown in Fig. 7b. This is because there are several images tested that do not follow the training process. As shown in Fig. 7b (Bill Gates face) should be included in the unknown people folder. This

was not done by the researcher, causing a mistake in identifying the name. Therefore, it can be concluded temporarily that this system still has errors in face identification, especially in facial images that did not follow the previous training process so the system does not recognize these faces. To improve the accuracy of the testing process, it is necessary to add training data for faces.

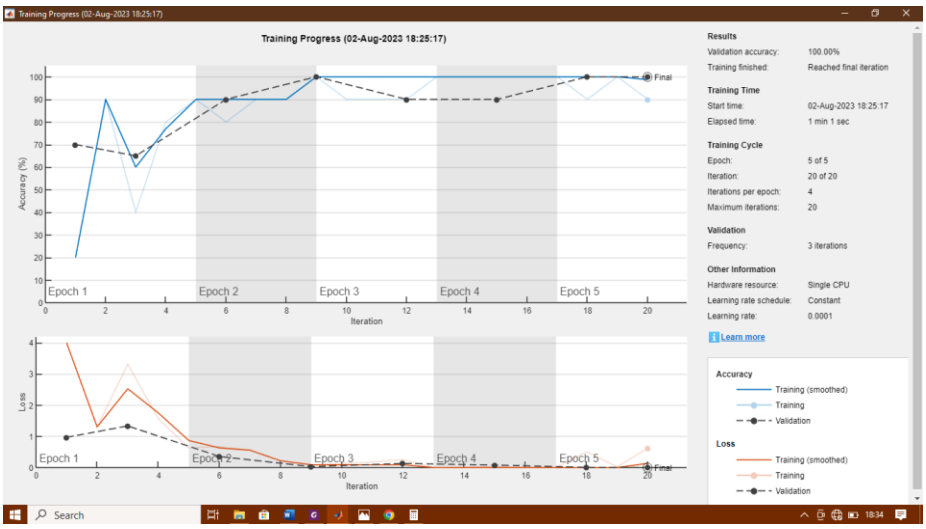
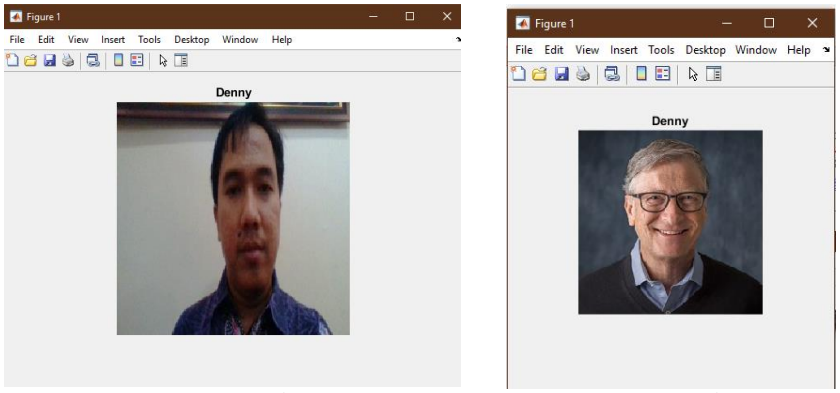


Fig 6. Training process using AlexNet



7a. Correct identification 7b. Wrong identification

Fig. 7 Example of personal identification test results using AlexNet

TABLE IV
CONFUSION MATRIX RESULTS

		Actual Values	
		Positive	Negative
Predicted Values	Positive	30	0
	Negative	2	8

The test results are shown in the form of Table IV confusion matrix. From the results of Table IV, the accuracy value can be calculated based on equations (1) and (2). Obtained an accuracy value of 95% and a sensitivity/recall of 94%. This is because there is an identification error in the name, where the image being

tested has not been included in the training class. The incorrectly identified image should be included in the face unrecognized category.

$$\text{Accuracy} = \frac{30+8}{30+0+2+8} x\ 100\% = 95\%$$

$$\frac{\boldsymbol{Sensitivity}}{\boldsymbol{Recall}} = \frac{30}{30+2} x\ 100\% = 94\ \%$$

V. CONCLUSION

The conclusion obtained from this study is that the CNN algorithm with the ALEXNET architecture is able to work well in recognizing faces. By using data of 400 facial images which are divided into 4 classes, the training accuracy value is 100%, the testing accuracy is 95%, and the test recall is 94%. The test image data carried out in this study only contained one face in each image/photo and had not been tested on photos with many faces. Suggestions for further research are to try to recognize many faces in one image/photo. Then add more training data, test data, classification class to get more real results.

ACKNOWLEDGMENT

The researcher would like to thank all parties who assisted in all part of this research. The researcher would like to thank all levels of Institute for Research and Community Service (LPPM) in Politeknik Negeri Madiun (PNM), all lecturers of the Computer Control Engineering study program and all research members who have fully assisted in completing the research and in accordance with the expected outcomes.

REFERENCES

[1] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.

[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, vol. 1, doi: 10.1109/cvpr.2001.990517.

[3] D. Hardiyanto and D. Anggun Sartika, "Optimalisasi Metode Deteksi Wajah berbasis Pengolahan Citra untuk Aplikasi Identifikasi Wajah pada Presensi Digital," *Setrum Sist. Kendali-Tenaga-Elektronika-Telekomunikasi-Komputer*, vol. 7, no. 1, pp. 107–116, 2018, doi: 10.36055/setrum.v7i1.3367.

[4] R. Wiryadinata, U. Istiyah, R. Fahrizal, P. Priswanto, and S. Wardoyo, "Sistem Presensi Menggunakan Algoritme Eigenface dengan Deteksi Aksesoris dan Ekspresi Wajah," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 6, no. 2, pp. 222–229, 2017, doi: 10.22146/jnteti.v6i2.319.

[5] R. Wiryadinata, R. Sagita, S. Wardoyo, and P. Priswanto, "Pengenalan Wajah pada Sistem Presensi Menggunakan Metode Dynamic Times Wrapping, Principal Component Analysis dan Gabor Wavelet," *J. Din. Rekayasa*, vol. 12, no. 1, pp. 1–8, 2016, [Online]. Available: http://dinarek.unsoed.ac.id/jurnal/index.php/dinarek/article/view/138.

[6] S. Wardoyo, R. Wiryadinata, and R. Sagita, "Sistem Presensi Berbasis Agoritma Eigenface Dengan Metode Principal Component Analysis," *J. Setrum*, vol. 3, no. 1, pp. 61–68, 2014, [Online]. Available: https://jurnal.untirta.ac.id/index.php/jis/article/viewFile/498/386.

[7] B. Falahkhi, E. F. Achmal, M. Rizaldi, and R. Rizki, "Perbandingan Model AlexNet dan ResNet dalam Klasifikasi Citra Bunga Memanfaatkan Transfer Learning Comparison of AlexNet and ResNet Models in Flower Image Classification Utilizing Transfer Learning," *Ilmu Komput. Agri-Informatika*, vol. 9, no. 1, pp. 70–78, 2022, doi: https://doi.org/10.29244/jika.10.1.1-9.

[8] M. E. Al Rivan and A. G. Riyadi, "Perbandingan Arsitektur LeNet dan AlexNet Pada Metode Convolutional Neural Network Untuk Pengenalan American Sign Language," *Komput. Terap.*, vol. 7, no. 1, pp. 53–61, 2021, doi: 10.35143/jkt.v7i1.4489.

[9] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015, doi: https://doi.org/10.1007/s11263-015-0816-y.

[10] S. Lu, Z. Lu, and Y.-D. Zhang, "Pathological brain detection based on AlexNet and transfer learning," *Comput. Sci.*, vol. 30, pp. 41–47, 2019, doi: https://doi.org/10.1016/j.jocs.2018.11.008.

[11] T. Shanthi and R. S. Sabeenian, "Modified Alexnet architecture for classification of diabetic retinopathy images," *Comput. Electr. Eng.*, vol. 76, pp. 56–64, 2019, doi: https://doi.org/10.1016/j.compeleceng.2019.03.004.

[12] M. Jiang, X. Fan, and H. Yan, "RetinaMask: A Face Mask detector," *Arxiv*, 2020.

[13] F. Sirait and Y. Yoserizal, "Pemanfaatan Raspberry Pi Sebagai Processor Pada Pendeteksian Dan Pengenalan Pola Wajah," *J. Teknol. Elektro*, vol. 7, no. 3, pp. 146–150, 2016, doi: 10.22441/jte.v7i3.892.

[14] M. M. Lambacing and F. Ferdiansyah, "Rancang Bangun New Normal Covid-19 Masker Detektor Dengan Notifikasi Telegram Berbasis Internet of Things," *Dinamik*, vol. 25, no. 2, pp. 77–84, 2020, doi: 10.35315/dinamik.v25i2.8070.

[15] N. T. N. Wahyudi, P. Endramawan, and ..., "Trainer Arduino Uno Sebagai Media Pembelajaran Mata Kuliah Dasar Pemrograman Pendidikan Teknik Elektro Universitas PGRI Madiun," *Jupiter (Jurnal Pendidik. Tek. Elektro)*, vol. 07, no. September, pp. 26–32, 2022, doi: http://doi.org/10.25273/jupiter.v7i2.14030.

[16] F. F. Kurniawan, P. Endramawan, and D. Hardiyanto, "Rancang Bangun Pengatur Kecepatan Motor DC Dengan PWM Berbasis Arduino Nano," *JUPITER (Jurnal Pendidik. Tek. Elektro)*, vol. 07, no. 02, pp. 9–16, 2022, doi: http://doi.org/10.25273/jupiter.v7i2.14028.

[17] Efanntyo and A. Rama Mitra, "Perancangan Aplikasi Sistem Pengenalan Wajah Dengan Metode Convolutional Neural Network (CNN) Untuk Pencatatan Kehadiran Karyawan," *J. Instrumentasi dan Teknol. Inform.*, vol. 3, no. 1, pp. 1–11, 2021.

[18] T. E. of Encyclopaedia, "Bill Gates American computer programmer, businessman, and philanthropist," *Encyclopedia Britannica*, 2023. https://www.britannica.com/biography/Bill-Gates.

[19] A. Pujara, "Concept of AlexNet: Convolutional Neural Network," *Analytics Vidhya*, 2023. https://medium.com/analytics-vidhya/concept-of-alexnet-convolutional-neural-network-6e73b4f9ee30.

[20] Wikipedia, "Alexnet," *Wikipedia the free encyclopedia*, 2023. https://en.wikipedia.org/wiki/AlexNet.

[21] S. Narkhede, "Understanding Confusion Matrix," *Towards Data Science*, 2018. https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62 (accessed Jul. 25, 2023).

[22] I. Kandel and M. Castelli, "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset," *ICT Express*, vol. 6, no. 4, pp. 312–315, 2020, doi: https://doi.org/10.1016/j.icte.2020.04.010.