JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 345 - 354 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.2067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

COMPARISON OF SGD, ADADELTA, ADAM OPTIMIZATION IN GENDER CLASSIFICATION USING CNN

Sartika Mandasari[1*], Desi Irfan[1], Wanayumini`[1], Rika Rosnelly[1]
[1]Faculty of Engineering and Computer Science, Computer Science Study Program, Main Potential University
email: [*]sartikamandasari12@gmail.com

Abstract: Gender classification is one of the most important tasks of video analysis. A machine learning-based approach was presented to identify male and female facial images with a data set of 2000 images taken from kaggles. This method plays a role in finding the weight value that gives the best output value. This study uses the most appropriate learning rate of each optimization method as a criterion for stopping training. The results showed that the Artificial Neural Network with Adam optimization produced the highest accuracy, which was 91.5% compared to the SGD and ADADELTA optimization methods. Deep Learning techniques that are applied extensively to image recognition used utilize Adam's optimizer method.

Keywords: artificial neural networks; adadelta; adam; gender; sgm;

Abstrak: Klasifikasi gender adalah salah satu tugas analisis video yang paling penting. Pendekatan berbasis machine learning disajikan untuk mengidentifikasi gambar wajah Pria dan Wanita dengan kumpulan data sebanyak 2000 gambar yang diambil dari kaggle. Metode ini berperan dalam menemukan nilai bobot yang memberikan nilai keluaran terbaik. Penelitian ini menggunakan learning rate yang paling sesuai dari masing-masing metode optimasi sebagai kriteria pemberhentian pelatihan. Hasil penelitian menunjukkan Jaringan Saraf Tiruan dengan optimasi Adam menghasilkan akurasi tertinggi yaitu 91,5 % dibandingkan dengan dengan metode optimasi SGD dan ADADELTA. Teknik Deep Learning yang diterapkan secara ekstensif pada pengenalan gambar yang digunakan memanfaatkan metode optimizer Adam.

Kata kunci: Adadelta; Adam; Jaringan Syaraf; *Gender;* Tiruan; SGM;

INTRODUCTION

The Live streaming application has attracted many users, which makes it possible to share videos of daily life with others[1]. This act of sharing constantly generates a large number of real-world videos and most of these videos take precedence over capturing human faces, so facial analysis in videos is becoming increasingly important in real-life applications for video content inspection and recommendation[2]. Classifying facial images in terms of biometric traits, such as gender, age, or ethnicity, has received a lot of attention in the Computer vision literature recently, especially in the context of video surveillance. Gender classification is one of the most important tasks of video analysis[3]. It would be advantageous if a computer system or machine could correctly classify a per-

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 345 - 354 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.2067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

son's gender[4]. For example, a mall buyer's surveillance camera system can be useful for knowing the customer's gender to create the right strategy, or a seller robot can use the right and intelligent approach to communicate with customers based on their gender[5].

Much of the work on gender classification is based on facial features and has proposed several valuable models, in which neural networks, supporting vector machines, and methods of increasing advertising are the most representative[6]. Moghaddam and Yang compared different gender classification methods on the FERET face database and showed that vector engine support has better recognition performance than other classifiers (such as nearest neighbor classifier, linear classifier)[7]. Deep Convolutional Neural Network (CNNs Trained on imagenet dataset) transfer learning works well in identifying male and female gender images and this is also inseparable from the optimization function[8]. In computer vison, CNN has been known as a powerful visual model as well as producing an accurate hierarchy of segmentation features. The model has also been known to make predictions relatively faster than other algorithms while maintaining competitive performance at the same time[9].

The Gradient Descent optimization method is often used for Artificial Neural Network (JST) training. This method plays a role in finding the weight value that gives the best output value[10]. The working principle of the Gradient Descent method is to reduce the Loss value by changing the parameter value step by step. Three optimization methods have been implemented, namely Stochastic Gradient Descent (SGD), ADADELTA, and Adam in the Artificial Neural Network system for the classification of arrhythmia data[11].

METHOD

At this stage of the study, a classification system for male and female sex in the wild was designed to determine accuracy using digital image processing methods. As figure 1 shows the block diagram of the system designed in this study[12].

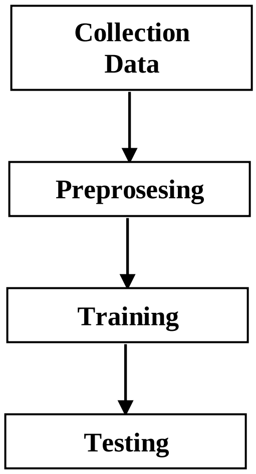

Image 1. System Diagram Blocks in General

In general, the system diagram block systematics shown in Figure 1 is:

1. Collection of Face imagery
2. The facial image is preprocessed with stages of resizing and data augmentation.
3. Training is a cit-ra learning process with the CNN method to obtain a training image model that will be stored in the database.
4. The test was performed with mod-el classified test imagery data from the trainer image using the CNN method to obtain the results and accuracy of the system.

The data used in this study are secondary data. The data is sourced from kaggle data, because kaggle data has been tested as a dataset[13]. A proven dataset will

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 345 - 354 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.2067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

increase accuracy during the training process. The criteria of the dataset are:

a. *.jpg extension color imagery
b. The face type consists of 2 classes, namely Male and Female.
c. The amount of data from each class is 1000 with a ratio of 80% training data and 20% testing data.
d. The size of the image type of the dataset varies.
e. Image results are taken from processes that vary in shape, size, lighting and angle of capture.

After the image collection process, pre-processing is carried out to optimize image quality, thereby facilitating and spurring the system's ability to identify objects [14]. The preprocessing stage of this system is divided into two stages, namely the original data and the augmentation data. In the original data, preprocessing is only done by resizing the image. Meanwhile, pre-processing augmentation data is carried out by resizing and augmenting data [15]. Figure 2 shows the flow chart of the original data preprocessing process as seen below:

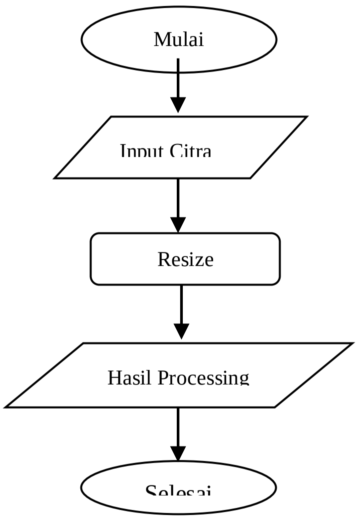

Image 2. Original Data Preprocessing Flow Chart

Figure 3 shows a flow chart of the augmentation data preprocessing process starting from inputting images, resizing, augmentation to processing results.

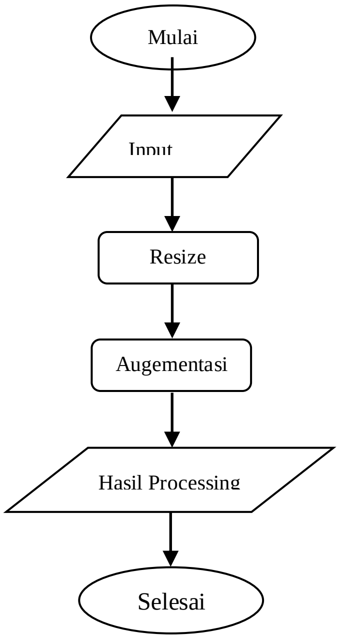

Image 3. Augmentation Data Preprocessing Flow Chart

The explanation of the working system of the image preprocessing flowchart is as follows:

1. Image Input is the stage of retrieving image data to be selected before processing.
2. Resize is the process of resizing the image both vertically and horizontally. In this process, resize the image from its original size to an image size of 64 x 64 pixels.
3. Augmentation is the process of processing image data by modifying image data. In this system, the augmentation stage carried out is Random Rotation, Random Horizontal flip. Preprocessing results are the output of images that have gone through the resize process and the augmentation stage.

At the training stage, the learning process is carried out on the image, which then the output results are in the

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 345 - 354 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.2067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

form of a model that will be stored for use in the testing process[16]. Model formation is a process of training data imagery training in identifying objects and categorized according to their class, according to figure 4.

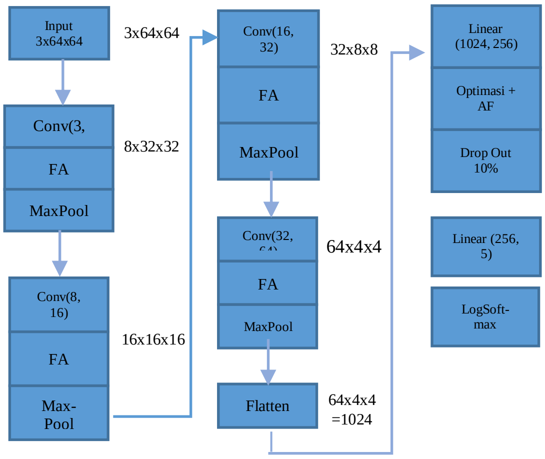

Image 4. Training System Stage Flow Diagram referring to LeNet-5

In this study, the method used is one of the branches of deeplearning algorithms, namely CNN, referring to the LeNet-5 architecture which is very popular and has been tested using 2 layers. In general, the training stage referring to LeNet-5 is shown in figure 4. The image input on the CNN model uses an image with a size of 64 x 64 x 3. Then the input image will be processed through a convolution process and a pooling process. In the first convolution using the number of kernels as many as 10 with a matrix of 3x3 with padding value = valid, ReLu activation is used in this convolution process as Non-Linearity. Then the pooling process is carried out, especially max pooling using a size of 2x2. Then in the second convolution stage using the number of kernels as many as 20 with a 5x5 matrix, still using the ReLU activation function with padding value = valid. Then continued with flatten, which is to change the output of the convolution process in the form of a matrix into a vector which will then be passed on the classification process using MLP (Multi Layer Perceptron) with the number of neurons in the hidden layer that has been determined. At this stage the SGD Optimization Algorithm, Adam and Adadelta will be applied to nodes for weight and bias optimization with the default Learning rate using the softmax activation function according to the desired number of classes, in this study 2 classes mean having 2 neurons. The class of the image is then classified based on the value of the neurons on the hidden layer by using the softmax activation function.

Figure 5 shows a flow chart of the test stage of the system. The testing stage is in the form of a sex classification process by testing test image data and comparing it with the training result model of training training image data stored in the database. The image data taken was 2000 for the original data and then 4000 augmented image data. The image that has been taken will be processed by the CNN algorithm until it then produces a system output in the form of gender class information.

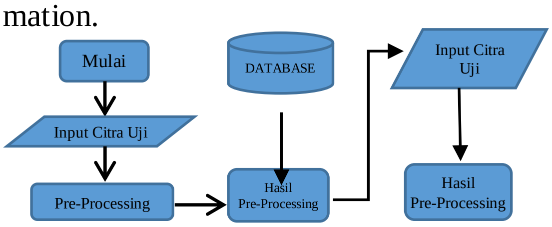

Image 5. System Testing Stage Flow Diagram

RESULTS AND DISCUSSION

In this section we describe the proposed solution as a selected Convolutional Network (ConvNet) Architecture and discuss related design options, evaluation methods and implementation aspects.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 345 - 354 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.2067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

The data set provided by Kaggle has 5000 train data and 1500 test data. All images vary in size by 96 dpi. However, this study used a modification in the number and size because it avoided the process of running data that was too long to become 800 data trains and 400 test data with a size of 64 x 64 pixels. A sample of male images can be seen in Figure 2 and a female sample can be seen in figure 3, as well as the distribution of data in figure 6 and 7 [17].

Image 6. Sample Images of males

Image 7. Sample Female Image

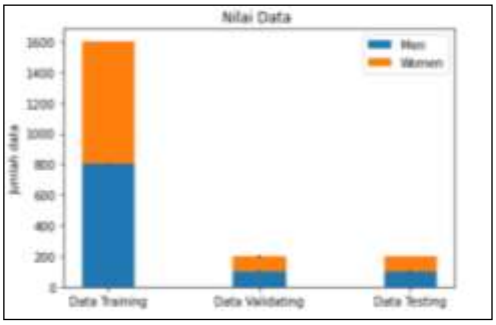

Image 8. Data Spread

Competition is a binary classification problem with the area under the ROC curve between the predicted probability and the observed target as an evaluation matrix, because we want the output to be as close as possible to the actual probability, so we use data augmentation i.e. transform before the image is processed. In addition, researchers used the NLL function (loss with Logsoftmax) as the last layer activation function.

To evaluate the model, researchers must optimize the dataset for the right optimizer between the adadelta, SGD, Nadam optimizer methods and Adam's method. In this experiment, researchers used one of the optimization functions, namely Nadam. The researcher observed a fairly good accuracy of 89.5% with a loss of 0.21 shown in figure 4, and the researcher concluded that the optimization method using Nadam was not suitable for this dataset, and we decided to use Adam's optimization method as a substitute.

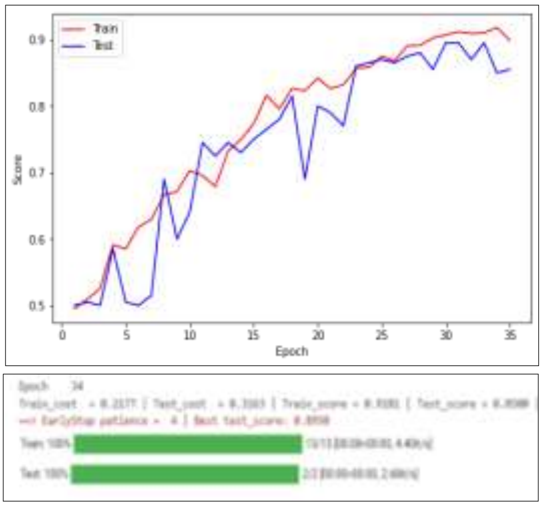

Image 9. Plot and Accuracy Results Using Nadam Optimizer

To have a basis that can be used as a comparison of how well the expected model is doing reasonably, the researchers used pre-trained VGG-19 table 1:

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 345 - 354 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.2067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Table 1. Pre-Trained VGG19

No	Layer	Output Shape
1	Batch Size	128
2	Crop size	64
3	Input Layer	3 x 64 x 64
4	nn.model	64 x 4 x 4
5	Global Average pooling 2d layer	3, 8, 3, 1, 1
6	Dropout	10 %

In all the experiments that the researchers have done, the only changes made are to change the model layer and modify the image size on the input.

Because the dataset is relatively small, researchers had to use learning transfer with a set of models using an experiment-focused CNN model using multiple optimized functions [17]. There are some general steps to take, the first step is to make a size modification to the image before it is used as input, this is necessary to avoid too long a process. In this case, the researcher uses the following sequence:

a. Global Average Pooling2D layer with sigmoid at convolution.
b. Dropout layer.
c. And the last layer is a Fully Connected Layer with an output size of 1 (for binary classification) and an activation function (softmax).

To maximize some training examples and improve model accuracy, researchers augmented the data through a number of random transformations that the data augmentation technique chose was: random rotation (150), resize crop scale = (80%) random horizontal flip. Furthermore it is expected that data augmentation should also avoid overfitting, as well as the ability to improve in model generalizations [18]. Researchers have experimented with several optimizer models from python libraries, namely Adam, Adadelta, Nadam and SGD on CNN models with reported results. In all the experiments that researchers have conducted, several different accuracy results have been obtained.

a. Adam

Adam Is a replacement optimization algorithm for stochastic gradient descent for training deep learning models. Adam combines the best properties of the AdaGrad and RMSProp algorithms to provide an optimization algorithm that can handle sparse gradients on noisy problems. In this optimization model, researchers get an accuracy mode of 91.5% with a loss of up to 0.1% with the best learning rate of 1.0 according to figure 10.

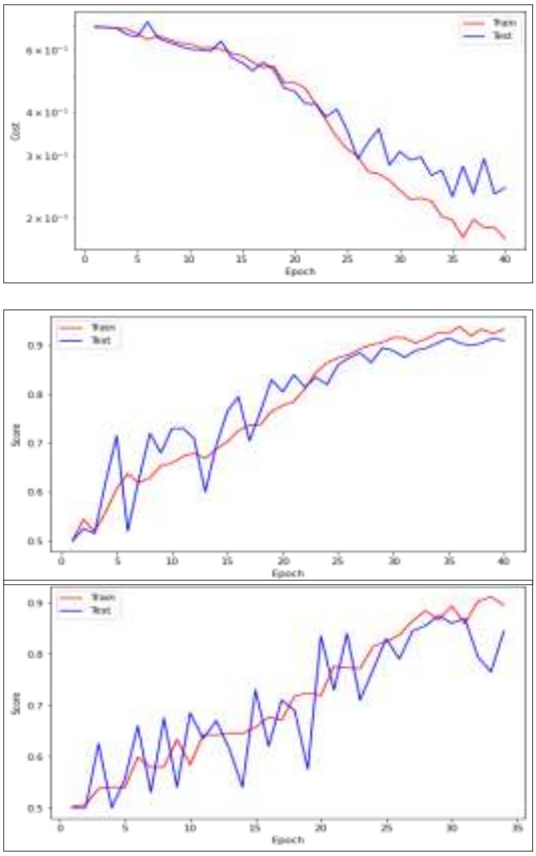

Image 10. Adam's Plot Cost and Score Results

DOI: https://doi.org/10.33330/jurteksi.v9i3.2067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

b. Adadelta

Adadelta Is a stochastic gradient lowering method that is based on adaptive learning rates per dimension to overcome. In the second model we get an accuracy of 0.90%, as seen in the following figure 11.

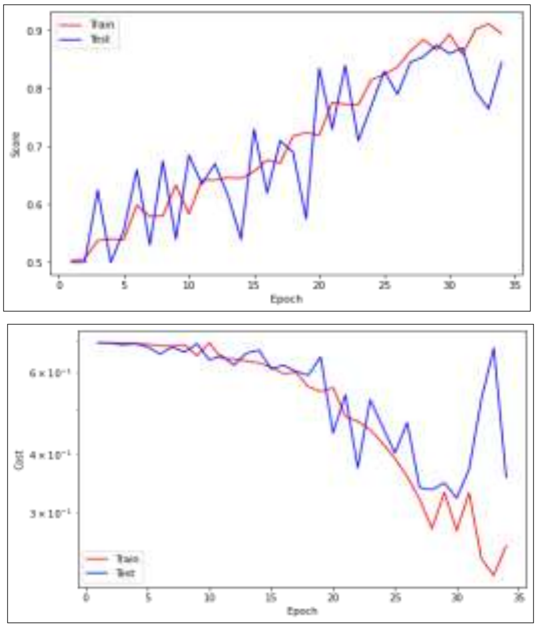

Image 11. Adadelta Plot Cost and Score Results

c. Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent (SGD) Is the solution to solve the issue of GD. SGD will update the weight without waiting for 1 epoch to finish. SGD uses a concept similar to batching by dividing the train data into batches [19]. The weight will be updated in each completed batch. In this optimizer method, we get an accuracy of 63% with a still high loss of 0.6 with the best learning rate of 0.01. This is no better than the two previous optimizer models, corresponding in figure 12.

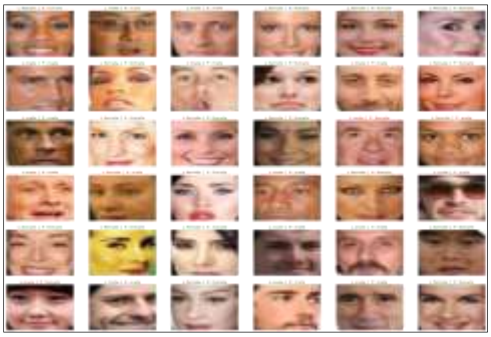

Image 12. Adadelta Plot Cost and Score Results

d. Data Visualization

To see how the model works and what exactly is learned, we chose to visualize intermediate activations consisting of the results of map features ejected by various convolution and unification layers in the network, given a specific optimizer input (the output of a layer is often called activation, the output of the activation function). Output the original image prediction data with predictions as shown in Figure 8, the model learned how to identify male and female genders using Adam's optimizer, especially on the face [20].

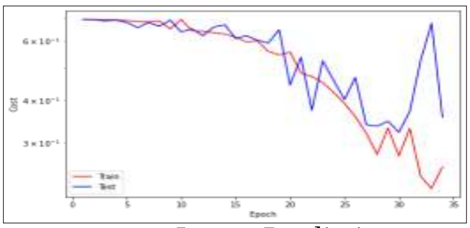

Image 13. Image Prediction

From figure 13 there are 36 predicted data and produced 32 correct predictions (green writing) and 4 wrong predictions (red writing). From the data, you can find the MSE with a confusion matrix, especially precision manually with the formula:

Precision = TP/(TP+FP) = 36/(36+4) = 36/40 = 0,9 setara 90%.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 345 - 354 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.2067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

The 36 predicted data can also be visualized as shown in figure 14 below:

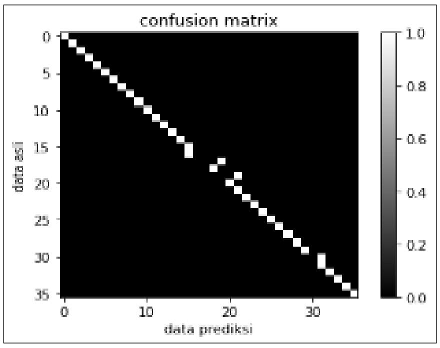

Image 14. Matrix Confusion Visualization

Table 2 shows the results of experiments that have been carried out using adam, adadelta, and SGD models in classifying gender.

Table 2. Experimental Results Adam

Model	Tes
Adam Epoch 31 Train_cost = 0.1945 | Test_cost = 0.2569 | Train_score = 0.9269 | Test_score = 0.9100 | ==> EarlyStop patience = 5 | Best test_score: 0.9100 ==> Execute Early Stopping at epoch: 31 |	Best test_score: 0.9100

Table 3. Experimental Results Adadelta, SGM

Model	Tes
Adadelta **Epoch 42** **Train_cost = 0.1908 |** **Test_cost = 0.4294 |** **Train_score = 0.9219 |** **Test_score = 0.7900 |** **==> EarlyStop patience = 5 |** **Best test_score: 0.9050** **==> Execute Early Stopping at epoch: 42 |**	Best test_score: 0.9050
SGD **Epoch 17** **Train_cost = 0.6926 |** **Test_cost = 0.6925 |** **Train_score = 0.5425 |** **Test_score = 0.5950 |** **==> EarlyStop patience = 5 |** **Best test_score: 0.6350** **==> Execute Early Stopping at epoch:17**	Best test_score: 0.6350

CONCLUSION

Based on the application of the method that has been carried out, the Adam Optimization method more accurately identifies gender, with an accuracy of 91.5%. There were 36 predicted data and produced 32 correct predictions and 4 false predictions. The application of this method can provide solutions in helping gender classification quickly, and more precisely compared to the SGD method, and Adadelta[21].

BIBLIOGRAPHY

[1] J. Ho *et al.*, “Imagen Video: High Definition Video Generation with Diffusion Models,” pp. 1–18, 2022, [Online]. Available: http://arxiv.org/abs/2210.02303

[2] F. D. Adinata and J. Arifin, “Klasifikasi Jenis Kelamin Wajah Bermasker Menggunakan Algoritma Supervised Learning,” *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 229, 2022, doi: 10.30865/mib.v6i1.3377.

[3] A. M. Rizki, G. E. Yuliastuti, E. Y. Puspaningrum, and A. Lina, “KLASIFIKASI JENIS KELAMIN BERDASARKAN CIRI FISIK,” vol. XVII, pp. 1–5, 2022.

[4] C. Neural, N. Cnn, R. Firdaus, and J. Satria, “Jurnal Computer Science and Information Technology (CoSciTech) Klasifikasi Jenis Kelamin Berdasarkan Gambar Mata Dengan Menggunakan Algoritma,” vol. 3, no. 3, pp. 267–273, 2022.

[5] Y. Dong, Q. Liu, B. Du, and L. Zhang, “Weighted Feature Fusion

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 345 - 354 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.2067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification," *IEEE Trans. Image Process.*, vol. 31, no. January, pp. 1559–1572, 2022, doi: 10.1109/TIP.2022.3144017.

[6] D. Irfan, R. Rosnelly, M. Wahyuni, J. T. Samudra, and A. Rangga, "Perbandingan Optimasi Sgd, Adadelta, Dan Adam Dalam Klasifikasi Hydrangea Menggunakan Cnn," *J. Sci. Soc. Res.*, vol. 5, no. 2, p. 244, 2022, doi: 10.54314/jssr.v5i2.789.

[7] D. Frenza and R. Mukhaiyar, "Aplikasi Pengenalan Wajah Menggunakan Metode Adaptive Resonance Theory (ART)," *Multidicsiplinary Res. Dev.*, vol. 3, no. 1, pp. 35–42, 2021, [Online]. Available: https://doi.org/10.31933/rrj.v3i3.392

[8] M. Ula, A. Faridhatul Ulva, M. Abdullah Ali, Y. Rilasmi Said, and S. Informasi, "Application of Machine Learning in Determining the Classification of Children'S Nutrition With Decision Tree," *J. Tek. Inform.*, vol. 3, no. 5, pp. 1457–1465, 2022, [Online]. Available: https://doi.org/10.20884/1.jutif.2022.3.5.599

[9] S. W. P. Listio, "Performance of Deep Learning Inception Model and MobileNet Model on Gender Prediction Through Eye Image," *Sinkron*, vol. 7, no. 4, pp. 2593–2601, 2022, doi: 10.33395/sinkron.v7i4.11887.

[10] B. Hardiansyah and A. Primasetya, "Sistem Deteksi Penggunaan masker (Face Mask Detection) Menggunakan Algoritma Deep Learning YOLOv4," vol. 2, pp. 313–318, 2023.

[11] S. H. Abdullah, R. Magdalena, and R. Y. N. Fu'adah, "Klasifikasi Diabetic Retinopathy Berbasis Pengolahan Citra Fundus Dan Deep Learning," *J. Electr. Syst. Control Eng.*, vol. 5, no. 2, pp. 84–90, 2022, doi: 10.31289/jesce.v5i2.5659.

[12] Bimrew Sendekie Belay, "No Title הכי קשה לראות את מה שבאמת *הארץ*לנגד העינים," vol. 7, no. 8.5.2017, pp. 2003–2005, 2022.

[13] HAMDANI MUBAROK, "Menggunakan Algoritma Convolutional Neural Network (CNN)," *Skripsi*, vol. 4, no. 2, pp. 89–96, 2019.

[14] 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, p. 107, 2018, doi: 10.36055/setrum.v7i1.3367.

[15] F. Mignacco and P. Urbani, "The effective noise of stochastic gradient descent," *J. Stat. Mech. Theory Exp.*, vol. 2022, no. 8, pp. 1–15, 2022, doi: 10.1088/1742-5468/ac841d.

[16] Y. Kong, X. Ma, and C. Wen, "A New Method of Deep Convolutional Neural Network Image Classification Based on Knowledge Transfer in Small Label Sample Environment," *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22030898.

[17] R. Rosnelly, L. Wahyuni, and E. Aditya, "Pelatihan Pengenalan Teknik Pengolahan Citra Digital

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 345 - 354 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.2067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Pada Bidang Medis Training Introduction to Digital Image Processing Techniques In The Medical Field,” *JUDIMAS J. Inov. Pengabdi. Kpd. Masy.*, vol. 3, no. 1, pp. 11–19, 2022, [Online]. Available: https://stmikpontianak.ac.id/ojs/index.php/judimas/article/view/1282

[18] K. C. Leowis, J. Raharjo, N. Ibrahim, U. Telkom, F. Detecion, and P. C. Analysis, “Rancang Bangun Sistem Pengenalan Wajah Diarea Publik Berbasis Video Menggunakan Metode Principal Component Analysis (Pca) Dan Viola Jones Design of the Public Face Recognition System Based on Video Using Principal Component Analysis (Pca) and Viola Jo,” vol. 8, no. 5, pp. 5448–5464, 2021.

[19] F. Tangguh and Y. Islami, “Analisis performa algoritma Stochastic Gradient Descent (SGD) dalam mengklasifikasi tahu berformalin,” *Indones. J. Data Sci.*, vol. 3, no. 1, pp. 1–8, 2022.

[20] M. Milano, G. Agapito, and M. Cannataro, “Application of CCTV Methodology to Analyze COVID-19 Evolution in Italy,” *BioTech*, vol. 11, no. 3, pp. 1–22, 2022, doi: 10.3390/biotech11030033.

[21] Z. Lou, W. Zhu, and W. B. Wu, “Beyond Sub-Gaussian Noises: Sharp Concentration Analysis for Stochastic Gradient Descent,” *J. Mach. Learn. Res.*, vol. 23, pp. 1–22, 2022.