Classification of Mask Use during a Pandemic using the CNN Algorithm with Voice Notifications

Ryan Gusti Nugraha\textsuperscript{1}, Ahmad Fauzi\textsuperscript{2}, Anis Fitri Nur Masruriyah\textsuperscript{3*}, Bayu Priyatna\textsuperscript{4}, Firman Nurdiansyah\textsuperscript{5}

\textsuperscript{1} Department of Informatics Engineering, Universitas Buana Perjuangan Karawang, 41361, Indonesia  
\textsuperscript{2} Department of Informatics Engineering, Universitas Buana Perjuangan Karawang, 41361, Indonesia  
\textsuperscript{3} Department of Informatics Engineering, Universitas Buana Perjuangan Karawang, 41361, Indonesia  
\textsuperscript{4} Department of Information System, Universitas Widyagama Malang, 65128, Indonesia

This article contributes to:

### Highlights:

- The Convolutional Neural Network method is used to distinguish faces wearing masks, masks that are not correct and not wearing masks.
- This model has a sound indicator and a notification to faces using the Convolutional Neural Network (CNN) algorithm method.

### Abstract

Various technologies were created to prevent the threat of the Covid-19 virus, which has spread in many countries including Indonesia. One of them is the use of masks in public places. With this in mind, this study aims to detect facial objects. Based on the Kaggle website, the object used for research is a human face in 2D form. This research consists of two stages, namely creating and testing a model. The model is a system that detects and classifies faces with masks, inappropriate masks and without masks. Then the model is tested for its accuracy. The result of thirty trials, the model has an accuracy of 99\% which is tested using a webcam in real time. This model has a sound indicator which is a notification to faces using the Convolutional Neural Network (CNN) algorithm method.

### Keywords: CNN, Covid-19, Identification Model

### 1. Introduction

The event of the Covid-19 pandemic, or coronavirus, spread throughout the world in early 2020. Almost the whole world is experiencing this viral pandemic \cite{1}, Covid-19 was first reported in December 2019 in Wuhan, China \cite{2}. Then this disease was identified by technological genome sequencing as a new form of coronavirus called acute respiratory syndrome or coronavirus 2 (SARS-CoV-2) and the disease was named coronavirus disease 2019 (Covid-19), WHO also designated this virus as a world pandemic \cite{3}. Since the publication of the WHO statement, the Indonesian government has issued policies to control the spread of Covid-19. This policy aims to break the chain of transmission of Covid-19 and was made based on policies from WHO, including washing hands, wearing masks, maintaining distance, staying away from crowds, reducing mobility (5M) \cite{4}. Problems occur because many people violate the 5M policy, one of which is wearing a mask. So this research created a system that can detect the use of masks, inappropriate masks and not...
on the face with sound notifications [5] [6] [7]. The goal is to be able to make people more comply with regulations and break the chain of transmission of Covid-19.

Previously, research was conducted by Tri septiana [8], identifying the use of objects in masks using the CNN algorithm. The system created is able to recognize the use of masks. Accuracy results obtained by 84.23%. Then Anirudh [9], conducted research on mask detection with the CNN algorithm and used a dataset of 1400 data which was divided into 700 data of people wearing masks and 700 not wearing masks. The final average value of the program's accuracy in detecting people's faces is 88.53%. Furthermore, research on the application of face recognition for detecting covid masks and body temperature using the convolutional neural network (CNN) method was conducted by Prasad et al [10] [11]. The research process began with the face recognition method and the CNN algorithm. The results obtained created a system that can detect the use mask and body temperature [12] [13].

The system that can detect the use of masks, inappropriate masks, and not using masks on the face uses the Convolutional Neural Network algorithm with the architectural preprocessing technique used, namely MobileNetV2 [14] [15]. In this research, several processes were carried out, such as collecting image data sets obtained from the Kaggle website. After the images were collected, preprocessing was carried out so that the images became uniform in size 224 × 224 pixels. Then, the data set resulting from the preprocessing was carried out by a split validation process by dividing the data set into 80 training data and 20 validation data, then augmentation was also carried out to increase the number of images such as rotation, zoom, cropping, sliding and flipping randomly but not eliminating the essence or essence of the data. Training is carried out with the tensorflow library and testing is carried out to obtain results from the images that have been trained. Tests on this system are carried out with several scenarios from several distances to find out whether all distances are detected, the accuracy obtained from model training is 99% and testing is 90% [16] [17].

2. Methods

The application of the Convolutional Neural Network method is used to distinguish faces wearing masks, inappropriate masks and not wearing masks. The Convolutional Neural Network method is a development of the Multi Layer Perceptron (MLP) method with the aim of processing two-dimensional digital image data [18]. Convolutional Neural Networks are included in the type of Deep Learning. There are two Convolutional Neural Network methods, namely feedforward to carry out classification, and backpropagation to carry out the learning stage.

![Convolutional Neural Network Stages](image)

In Figure 1, the stages of the CNN method are divided into two main layers, namely feature learning, and classification. At the feature learning stage, it has a feature extraction layer which consists of a convolution layer and a pooling layer. The convolution layer will carry out the process of moving a convolution filter of a certain size into an image. The result of a convolution is used as input to produce a feature representation. In convolution, a stride shift is also carried out at each location that has a matrix multiplication and sums the results into feature data. Then apply the ReLU (Rectifier Linear Unit) activation function to get the output from the convolution and make the output non-linear. Then the process of the pooling layer or pooling layer, which functions to reduce the spatial size of the convolution features, thus reducing the computation needed to speed up computation and during the training process is faster. The pooling used in this study is average pooling which aims to determine the average value. Next, the Classification stage classifies each neuron that has been obtained and extracted in feature learning. The Classification stage consists of flatten, fully connected, and softmax. In the flatten stage, converting the feature map into a vector will be used as a combined input. Furthermore, at the fully connected layer stage where all activated neurons from the previous layer will be connected to all neurons in the next layer. The last step, namely softmax (finding the maximum value in a classification).
Sourced from the Kaggle website with a total of 8,982 images divided into three classes with 2994 images wearing masks, 2994 images using the wrong masks and 2994 images not using masks. Then, the data that has been collected is stored in a structured folder. Next, the preprocessing stage is carried out on the image by changing the size to 224x224 pixels. The initial stage is creating a file in MPEG-1 Audio Layer 3 (MP3) format containing sound notifications. By using the text-to-speech method or converting a text into sound. This process is assisted by using the Google Text-to-Speech (gTTS) library, which will output in the form of an MP3 file that is used when conducting mask detection tests [19] [20].

The preprocessing result data set will be used for the split validation process by dividing 80% of the data for training and 20% for testing. After that, data augmentation will be carried out, such as datasets that are rotated, zoomed, cropped, shifted and reversed randomly. The purpose of this process is to reproduce the dataset as well as increase the accuracy of the model. The results of this process will be used as model training using the tensorflow library. Tests were carried out with several scenarios from different distances; at a distance of three meters this mask detection system could not detect it. Meanwhile, at other distances, all are seen. In the testing process, the confusion matrix method is used, with the following Equation 1.

\[
\frac{TP + TN}{TP + FP + TN + FP} \times 100\%
\]

(1)

Accuracy is described by the accuracy value of a model in classifying objects correctly. True Positive (TP) is an interpretation in predicting that a positive is true, and True Negative (TN) is an interpretation in predicting a negative is true.

### 3. Results and Discussion

The data set that has been obtained from the Kaggle website is then created folders with the names "with_masker", "mask_false", and "without_masker", these folders will later be used at the preprocessing stage so that the size of all images is uniform or commonly called normalization to 224x224 pixels as in Figure 2. Then it is converted into an array as shown in Figure 3, because computers do not have human-like capabilities that are able to recognize an object. The computer can only interpret the data included in the image.
In Figure 4, the next step is to add the gTTS library then create a variable to initialize gTTS and there are two mandatory parameters that must be filled in, namely the text you want to convert into sound, for example “You are not wearing a mask” and the output language you want to use, in this case it is Indonesian. Next is to execute the command to save the output followed by the desired file name.

In the training process that uses data from split validation and augmentation, the results obtained accuracy can be seen in Figure 5 by obtaining an accuracy of 99%. In this figure there is precision or accuracy which is the ratio of correct predictions (positive and negative) to the entire data, then recall or ratio of successful positive predictions to the total number of correct positive observations. Then, the f1-score is a comparison of the average precision and recall weights.

After the training process is carried out, it will produce the best CNN model; then, it will be tested. The goal is to determine the performance of the model. In the test process, a test was carried out on video data (real time) via a webcam. The prediction results from video data in real time using a webcam can be seen in Figure 6. The detected face is not wearing a mask with a bounding box accuracy of 97.94%, Figure 7 is detected using the wrong mask with a bounding box accuracy of 95.73% and Figure 8 is detected using a mask with bounding box accuracy 99.84%, each trial scenario is followed by a pre-generated sound notification.
Furthermore, testing from various distances ranging from 30 centimetres to 3 meters was carried out to see the results of detecting masks and not using masks, whether they were successful or not.

<table>
<thead>
<tr>
<th>No.</th>
<th>Distance</th>
<th>Wearing a Mask</th>
<th>Wrong Mask</th>
<th>Not Wearing Masks</th>
<th>Detect System</th>
<th>Sound Notifications</th>
<th>Reality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30 cm</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Green Bounding Box Detected Mask</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>False Mask Detected Yellow Bounding Box</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Red Bounding Box Mask Detected</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td>2</td>
<td>60 cm</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Green Bounding Box Detected Mask</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>False Mask Detected Yellow Bounding Box</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Red Bounding Box Mask Detected</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td>3</td>
<td>90 cm</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Green Bounding Box Detected Mask</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>False Mask Detected Yellow Bounding Box</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Red Bounding Box Mask Detected</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td>4</td>
<td>1,2 m</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Green Bounding Box Detected Mask</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>False Mask Detected Yellow Bounding Box</td>
<td>✓</td>
<td>accordance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Red Bounding Box Mask Detected</td>
<td>✓</td>
<td>accordance</td>
</tr>
</tbody>
</table>
4. Conclusion

Based on the results of testing and analysis of the mask detection system using the Convolutional Neural Network (CNN) algorithm, this study has concluded that the system has succeeded in building a face identification model that uses masks, but masks are incorrect. They do not use masks with the Convolutional Neural Network algorithm. The system has successfully tested facial identification using masks, inappropriate masks and not using masks with the Convolutional Neural Network algorithm with several scenarios. The model made has an accuracy of 99%. Testing the system with several scenarios produces 90% accuracy. As for the suggestions in this study, I want you to know that I can develop into the internet of things version and make the system detectable from any distance.
Authors' Declaration

Authors' contributions and responsibilities – The authors made substantial contributions to the conception and design of the study. The authors took responsibility for data analysis, interpretation, and discussion of results. The authors read and approved the final manuscript.

Funding – No funding information from the authors.

Availability of data and materials – All data are available from the authors.

Competing interests – The authors declare no competing interest.

Additional information – No additional information from the author's.

References


