Journal of Theoretical and Applied Information Technology
15th August 2023. Vol.101. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FACIAL RECOGNITION MODEL AS EMPLOYEE ATTENDANCE USING MODERN DEEP LEARNING WITH CONVOLUTIONAL NEURAL NETWORK METHOD

HALIM MAULANA[1] , FERDY RIZA[2], AL-KHOWARIZMI[3]

[1,2]Department of Information System, Universitas Muhammadiyah Sumatera Utara, Indonesia
[3]Department of Information Technology, Universitas Muhammadiyah Sumatera Utara, Indonesia
E-mail: [1]halimmaulana@umsu.ac.id

ABSTRACT

Development of technology is developing very quickly, thus providing many benefits, especially in the field of Information Technology. Demand for services with the use of technology is increasingly needed by various industrial fields, especially the impact generated by Covid-19 has resulted in the application of technology being required in various fields to minimize physical contact. The conventional attendance process by touching the attendance device/devices can be replaced by using face recognition technology. Face Recognition technology uses Deep Learning discussion. Using this technology, without needing to touch employees, they can record attendance by looking into the camera. This study uses the Convolutional Neural Network (CNN) algorithm. The programming language used in this program is Python. The process of making this application with the stages of making Face Recognition, namely image acquisition, preprocessing, extraction, classification, and identification of image data. The dataset is separated into 3 stages of data, namely train data, validation data, and test data.
Keywords: *Deep Learning, Machine Learning, TensorFlow, Python, Facial Recognition*

1. INTRODUCTION

Technology is developing very rapidly in this era, thus providing many benefits in various fields, especially in Information Technology. Information technology is also widely used to assist humans in completing various kinds of work. Information technology is also a means to improve human capabilities and an instrument of change [1,2]. One of the information technologies that is currently developing is face recognition. Face Recognition is a technology method with a face recognition process that is applied to existing technology [3]. These technologies such as cameras, computers, smartphones, and so on. This technology has many things that can be used to help various kinds of human work. Every day, humans do all activities in a conventional way, such as door locks that still use conventional keys, and attendance still uses fingerprints and even excel tables. Face Recognition technology can integrate automatic technology with conventional ones such as absences using a face scanner and access to open doors using a facial scanner. This is interesting to design and realize because currently the technology for security of a door is still using conventional security systems. Conventional security sometimes has some drawbacks such as requiring conventional (physical) keys and access cards to be carried. If the conventional keys have problems such as broken keys or left cards, then people cannot open the door. In some places such as malls and offices, doors that require sensors or access are often found. Especially in offices, every visitor must have an access card to be able to open the door. In its use, the access card is given access to open certain doors. After that, the access card is usually given to visitors with frequent arrivals or company employees. By using face recognition, employees do not need to use access cards because their faces have been registered into the company's dataset to be given door access. Face detection is used because sometimes the face image taken by the camera is mixed with other objects around it that do not want to be detected. The method commonly used to find facial patterns in an image is the Haar Feature-Based Cascade Classifier method. This method is often used because it can capture facial patterns and is able to remove areas in the image that do not have facial patterns [4,5]. To design and make Face Recognition Machine

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Learning is used with a supervised learning model. Machine Learning is a part of artificial intelligence that is used to replace or imitate human behavior to solve problems. While supervised learning is a learning technique in Machine Learning that makes a function from the data provided [6,7]. In the Face Recognition process, the Convolutional Neural Network (CNN) algorithm is also used. Basically, the CNN algorithm is an artificial neural network architecture that is more effective for image classification. The main concept of CNN itself is in its convolution operation, where an image will be extracted for each feature to form several patterns that will be easier to classify [8]. This technique can make the image learning function more efficient to implement.

The development of Machine Learning has now been facilitated by the many libraries and Application Program Interfaces (APIs). Libraries for Machine Learning are usually based on the Python programming language. The use of libraries in python uses TensorFlow to express a mathematical problem in Deep Learning. Deep learning is one of the fields of Machine Learning that utilizes neural networks and adds suggestions for implementing problems with large datasets [9]. TensorFlow has a feature to train models using the Graphics Processing Unit (GPU). Another library used by Deep Learning is Kera's. Kera's uses several features from TensorFlow to create an artificial neural network because Kera's already provides several basic CNN models that have been optimized to facilitate Deep Learning.

In this study, a review of previous studies related to Face Recognition and Convolutional Neural Network (CNN) has been carried out. There are several related studies that have been studied, such as research in the journal Sepritahara, where research was carried out for Face Recognition using the Hidden Markov Model (HMM) method [10]. This method has several steps carried out such as labeling, codebook and HMM database training and facial image recognition. In the labeling process, labels are made for each face image. Each face image will be grouped under one label according to the name that corresponds to the dataset. The second step is the codebook. This step performs the creation of a codebook on the formed labels. The next step is to create a training database for the HMM model from the labels and codebooks that have been formed previously. In the journal Huang et al, the research entitled Densley Connected Convolutional Network uses the part-stacked CNN method [11][12]. This method uses a dataset and includes each part of an image for comparison. This process breaks up an image and picks up each piece. This method has a faster image processing process than other methods, so this method is best used in real time. The drawback of this method is its accuracy. The accuracy result of this method is no better than the CNN bilinear method which does not perform part attention on the image. The architecture used in the part-stacked CNN method is the cafenet architecture. The accuracy obtained by this method is 76.63%. In the journal Yang et al this study used the Two Stage CNN algorithm [13]. Before doing two-stage CNN, this program performs a multi-scale algorithm. This step is intended as a broad shooting process. Getting each face image multi-scale results in capturing multiple faces in one camera. After the multi-scale process is obtained, a two-stage CNN step is carried out. This step serves to separate several captured images according to the typical. After separating the images according to their characteristics, a classification of images can be obtained that can be identified according to the stored dataset or database. This process obtains an accuracy of 70% [14].

Three architectural steps were applied. The first step is the Rectified Linear Unit (ReLU). This architectural model eliminates vanishing gradients in the image to separate the gradients in the image for easier classification. The second step is the Max Pooling Layer. In this step, the second layer captures the results of the first layer that has gone through the Rectified Linera Unit process so that better image details will be obtained. The third step is the Fully Connected (FC) process. In this process all layers are combined to get results that can be classified according to the stored dataset. The accuracy obtained is 80% [15,16]. In the journal Fu et al, the research was conducted using the CNN Recurrent Attention method. This method combines several scales of the design used. The results obtained are more accurate than the results of the CNN bilinear algorithm. The percentage of accuracy obtained from the use of CNN's Recurrent Attention is 85.3% [17,18]. Face Recognition Research for Bank Employee Door Access Using Machine Learning with the Convolutional Neural Network (CNN) Method uses the VGG16 architectural model. VGG16 is a CNN architecture with the input used in the form of an RGB image measuring 224 × 224 pixels. There are 2 types of convolutional layers used in this architecture, namely convolutional layers with a filter size of 3x3 (conv3) and a filter size of 1x1 (conv1). Convolutional layer sizes used vary, namely 64x64, 128x128, 256x256, and 512x512. This model is the best in localization and classification. The CNN method used in this

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

application is bilinear CNN which is one of the best methods for now so that it can still get better accuracy results.

2. METHOD

In this study, researchers classified images of several employees using the Deep Learning method and the Convolutional Neural Network (CNN) algorithm. The main process in making this model begins with the data training process. This process aims to form a model that will be used for testing data testing. The parameter to measure the success rate of the model is the accuracy value. The value of model accuracy can be determined by testing using data testing. The training process uses Keras packages in python with tensorflow back-end. Keras is one of the modules created by Google to facilitate research on neural networks and is able to run on tensorflow, theano, MXNet.

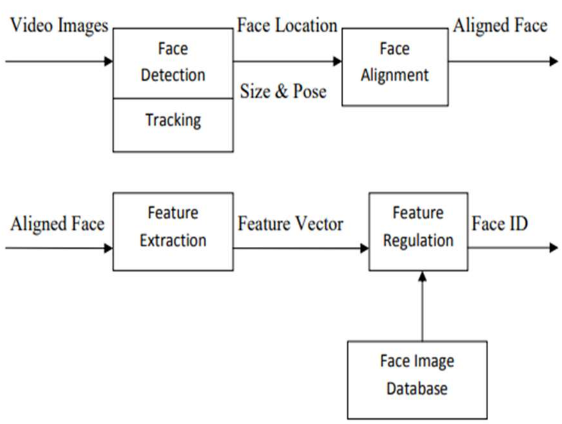

Figure 1. Face Recognition Process

The face image is represented as an array of pixels with high dimensions. Face recognition, and computer vision research in general, have observed a growing interest in techniques that apply algebra and statistical features to perform the extraction and analysis of these types of cases. Computer analysis of facial images is influenced by visual signals (light reflected on the surface of the face) stored by a digital sensor as an array of pixel values. This pixel value stores the color or just the light intensity. A pixel array of m x n face images can be stored in the form of a trick (eg vector) in a dimensional image space by writing the pixel values in a fixed order. The main problem with multidimensional data is its dimensionality, the number of coordinates needed to specify a data point. So the results of face recognition will be in the form of information that is known or not as a face by previously comparing it with information from a known face. This face recognition process has problems with lighting, camera position, camera parameters, and noise obtained in an image.

Convolutional Neural Network (CNN) is the development of a multilayer perceptron (MLP) which is designed to process two-dimensional data in the form of images. CNN is included in the type of Deep Neural Network because of the high network depth and widely applied to image data. Basically,y image classification can be used with MLP, but the MLP method is not suitable for use because it does not store spatial information from the image data and assumes each pixel is an independent feature, resulting in poor results. The initial research that underlies the CNN findings was first carried out by Hubel and Wiesel (Hubel & Wiesel, T, 1968) regarding the visual cortex of the cat's sense of sight. Technically, CNN is an architecture that can be trained and consists of several stages. The input (input) and output (output) of each stage is composed of several arrays which are commonly called feature maps. Each stage consists of three layers, namely convolution, activation function layer, and pooling layer. The following is a Convolutional Neural Network architecture network.

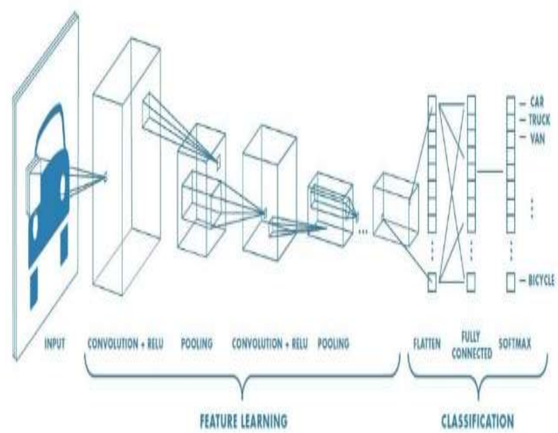

Figure 2. General Architecture Of CNN

At this stage perform convolution operations on the output of the previous layer. This layer is the main process that underlies the CNN architecture network. Convolution is a mathematical term in which the application of one function to the output of another function is repeated. The convolution operation is an operation on two real-valued argument functions. This operation applies the output function as a Feature Map of the input image. These inputs and outputs can be viewed as two real-valued arguments. The convolution operation can be written as follows:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$s(t) = (x*t)(t) = \sum_{\alpha=}^{\infty} x(\alpha) * w(t-\alpha)$

S(t) = Function Results of Convolution Process

X = Input

W = weight (kernel)

$s(i,j) = (K*I)(i,j) = \sum_{\infty}\sum_{\infty} I(i-m, j-n)K(m,n)$

$s(i,j) = (K*I)(i,j) = \sum_{\infty}\sum_{\infty} I(i+m, j+n)K(m,n)$

Based on the two equations above, it is a basic calculation in the convolution operation, where i and j are a pixel of the image. The calculation is cumulative and occurs when K is the kernel, then I is the input and the kernel is reversible relative to the input. As an alternative, the convolution operation can be seen as a matrix multiplication between the input image and the kernel where the output is calculated with the dot product. In addition, the determination of the output volume can also be determined from each layer with hyperparameters. The hyperparameters used in the equation below are used to calculate the number of activation neurons in one output. Pay attention to the following equation :

$$(W - F + 2P)/(S + 1)$$

CNN model training. Generally, CNN has 2 stages, namely the feature learning and classification stages. Image input on the CNN model uses an image size of 64x64x3. The number three is an image that has 3 channels, namely Red, Green, and Blue (RGB). The input image will then be processed first through the convolution process and the pooling process at the feature learning stage. The number of convolution processes in this design has two layers of convolution. Each convolution has a different number of filters and kernel sizes. Then the flattening process is carried out or the process of changing the feature map resulting from the pooling layer into vector form. This process is commonly referred to as the fully Connected layer stage. The following is the design of the CNN architecture in this study as shown in Figure 3.

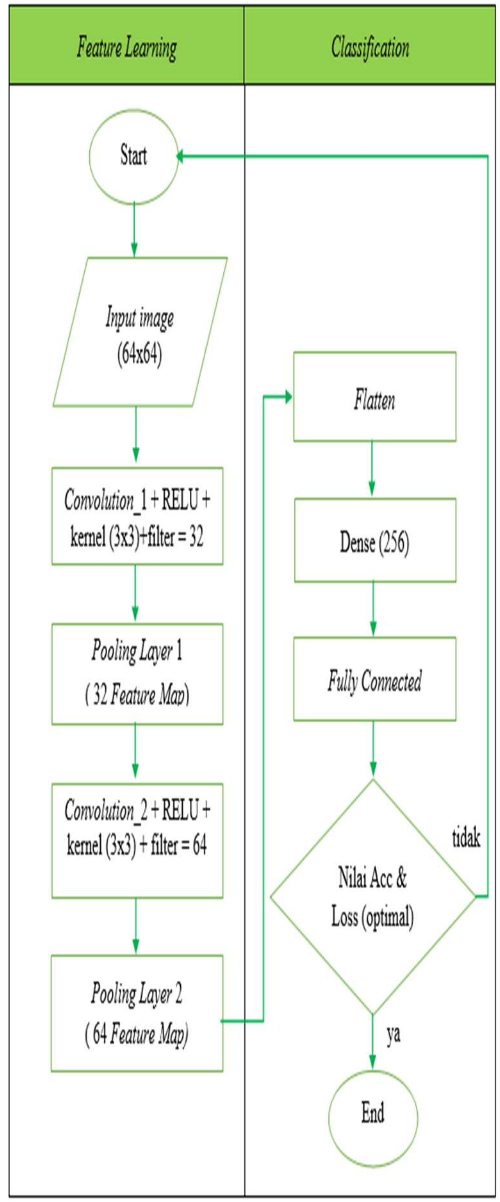

Figure 3. Flow Chart Model Research

Based on the table above, it is explained that there are two stages in the Convolutional Neural Network architecture, namely Feature Learning and classification. Feature learning is a technique that allows a system to run automatically to determine the representation of an image into features in the form of numbers that represent the image. The Classification stage is a stage where the results of feature learning will be used for the classification process based on the predefined subclasses. If the flow chart above is converted into an image, it can be seen as shown in Figure 4 below:

Journal of Theoretical and Applied Information Technology
15th August 2023. Vol.101. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

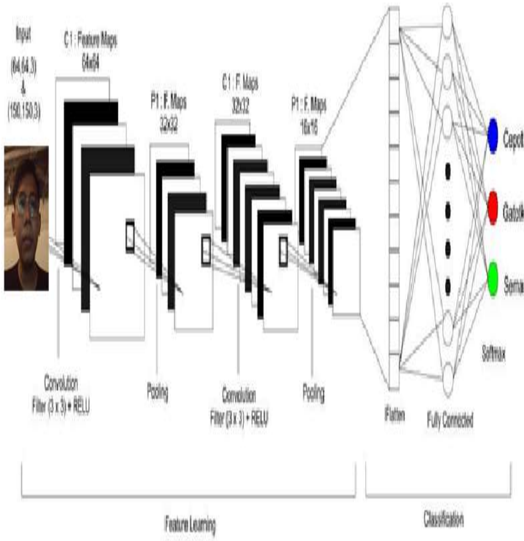

Figure 4. Architecture Design

In the first convolution, it uses 32 filters and a kernel with a 3x3 matrix. Then the pooling process is carried out using a pooling size of 2x2 with a mask shift of two steps. Then in the second convolution stage by using the number of filters is as many as 64 and the kernel with a 2x2 matrix. Then proceed with flattening, which is changing the output of the convolution process in the form of a matrix into a vector which will then be forwarded to the classification process using MLP (Multi-Layer Perceptron) with a predetermined number of neurons in the hidden layer. The class of the image is then classified based on the value of the neuron in the hidden layer using the SoftMax activation function. Figure 4. is a network architecture in the training process to produce an optimal model. This study uses image input with a size of 64x64x3, the aim is to compare the accuracy values based on the size of the image. The architecture above can be explained as explained below:

1. The first convolution process uses a 3x3 kernel and a total of 32 filters, this convolution process is a combination process between two different matrices to produce a new matrix value. After the convolution process, an activation function is added, namely RELU (Rectified Linear Unit). This activation function aims to change the negative value to zero (removing negative values in a convoluted matrix). The results of this convolution have the same size, namely 64x64 because during the convolution process the padding value of 0 is used.
2. Pooling process. Pooling is a reduction in the size of the matrix by using a pooling operation. Pooling process. Basically, the pooling layer consists of a filter with a certain size and which will alternately shift over the entire feature map area. This research uses max pooling to get the new matrix value resulting from the pooling process. Based on the pooling results, a new matrix of 32x32 size is generated using a 2x2 pooling kernel. The way max-pooling works is to take the maximum value based on the shift in the kernel as much as the stride value is 2.
3. The seconconvolutionon process is to continue the results of the first pooling process, namely by inputting an image matrix of 32x32 with a number of filters as many as 64 filters and with a kernel size of 3x3. This second convolution process uses the RELU activation function.
4. The next process enters the second pooling process, this process is almost the same as the first hang pooling process, but there is a difference in the final output value of the matrix. The resulting output has an image size of 16x16.
5. Next Flatten or fully connected. At this stage, only one hidden layer is used in the MLP (Multi-LayerPerceptron) network. Flatten here converts the output pooling layer into a vector. Before performing the classification, process, or predicting images, in this process, the Dropout value is used. Dropout is a neural network regulation technique with the aim of selecting some neurons at random and will not be used during the training process, in other words, the neurons are discarded randomly. The purpose of this process is to reduce overfitting during the training process.
6. The last process is to use Softmax function activation. This function is specifically used in the classification method of multinomial logistic regression and multiclass linear discriminant analysis.

Based on the description of explanation of the network architecture above, the architecture is used for the training process. So that from the training process a model of the architecture is obtained. The following model is formed:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1. Cnn Model

No	Nama	Size	Parameter
0	Input	64*64*3	0
1	Conv2d_1	(64+(2*1)-(3-1)) = 64*64*32	((3*3*3)+1)*32 = 896
2	MaxPool_1	32*32*32	0
3	Conv2d_2	(32+(2*1)-(3-1)) = 32*32*64	((3*3*32)+1)*64 = 18496
4	MaxPool_2	16*16*32	0
5	Flatten	16384	0
6	Dense	256	(16384*256)+256 = 4194560
7	Output	3	(256+1)*3 = 771
	Total		4.214.723

The picture above is a model formed from the results of the training. To calculate the input into the convo used the formula "input_size + 2*padding - (filter_size -1)". The total parameters formed from the model are 4,214,723 neurons.

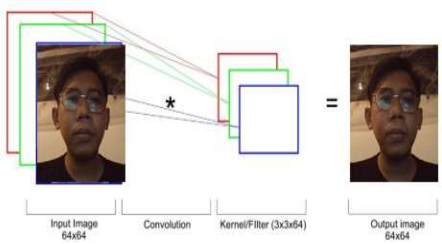

Figure 5. Convolution Process

Convolution is the process of combining two series of numbers to produce a third series of numbers. If implemented, the numbers in this convolution are in the form of an array matrix. At the input, the image has a pixel size of 64x64x3, this shows that the pixel height and width of the image are 64 and the image has 3 channels, namely red, green, and blue or commonly referred to as RGB. Each pixel channel has a different matrix value. The input will be in convo with the filter value that has been determined. A filter is another block or cube with a smaller height and width but the same depth that is swept over the base image or original image. Filters are used to determine what pattern will be detected which is then convoluted or multiplied by the value in the input matrix, the value in each column and row in the matrix depends on the type of pattern to be detected. The number of filters in this convo is 64 pixels with a kernel size (3x3), this means that the image generated from the convolution results will be as many as 64 map features. In order to better understand the workings of the convolution process, the researcher will use a sample matrix in the input image. Because the input image has a pixel size of 64x64, the researcher only takes some of the matrix values which will be sampled in the convolution process.

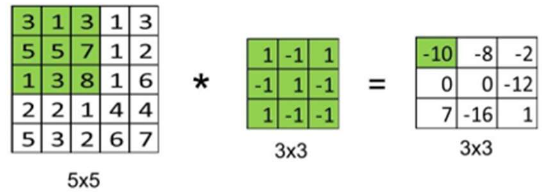

Figure 6. Convolution Computing Process

Figure 6 shows the convolution process using a kernel size of 3x3, using stride 1. Stride here means that the number of kernel shifts to the input matrix is one. If visualized as follows:

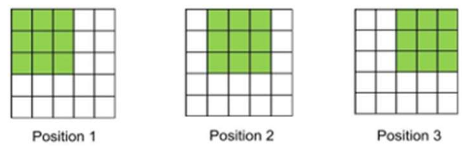

Figure 7. Kernel Position

Figure 6. shows the calculation of the dot product in the convolution process where a 3x3 kernel size starts on the left side. This process is called a sliding window. However, in this study, a padding value of 1 is given, namely the addition of a 0 value around the input matrix value so that the input and output have the same matrix value, so as not to reduce the information in the image. This process is carried out from the top left corner to the bottom left corner. The dot product calculation can be seen as follows:

a. $Position\ 1 = (3x1) + (5x(-1) + (1x1) + (1x(-1)) + (5x1) + (3x(-1)) + 3x1) + (7x(-1)) + (8x1) = -10$
b. $Position\ 2 = (1x1) + (5x(-1) + (3x1) + (3x(-1)) + (7x1) + (8x(-1)) + (1x1) + (1x(-1)) + (1x1) = -8$
c. $Position\ 3 = (3x1) + (7x(-1) + (8x1) + (1x(-1)) + (1x1) + (1x(-1)) + (3x1) + (2x(-1)) + (6x1) = -2$
d. $Position\ 4 = (5x1) + (1x(-1) + (2x1) + (5x(-1)) + (3x1) + (2x(-1)) + (7x1) + (8x(-1)) + (1x1) = 0$
e. $Position\ 5 = (5x1) + (3x(-1) + (2x1) + (7x(-1)) + (8x1) + (1x(-1)) + (1x1) + (1x(-1)) + (4x1) = 0$
f. $Position\ 6 = (7x1) + (8x(-1) + (1x1) + (1x(-1)) + (1x1) + (4x(-1)) + (2x1) + (6x(-1)) + (4x1) = -12$
g. $Position\ 7 = (1x1) + (2x(-1) +$

$(5x1) + (3x(-1)) + (2x1) + (3x(-1)) + (8x1) + (1x(-1)) + (2x1) = 7$

h. $Position\ 8 = (3x1) + (2x(-1) + (3x1) + (8x(-1)) + (1x1) + (2x(-1)) + (1x1) + (4x(-1)) + (6x1) = -16$

i. $Position\ 9 = (8x1) + (1x(-1) + (2x1) + (1x(-1)) + (4x1) + (6x(-1)) + (6x1) + (4x(-1)) + (7x1) = 1$

Then before proceeding to the pooling layer process, to eliminate negative values in the results, the network architecture uses ReLU (rectified linear unit) activation after the convolution process. The function of this activation is to do a "threshold" from 0 to infinity. The value in the negative convolution result will be changed with this activation to zero and the others to infinity.

3. RESULT AND DISCUSSION

3.1 Pooling Process

Pooling is a reduction in the size of the matrix by using a pooling operation (merging). The method used in this pooling process uses max pooling. Max-pooling is one of the common methods used by researchers related to deep learning research. A study conducted by Dominik Scherer et al (Scherer, 2010) showed that the use of the max pooling method was superior to the sub-sampling method. the use of this method is one of the best methods in the pooling process. The following is an overview of the pooling process:

$$\begin{bmatrix} 10 & 8 & 2 \\ 0 & 0 & 12 \\ 7 & 16 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 12 \\ 16 & 16 \end{bmatrix}$$

Figure 8. Pooling Process

This pooling process uses a size of 2x2 with stride 1 where the number of kernel shifts to the input matrix is one. In this pooling process, the max-pooling method is used, where the window will shift according to its size and stride to get the maximum value. It can be seen in Figure 5.5 that the output of this process has the maximum value taken from the convoluted map feature matrix. The result of the max-pooling is 2x2. Next is the Fully Connected Layer. This process aims to transform the data dimensions so that the data can be classified linearly.

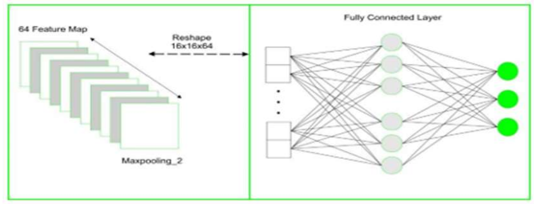

Figure 9. Process Fully Connected Layer

Figure 9 is the process of converting the result of the max-pooling map feature into flatten or vector. In this process the input matrix value from the previous layer will be converted into a vector. This process is the same as the MLP (Multilayer Perceptron) Process. These networks generally use fully connected layers where each pixel is considered a separate neuron. In this process, the "dropout" method is usually applied. This method aims to disable some edges connected to each neuron to avoid overfitting. After that the last process is classification. In this process, activation of the SoftMax function is used. This activation will help the MLP to classify the input against its target.

3.2 Result Training Model

After going through several processes in the Convolutional Neural Network (CNN) algorithm, the results of training and validation are obtained. This process uses a total of 20 epochs, the learning rate value is 0.001. The following is a graph of the results of training using a tensorboard:

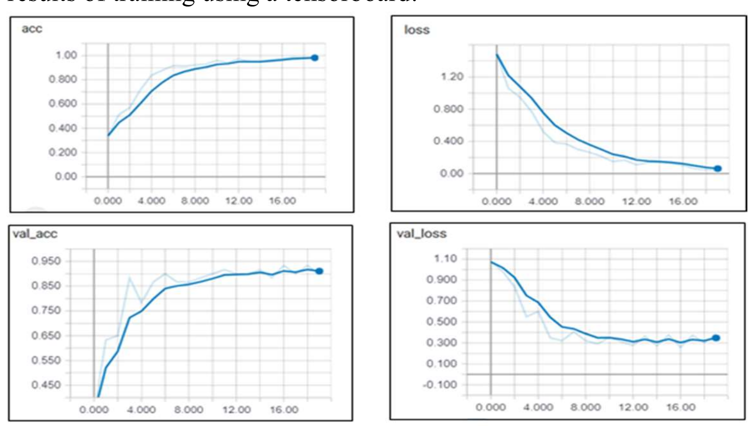

Figure 10. Training Graph

The more epochs, the longer the time needed for model training. Then the accuracy of data validation reaches 90% with a loss value of 0.2972.

Table 2.Confusion Matriks

Matriks		Predict Class		
		Halim	Alko	Ferdy
Actual Class	Halim	19	0	1
	Alko	2	18	0
	Ferdy	1	0	19

Based on table 1 above, the prediction results from the model on the new data testing data show good results. Predictions for Halim are classified into Halim, this means that the classification of the image is correct with a total of 19 correct trials. Predictions on Alko are correctly classified as Alko as many as 18 and missing data from the input Alko are classified as Halim as much as 2 data. Then the last one is Ferdy's prediction is classified as correct as Ferdy as much as 19 and missing data from Ferdy's input is classified as Halim as much as 1 data. The calculation of the accuracy of the entire matrix above is as follows:

$$Overall\ Accuracy = \frac{TTP_{all}}{Total\ Number\ of\ Testing\ Entries}$$

$$Overall\ Accuracy = 56/60 = 93\ \%$$

So, the accuracy generated by the model with an image input of 64x64 pixels, the learning rate value of 0.001 and the number of samples testing 60 data obtained an accuracy value of 93%. To determine the best model, the best value of the parameters in the CNN model must be sought. The parameters in question are the effect of the number of epochs, the effect of the size of the input image, the effect of the number of data trains, the effect of the data scenario, the size of the kernel, and the learning rate. The purpose of determining the parameters of this model is to compare which model is the best by paying attention to the parameter values.

Tabel 3 Accuracy Based on Epoch

Epoch	Accuracy Validatio	Loss Validatio	Time (Seconds
20	91%	0.2003	147
30	90%	0 2018	185
50	93%	0.1924	313
100	97%	0.1818	636

Based on table 2 above, using a learning rate value of 0.001, the accuracy is quite high, reaching 97%. If seen from the table, it can be concluded that the closer to the 100-epoch value used, the higher the accuracy of the testing results. But when more than one hundred epochs are added, the accuracy will decrease. This can be caused by the number of epochs that are too many can also be influenced by the number of datasets.

4. CONCLUSION

This study shows that the accuracy model has been running optimally with using Deep Learning on the CNN algorithm to detect employee's faces. The object being tested is that it can be implemented for employee attendance. The use of the latter value produces a fairly good level of accuracy, but the use of a learning rate of this magnitude is quite slow in reducing the loss value, so the graph of the loss value will experience slow convergence. Data Analysis Constraints In conducting data analysis and the training process that is carried out, there are often some errors in the algorithm that are often wrong in reading the results.

REFERENCES:

[1] H. Maulana and A.-K. Al-Khowarizmi, "Analysis of the effectiveness of online learning using EDA Data Science and Machine Learning," Sinkron, vol. 7, no. 1, pp. 222–231, 2022. doi:10.33395/sinkron.v7i1.11206

[2] [1] H. Maulana and Al-Khowarizmi, "Analyze and designing low-cost network monitoring system using Icinga and Raspberry Pi," IOP Conference Series: Earth and Environmental Science, vol. 704, no. 1, p. 012038, 2021. doi:10.1088/1755-1315/704/1/012038

[3] M. D. Nasution, A.-K. Al-Khowarizmi, and H. Maulana, "Optimization of faster R-CNN to detect SNI masks at mandatory mask doors," JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING, vol. 6, no. 2, pp. 602–611, 2023. doi:10.31289/jite.v6i2.8128

[4] R. Yustiawati et al., "Analyzing of different features using haar cascade classifier," 2018 International Conference on Electrical Engineering and Computer Science (ICECOS), 2018. doi:10.1109/icecos.2018.8605266.Sun, X.Wang, X.Tang, Deep Convolutional network Cascade for facial point detection in Computer Vision and Pattern Recognition, 2013 IEEE conference.

[5] A. Rahmatulloh, R. Gunawan, H. Sulastri, I. Pratama, and I. Darmawan, "Face mask detection using haar cascade classifier algorithm based on internet of things with telegram bot notification," 2021 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS), 2021. doi:10.1109/icadeis52521.2021.9702065

[6] A. Khowarizmi, Akhm, M. Lubis, and A. R. Lubis, "Classification of Tajweed Al-Qur'an on Images Applied Varying Normalized Distance Formulas," no. 3, pp. 21–25, 2020, doi: 10.1145/3396730.3396739

[7] A. R. Lubis, S. Prayudani, M. Lubis, and Al-Khowarizmi, "Decision making in the tea

leaves diseases detection using mamdani fuzzy inference method," Indones. J. Electr. Eng. Comput. Sci., vol. 12, no. 3, 2018, doi: 10.11591/ijeecs.v12.i3.pp1273-1281

[8] D. Pallavi and T. P. Anithaashri, "Novel predictive analyzer for the intrusion detection in student interactive systems using convolutional neural network algorithm over artificial neural network algorithm," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2022. doi:10.1109/icac3n56670.2022.10074027.

[9] V. Bittal, V. Jagdale, A. Brahme, D. Deore, and B. Shinde, "Multifarious face attendance system using machine learning and Deep Learning," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), 2023. doi:10.1109/iciccs56967.2023.10142759

[10] M. M. Gavrikov, A. Y. Mezentseva, and R. M. Sinetsky, "Heuristic techniques for constructing hidden Markov models of Stochastic Processes," 2023 International Russian Smart Industry Conference (SmartIndustryCon), 2023.doi:10.1109/smartindustrycon57312.2023.10110792

[11] C. Dong, R. Wang, and Y. Hang, "Improved facial expression recognition algorithm based on convolutional neural network," 2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT), 2023. doi:10.1109/iccect57938.2023.10140646

[12] C. Dong, R. Wang, and Y. Hang, "Improved facial expression recognition algorithm based on convolutional neural network," 2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT), 2023. doi:10.1109/iccect57938.2023.10140646

[13] H. Jun, L. Shuai, S. Jinming, L. Yue, W. Jingwei and J. Peng, "Facial Expression Recognition Based on VGGNet Convolutional Neural Network," 2018 Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 4146-4151, doi: 10.1109/CAC.2018.8623238.

[14] P. Muzakirlaikhkhan, A. Gayathri and I. M. Arumugam, "Classification of Abhorrence Facial Expressions Using Convolutional Neural Network with Comparison with Recurrent Neural Networks for Better Accuracy," 2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), Karachi, Pakistan, 2022, pp. 1-7, doi: 10.1109/MACS56771.2022.10023318.

[15] M. Defaoui, L. Koutti, M. El Ansari, R. Lahmyed and L. Masmoudi, "PedVis-VGG-16: A Fine-tuned deep convolutional neural network for pedestrian image classifications," 2022 9th International Conference on Wireless Networks and Mobile Communications (WINCOM), Rabat, Morocco, 2022, pp. 1-6, doi: 10.1109/WINCOM55661.2022.9966465.

[16] M. F. Haque, H. -Y. Lim and D. -S. Kang, "Object Detection Based on VGG with ResNet Network," 2019 International Conference on Electronics, Information, and Communication (ICEIC), Auckland, New Zealand, 2019, pp. 1-3, doi: 10.23919/ELINFOCOM.2019.8706476.

[17] A. Verma, P. Singh and J. S. Rani Alex, "Modified Convolutional Neural Network Architecture Analysis for Facial Emotion Recognition," 2019 International Conference on Systems, Signals and Image Processing (IWSSIP), Osijek, Croatia, 2019, pp. 169-173, doi: 10.1109/IWSSIP.2019.8787215.

[18] L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang and T. Zhang, "Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts," 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Orlando, FL, USA, 2017, pp. 1318-1321, doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.213.