Applied Information System and Management (AISM)
Volume 6, (2) 2023, p. 77-82
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v6i2.31396

Mask Detection App Uses Haar Cascade and Convolutional Neural Network to Alert Comply with Health Protocols

Cahya Rahmad[1], Nurfaidah[2], Supriatna Adhisuwignjo[3], Mamluatul Hani'ah[4]

***Abstract*—This study aims to identify the face of a person whether wearing a mask or not wearing a mask accompanied by an appeal to the importance of wearing a mask. The contribution of this paper to science is to provide an overview of the results of accuracy, precision, recall used by the method used with data that can be accessed by many people, so that it can be developed further or can be compared. This system uses two techniques, namely the classification of whether a person is wearing a mask or not using the Convolutional Neural Network (CNN) model. The architecture used is DenseNet-12 to detect human face objects. The data used has a total of 2332 data sets, 200 of which were retrieved manually as research objects, and the rest were obtained from Kaggle. All data is evaluated using the camera in real-time. The test results show that testing scenario one has the highest score with an accuracy of 85% while testing scenario two gets results of 80%, the precision value in testing scenario one gets results of 75%, and testing scenario two has results of 88%. Scenarios 1 and 2 also have the same recall value of 100%. Based on the data analysis, it can be concluded that the use of the Haar Cascade approach and the Convolutional Neural Network with the DenseNet-121 architecture produces good performance in the case of real-time detection of masked and non-masked facial objects.**

***Index Terms*—Convolutional neural network, haar cascade, mask detection.**

I. INTRODUCTION

In early March 2020 the Covid-19 virus first spread in Indonesia, which became the center of the spread for the first time in Wuhan China in December 2019. The Covid-19 virus had an impact on all sectors, both health and the economy [1]. When an infected person with symptoms coughs, sneezes, talks, or breathes in, small liquid particles can move or spread through the nose or mouth and infect other people. When someone is in close proximity to someone who has been infected with Covid-19, the virus has the potential to spread through air contact in closed spaces. Because people must keep their distance from one another, the COVID-19 pandemic has altered every facet of people's life [2]. The government encourages the use of masks, washing hands, and avoiding crowds as part of efforts to reduce the transmission rate of Covid-19. Until now, the use of masks is still mandatory to reduce the transmission rate of Covid-19. However, many people still use masks that do not comply with these regulations. So, a new step is needed to overcome this problem and have a deterrent effect. Effective and efficient use of technology can make it easier to find faces that are not wearing masks. Because protecting yourself when you are around sick people, whether they are showing symptoms or not, is the main goal [3]. This paper aims to identify whether a person is wearing a mask or not by using haar cascade face detection and detecting mask use by using the CNN DenseNet architecture. If it is detected that someone is not wearing a mask, an alarm will sound automatically.

II. RELATED WORK

The research conducted by [4] had 100% successful functional test results for detecting all data used using Haar cascade detection. This test gives the best results using a scale factor of 1.2, with an accuracy rate of 88.7%, and the worst results when a scale factor is 1.3, with an accuracy rate of 44.9%. The results of the detection process are strongly influenced by the position of the object which must be perpendicular to the image source.

Research by [5], used a Convolutional Neural Network (CNN) in this study to detect masks using techniques and models which include aspects of validation, detection, and tracking using Raspberry Pi, and monitoring using CCTV cameras and producing promising accuracy. Another study was also carried out by [6] using the CNN method, to be able to identify medical masks. The accuracy of the Convolutional Neural Network model in the 100 samples of data analyzed has an average accuracy of 97%. The average of 0.563271 seconds is determined from the 10 data tested and the execution time. From previous studies, the method used in this paper can add results using the haar cascade method for detecting faces and the CNN method with another architecture, namely DenseNet-121. Another research by [7] "Detection User Face mask use method Haar cascade As Prevention COVID", This study found test results on the program can detect the use of the Haar

Received: 6 March 2023; Revised: 21 April 2023; Accepted: 2 June 2023

[1]C. Rahmad, Department of Information Technology, State Polytechnic of Malang, Indonesia (e-mail: cahya.rahmad@polinema.ac.id).
[2]N. Nurfaidah, Department of Information Technology, State Polytechnic of Malang, Indonesia (e-mail: faidahn363@gmail.com).
[3]S. Adhisuwignjo, Department of Electronic Engineering, State Polytechnic of Malang, Indonesia (e-mail: supriatna@polinema.ac.id).
[4]M. Hani'ah, Department of Information Technology, State Polytechnic of Malang, Indonesia (e-mail: mahluatulhaniah@polinema.ac.id).

cascade masks in real time using an external camera with the highest degree of accuracy which is equal to 93.33% with a distance of 40 cm with high light intensity. This research also described the drawbacks of the program i.e farther too the level is getting lower light intensity can be reduced the ability program in doing detection. Also Another research by [8] "Detection Use Mask Face With The Method Of Neural Convolutional Network" Test results in this study using the CNN model against the model that has been trained the number of epochs is 100 epoch. The training data 1400 used the image is divided into 700 images using masks and 700 images that do not use a mask. Testing on the CNN model uses architecture MobileNetV2 already trained to detect faces and obtained an accuracy of 88.53% and a classification exactly 84.45%.

III. RESEARCH METHOD

This study uses DenseNet as an architecture to build a mask and non-mask classification model. The DenseNet that will be used in this research is DenseNet-121 which is proposed to have a 94% accuracy performance when compared to other pre-trained CNN models. DenseNet has the advantage of encouraging the use of return features to strengthen feature propagation, and has fewer parameters [9].

A. *Haar Cascade*

The Viola-Jones algorithm which is known to have a high level of detection is used to detect objects with the Haar Cascade Classification Model, especially in the identification of Faces, Noses, Mouths, and Eyes. Because this method considers only the number of pixels in a square and not the individual pixel values of an image, it has the advantage of being faster computationally. By using the four main methods of Integral Image, Adaboost learning, Cascade Classifier, and the Haar-like feature, to identify objects in photos [10]. Figure 1 shows the features that are generally used in the Haar cascade method.

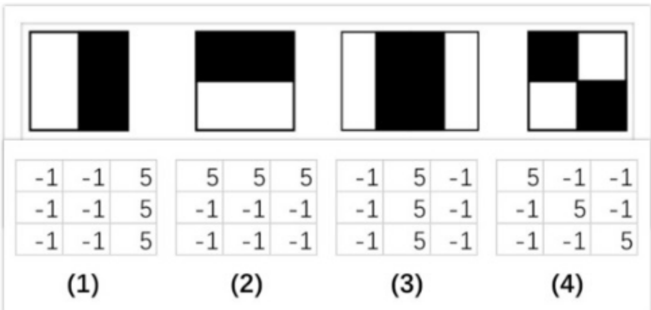

Fig. 1. Features in the Haar cascade

Figure 1 shows the Haar features consist of Edge Features, Line Features, and Four-rectangle features.

B. *DenseNet*

Dense Convolutional Network (DenseNet), or a network in which every other layer or block is connected in a feed-forward manner. DenseNet has some attractive advantages, including eliminating the problem inherent in gradients, better implementation of features, encouraging reuse of used features, and a greatly reduced number of parameters. The DenseNet architecture is depicted in Figure 2.

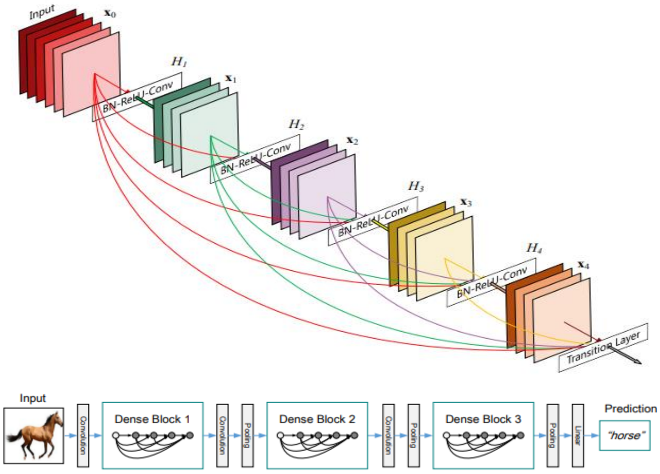

Fig. 2. DenseNet architecture

In Fig. 2, batch normalization, ReLU activation, and convolution with a 3x3 filter are all used in the layer composition for each image of the dataset used. To reduce overfitting during the training process, each block receives input in the form of a matrix based on image pixels before entering the batch normalization procedure. ReLU activation is used to change the x value to 0 if the x value is negative, while the x value is maintained if the value is not less than 0 [11]. Convolution with a 3x3 filter processes an image matrix that has undergone the ReLU activation operation. The result of this multiplication is the matrix value that has been previously processed. The architecture used in this paper is DenseNet-121 which uses 4 blocks, depicted in Figure 3.

Layers	Output Size	DenseNet-121
Convolution	112 × 112	
Pooling	56 × 56	
Dense Block (1)	56 × 56	[1 × 1 conv, 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	
	28 × 28	
Dense Block (2)	28 × 28	[1 × 1 conv, 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	
	14 × 14	
Dense Block (3)	14 × 14	[1 × 1 conv, 3 × 3 conv] × 24
Transition Layer (3)	14 × 14	
	7 × 7	
Dense Block (4)	7 × 7	[1 × 1 conv, 3 × 3 conv] × 16
Classification Layer	1 × 1	

Fig. 3. DenseNet-12 architecture

In DenseNet which uses 4 blocks like DenseNet121, the layers in the dense block consist of batch normalization, ReLU activation, and convolution with a 3x3 pixel filter. The layer between the 2 adjacent blocks is called the transition layer and the feature size changes through convolution and pooling using average pooling techniques. For the classification stage using global average pooling. The next stage is softmax activation.

C. *Convolutional Neural Network*

Artificial Neural Network (ANN) has developed into a Convolutional Neural Network (CNN). CNN Having tens to hundreds of layers make up the CNN network architecture.

Applied Information System and Management (AISM)
Volume 6, (2) 2023, p. 77-82
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v6i2.31396

CNN produces output in a certain layer after processing the image through the network layer. Learning is done at each layer. The output of each layer is used to create the next layer. Layers create basic qualities like color, brightness, and edges at the start of the grid. CNN is an example of deep learning, and deep learning is a subset of machine learning that teaches computers to do tasks that humans often do. as a result of the training data procedure [12].

The convolutional neural network is a feed-forward artificial neural network modeled after the structure of an animal's visual cortex. They can be used for a variety of purposes, including image and video recognition, recommendation generation, and natural language processing. CNN, as demonstrated, is composed of two major components: convolutional layers and pooling layers. A CNN's most significant component is its convolutional layer. It generates feature maps by determining the dot product between a local area and a filter in the feature maps that were received. Each feature map is followed by a nonlinear function, such as the Rectified Linear Unit (ReLU) nonlinearity, which is popular because of its ease of calculation. The pooling layer minimizes the sample count [13].

D. *Confusion Matrix*

Confusion Matrix (CM) is one of the methods used to evaluate classification method [14]. The terms contained in CM] are listed in Table 1. True Negative (TN) values are data that are correctly classified as false negative outputs. True Positive (TP) is data that is correctly classified as a true positive output. False Positive (FP) is data that is classified incorrectly if the output is positive or true. False Negative (FN) is data that is classified incorrectly [15].

Table 1.
Confusion Matrix Terms

Prediction Class (Observations)	Actual Class (Expectation)	
	Positive	Negative
Positive	(TP)	(FP)
Negative	(FN)	(TN)

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

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Equation (1) is the calculation of the average accuracy value to show the level of effectiveness per class of a classification. Equation (2) is the calculation of the average precision value, that is, how much data is classified as a result of how much data is correct. Between the actual value with the prediction given by the system. While (3), which is a calculation of the recall value of all data, is correct how much data comes out in the classification results? Recall evaluation is used if you prefer a False Positive value over a False Negative [16]. *TP*, *FP*, *FN*, and *TN* are described as follows:

TP = True Positive is the amount of positive data that is classified as a positive value
FP = False Positive is the amount of negative data that is classified as a negative value
FN = False Negative is the amount of positive data that is classified as negative
TN = True Negative is the amount of negative data that is classified as a negative value.

E. *Data Collection Techniques*

The data used for implementing the application to detect someone wearing a mask or not wearing a mask uses digital imagery, searching and collecting data is taken directly using a smartphone camera [17], searching on internet media (www.kaggle.com), and obtained by recording a video. An example dataset is shown in Fig. 4. A total of 2232 images will later be divided into training data and test data, the details of which are described in Table 2.

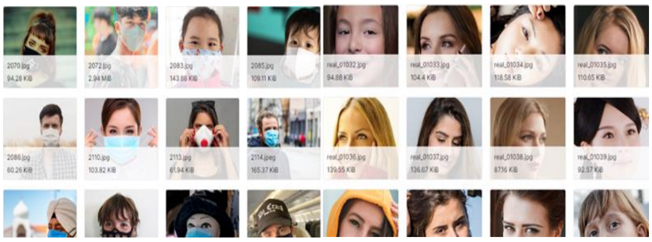

Fig. 4. Sample of dataset

Table 2.
The Datset Used

Source	Image Classification	Description	Number of Images
Dataset 1: search manually by researchers	with_mask dan without_mask	Face with a mask and without a mask.	200
Dataset 2: from the site (Kaggle.com)	with_mask dan without_mask	Face with a mask and without a mask	2132
		Total	2332

Table 2 shows the dataset used, the information in the table contains the division of dataset 1 which was collected manually by the researcher with a total of 200 data, and dataset 2 taken from the kaggle.com website with a total of 2132 data, so the total data is 2332.

F. *Design System*

This mask-wearing detection system has a system design that will be made, in the form of a face detection or recognition process first by providing image input and video input obtained from the camera [18]. After the video is obtained, it will be followed by the recognition process of whether there is a facial object. When it is detected as a face object, the next process is face mask recognition with DenseNet. The mask detection system is shown in Fig. 5, starting from the input image resulting from data known to be a human face object, then it will be processed using the

Haar cascade method, if a face is detected then a face will be tagged and if it is not detected then the input image results from object detection again [19]. After giving a bounding box or marking a face that is detected as a face then cropping is carried out and processed and classified using CNN, if the classification of the nose and mouth is not detected then a person is considered to be wearing a mask, and if the mouth and nose are detected then someone is not wearing a mask, then a warning will be given the vote to immediately wear a mask. The input of this system is an image that has a face in it, and the output is whether the person is wearing a mask or not, if it is detected that they are not wearing a mask, an alarm will sound.

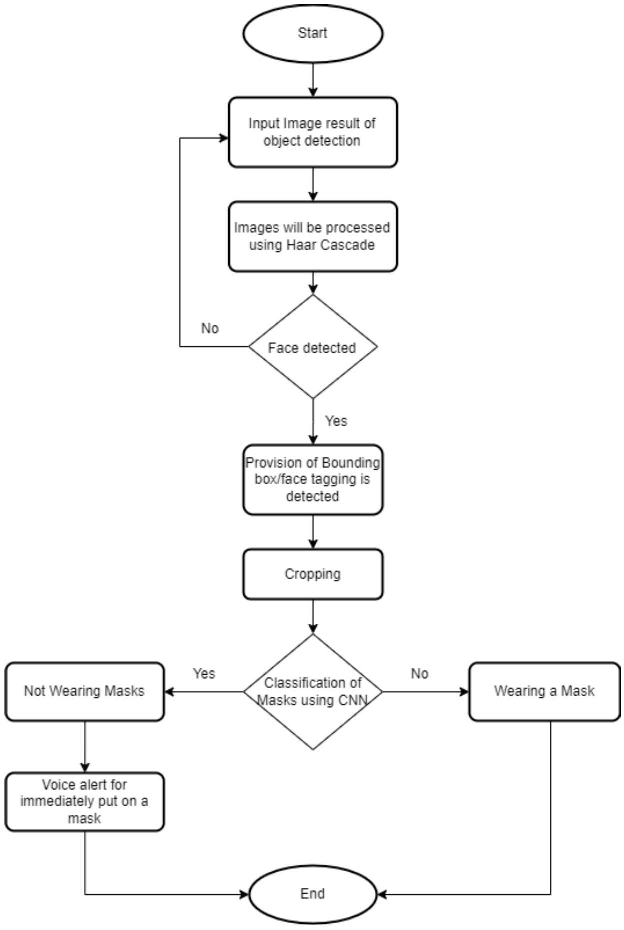

Fig. 5. Mask detection system flowchart

Table 3 is a description of the system alarm on the mask detection system. Description of the detection system, which explains the alarm function when there is an undetected facial object wearing a mask, the alarm will automatically sound and if an object is detected wearing a mask, the alarm will not sound [20].

Table 3.
Description of The Detection System

Alarm Sound	The system detects faces without masks	Alarm goes off
	The system detects faces using masks	The alarm doesn't sound

G. *Trial System*

Testing this system based on data with calculation accuracy begins with testing the results of facial object detection training, by measuring the performance of the DenseNet-121 architecture using the confusion matrix, namely accuracy, precision, and recall using the following equation. The Accuracy, Precision, and Recall formulas are shown in (1)-(3).

IV. RESULT

A. *Accuracy Testing*

Based on testing of the scenarios that have been carried out with 2 scenarios, to ensure these results can be calculated to find out the percentage based on the number of correct data and the amount of wrong data based on scenario testing. Table 4 and Table 5 below are calculations of accuracy, precision, and recall from testing scenario 1 and 2.

Table 4.
Scenario Testing Results 1

Scenario 1		Manual Detect Masked	Manual Detect Not Masked
System Detect	Masked	3 (TP)	1 (FP)
	Not Masked	0 (FN)	3 (TN)

$$Accuracy = \frac{3+3}{3+1+3+0}$$
$$= \frac{6}{7}$$
$$= 0.85 \times 100\%$$
$$= 85\%$$

$$Precision = \frac{3}{3+1}$$
$$= \frac{3}{4}$$
$$= 0.75 \times 100\%$$
$$= 75\%$$

$$Recall = \frac{3}{3+0}$$
$$= \frac{3}{3}$$
$$= 1$$

Table 5.
Scenario 2 Test Results

Scenario 2		Manual Detect Masked	Manual Detect Not Masked
System Detect	Masked	8 (TP)	1 (FP)
	Not Masked	0 (FN)	1 (TN)

$$Accuracy = \frac{8}{8+1+1+0}$$
$$= \frac{8}{10}$$
$$= 0.8\ x\ 100\%$$
$$= 80\ \%$$

$$Precision = \frac{8}{8+1}$$
$$= \frac{8}{9}$$

P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v6i2.31396

$$= 0.88 \, x \, 100\%$$
$$= 88 \%$$

$$Recall = \frac{8}{8+0}$$
$$= \frac{8}{8}$$
$$= 1$$

In the scenario that has been carried out, mask detection affects the condition of the object to be detected, the condition in question is when the object is directly in front of the camera and the object is at an angle that is not caught by the camera. Detection also affects the light around the object, if the lighting is lacking or dim then the detection will also have an effect, and if the light is good or bright then the system will also be better for detecting. The type of mask and the color of the mask affects the detection time, for example, on detection with 30 degrees to the right or to the left the blue 4D Premium mask type is better at detecting it while the white duckbill mask type does not detect well when detecting with 30 degrees to the right nor to the left.

The results of the accuracy of testing scenarios 1 and 2 have in common that they both get an accuracy of 85% with the remaining 15% being a value outside the calculation of accuracy or a data failure value which means the system can detect well on facial objects wearing masks and not wearing a mask, because in the test scenarios that have been carried out the system can correctly detect the detected object. because in the first scenario the system was able to detect 3 objects using masks and 3 objects not using masks according to the original object. However, there was 1 object that detected wearing a mask even though the object was not wearing a mask. in the second scenario the system was able to detect 8 objects using masks and 1 object not using a mask according to the original object. However, there was 1 object that detected wearing a mask even though the object was not wearing a mask. so that way, the system will not cause false alarms.

V. Conclusion

The system created using haar cascade face detection and detection using a mask or not using CNN with the DenseNet-121 architecture can properly detect the use of a mask on someone's face. by using the Haar Cascade and CNN methods in real-time.

Testing scenario 1 gets an accuracy of 85% while scenario 2 gets 80%. The precision value in scenario 1 testing gets 75% results, and scenario 2 testing gets 88% results. Recall values from scenarios 1 and 2 both get a value of 100%. The limitation of the problem in this paper is that the test data used are 7 image data in scenario 1 and 10 image data in scenario 2. So the same amount of data is needed in both scenarios so that the research results in both scenarios have more definite numbers when compared to each other. Work that can be added to previous research is to add test data and strengthened it with k-fold cross-validation to get more accurate results. well, as well as with more test data. The use of other architectures can also be tested with existing data so that the results can be compared. Future research is to improve detection methods that produce higher accuracy, install the best algorithms on portable hardware such as the Raspberry Pi 4, and place devices in areas where masks are required, such as hospitals.

References

[1] R. T. Puteri dan F. Utaminingrum, “Deteksi Kantuk Menggunakan Kombinasi Haar Cascade dan Convolutional Neural Network,” *J-PTIIK*, vol. 4, no. 3, pp. 816–821, Jun 2020.

[2] Y. Religia, M. Ekhsan, M. Huda, and A. D. Fitriyanto, “TOE Framework for E-Commerce Adoption by MSMEs during The COVID-19 Pandemic : Can Trust Moderate ?,” *Appl. Inf. Syst. Manage.*, vol. 6, no. 1, pp. 47–54, 2023, doi: 10.15408/aism.v6i1.30954.

[3] R. P. Sidik and E. Contessa Djamal, "Face Mask Detection using Convolutional Neural Network," *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, Depok, Indonesia, 2021, pp. 85-89, doi: 10.1109/IC2IE53219.2021.9649065.

[4] G. A. Anarki, K. Auliasari, and M. Orisa, “Penerapan Metode Haar Cascade pada Aplikasi Deteksi Masker,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 179–186, March 2021.

[5] K. Suresh, M. B. Palangappa, and S. Bhuvan, “Face Mask Detection by using Optimistic Convolutional Neural Network,” in *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, Jan. 2021, pp. 1084–1089. doi: 10.1109/ICICT50816.2021.9358653.

[6] B. Dimas Nugraraga, H. Fitriyah, and D. Syauqy, “Deteksi Orang Bermasker Medis Menggunakan Metode Convolutional Neural Network Berbasis Raspberry Pi,” *J-PTIIK*, vol. 5, no. 6, pp. 2600-2609, Jun 2021.

[7] F. L. Ahmad, A. Nugroho, and A. F. Suni, “Deteksi Pemakai Masker Menggunakan Metode Haar Cascade Sebagai Pencegahan COVID 19,” *Edu Elektr. J.*, pp. 13–18, 2021, doi: 10.15294/eej.v10i1.47861.

[8] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 6999-7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[9] N. Awalia, “Identifikasi Penyakit Leaf Mold Pada Daun Tomat Menggunakan Model Densenet121 Berbasis Transfer Learning,” *J. Ilm. Ilmu Komput.*, vol. 8, no. 1, pp. 49–52, 2022, doi: 10.35329/jiik.v8i1.212.

[10] M. Syarif, “Deteksi kedipan mata dengan haar cascade classifier dan contour untuk password login sistem,” *Techno.COM*, vol. 14, no. 4, Nov pp. 242-249, 2015.

[11] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

[12] W. S. Eka Putra, “Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101,” *J. Tek. ITS*, vol. 5, no. 1, 2016, doi: 10.12962/j23373539.v5i1.15696.

[13] M. B. Tamam, M. Walid, J. Freitas, and A. Bernardo, “Classification of Sign Language in Real Time Using Convolutional Neural Network,” *Appl. Inf. Syst. Manage.*, vol. 6, no. 1, pp. 39–46, 2023, doi: 10.15408/aism.v6i1.29820.

[14] A. Rohim, Y. A. Sari, and Tibyani, “Convolution neural network (cnn) untuk pengklasifikasian citra makanan tradisional,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, pp. 7038–7042, 2019, [Online]. Available: http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5851/2789.

[15] M. Fennell, C. Beirne, and A. C. Burton, “Use of object detection in camera trap image identification: Assessing a method to rapidly and accurately classify human and animal detections for research and application in recreation ecology,” *Glob. Ecol. Conserv.*, vol. 35, p. e02104, 2022, doi: 10.1016/j.gecco.2022.e02104.

[16] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLoS One*, vol. 10, no. 3, p. e0118432, 2015, doi: 10.1371/journal.pone.0118432.

[17] A. Waheed and J. Shafi, "Successful Role of Smart Technology to Combat COVID-19," in *Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC 2020*, 2020, pp. 772–777. doi: 10.1109/I-SMAC49090.2020.9243444.

[18] B. A. Kumar and M. Bansal, "Face Mask Detection on Photo and Real-Time Video Images Using Caffe-MobileNetV2 Transfer Learning," *Appl. Sci*., vol. 13, no. 2, p. 935, 2023, doi: 10.3390/app13020935.

[19] D. Tyas Purwa Hapsari, C. Gusti Berliana, P. Winda, and M. Arief Soeleman, "Face Detection Using Haar Cascade in Difference Illumination," in *Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, iSemantic 2018*, 2018, pp. 555–559. doi: 10.1109/ISEMANTIC.2018.8549752.

[20] L. Zhu, T. D. Ha, Y. H. Chen, H. Huang, and P. Y. Chen, "A Passive Smart Face Mask for Wireless Cough Monitoring: A Harmonic Detection Scheme With Clutter Rejection," *IEEE Trans. Biomed. Circuits Sys*t., vol. 16, no. 1, pp. 129–137, 2022, doi: 10.1109/TBCAS.2022.3148725.