Face Race Classification using ResNet-152 and DenseNet- 121

JASMAN PARDEDE, SYAFIQ SALIM KLEB

Department of Informatics, Institut Teknologi Nasional Bandung, Indonesia
Email: jasman@itenas.ac.id

ABSTRAK

Penelitian ini bertujuan untuk membandingkan hasil kinerja arsitektur ResNet-152 dan DenseNet-121 dalam mengklasifikasikan wajah berdasarkan ras. Ras yang diklasifikasikan terdiri dari 4 kelas: Putih, Hitam, India, dan Asia. Penelitian ini menggunakan parameter batch_size 32, optimizer, dan learning rate untuk meningkatkan kinerja model. Dua optimizer yang dibandingkan, yaitu Adam Optimizer dan Nadam Optimizer. Nilai learning rate yang dipertimbangkan adalah 0,0001 dan 0,001. Berdasarkan hasil percobaan klasifikasi wajah berdasarkan ras, baik arsitektur ResNet-152 maupun DenseNet-121 mencapai kinerja akurasi dan recall yang sama yaitu 0,788. Kinerja terbaik untuk presisi dan f1-score adalah ResNet-152. ResNet-152 memiliki presisi dan f1-score yang lebih baik masing-masing sebesar 0,376% dan 0,252%, jika dibandingkan dengan DenseNet-121. Oleh karena itu, ResNet-152 memiliki kinerja terbaik jika dibandingkan DenseNet-121 dalam klasifikasi wajah berbasis ras.

Kata kunci*: kinerja, ResNet-152, DenseNet-121, wajah, klasifikasi*

ABSTRACT

This study aims to compare the performance results of the ResNet-152 and DenseNet-121 architectures for classifying faces based on race. The classified race consists of 4 classes i.e.: White, Black, Indian, and Asian. The study used a batch size of 32, an optimizer, and a learning rate to improve model formation performance. Two optimizers are being compared, namely Adam Optimizer and Nadam Optimizer. The learning rate values considered are 0.0001 and 0.001. Based on the results of facial classification experiments based on race, both the ResNet-152 and DenseNet-121 architectures achieve the same accuracy and recall performance, namely 0.788. The best performance for precision and f1-score is ResNet-152. The ResNet-152 has better Precision and F1-Score of 0.376% and 0.252% respectively, in case it's compared to DenseNet-121. Hence, it can be inferred that ResNet-152 surpasses DenseNet-121 in delivering superior performance outcomes for racial-based facial classification.

Keywords*: performance, ResNet-152, DenseNet-121, face, classification*

1. INTRODUCTION

Facial recognition technology that continues to innovate encourages developments in the anthropology world, especially in the forensics field. The development of human facial identification also influences the progress of biometric-based security systems. Facial recognition is a biometric-based identification system that has high accuracy, from a faster and more accurate authentication system. **(Dewi et al., 2019)**

Image classification is the process of classifying image pixels into categories of the same type or previously determined. Then, this image classification began to be developed into the science of biometric systems. A biometric system is a system that uses physical characteristics or human behavior to identify a person's identity using a computer **(Jagtap et al., 2023)**. Physical characteristics are related to a person's body shape, such as fingerprints, skin color, palms, smell, etc.

Every living creature, such as humans, animals, and plants, has races. Race is a classification system that classifies a group of living creatures from each other. In humans, there are four main racial types, namely Caucasian, Mongoloid, Negroid, and Australoid. In the early 20th century, this term was often used in a biological sense to designate genetically diverse human populations, with members having the same phenotype (outer appearance). **(Setyadi & Sutanto, 2023)**

As time progresses, the human population becomes increasingly heterogeneous, making it difficult to distinguish one human race from another just through their specific racial characteristics. Race is one of the methods used for the classification system that humans use to determine the large number of different populations or groups based on phenotypic characteristics, geographic origin, appearance, and genetic ethnicity **(Dewi et al., 2019)**. Human race identification also has an important role in the criminal field, namely it can help identify suspects, fugitives, or victims of crimes.

Several studies on race image classification use the Discrete Wavelet Transform (DWT) method and Learning Vector Quantization (LVQ) classification **(Dewi et al., 2019)**, then use the Histogram of Oriented Gradient (HOG) and Linear Discriminant Analysis (LDA) methods, using the CNN method by **(Rahman et al., 2020)** and **(Abdulwahid, 2023)**, including using Resnet50. **(Setyadi & Sutanto, 2023)**

One of the CNN architectures that includes a deep network is the Residual Network (ResNet). The advantage of this architecture is that it overcomes the degradation problem by introducing identity blocks and convolution blocks. This block allows the network to learn differences in input features, so training is easier and performance does not decrease even if the network is deeper. However, ResNet can be complicated in some deep learning frameworks because of the many layers that need to be arranged correctly **(Phang, 2021)**. Then another CNN architecture used for classification is Dense Connected (DenseNet). The advantage of DenseNet is that it can maximize the use of information and reduce the risk of vanishing gradients. Because DenseNet has dense connections connecting all layers, DenseNet has higher computational requirements to calculate features at each layer. **(Hidayati & Liliana, 2021)**

ResNet and DenseNet are used to improve accuracy and efficiency in image classification tasks. These networks overcome the degradation problem faced by deeper neural networks and are able to learn better feature representations. With their ability to understand more complex

relationships in image data, ResNet and DenseNet have been successful in a variety of image classification tasks, including object classification, object detection, and pattern recognition.

This study compares the ResNet-152 and DenseNet-121 architectures for face classification based on race. In previous research, only one of the two architectures was used. The classification model formed from the architecture is expected well-to-do to determine the racial classes i.e.: White, Black, Indian, and Asian. The hope is that this research will yield insights into which architecture and optimizer are most recommended for this particular task of facial race classification.

2. RESEARCH METHODOLOGY

In this research, two main methods are utilized: ResNet-152 and DenseNet-121. These two methods will be compared to assess the performance of the best-model.

2.1 Block Diagrams
This research is divided into several steps as stated in the block diagram in Figure 1. Based on Figure 1, the process of implementing the CNN method using Resnet-152 and DenseNet-121 starts from the data input process and reaches the final step which produces the output performance of the model.
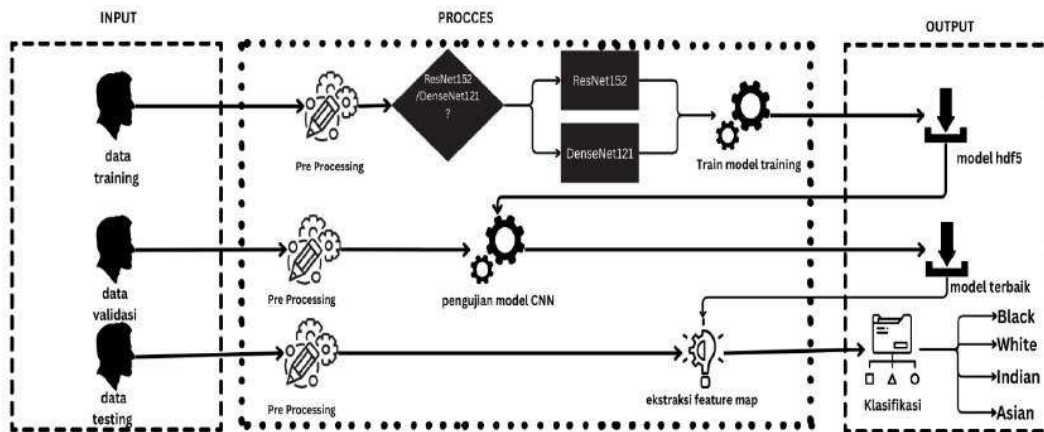
Figure 1. System design block diagram

In Figure 1 are the stages in the ResNet-152 and DenseNet-121 architecture research for face classification based on race. In the training and validation process input is a dataset of facial images with 4 (four) labels, namely Black, White, Indian, and Asian. In the input, there is training data and validation data where the data from the dataset is split by 10% to get validation data. The training process is carried out using training data and validation data. The training data was resized to 224x224 pixels and normalized after resizing. The training data images are processed using the ResNet-152 and DenseNet-121 models which consist of Max Pooling operations, ReLU activation, Softmax activation, and convolution. The DenseNet-121 architecture with 121 layers has a dense block and transition layer process which is followed by another dense block and transition layer. The ResNet-152 architecture has a residual block and transition layer process which is followed by another residual block and transition layer. In the end, this model has a classification layer that produces output in the form of a model in h5 format.

At the validation data stage, the model is tested using validation data to find the best weights. The best weight will be used for the identification process in testing data. The testing data image also underwent pre-processing with the same size, namely 224x224 pixels and was normalized. Then, the features in the image are taken using the best weights generated from the training data. The final output of the model formed is a face classification based on race with labels Black, White, Indian, and Asian.

2.2 Dataset
The data used is the Kaggle dataset obtained from the website, namely https://www.kaggle.com/datasets/jangedoo/utkface-new which contains facial images with four labels, namely: White, Black, Indian, and Asian. The dataset is divided into three parts, namely training data, testing data, and validation data with a ratio of 80:10:10 with the total dataset used being 1600 image data, as stated in Table 1.

Table 1. Dataset distribution

Classes	Training	Validation	Testing
Black	320	40	40
White	320	40	40
Indian	320	40	40
Asian	320	40	40

In the dataset, there are 4 classes, namely Black, White, Indian, and Asian. The following is an example dataset of the four classes shown in Figure 2.
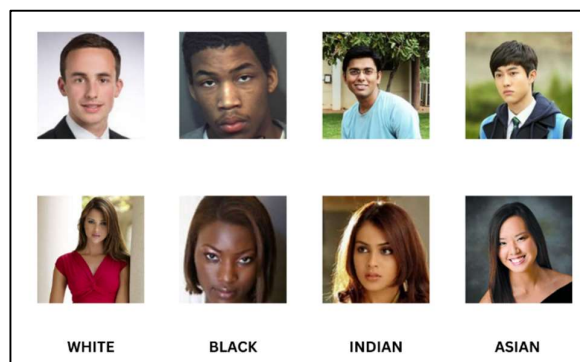
Figure 2. Example of the facial race image

2.3 Pre-processing
Pre-processing is the initial stage in processing image data before it is run through an algorithm or machine learning model. The aim of image pre-processing is to prepare image data so that it is easier to process and produces better results. **(Yenusi et al., 2023)**

2.3.1 Resize
The image that has been input goes into the resizing process. The resizing process aims to change each image's size so that it has the same number of pixels. In this study, the image changed to a size of 224x224 pixels.

2.3.2 Augmentation
Augmentation is a process that aims to enrich an image by creating a new image from an existing image. The aim is for the image data to have data variations to produce additional training data and data that has been adjusted to suit needs **(Aditama et al., 2023)**. The

augmentation process carried out in this research is rotating the image, inverting the image horizontally and vertically, shifting the image horizontally, shifting the angle (shear), and enlarging (zooming) the image.

2.4 Convolutional Neural Network

CNN is a convolutional neural network, which is a type of neural network architecture specifically designed to process grid data such as image data or other spatial data. This architecture is inspired by how human vision works and has become a key component in advances in the field of image processing and pattern recognition **(Dewi & Ismawan, 2021)**. CNN consists of several operational layers, namely: convolution, non-linear activation, and pooling **(Prasetyo & Ichwan, 2021)**. Convolution layers are the core of CNN and function to extract important features from input data. Each convolution layer consists of the number of filters (kernels) that will be applied to the image in a shifting manner to produce a feature map. **(Pratama et al., 2022)**
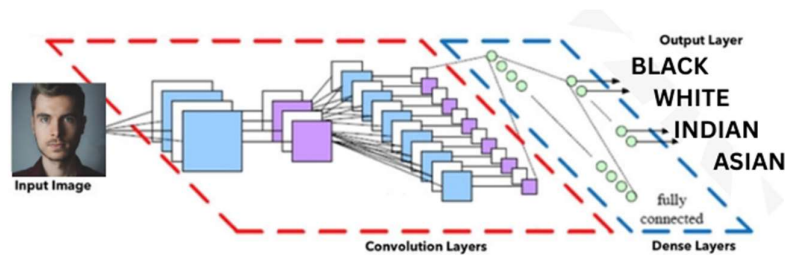
Figure 3. CNN architecture

As shown in Figure 3, a CNN consists of several layers, including an input layer, an output layer, and several hidden layers. The architecture often used in CNN is LeNet-5, which was developed specifically for analyzing images. In a CNN, image data is passed through layers of convolution, non-linear activation, and pooling to extract important features. Next, these features will be connected through hidden layers that perform mathematical calculations to study and classify images with high accuracy. In the end, the output layer produces predictions or results from image analysis carried out by CNN. **(Wahyuddin et al., 2023)**

2.5 Transfer Learning

Transfer learning is a technique used in deep learning to take a pre-trained model and apply it to a different problem **(Aditama & Haryanti, 2023)**. The trained model has been tested and proven to be able to handle large and general data well, so it can help increase the accuracy of the new model. In transfer learning, adjustments are usually made to the final layers of the model that are used to build a new model. These adjustments can take the form of changing some parameters or adding new layers.

2.6 Residual Network (ResNet)

ResNet-152 is a convolutional neural network model. ResNet-152 is widely used for various computer vision tasks, including image classification. It is a variation of the ResNet architecture introduced to address the problem of vanishing gradients in deep neural networks **(Erwandi & Suyanto, 2020)**. The ResNet-152 architecture is represented in Figure 4.
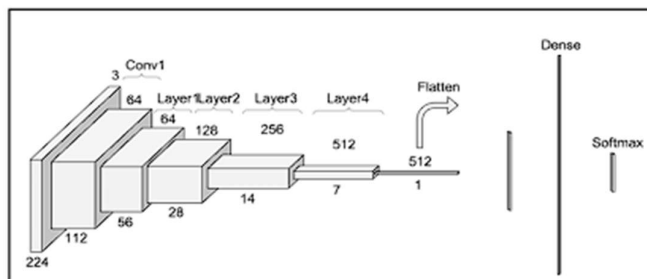
Figure 4. ResNet-152 architecture

2.7 Densely Neural Network (DenseNet)

Densely Neural Network, also known as DenseNet, is an architecture that connects each layer with all other layers **(Pardede & Putra, 2020)**. This Dense Block is part of DenseNet which functions to receive input from all previous layers in the block, which allows a more direct and dense flow of information through the network **(Putra & Bunyamin, 2020)**. The DenseNet-121 architecture is represented in Figure 5.
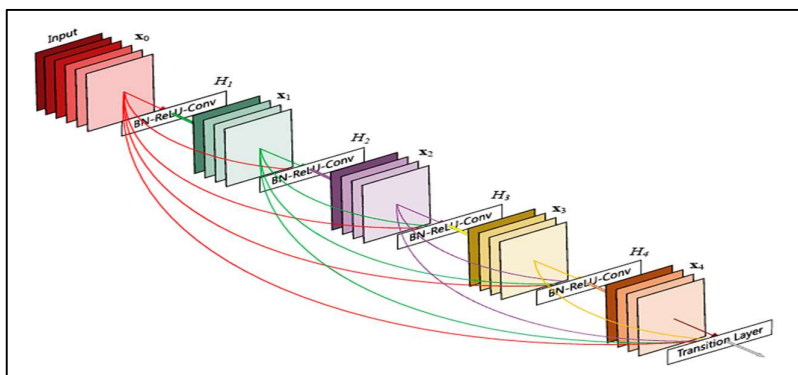

Figure 5. DenseNet-121 architecture

2.8 Model Evaluation

At this stage, the model will be evaluated to determine the performance it produces. This stage is measured using accuracy, precision, f1-score, and recall values. In the process of measuring the performance of the classification model, a confusion matrix is used which contains several general terms. These terms are as follows: True Positive (TP) is positive data that is predicted correctly by the model. True Negative (TN) is negative data that is predicted correctly by the model. False Positive (FP) is negative data that should be predicted correctly but is actually predicted as positive data by the model. Meanwhile, False Negative (FN) is positive data that should be predicted correctly by the model but turns out to be predicted as negative data **(Pardede & Putra, 2020)**. The formula for calculating accuracy, precision, recall, and f1-score can be calculated using Equation (1) to Equation (4).

Accuracy is the percentage of test data that is successfully classified correctly into the correct class. The accuracy value can be seen in Equation (1).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

Precision is a measure that describes how accurate a model is in predicting positive events from a series of prediction activities. The precision value can be calculated with Equation (2).

$$precision = \frac{TP}{TP+FP} \tag{2}$$

Recall is a measure of completeness in the form of the percentage of positive tuples that are successfully classified as positive classes by the model. The recall value can be calculated with Equation (3).

$$recall = \frac{TP}{TP+FN} \tag{3}$$

F1-score is an evaluation calculation in information retrieval (retrieval) that pays attention to recall and precision values. The f1-score can be calculated with Equation (4).

$$f1 - score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \tag{4}$$

3. RESULTS AND DISCUSSION

System performance testing is carried out to see the confusion matrix and measure accuracy, precision, recall, and f1-score. In testing the performance of this system, testing data contained in the dataset is used, where there are 4 (four) classes, namely: Black, White, Indian, and Asian. In the research there are several hyper-parameters, one of which is used, namely: epoch value, learning rate, early stopping, and optimizer. The epoch values used are 150 and 100. Early stopping with a patience value of 10. Learning rate values are 0.0001 and 0.001. Meanwhile, the optimizers used are the Nadam and Adam optimizers, as stated in Table 2.

Table 2. Comparison using hyperparameter

Architecture	Epoch	Learning rate	Optimizer	Accuracy	Loss	Val. accuracy	Val. loss
ResNet152	150	0.0001	Adam	0.9609	0.1132	0.7563	0.8228
ResNet152	150	0.0001	Nadam	0.9734	0.0950	0.8062	0.8052
ResNet152	48	0.001	Adam	0.8375	0.4767	0.7875	0.7801
ResNet152	29	0.001	Nadam	0.8258	0.4769	0.7750	0.6195
DenseNet121	150	0.0001	Adam	0.9445	0.1687	0.7812	0.7990
DenseNet121	150	0.0001	Nadam	0.9508	0.1418	0.8000	0.7004
DenseNet121	43	0.001	Adam	0.8727	0.3287	0.7563	0.7258
DenseNet121	21	0.001	Nadam	0.8297	0.4398	0.7188	0.7394

In the implementation of the CNN architecture for face classification based on race, it can be seen from the accuracy and loss values of the batch size parameters used to achieve the optimal state. The training results using batch size 32 with epoch 150 and Nadam Optimizer from the ResNet-152 architecture can be seen in Figure 6 and Figure 7.
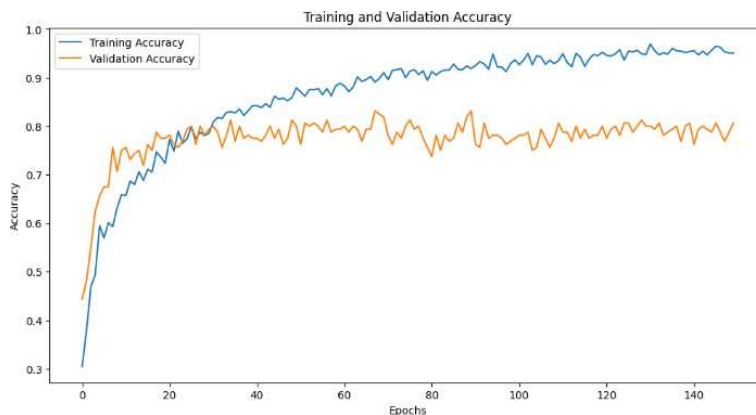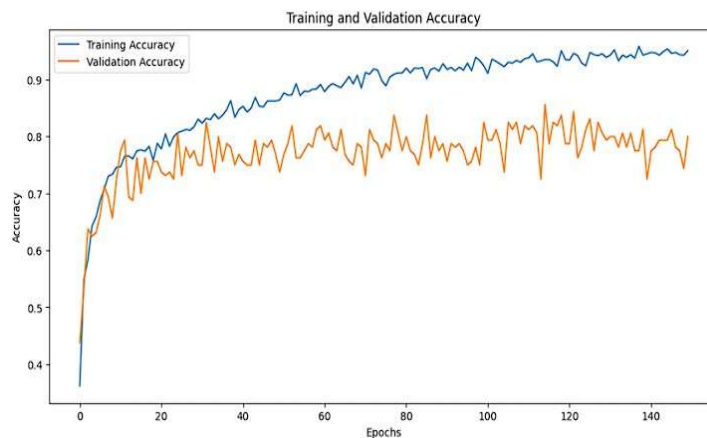
**Figure 6. ResNet-152 accuracy test results**
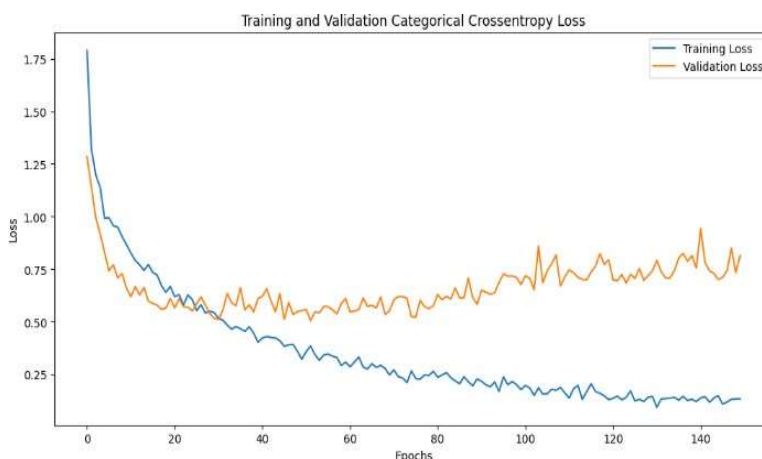
Figure 7. ResNet-152 loss test results

The ResNet-152 testing results can be seen in Figure 6 and Figure 7. The training results using batch size 32 with epoch 150 from the DenseNet121 architecture with the Adam optimizer can be seen in Figure 8 and Figure 9. The DenseNet-121 test results can be seen in Figure 8 and Figure 9. While, the confusion matrix results from Resnet152 with 150 epochs and the Adam optimizer for classes can be seen in Figure 10.


**Figure 8. DenseNet-121 accuracy test results**
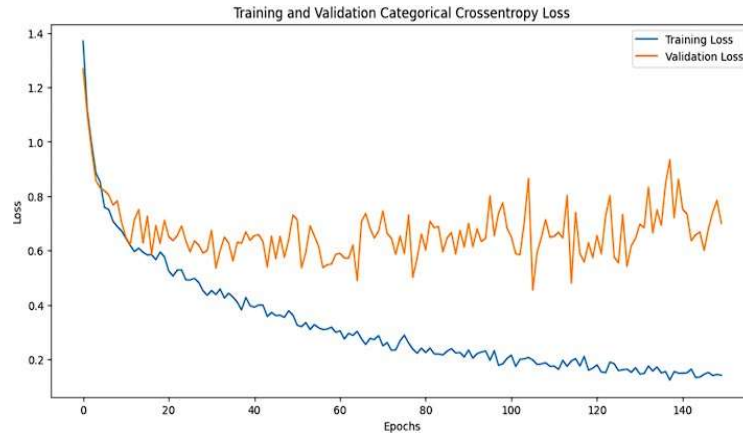
Figure 9. DenseNet-121 loss test results

As can be seen from Figure 10, we can observe the comparison of confusion matrix values for each class trained by both architectures. The ResNet-152 architecture shows the highest precision value for the Asian class, while DenseNet-121 achieves the highest precision for the Black class, ResNet-152 for the Indian class, and DenseNet-121 for the White class. Regarding recall values, ResNet-152 achieved the highest recall for the Asia class, ResNet-152 for the Black class, DenseNet-121 for the India class, and DenseNet-121 for the White class. Finally, the highest f1-score values were achieved by ResNet-152 for the Asia class, DenseNet-121 for the Black class, ResNet-152 for the India class, and DenseNet-121 for the White class.
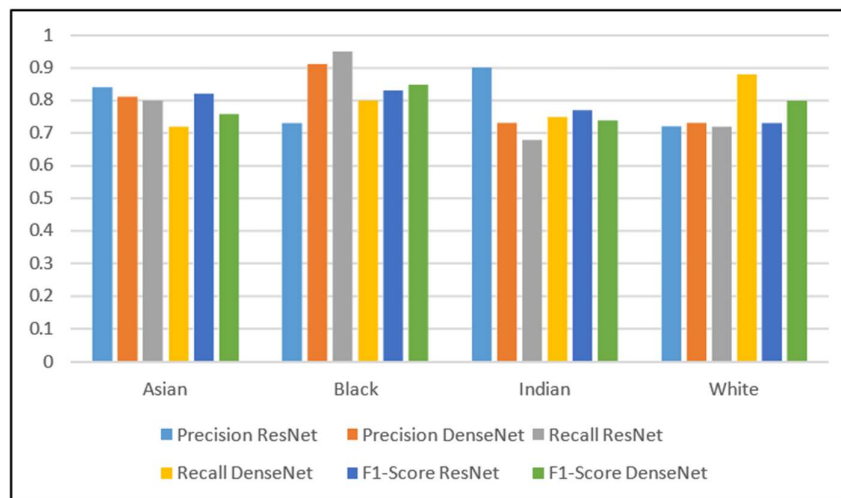
Figure 10. Performance of ResNet-152 and DenseNet-121
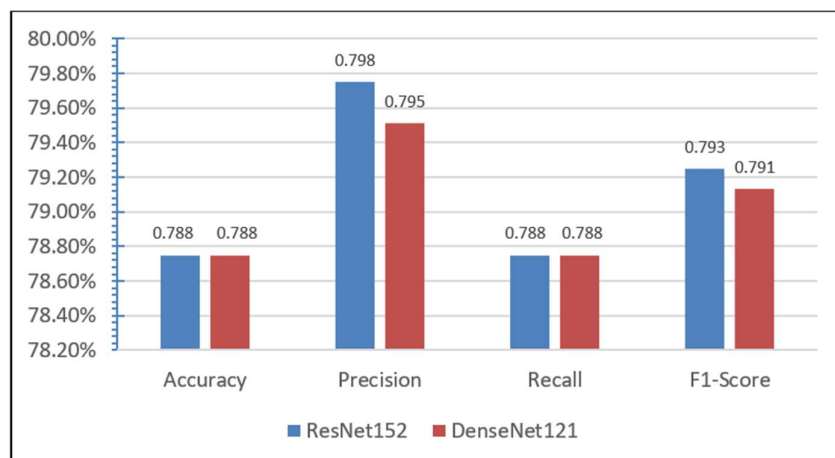
Figure 11. Performance comparison of ResNet-152 and DenseNet-121

From Figure 11, it shows that the two architectures, ResNet-152 and DenseNet-121, have the same level of accuracy, namely 0.788. However, when checking the precision values, ResNet-152 achieved a precision score of 0.798, while DenseNet-121 achieved a precision score of 0.795. Apart from that, the recall value for both architectures is 0.788. The f1-score values are also in line, with ResNet-152 having a higher value of 0.793, while DenseNet-121 achieved a value of 0.791.

4. CONCLUSION

In the race-based face classification task, both ResNet-152 and DenseNet-121 architectures show the same accuracy rate of 0.788. Although there are differences in precision values, ResNet-152 has an advantage with a value of 0.798 compared to DenseNet-121 with a value of 0.795. The recall value for both architectures is the same, namely 0.788. Meanwhile, for the f1-score value, there are differences between the two architectures. ResNet-152 has a higher score compared to DenseNet-121. The f1-score values of ResNet-152 and DenseNet-121 are 0.793 and 0.791 respectively. Therefore, it can be concluded that ResNet-152 provides better performance results than DenseNet-121.

REFERENCES

Abdulwahid, A.A. (2023). Classification of Ethnicity Using Efficient CNN Models on MORPH and FERET Datasets Based on Face Biometrics. *Applied Sciences*, *13*(7288), 1-29. doi: 10.3390/app13127288

Aditama, M. F., & Haryanti, M. (2023). Sistem Pengenalan Dan Verifikasi Wajah Menggunakan Transfer Learning Berbasis Raspberry PI. *Jurnal Teknologi Industri*, *12*(1), 30–39. doi: 10.35968/jti.v12i1.1045.

Dewi, A., Hidayat, B., & Arif, J. (2019). Identifikasi Ras Manusia Berdasarkan Citra Wajah Berbasis Discrete Wavelet Transform Dan Learning Vector Quantization-Neural Networks. *Prosiding SENIATI*, (pp.7–13). doi: 10.36040/seniati.v5i4.1157.

Dewi, N., & Ismawan, F. (2021). Implementasi Deep Learning Menggunakan CNN Untuk Sistem Pengenalan Wajah. *Faktor Exacta*, *14*(1), 34-43. doi: 10.30998/faktorexacta.v14i1.8989

Erwandi, R., & Suyanto, S. (2020). Klasifikasi Kanker Payudara Menggunakan Residual Neural Network. *Indonesian Journal on Computing (Indo-JC)*, *5*(1), 45–52. doi: 10.21108/indojc.2020.5.1.373.

Hidayati, N., & Liliana, D. (2021). *Modifikasi Arsitektur DenseNet121 dengan Transfer Learning Untuk Deteksi Penyakit Tanaman Jagung Berdasarkan Citra Daun*. Tesis: Magister Ilmu Komputer, Universitas Nusa Mandiri, 1–78.

Jagtap, V., et al. (2023). Biometric Face Recognition System using Deep Dream and CNN. *International Journal of Intelligent Systems and Applications In Engineering*, *11*(10s), 145-154.

Pardede, J., & Putra, D. A. L. (2020). Implementasi DenseNet Untuk Mengidentifikasi Kanker Kulit Melanoma. *Jurnal Teknik Informatika Dan Sistem Informasi*, *6*(3), 425–433. doi: 10.28932/jutisi.v6i3.2814

Phang, B. C. (2021). Implementasi Metode Convolutional Neural Network Pada Pengenalan Wajah Menggunakan Pre-Trained ResNet. Retrieved from http://e-journal.uajy.ac.id/24395/%0Ahttp://e-journal.uajy.ac.id/24395/1/1507082391.pdf

Prasetyo, R. R. E., & Ichwan, M. (2021). Perbandingan Metode Deep Residual Network 50 dan Deep Residual Network 152 untuk Deteksi Penyakit Pneumonia pada Manusia. *MIND Journal*, *6*(2), 168–182. doi:10.26760/mindjournal.v6i2.168-182

Pratama, R. A., Achmadi, S., & Auliasari, K. (2022). Penerapan Metode Convolutional Neural Network Pada Aplikasi Deteksi Wajah Pengunjung Perpustakaan. *JATI* (*Jurnal Mahasiswa Teknik Informatika*), *6*(1), 253–258. doi: 10.36040/jati.v6i1.4517

Putra, A. K., & Bunyamin, H. (2020). Pengenalan Simbol Matematika dengan Metode Convolutional Neural Network (CNN). *Jurnal Strategi*, *2*(2), 426–433.

Rahman, R. A., Arifianto, A., & Ramdhani, K. N. (2020). Klasifikasi ras berdasarkan citra wajah menggunakan convolutional neural network. *E-Proceedings of Engineering*, *7*(1), 2526–2536.

Setyadi, A. D., & Sutanto, F. A. (2023). Klasifikasi ras manusia menggunakan metode convolutional neural network berbasis telegram bot. *JATISI* (*Jurnal Teknik Informatika dan Sistem Informasi*), *10*(2), 1–8. doi: 10.35957/jatisi.v10i2.4814.

Wahyuddin, K., Wahiddin, D., & Kusumaningrum, D. S. (2023). Sistem Deteksi Wajah Keamanan Pintu Menggunakan Metode Convolutional Neural Network (CNN) Berbasis

Arduino. *Scientific Student Journal for Information, Technology and Science*, *4*(1), 15–23.

Yenusi, Y. N., Trihandaru, S., & Setiawan, A. (2023). Perbandingan Model Convolutional Neural Network pada Klasifikasi Wajah Orang Papua dan Etnis Lainnya. *JST* (*Jurnal Sains Dan Teknologi*), *12*(1), 261–268. doi: 10.23887/jstundiksha.v12i1.46861.