Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 5 (2022) 790 - 796 ISSN Media Electronic: 2580-0760

Classification of Face Mask Detection Using Transfer Learning Model DenseNet169

Lidya Fankky Oktavia Putri[1], Ahmad Junjung Sudrajad[2], Vinna Rahmayanti Setyaning Nastiti[3]
[1,2,3]Informatics, Faculty of Engineering, University of Muhammadiyah Malang
[1]fankkylid22@gmail.com, [2]junjungahmad@gmail.com, [3]vinastiti@umm.ac.id*

Abstract

COVID-19 has become a threat to the world because it has spread throughout the world. The fight against this pandemic is becoming an unavoidable reality for many countries. The government has set policies on various transmission prevention efforts. One of these efforts is for everyone to wear masks in order to break the transmission chain. With such conditions, the government must continue to monitor so that people can apply the appeal in their daily lives when participating in outdoor activities. The present time involves new problems in so many fields of information technology research, especially those related to artificial intelligence. The purpose of this study is to discuss the classification of face image detection in people who wear masks and do not wear masks. designed using the Convolutional Neural Network (CNN) model and built using the transfer learning method with the DenseNet169 model. The model used is also combined with the DenseNet169 transfer learning method and the fully connected layer model architecture, so as to optimize the performance test in the evaluation. These models were trained under similar conditions and evaluated on benchmarks with the same training and validation images. The result of this research is to get an accuracy value of 96% by combining the two datasets. This dataset is the same as previous research; the number of datasets is 8929 images.

Keywords: *classification, transfer learning, COVID-19, DenseNet169, face mask*

1. Introduction

Covid-19 is a virus that emerged in Wuhan in December 2019. The virus has spread throughout the region. The virus has been designated as a plague because it has attacked all countries on Earth and has become a global threat [1]. The COVID-19 virus can spread easily through droplets coming out of the mouth or nose of someone who has been infected by sneezing or coughing [2]. Droplets containing the virus can be attached to objects that are often touched by humans, so that the virus can be transferred to other human bodies that are still in good health [3]. Health is the most important point in the post-pandemic recovery prioritization process because health and economics are interrelated [4]. With the occurrence of disease outbreaks such as COVID-19, it has an impact on all communities, without exception, in different parts of the country. The impact of this virus results in several losses, especially in terms of economy, because people are restricted from doing outdoor activities and are restricted from doing activities that are crowded. Therefore, governments around the world have issued policies on various transmission prevention efforts. One of the efforts that has been agreed upon by the world is the use of masks by the entire community [5]. The use of masks in the community can be likened to a motorcyclist who is careful when driving his vehicle on the road or to the use of seat belts for car drivers [6]. Although the COVID-19 pandemic will turn into an endemic, life in society still requires significant care to avoid all the bad possibilities that will occur.

In responding to the pandemic's change to endemic, a policy is needed from the government, namely, monitoring the community. This monitoring must continue to be applied by the government to people who do not use masks by applying information technology to identify people who use masks and do not use masks through the use of artificial intelligence. One of the applications of artificial intelligence is using image processing methods and pattern recognition technology. Within the scope of artificial intelligence, there is a machine learning method that has been developed into

Accepted: 25-07-2022 | Received in revised: 10-09-2022 | Published: 31-10-2022

a deep learning method. The deep learning method is better when compared to the machine learning method [7]. There are many algorithms that can be implemented, such as Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM), Recurrent Neural Networks (RNN), Generative Adversarial Networks (GAN), Multilayer Perceptrons (MLP), and others.

This research will use the Convolutional Neural Network (CNN) algorithm, which is one of the models of the Deep Learning method [8]. CNN is a widely applied method to obtain feature information on an image. This method has high accuracy for obtaining complete feature information on an image through the training process of the training image [9]. For image classification, this CNN model is often used for classification of a large number of images [10]. CNN has significant conceptual frameworks, such as weight sharing, perception, and domain sampling space, to guarantee relative displacement, distortion, and scaling characteristics [10]. CCN has also undergone many developments in various fields. One of the results of these developments is the ImageNet dataset [11], which produces a pre-trained machine learning model. Transfer learning incorporates the pre-trained network [12]. The benefit of transfer learning is that it can learn data well even though the training data is limited [12].

The classification of facial images of people who use masks and people who do not use masks is discussed in research [13]. Research [13] uses datasets from the Github open repository (https://github.com/prajnasb/observations/tree/master/experiments/data) and implements the use of four transfer learning methods, including MobileNetV2-SVM resulting in an accuracy value of 97.11%, VGG19-SVM resulting in an accuracy value of 96.65%, MobileNetV2-KNN resulting in an accuracy value of 94.92%, and XCeption-SVM resulting in an accuracy value of 94.56%. Of the four methods developed again, there are two that are considered the most optimal in the evaluation model on the precision and recall values, namely MobileNetV2-SVM, which produces a recall value of 94.84% and a precision value of 95.08%, while VGG19-KNN produces a recall value of 90.09% and a precision value of 91.3%.

Research [14] also discusses case studies of facial image classification using masks and not using masks. However, research [14] uses datasets from Kaggle (https://www.kaggle.com/datasets/omkargurav/face-mask-dataset) and uses several transfer learning methods, namely CNN-MobileNetV2 with an accuracy result of 0.981, CNN-DenseNet201 with an accuracy result of 0.987, CNN-VGG16 with an accuracy result of 0.582, and CNN-Xception with an accuracy result of 0.988.

The contribution of the research to be designed is to develop from research [13][14] using datasets sourced from Github and Kaggle, but the method used is different, namely using DenseNet169 as a Transfer Learning architecture. DenseNet169 is also combined with a fully connected layer architecture for the classification of facial images of people who use masks and do not use masks. Then the two architectures are combined with the aim of knowing the level of accuracy value and evaluation value using the dataset, and they are compared with the methods used in previous studies.

2. Research Methods

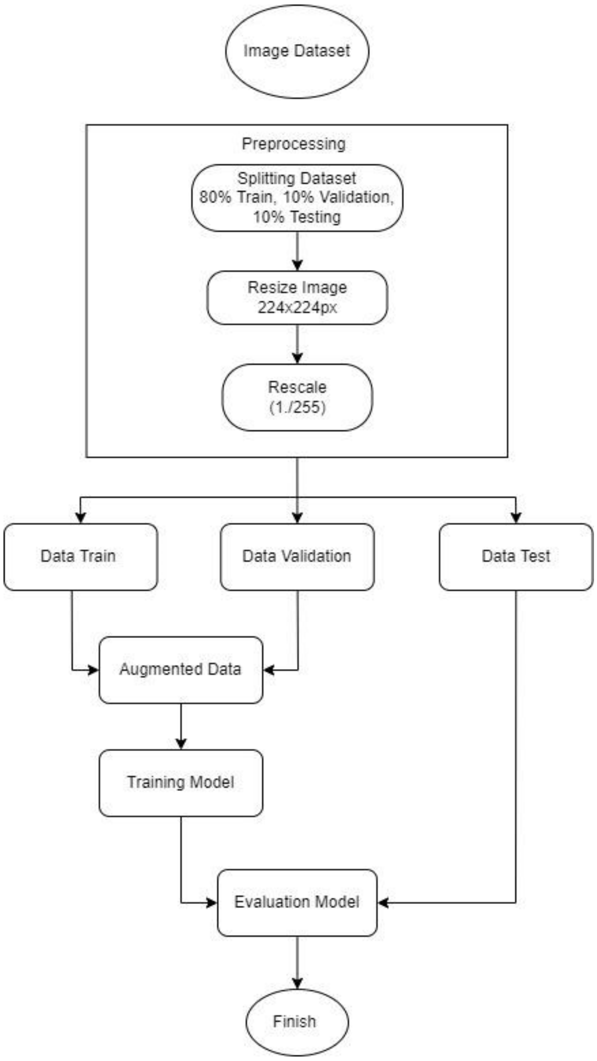

Figure 1. Research Methods

The research stages at this time can be seen in Figure 1, which shows that there is a research flow from beginning to end. There are several processes, namely data processing, where at this stage there is a division of data into training, validation, and testing data. After that, the data is resized into pixels and given a rescaling

DOI: https://doi.org/10.29207/resti.v6i5.4336

approach. After that, the train data and validation data will perform the augmentation process.

Then, after the augmentation process is complete, proceed with the process of creating and training the model, which, at the end of the process, will continue with the evaluation process. In this evaluation process, the data from the testing data will be processed, and the existing model process from the training and validation data will be applied.

2.1 Dataset

In this research stage, it will explain the use of datasets. The dataset used is the same dataset from previous research [13][14]. This dataset is also available on the open repository site, namely Github, which can be accessed at the link (https://github.com/prajnasb/observations/tree/master/experiements/data) and Kaggle, which can also be accessed at the link (https://www.kaggle.com/datasets/omkargurav/face-mask-dataset). Datasets originating from Github have a total of 1376 images consisting of two classes, namely With Mask and Without Mask, and datasets originating from Kaggle have a total of 7553 images consisting of two classes, namely Mask and No Mask, without any division of data groups. A clearer distribution of data can be seen in Table 1 below.

Table 1. Dataset Distribution

Group	Class	Quantity
Data Trainig	WithMask	3532
	WithoutMask	3611
Data Validasi	WithMask	441
	WithoutMask	451
Data Testing	WithMask	442
	WithoutMask	452

After getting the dataset from the Github repository, the dataset is immediately divided by the rules of 80% training data, 10% validation data, and 10% test data. Figure 2 below is a sample image from the dataset.

Figure 2. Dataset Image Sample

2.2 Datasets Processing

The process of dividing the dataset to be used into several stages, including resizing the image and providing augmentation to the training data and validation data, Meanwhile, the test data is only given a processing in the form of resizing the image, namely $rescale\ = 1/225$ and $resize\ = 224{,}224$. The use of data augmentation is a process of enriching training data, which aims to avoid overfitting [15].

The data augmentation process has a preprocsessing_function parameter and also uses ImageDataGenerator as the augmentation parameter function class. The use of data augmentation functions in this processing comes from the TensorFlow library. The methods used in the function parameters in the augmentation process are (1) $rescale\ = 1/225$, (2) $rotation\_range\ = 40$, $width\_shift\_range\ = 0.2$, (3) $height\_shift\_range\ = 0.2$, (4) $shear\_range\ = 0.2$, (5) $zoom\_range\ = 0.2$, (6) horizontal_flip = “True”, and finally (7) fill_mode = “Nearest”.

Figure 3. Sample Augmented Image Result

2.3 Model Architecture

This research applies the transfer learning method to model building. The selected model is DenseNet169, coupled with the addition of layer architecture in the Fully Connected Layer section. In Figure 4 below, there are 2 architectures carried out: the first pre-trained architecture of the DenseNet169 model is used as a feature extraction layer located in the Densnet169 input layer, and the second fully connected layer architecture is located after the pre-trained architecture to the end.

This created architecture, which can be seen in Figure 4, shows the arrangement of layers that form an architecture, namely: (1) Densnet169 input layer, (2) Flatten layer, (3) Dense layer using “relu” activation of 256, (4) BatchNormalization layer, and (5) Dropout layer of 0. 6: (6) Dense layer using “relu” activation of 512; (7) BatchNormalization layer; (8) Dropout layer of 0.6; (7) Dense layer using “relu” activation of 512; (8) BatchNormalization layer; (9) Dropout layer of 0.6; and the last layer, (10) Dense layer using “sigmoid” activation with 2 class outputs.

This architecture also uses ReLu activation in the Fully Connected Layer architecture where ReLu activation aims to generalize the value generated by the convolutional layer [16]. While the use of optimizers in this architecture is to use the "Adam" optimizer technique, which aims to be able to influence the weight value in the model used [17], The use of binary_crossentropy as a loss function in this case aims to facilitate the training process and also solve many classification problems at the same time [18].

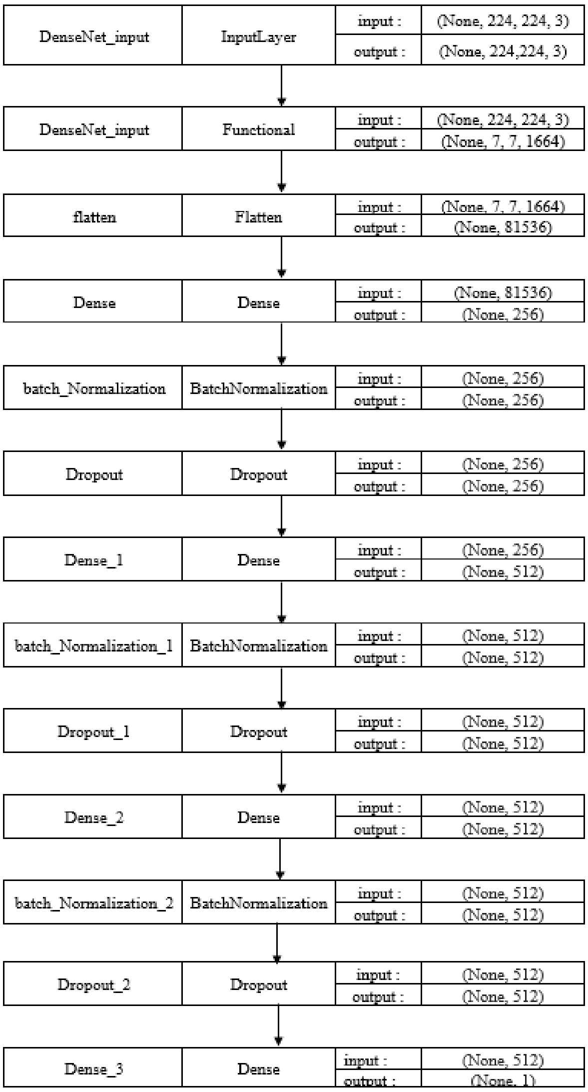

Figure 4. Proposed Methods

2.4 Test Scenario

This dataset is divided into two classes, namely With Mask and Without Mask. There are approximately 8929 image data datasets used. The data is then divided into 3 parts, namely training data, validation data, and testing data. The distribution of data can be seen in Table 1. The data that will be processed in testing the method is only training and validation data, which will produce h5 files. The file will be used to run the evaluation process using testing data as a model load process.

This dataset test also applies the callback function, with the aim of saving time in the training process [19] and the application of the EarlyStopping technique in the epoch process or model training iteration. The application of callback functions in this scenario is used when the training process uses two callback functions. The first callback function has the purpose of monitoring the val_loss value, the loss index level of the validation data.

The use of the second callback function has a role for storing the training model when it finishes training each epoch with the val_accuracy value parameter. The use of the val_accuracy parameter aims to monitor the increase in the accuracy index value for each epoch process on validation data. The epoch model used utilizes EarlyStopping which aims to alleviate computational problems and to execute several experiments performed [20]. The use of EarlyStopping in this scenario uses the val_loss parameter as in the callback function with a patience amount of 25. This patience serves as a determinant of when the epoch amount stops the training process if there is no improvement in the intended parameters.

3. Results and Discussions

This stage explains the results of the discussion related to the proposed method, with a discussion of data processing and the final discussion in the form of evaluating the values of accuracy, recall, precision, and F1-Score. In the final result of the evaluation value, namely the accuracy value, which will be compared with previous research using the same dataset and in terms of the same case study, namely the classification of people's faces using masks and not using masks.

3.1 Dataset Processing

The first stage is to get a dataset that has been provided from previous research. After that, it is processed first before being given model training. The earliest data processing is done by dividing the dataset distribution into 80% training data, 10% validation data, and 10% test data. The results of the dataset division used can be seen in Table 1.

After the dataset is divided into training, validation, and test data, the dataset will be given the next processing approach in the form of augmentation. In this augmentation process, it will produce duplicate data that is different from the initial data by changing the image to rotate by 40 degrees as in Figure 3. In data processing, it is also given a change in image size to 224×224 pixels. In this augmentation process, only training and validation data are used, while test data is not given an augmentation approach and only given a change in image size to 224×224 pixels.

After the data is ready to be used for model training, the model training process will be given using two functions, namely callback using the EarlyStopping function and callback using the ModelCheckpoint function. The first callback uses the EarlyStopping feature of the Tensorflow library; in this first model training, it uses the value of the vall_loss matrix as a trigger to terminate the training process if it has not changed for 25 iterations of model training. The second

DOI: https://doi.org/10.29207/resti.v6i5.4336

callback is done by using the ModelCheckpoint function feature from the same library as the first callback function but using the vall_accuracy matrix value as a parameter for the stopping value of the model training process. Furthermore, the model will be trained with 100 training iterations (epochs) and also apply the two callback functions above.

3.2 Graph Evaluation

As seen in Figure 5, the model training process using training data has a blue line color and validation data has an orange line color, resulting in two accuracy values from the data used, namely the accuracy of the training data and the accuracy of the validation data. In Figure 5, it can be seen that the accuracy value of the validation data gets a value above 0.98 at epoch 0 and gets the best accuracy value at the 34th epoch with a value of 0.99769. While the accuracy value of the training data gets an accuracy value that develops according to the number of epochs running, the final result is a best accuracy value of 0.9944 at the 48th epoch.

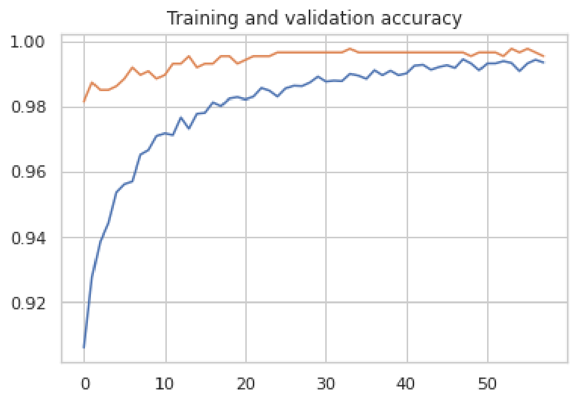

Figure 5. Graph of Model Accuracy

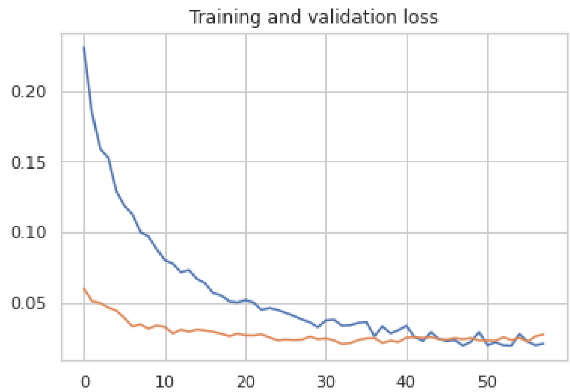

Figure 6. Loss Model Graph

In Figure 6, it can be seen that the loss graph of the model made on the validation data, namely the orange line, has reached 0.05 when training the model at the 5th epoch and experienced the best increase in loss at the 33rd epoch with a value of 0.0202, while the train data has a loss value that is getting smaller following the magnitude of the epoch, where the loss value reaches a number below 0.0499 at the 20th epoch and gets the best loss value of 0.190 at the 48th epoch. This can be seen in the blue line in Figure 6.

3.3 Accuracy, Recall, Precision, and F1-Score

At this stage, I will explain the results of the model evaluation by applying the test data in Table 1 to the h5 file model results. The evaluation result parameter has four value points, namely accuracy, recall, precision, and F1-Score. As can be seen in Figure 7, the result of the score evaluation using the confusion matrix model from the SkLearn library, which aims to measure the performance level of the proposed method model stored in the h5 file.

With the evaluation of the model through a confusion matrix, it is intended to measure the level of work of a machine learning method to find out how many values will be issued from correct or incorrect data predictions. The number of predictions comes from the total amount of data to be tested. The h5 file will be used to run the evaluation process using testing data and to find out what the validity value of each correct or incorrect prediction is from the total test data that will be tested.

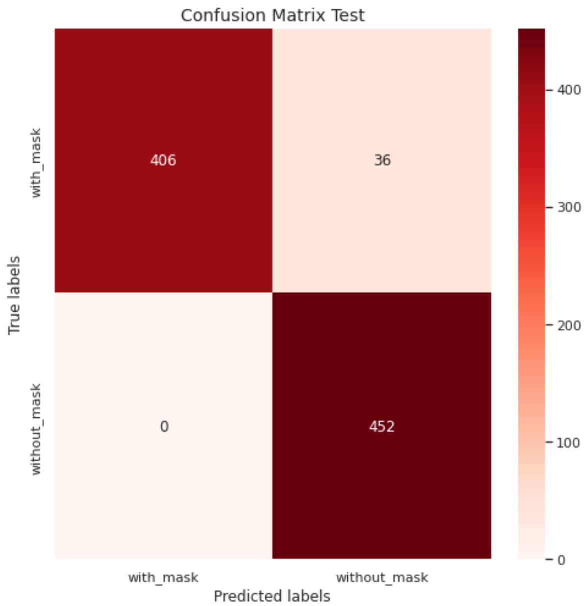

Figure 7. Confusion Matrix Model

In Figure 7 above, it can be seen that the prediction results of the model built get a validity value in the prediction of each class, namely with_mask and without_mask. Where the result of the accuracy rate value is 96% on the with_mask label. These results found false predictions as many as 36 images in the test data when determining the classification of images using masks, while the classification of images not using masks predicted correctly as a whole.

These results are parameterized by the precision, recall, and F1-Score of the evaluation values. The value of the model results from previous research can be interpreted as a measure of positive values in images that are correctly predicted, with a total number of positive predictions. The F1-Score value can be interpreted as a measure that has a value to offset the value of precision and recall evaluations. While the recall value can be interpreted as a measure of positive classes formed from the outcomes of all positive predictions, The three values produce a value of 0.96 on each average aspect, as can be seen in Figure 8.

```
              precision    recall  f1-score   support

   with_mask       1.00      0.92      0.96       442
without_mask       0.93      1.00      0.96       452

    accuracy                           0.96       894
   macro avg       0.96      0.96      0.96       894
weighted avg       0.96      0.96      0.96       894
```

Figure 8. Precision, Recall, and F1-Score Value

Previous research also used several transfer learning methods as training models in the classification of face images using masks and not using masks. The models used in the first previous research were MobileNetV2-SVM with an accuracy value of 97.11%, VGG19-SVM with an accuracy value of 96.65%, MobileNetV-KNN with an accuracy value of 94.92%, and Xception-SVM with an accuracy value of 94.56%, which came from previous research. Then the second previous research is done by using several models, namely CNN-MobileNetV2 with an accuracy result of 0.981, CNN-DenseNet201 with an accuracy result of 0.987, CNN-VGG16 with an accuracy result of 0.582, and CNN-Xception with an accuracy result of 0.988.

Table 2. Model Results from previous research [13][14]

Author	Model	*Accuracy*
(Abdellah Oumina, 2020) [13]	MobileNetV2-SVM VGG19-KNN MobileNetV2-KNN Xception-SVM	0.971 0.965 0.949 0.945
(Muhammad Farid, 2021) [14]	CNN-MobileNetV2 CNN-DenseNet201 CNN-VGG16 CNN-Xception	0.981 0.987 0.582 0.988

The results of the Precision, Recall, and F1-Score values of the designed model get a value of 96% in all three aspects. In research [13][14], the results of model evaluation can be seen in Table 2. A comparison of the results of model evaluation between the proposed method and previous research can be seen in Table 2 above.

Table 3. Proposed Method and Comparison between models

Dataset	*Model*	*precision*	*recall*	*akurasi*
Github	MobileNetV2-SVM [13]	0.97	0.94	0.95
	VGG19-KNN [13]	0.96	0.90	0.91
Kaggle	CNN-MobileNetV2 [14]	0.98	0.98	0.98
	CNN-DensNet201 [14]	0.98	0.98	0.98
	CNN-VGG15 [14]	0.58	0.58	0.58
	CNN-Xception [14]	0.98	0.98	0.98
Table 1	**CNN-DenseNet169** (Proposed Method)	**0.96**	**0.96**	**0.96**

4. Conclusion

The author examines the classification of facial images of people who use masks and those who do not use masks using the pre-trained Deep CNN method. The results of this study use the DenseNet-169 model as a pre-trained Deep CNN model by combining the Fully Connected Layer architecture, which can provide the results of the accuracy performance value in the classification report of 96% using testing data. The increase in value can occur due to the use of the DenseNet169 model and the addition of the Fully Connected Layer architecture designed, namely: (1) DenseNet169 Input layer, (2) Flatten layer, (3) Dense layer using “relu” activation of 256, (4) BatchNormalization layer, (5) Dropout layer of 0.06, (6) Dense layer using “relu” activation of 512, (7) BatchNormalization layer, (8) Dropout layer of 0.6, (7) Dense layer using “relu” activation of 512, (8) BatchNormalization layer, 9) Dropout layer, and the last layer using "sigmoid" activation with 2 class outputs. The use of the architecture model or proposed method produces an accuracy value of 0.96 by combining the two datasets in the case of face classification using masks and not using masks. For further research, it is recommended to use a more varied dataset related to the shape or model of the mask and a larger number of images, so as to optimize the research generated in the application of various machine learning methods.

Reference

[1] Melly Damara Chaniago, Amellia Amanullah Sugiharto, Qhistina Dyah Khatulistiwa, Zamah Sari, and Agus Eko, “Covid-19 Detection Using Convolutional Neural Networks (CNN) Classification Algorithm,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 2, pp. 190–197, 2022, doi: 10.29207/resti.v6i2.3823.

[2] R. Wihandika, “Deteksi Masker Wajah Menggunakan Metode Adjacent Evaluation Local Binary Patterns,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 705–712, 2021, doi: 10.29207/resti.v5i4.3094.

[3] M. Ikbal, S. Andryana, and R. T. Komala Sari, “Visualisasi dan Analisa Data Penyebaran Covid-19 dengan Metode Klasifikasi Naïve Bayes,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 5, no. 4, p. 389, 2021, doi: 10.35870/jtik.v5i4.233.

[4] C. Wang *et al.*, “The Association Between Physical and Mental Health and Face Mask Use During the COVID-19 Pandemic: A

Comparison of Two Countries With Different Views and Practices," *Front. Psychiatry*, vol. 11, no. September, pp. 1–13, 2020, doi: 10.3389/fpsyt.2020.569981.

[5] J. T. Atmojo *et al.*, "Penggunaan Masker Dalam Pencegahan Dan Penanganan Covid-19: Rasionalitas, Efektivitas, Dan Isu Terkini," *Avicenna J. Heal. Res.*, vol. 3, no. 2, pp. 84–95, 2020, doi: 10.36419/avicenna.v3i2.420.

[6] P. P. Siregar, R. Sutan, and C. Mourisa, "COVID-19 dan Penggunaan Masker Muka: Antara Manfaat dan Resiko," *J. Implementa Husada*, vol. 1, no. 3, pp. 221–231, 2020.

[7] N. Audebert, B. Le Saux, and S. Lefevre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, 2019, doi: 10.1109/MGRS.2019.2912563.

[8] A. TiaraSari and E. Haryatmi, "Penerapan Convolutional Neural Network Deep Learning dalam Pendeteksian Citra Biji Jagung Kering," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 265–271, 2021, doi: 10.29207/resti.v5i2.3040.

[9] R. Prathivi, "Optimasi Model TL-CNN Untuk Klasifikasi Citra CIFAR-10," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 3, pp. 717–722, 2017.

[10] Luqman Hakim, Z. Sari, and H. Handhajani, "Klasifikasi Citra Pigmen Kanker Kulit Menggunakan Convolutional Neural Network," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 379–385, 2021, doi: 10.29207/resti.v5i2.3001.

[11] S. Kornblith, J. Shlens, and Q. V. Le, "Do better imagenet models transfer better?," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 2656–2666, 2019, doi: 10.1109/CVPR.2019.00277.

[12] W. Setiawan, M. I. Utoyo, and R. Rulaningtyas, "Transfer learning with multiple pre-trained network for fundus classification," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 18, no. 3, pp. 1382–1388, 2020, doi: 10.12928/TELKOMNIKA.v18i3.14868.

[13] A. Oumina, N. El Makhfi, and M. Hamdi, "Control the COVID-19 Pandemic: Face Mask Detection Using Transfer Learning," *2020 IEEE 2nd Int. Conf. Electron. Control. Optim. Comput. Sci. ICECOCS 2020*, 2020, doi: 10.1109/ICECOCS50124.2020.9314511.

[14] M. F. Naufal and S. F. Kusuma, "Pendeteksi Citra Masker Wajah Menggunakan CNN dan Transfer Learning," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 6, 2021, doi: 10.25126/jtiik.2021865201.

[15] F. J. P. Montalbo and A. A. Hernandez, "Classification of fish species with augmented data using deep convolutional neural network," *2019 IEEE 9th Int. Conf. Syst. Eng. Technol. ICSET 2019 - Proceeding*, no. December, pp. 396–401, 2019, doi: 10.1109/ICSEngT.2019.8906433.

[16] Y. Achmad, R. C. Wihandika, and C. Dewi, "Klasifikasi emosi berdasarkan ciri wajah wenggunakan convolutional neural network," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 11, pp. 10595–10604, 2019.

[17] C. R. Rahman *et al*., "Identification and recognition of rice diseases and pests using convolutional neural networks," *Biosyst. Eng.*, vol. 194, pp. 112–120, 2020, doi: 10.1016/j.biosystemseng.2020.03.020.

[18]. Usha Ruby Dr.A, "Binary cross entropy with deep learning technique for Image classification," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 5393–5397, 2020, doi: 10.30534/ijatcse/2020/175942020.

[19] Z. Ren, Y. Zhang, and S. Wang, "A Hybrid Framework for Lung Cancer Classification," *Electronics*, vol. 11, no. 10, p. 1614, 2022, doi: 10.3390/electronics11101614.

[20] J. Dodge, G. Ilharco, R. Schwartz, A. Farhadi, H. Hajishirzi, and N. Smith, "Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping," 2020, [Online]. Available: http://arxiv.org/abs/2002.06305

DOI: https://doi.org/10.29207/resti.v6i5.4336