Sriwijaya Journal of Informatic and Applications
Vol. xx, No. xx, Month Years, pp. xx-xx
ISSN xxxx-xxxx, DOI:xxxxx

Additive Angular Margin Loss for Masked-Face Image Recognition

Ferza Reyaldi [a,1,*], Muhammad Fachrurrozi [a,2], Muhammad Qurhanul Rizqie [a,3]

[a] Informatics Engineering, Faculty of Computer Science, Universitas Sriwijaya, Indonesia
[1*] reyaldiferza@gmail.com; [2] mfachrz@unsri.ac.id; [3] qurhanul.rizqie@ilkom.unsri.ac.id
* corresponding author

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
Additive Angular Margin Loss
Convolutional Neural Network
Masked Face Recognition
Real-World Masked Face Dataset

ABSTRACT

The identity of a person is difficult to recognize due to partial occlusion of some faces. The partial occlusion obscures the important features that are characteristic at recognition. One of the cases of occlusion on the face is a mask. This research uses Convolutional Neural Network (CNN) and Additive Angular Margin Loss (ArcFace) to recognize masked faces. ArcFace works to find features that are highly discriminatory on the face. Real-World Masked Face Data (RMFD) of 90,000 face images without masks and 5000 face images with masks are used as training data and test data. The ArcFace model with InceptionV3 produces the best value compared to the ArcFace model with ResNet50 and the model without ArcFace. The results of the InceptionV3 ArcFace test showed a significant increase in accuracy and validation values of 36.33% and 12.65%.

1. Introduction

Face recognition is one of the preferred biometric recognition solutions because it does not require direct contact with the user and achieves a high degree of accuracy. However, the mandatory wearing of face masks began to be enforced in public places during the Covid-19 Pandemic which aims to keep the pandemic under control [1]. This causes partial occlusion of the face due to the unavoidable use of clothing accessories in the form of masks [2]. This particular problem is a major challenge in the field of facial recognition because the available facial features are decreased [3].

CNN is suitable for classification problems because CNN can extract and study features in data automatically [4]. However, the CNN architecture requires quite high computational power due to the use of filters in the form of a 2-dimensional matrix in feature extraction in data [5].

Mandal et al. [6] conducted research on masked facial recognition using the CNN ResNet-50 pre-trained model. The results showed that the accuracy rate of the face recognition model with a mask was 47.9%, much lower than the accuracy rate of the face recognition model without a mask which reached 89.7%.

Additive Angular Margin Loss is a loss function that allows to obtain very discriminatory features for face recognition. The Additive Angular Margin Loss has a clear geometric interpretation because of its exact correspondence to the geodesic distance on the hypersphere [7].

In this study, ArcFace is implemented on the CNN architecture which will be trained to increase the discriminatory level of features on limited masked facial images.

2. Literature Study / Hypotheses Development

a. Convolutional Neural Network

Convolutional Neural Network (CNN) is one of the most effective types of artificial neural networks. CNN has demonstrated many advantages in various applications such as image classification, object recognition, object retrieval, and object detection. CNN can consist of many cascaded layers. These tiered layers function to control the level of shift, scale, and distortion. The types of tiered layers include input layer, convolutional layer, subsampling layer, full-connected layer, and output layer [8].

In practice, there are many CNN pre-trained models that can be used to solve classification problems. AlexNet, AlexNetOWTBn, GooLeNet, Overfeat, VGG are some of the most commonly used examples of the CNN model architecture. These architectures use multiple convolution layers. However, this causes new problems to arise, such as the difficulty of network optimization, the problem of disappearing gradients, and the problem of degradation. A new idea that can be offered for this problem is Residual Network (ResNet). ResNet has the advantage of solving complex tasks and also increasing detection accuracy. ResNet tries to overcome the difficulties in the deep CNN training process, saturation and decreased accuracy. In research, we use the ResNet-50 architecture. As the name implies, ResNet-50 uses 50 residual layers [9]. Fig. 1 below shows the architecture of ResNet50.

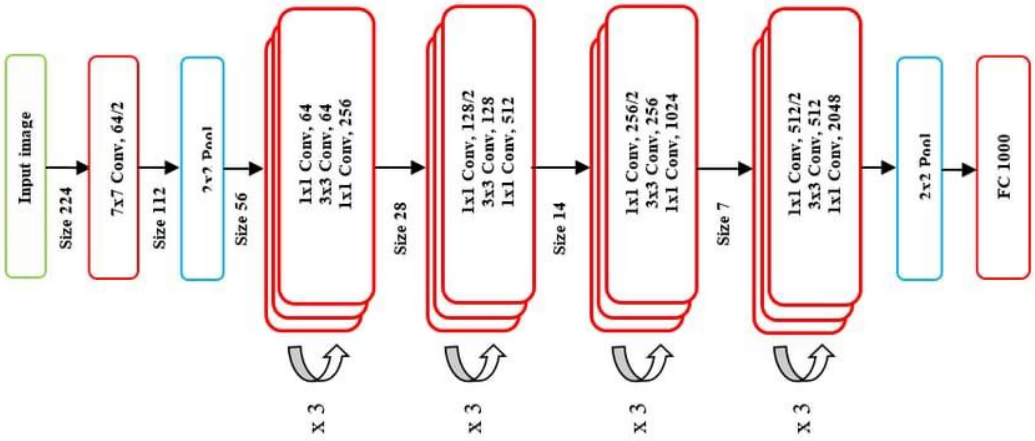


Fig. 1. ResNet-50 Architecture [9]

Apart from ResNet-50, another new Idea that can be offered is InceptionV3. The InceptionV3 model is a pre-trained model that has superior performance in object recognition when compared to its predecessor, GoogleNet (InceptionV1). The InceptionV3 model includes three parts: the basic convolutional block, the enhanced Inception module, and the classifier [10]. The CNN architecture of the InceptionV3 model consists of 48 layers [11]. Fig. 2 below shows the architecture of InceptionV3.

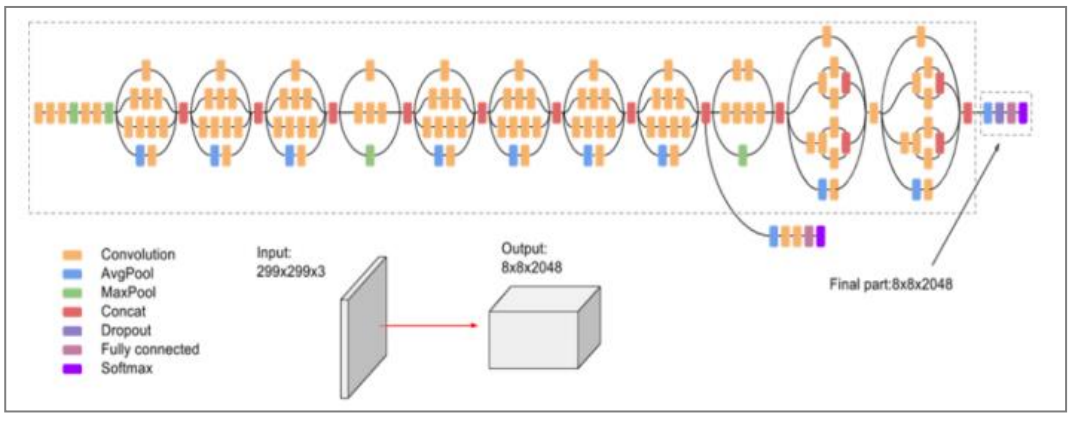


Fig. 2. InceptionV3 Architecture [11]

b. Additive Angular Margin Loss

Additive Angular Margin Loss or also called ArcFace, is a type of loss function that can be used in building CNN models. ArcFace makes it possible to get very discriminatory features for face recognition. ArcFace has a clear geometric interpretation due to its exact correspondence to geodesic distances on the hypersphere.

The benefits of using ArcFace can be summarized in four ways, namely engaging, effective, easy, and efficient. Engaging, ArcFace directly optimizes geodesic boundary spacing. Effectively, ArcFace achieves state-of-the-art performance on ten facial recognition benchmarks including large scale image and video data sets. Easy, ArcFace only requires a few lines and is easy to implement on computational graph-based deep learning. Efficient, ArcFace only adds negligible computational complexity during training [7].

Mathematically, ArcFace calculations can be shown by the formula in (1).

$$L_3 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))}+\sum_{j=1,j\neq y_i}^{n}e^{s\cos\theta_j}} \quad (1)$$

Where are L_3 represents ArcFace, N represents the batch size, m represents the angular margin penalty and y_i denotes the y_i class. m equals to the geodesic distance margin in the normalized hypersphere.

c. Metric Evaluation

Evaluation of the classification model can be done using the Confusion matrix. The confusion matrix is used as a visualization tool for model evaluation in supervised learning problems. The matrix column represents the predicted class and the row represents the actual class [12]. Fig. 3 shows the visualization of the confusion matrix.

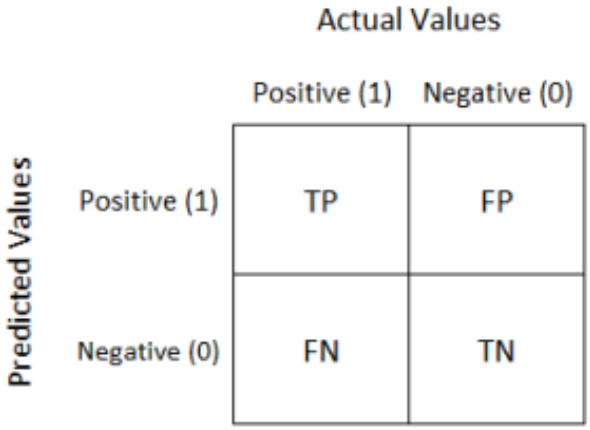


Fig. 3. Confusion Matrix [12]

The confusion matrix calculation produces four outputs that can be used as benchmarks to measure model performance, namely accuracy, recall, precision, and F-1 Score [13].

1. Accuracy

Accuracy is the number of models that correctly predict a class (true positives and true negatives) divided by the total of all predicted results. Mathematically, Accuracy is formulated as in formula (2).

$$acc = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

2. Recall

Recall is a calculation of conditions when the actual class is positive, how often the model predicts positive. Mathematically, recall is formulated as in formula (3).

$$r = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

3. Precision

Precision is the result of calculating when the model predicts positively, how often the prediction is true. Mathematically, precision is defined as in formula (4).

$$p = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

4. F-1 Score

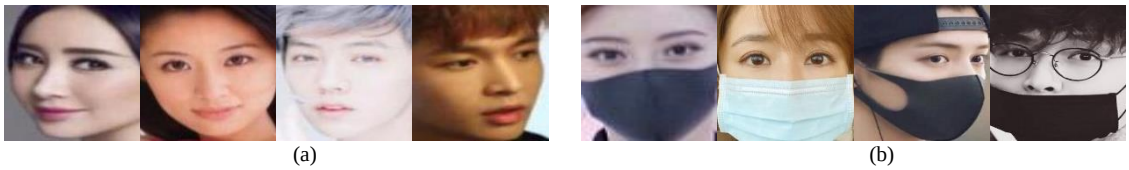
F-1 Score is the harmonic average value of precision and recall. Mathematically, the F-1 score is formulated as in the formula, which is the number of models that correctly predict a class (true positive and true negative) divided by the total of all predicted results. Mathematically, Accuracy is formulated as in formula (5).

$$FM = 2 \times \frac{p \times r}{p+r} \times 100\% \quad (5)$$

3. Methodology

a. Data Acquistion

The type of dataset used in the study of the masked face image classification model is secondary data. The dataset is the Real-World Masked Face Dataset (RMFD). RMFD consists of 525 subjects (classes) which include 90,000 face images without masks and 5,000 face images with masks. The size of the images available in the dataset used varies. Fig. 4 shows sample images from the unmasked face dataset and the masked face dataset.



(a) (b)

Fig. 4. Sample Data of Real-World Masked Face Dataset RMFD [14]
(a) Sample of Unmasked Face Data, (b) Sample of Masked Face Data

The research was conducted using only 15 classes. The selection of the 15 classes is based on the minimum number of images of 20 images in that class. Then resizing is done, the image data is equated to the resolution ratio to 180 x 180 using the help of the TensorFlow library. After that, the Train-test split process was carried out, the dataset was divided into 2 parts, namely training data and test data. Comparison of the amount of training data and test data is 7:3, 70% for training data and 30% for test data.

b. Proposed Model

There are 2 architectures proposed in this study, namely ResNet50-ArcFace and InceptionV3-ArcFace.

The ResNet50-ArcFace architecture is almost the same as the ResNet50 architecture. The difference is that the last layer of ResNet50 is removed, replaced with the ArcFace layer. Between ResNet50 and ArcFace layers are inserted 3 layers. The 3 layers sequentially include, dropout layer, flatten layer, and dropout layer. The ResNet50-ArcFace architecture is shown in Fig. 5.

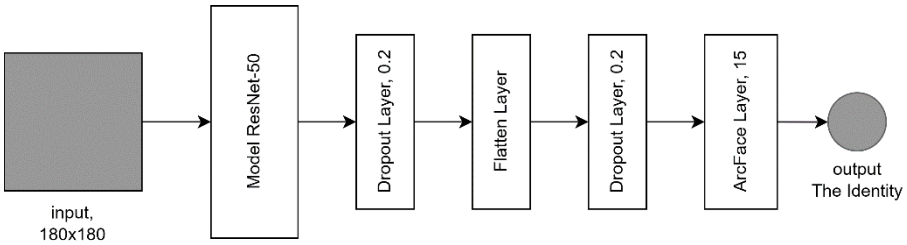


Fig. 5. Proposed ResNet50-ArcFace Architecture

As with the previously proposed architecture, the InceptionV3-ArcFace Architecture makes the same modifications. The difference with the architecture of the InceptionV3 model is that the last layer of InceptionV3 is removed, replaced with the ArcFace layer. Between InceptionV3 and ArcFace layers are inserted 3 layers. The 3 layers sequentially include, dropout layer, flatten layer, and dropout layer. The architecture of InceptionV3-ArcFace is shown by Fig. 6.

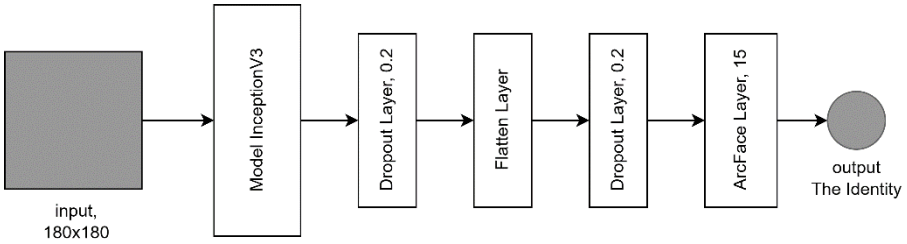


Fig. 6. Proposed InceptionV3-ArcFace Architecture

The dropout layer added to the architecture shown in Fig. 5 and Fig. 6 to reduce the number of connected neurons. This aims to prevent overfitting when doing model training. In addition, a flatten layer is also added to the proposed architecture. It aims to change the dimensions of the input which has many dimensions to only one dimension. The ArcFace layer that is used functions as a classifier, replacing the dense layer that is usually used with softmax activation. The ArcFace layer is given an argument worth 15. This value indicates the number of classes in the dataset used.

c. Research Configuration

Table 1 shows the parameter configurations carried out in the study. These configurations conform to the configurations made by [6]. This is done so that the performance comparison between the proposed model and previous research is valid.

Table 1. Parameter Configuration

Parameter	Value
Optimizer	Adam
Learning rate	0.0016
Batch Size	32
Loss	Categorical Crossentropy
Epoch	25

4. Result and Discussion

Each variation of the CNN model configuration has different training and validation performance in the classification of masked face images. Table 2 below shows a comparison of model performance based on training data and validation data.

Table 2. Table of Train-Validation Accuracy and Loss

Dataset	Model	Training		Validation	
		Accuracy	Loss	Accuracy	Loss
Masked	ResNet-50	60,05%	47,91%	1,5005	2,4092
Masked	1: ResNet-50 + ArcFace	12,06%	11,21%	11,3264	8,9414
Masked	**2: InceptionV3 + ArcFace**	**96,38%**	**60,56%**	**0,1357**	**2,3069**
Masked + Unmasked	3: ResNet-50 + ArcFace	12,60%	18,11%	10,3623	6,0141
Masked + Unmasked	4: InceptionV3 + ArcFace	61,84%	44,47%	6,5041	12,7073

Based on Table 2, Model 2 has the best accuracy and loss values. Model 2, a scenario that uses the CNN architecture model InceptionV3 with ArcFace, has better accuracy and loss than the reference model, Model ResNet-50 [6]. This is because the use of ArcFace is able to enlarge the boundary distance between classes, so that the model features of each class produce high classification accuracy. Model 2 has a better training data accuracy of 36.33% than the reference model and has a better validation data accuracy of 12.65% compared to the reference model. Then, the loss obtained by Model 2 on the training data and validation data is also better than the reference model. The loss values of Model 2 for the training data and validation data are 0.1357 and 2.3069, respectively.

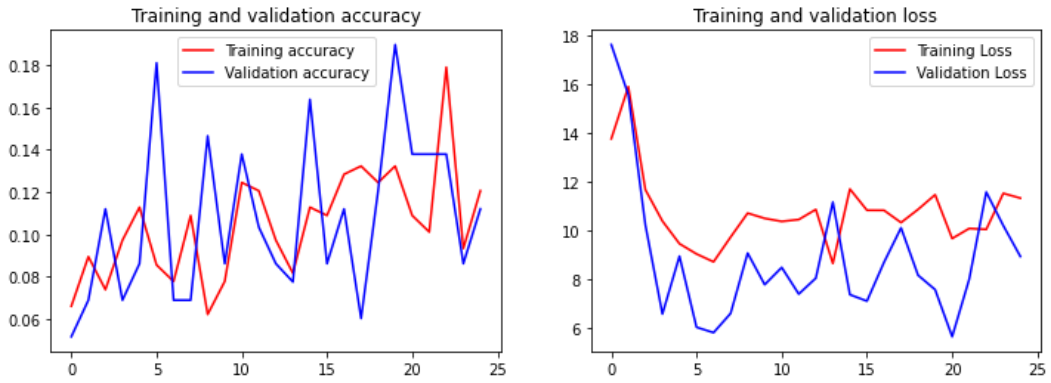


Fig. 7. Accuracy and Loss of Model 1

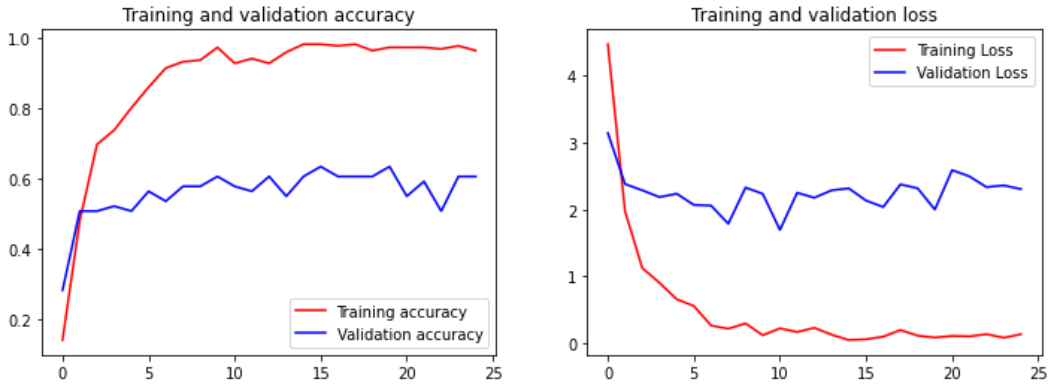


Fig. 8. Accuracy and Loss of Model 2

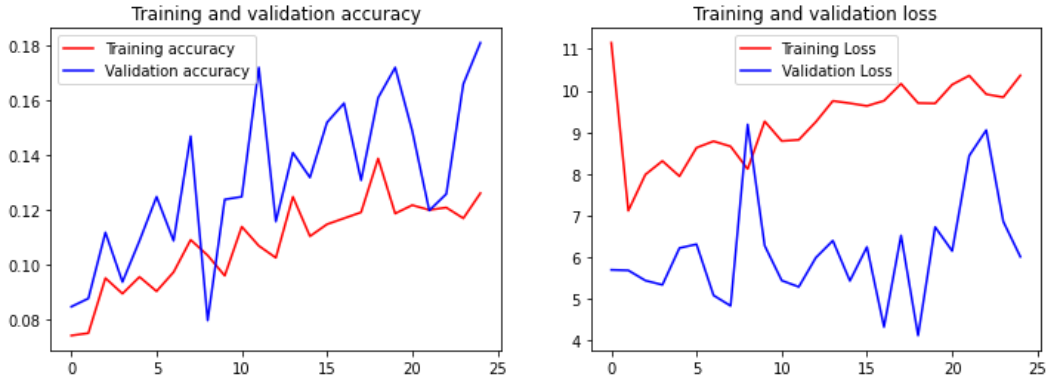


Fig. 9. Accuracy and Loss of Model 3

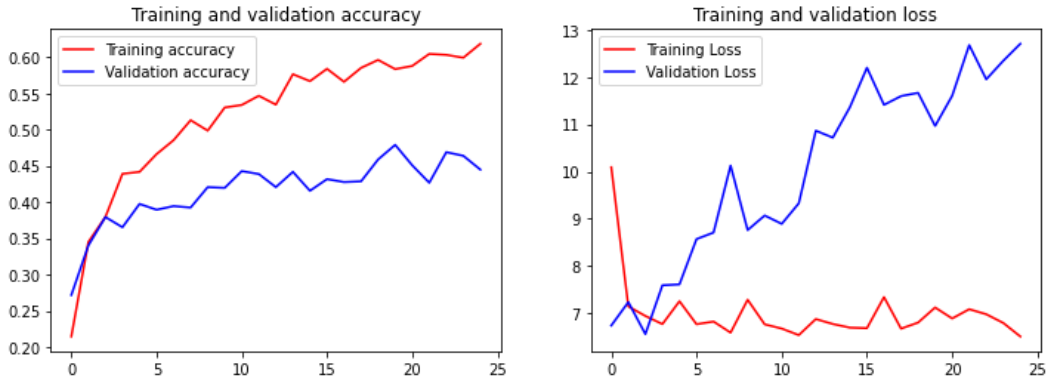


Fig. 10. Accuracy and Loss of Model 4

Based on the accuracy and loss graphs in Fig. 7 to Fig. 10 above, Fig. 8, which is the accuracy and loss curve of Model 2, shows the best accuracy and loss curve compared to the graphs of other models. Fig. 7 and Fig. 9 shows that the model is underfitting, because it shows the accuracy and loss values that do not have a significant and/or stable increase. Accuracy and loss of model 1 and model 3 Then, Fig. 10 shows model 4 overfitting. This is shown in the loss curve on the validation data which is increasing at each epoch. Although model 2 has a graph that also shows overfitting, but model 2 is better than model 4. This is because the validation loss in model 2 does not continue to increase every epoch like model 4, but the loss is static, showing the same properties as loss on training data.

Table 3. Table of Metric Evaluation Result

Dataset	Model	Accuracy	Recall	Precision	F-1 Score
Masked	ResNet-50	0,4791	0,4719	0,4613	0,4473
Masked	1: ResNet-50 + ArcFace	0,1358	0,6471	0,1467	0,2392
Masked	**2: InceptionV3 + ArcFace**	**0,7901**	**0,9142**	**0,8533**	**0,8827**
Masked + Unmasked	3: ResNet-50 + ArcFace	0,1111	0,4500	0,1323	0,2044
Masked + Unmasked	4: InceptionV3 + ArcFace	0,7283	0,8428	0,8676	0,8550

Based on the data shown in Table 3, model 2 also has a better value compared to model [6] based on the evaluation metrics used. Model 2 has a test accuracy of 0.7901, 0.3110 better than the reference model. Then, Model 2 recall is 0.9142, 0.4423 better than the reference model. Model 2 has a precision of 0.8533, 0.3920 better than the reference model. In the measurement using the evaluation metric f-1 score, Model 2 has a value of 0.8827, which is also better than the reference model, the difference in value is 0.4354. Therefore, Model 2 is proven to have better overall performance compared to the reference model based on evaluation metrics.

5. Conclusion

Based on the results and analysis of the research that has been done, the researcher can draw several conclusions:

1. Masked-face image recognition system software using various CNN models, namely InceptionV3 and ResNet-50 with additive angular margin loss has been successfully developed and can classify masked face image identities.
2. The best model produced is the InceptionV3 model using the additive angular margin loss (ArcFace). ArcFace is able to increase the boundary distance between classes, so that the model features of each class produce high classification accuracy. This model has a greater accuracy of 12.65% compared to previous research models without using ArcFace.

References

[1] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Self-restrained Triplet Loss for Accurate Masked Face Recognition," Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.01716

[2] D. Montero, M. Nieto, P. Leskovsky, and N. Aginako, "Boosting Masked Face Recognition with Multi-Task ArcFace," Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.09874

[3] W. Hariri, "Efficient masked face recognition method during the COVID-19 pandemic," *Signal, Image Video Process.*, vol. 16, no. 3, pp. 605–612, Apr. 2022, doi: 10.1007/s11760-021-02050-w.

[4] B. S. Chandra, C. S. Sastry, S. Jana, and S. Patidar, "Atrial fibrillation detection using convolutional neural networks," *Comput. Cardiol. (2010).*, vol. 44, pp. 1–4, 2017, doi: 10.22489/CinC.2017.163-226.

[5] S. Nurmaini *et al.*, "Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks," *Futur. Gener. Comput. Syst.*, vol. 113, pp. 304–317, 2020, doi: 10.1016/j.future.2020.07.021.

[6] B. Mandal, A. Okeukwu, and Y. Theis, "Masked Face Recognition using ResNet-50," Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.08997

[7] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition." [Online]. Available: https://github.com/

[8] A. Alzu'bi, F. Albalas, T. Al-Hadhrami, L. B. Younis, and A. Bashayreh, "Masked face recognition using deep learning: A review," *Electronics (Switzerland)*, vol. 10, no. 21. MDPI, Nov. 01, 2021. doi: 10.3390/electronics10212666.

[9] I. Z. Mukti and D. Biswas, "Transfer Learning Based Plant Diseases Detection Using ResNet50," *2019 4th Int. Conf. Electr. Inf. Commun. Technol. EICT 2019*, no. December, pp. 1–6, 2019, doi: 10.1109/EICT48899.2019.9068805.

[10] C. Lin, L. Li, W. Luo, K. C. P. Wang, and J. Guo, "Transfer learning based traffic sign recognition using inception-v3 model," *Period. Polytech. Transp. Eng.*, vol. 47, no. 3, pp. 242–250, 2019, doi: 10.3311/PPtr.11480.

[11] G. Jignesh Chowdary, N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Face Mask Detection Using Transfer Learning of InceptionV3," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12581 LNCS, pp. 81–90, 2020, doi: 10.1007/978-3-030-66665-1_6.

[12] A. Novandya, "Penerapan Algoritma Klasifikasi Data Mining C4.5 pada Dataset Cuaca Wilayah Bekasi," *KNiST*, pp. 368–372, 2017.

[13] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.

[14] Z. Wang *et al.*, "Masked Face Recognition Dataset and Application," Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.09093