Age Classification of Moviegoers Based on Facial Image Using Deep Learning

Abba Suganda Girsang, Dewa Bagus Gde Khrisna Jayanta Nugraha

Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

Abstract **— The number of moviegoers in Indonesia continues to rise year after year until 2019. However, due to the COVID-19 pandemic, most Indonesian cinemas were closed in early 2020. Moviegoers are increasingly turning to digital platforms to watch films. Based on the films shown, they can be divided into three categories: films for children, films for adolescents, and films for adults. A system that can automatically classify the faces of the audience based on their age category is required. Using Deep Learning, this study aims to classify the audience's age based on facial photos. The first stage involves collecting data from three datasets: All-Age-Face, FaceAge, and FGNET, which are then combined and relabeled based on age group. Preprocessing and hyperparameter testing were also performed. Finding the best learning rate and bottleneck layer is the goal of hyperparameter testing. The training process employs learning rete and the two best bottleneck layers with six models, namely MobileNet, MobileNetV2, VGG16, VGG19, Xception, and ResNet101V2. Global Average Pooling was added at the end of the layer in each model. The MobileNet model on two bottleneck layers yielded the best testing accuracy value of 85.44 percent in this study.**

DOI: 10.18421/TEM113-52
https://doi.org/10.18421/TEM113-52

Corresponding author: Abba Suganda Girsang,
Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia.
Email: agirsang@binus.edu

Received: 23 May 2022.
Revised: 15 August 2022.
Accepted: 21 August 2022.
Published: 29 August 2022.

Keywords **— age classification, image classification, deep learning, hyperparameter tuning.**

1. Introduction

From early 2020 to the present, there has been a swift trend in the way people watch movies, with the majority of them shifting to digital streaming services such as Netflix, Disney Plus, HBO Max, and others. This is because of the COVID-19 pandemic, which has forced the closure of most Indonesian theatres for health reasons. Children, adolescents, and adults are the three age groups of moviegoers.

A classification system based on the faces of the audience is required so that films displayed in cinemas and on various digital platforms can be watched according to their age category. Deep learning is used in the strategy suggested in this paper. Deep learning has been shown to address various difficulties of digital image grouping with high accuracy.

Previously, Agbo-Ajala & Viriri published a paper titled Face-Based Age and Gender Classification Using Deep Learning Model, which investigated age classification based on face image and deep learning [1]. CNN method comprising four convolutional layers, two fully connected layers, and a dropout layer is used in this study. A confusion matrix is used to calculate the accuracy of the classification findings, which is 84.8 %. Then [2] researched facial image age classification using the Artificial Neural Network method and the Gabor Filter. The confusion matrix was used to calculate the accuracy of the classification findings, which is 83 %.

There are a wide variety of deep learning models available in the Keras library. Moreover, Xception, MobileNet, MobileNetV2, VGG16, VGG19, and ResNet101V2 are among the models proposed in this study. Some of the best layers in each model are determined by testing the bottleneck layer, as done by [3] in his paper Classification of C2C e-Commerce Product Images Using Deep Learning Algorithm. In each model, Global Average Pooling will be added to the final layer. This has been shown to reduce validation loss, as demonstrated by [4] in his study titled Batik Classification Using Deep Learning.

2. Related Work

Age classification based on face is one of the classification methods used to identify individual age groups based on facial features. By using a database of previously stored facial images, this face-based age recognition system can identify a person's age in a database. A person's age and gender can be categorized by visual observation of images, but it is difficult to make visual observations on a computer. Age recognition can be identified by analyzing the forehead area, eyelid area and others [5] .

Qawaqneh, Malluoh, & Buket D proposed a model to handle the age estimation problem based on a person's face image using a deep CNN dubbed VGG-Face and then trained on a big database for facial recognition VGG Face CNN has been tweaked and well-designed to handle the age estimation problem. The experiment was repeated twice, with the first time utilizing the GoogLeNet model with the ILSVRC ImageNet database, which had facial data ranging from 0 to 60 years old, with eight classifications in each age group, yielding an accuracy of 45.07 percent. The same experiment was repeated, this time with the VGG-Face CNN model, yielding a superior accuracy result of 59.90% [6].

Anand et al. suggested a method for estimating age based on heterogeneous networks trained on non-ideal samples to extract strong features from non-ideal facial photos. They performed feature-level fusion of data derived by a CNN set in this study, reduced the size of the obtained feature set, and assessed age using the Feed-Forward Neural Network (FFNN). They devised a dimensionality reduction method utilizing a public face dataset of low-quality facial photos acquired under uncontrolled situations to achieve resistance to non-ideal conditions. The obtained results show that this method performs satisfactorily for non-ideal images obtained in unrestricted scenarios [7].

Melangi used the Artificial Neural Network and Gabor Filter algorithms to research age classification based on facial photos. The FG-NET dataset is used in the categorization process, which is divided into three classes: children, adolescents, and adults. A confusion matrix is used to calculate the accuracy of the classification findings, which yields an accuracy of 83 % [2].

Face-Based Age and Gender Classification Using Deep Learning Model was carried out by Agbo-Ajala and Viriri. This study makes use of the IMDB-WIKI dataset, which contains 26,000 facial photos divided into six age groups. Face Detection and Landmark Detection are the methods employed. The dataset is then trained using the CNN algorithm, which has four Convolutional layers, two fully connected layers, and a dropout layer. A confusion matrix is used to calculate the level of accuracy of the classification findings, which yields a value of 84.8 percent [1].

3. Theory and Methods

A. Deep Learning

According to Fei at al. [8], this theory is contained in a book published at Stanford University with the title "Deep Learning," namely "CS231n: Convolutional Neural Networks for Visual Recognition," where this method has been widely used to perform image detection, image classification, and image prediction. This is consistent with Yann research (review) named Deep Learning, which explains that deep learning is a machine learning application that can be used to recognize things like objects in photos, match news articles, speeches to text, and analyze user sentiment from social media.

Deep learning is a representation-learning method with different levels of depiction [9]. Representation-learning is a set of approaches that allows machines to take in raw data and automatically find the representations required to detect or classify items This strategy is achieved by building simple yet non-linear modules, every of which translates a depiction at one level into a portrayal at a greater level (i.e. starting with raw input to an abstract level). Deep learning is a feature layer that is learned from data through learning techniques. Using enough transformation compositions, a very complex function may be examined. Since standard machine learning techniques have limited ability to analyze natural data in the form of raw data, this method cannot be built by an engineer or someone with specific capabilities. Then, in a few decades, this machine learning technique will necessitate careful engineering to create feature extractors that can convert raw data into mature data (such as image pixel values converted into internally representative data or feature vectors derived from learning subsystems) where Classifiers in this technique can classify or detect patterns in the input.

B. Transfer Learning

The availability of enormous amounts of data, such as ImageNet, is one of the reasons behind deep learning's appeal. However, having such a vast number of datasets is usually difficult. As a result, it will be difficult to carry out the training process for a model from the start. Transfer learning is one option for dealing with this scenario. Transfer learning provides a promising alternative that makes use of a pre-trained deep convolutional neural network which is already trained by another dataset [10].

Although transfer learning provides one solution to the problem of dataset availability, it is equally critical to identify how many datasets are required so that the transfer learning process yields accuracy that is equal to or greater than the accuracy score of the original model. Soekhoe et al. used two types of datasets, tiny-imagenet, and multiplaces2, to research how to transfer learning and how many datasets are suitable for a transfer learning procedure [11]. Several experiments were carried out by reducing the number of datasets, specifically for tiny-imagenet from 500 to 50 and for multiplaces2 from 1000 to 500. Various outcomes for a series of experiments, such as training a new target class from scratch, training all layers, or freezing several layers in a network, are reported in his paper. His findings show that freezing the weights of some initial levels (3 initial layers) of the network produces superior outcomes for small target datasets. Furthermore, a small dataset of 300 images per class from tiny-imagenet and 900 images per class from multiplaces2 yields the best results.

The transfer learning technique that will be employed in this study will be to use some layers that were learned using the ImageNet dataset, then transfer and fine-tune the model used for the new problem, namely age classification based on the facial picture. This is referred to as network-based deep transfer learning in the research [12]. Figure 1 depicts a network-based deep transfer learning diagram.

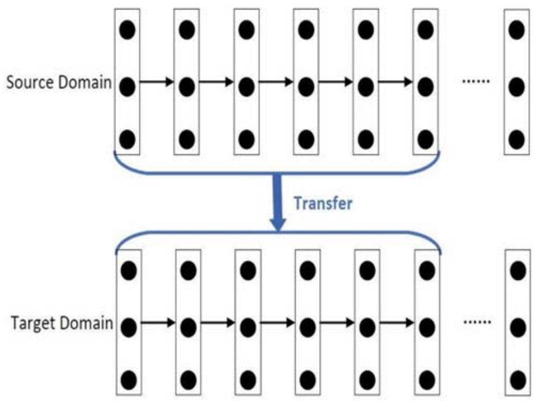

Figure 1. Diagram Network-based Deep Transfer Learning [12]

C. Xception

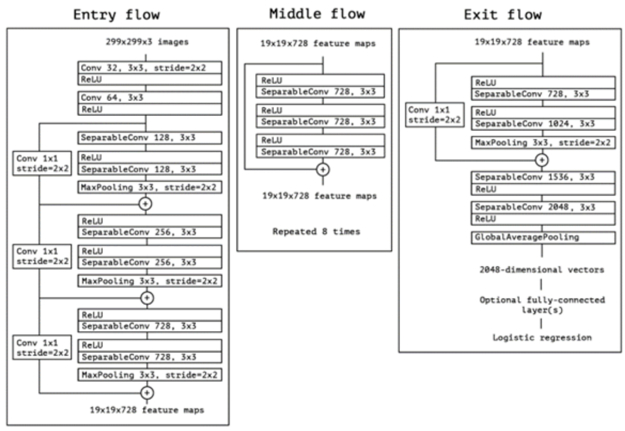

Figure 2. Xception Architecture

Xception or Extreme Inception was established in 2016 by Google's Francois Chollet as an upgrade to the Inception module and its design with a more basic and elegant architecture that is effective. The following is an illustration of the Xception Net architecture as shown in Figure 2.

D. MobileNet

MobileNet, also known as MobileNetV1, is a model with a small number of parameters and a compact resulting model. MobileNetV1 is a model architecture designed for application development on mobile devices or other devices with limited hardware resources by lowering the model's size and complexity through the use of depthwise separable convolutions. When compared to traditional convolutions, using depthwise separable convolutions on MobileNetV1 reduces the number of parameters by more than seven times, with just a 1% drop in accuracy on ImageNet [13]. Figure 3 depicts an image of depthwise separable convolutions, specifically a block layer composed of depthwise and pointwise convolutions, followed by batch normalization and ReLU nonlinearity. The following is an architectural image of depthwise separable convolutions.

The layer blocks are then stacked one on top of the other to construct the MobileNetV1 architecture. MobileNetV1 comprises 87 layers based on the hard library, separated into 13 blocks, and does not include the classification layer, which has features based on the original dataset, namely ImageNet.

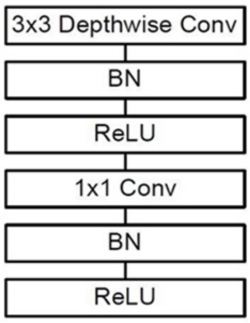

Figure 3. Depthwise Separable Convolution [13]

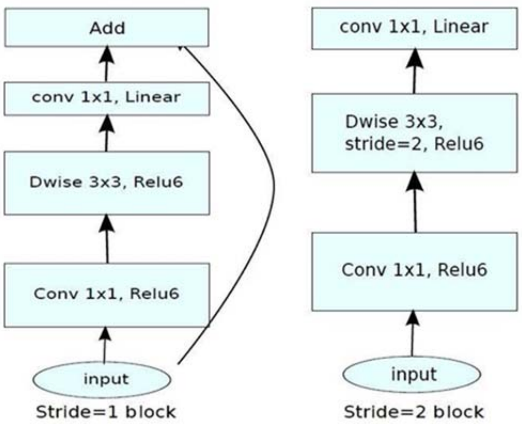

Figure 4. Residual Block of MobileNetV2 [14]

E. MobileNetV2

MobileNetV2 is a development model of MobileNetV1. In an image classification experiment using ImageNet, MobileNetV2 outperforms MobileNetV1 with fewer parameters *[14]*. As shown in Figure 4, there are two types of blocks in MobileNetV2. The residual blocks with stride 1 and stride 2 are shown below, and two blocks are arranged to form the MobileNetV2 architecture.

It is built on the MobileNetV2 hardware library and has 155 layers grouped into 16 blocks.

F. VGG16

The VGG16 deep learning model employs a convolution layer on a tiny (33) convolutional filter specification. The depth of the neural network can be raised by adding more convolutional layers based on the size of the convolutional filter.

As a result, the CNN model outperforms earlier CNN models in terms of accuracy. The VGG16 model includes 19 layers, 16 of which are convolutional and 3 of which are completely linked. VGG16's architecture is shown below.

G. VGG19

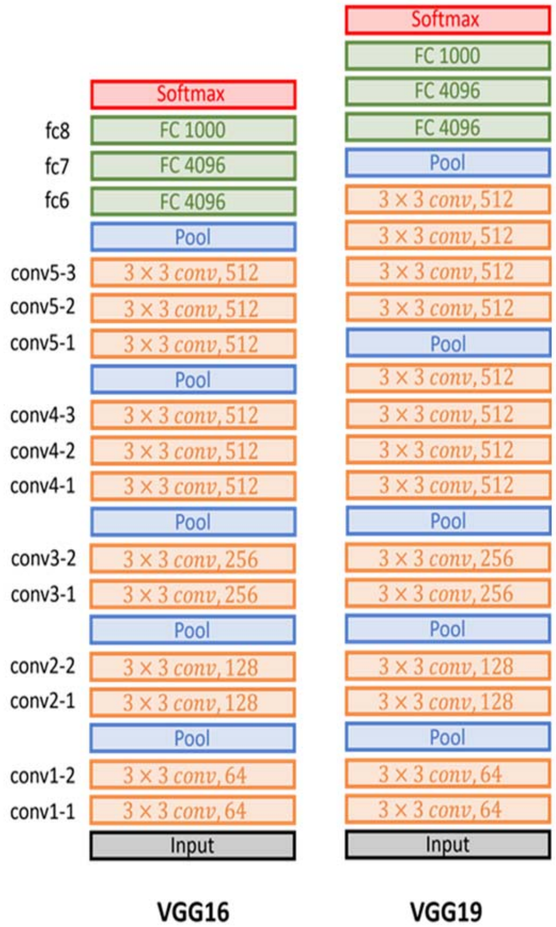

Figure 5. Comparison of VGG-16 dan VGG-19 Architecture

VGG-19 is a deep convolutional neural network with 19 layers. Below is a network version that has been pre-trained on over a million photos from the ImageNet database. The trained network can classify pictures into 1000 distinct item categories, such as keyboards, mice, pens, and animals. As a result, the network has looked at feature-rich representations for a wide range of pictures. The network's picture input size is 224x224. The VGG-19 and VGG-16 architectures are compared as shown in Figure 5.

H. ResNet101V2

The residual network created by He at al. [15] was the ILSVRC 2015 winner. Skip connections and the usage of very large batch normalization are unique elements of this architecture. They proposed modified residual network architecture as shown in Figure 6, which part (a) shows the original (ResNetV1), and part b shows the modified (ResNetV2).

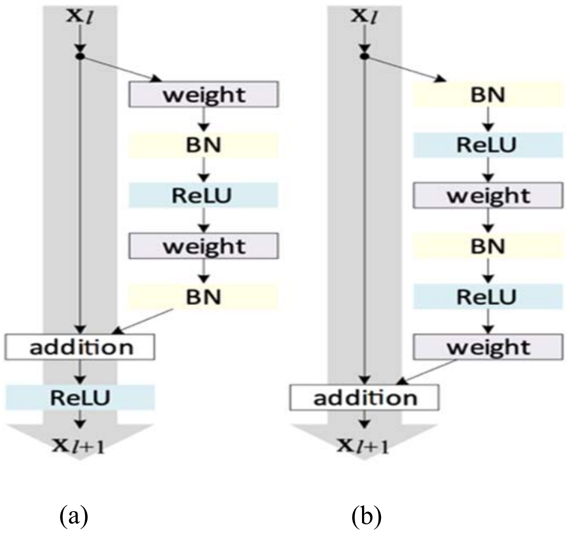

Figure 6. Architecture ResNetV1(a) and ResNetV2 (b) [16]

4. Proposed Method

The accumulation of public datasets AllAgeFace, FG-NET, and FaceAge, which will be separated into three parts: training data, validation data, and testing data, will be the first step in this work. Labelling will be done according to age categories, which are as follows: 0-12 child, 13-17 adolescent, and 18-77 adult. The data is then pre-processed by labelling and scaling it to 224x224 pixels. The data is then encoded and normalized to 0-1. An experiment was carried out prior to training to determine the value of the learning rate with steady performance. Following that, the training process begins by determining the model to be utilized as well as the parameter list to be decided, such as batch size and the number of training epochs.

The previously developed training and validation data will be used in the training process. During the training phase, the accuracy will be calculated using a loss function, cross-entropy category, and training time calculation. VGG16, VGG19, MobileNet, MobileNetV2, Xception, and ResNet101V2 are the deep learning models presented in this work. As shown in Figure 7, the training procedure will be repeated 12 times with different architectures. Following training, each model is tested using distinct data, referred to as testing data. The testing findings will be examined with a cofusion matrix and then evaluated by comparing test results from one model to another. This research makes use of the open-source libraries TensorFlow and Keras.

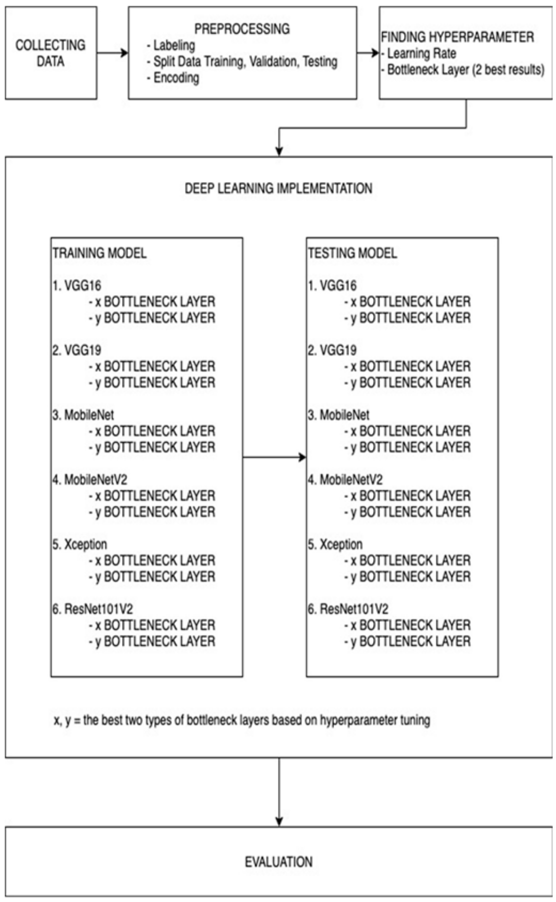

Figure 7. Proposed Method

5. Experiment Results

A. Datasets

The form of the dataset in this study is an image of a person's face with an age label ranging from 0 to 77 years. The facial photos were gathered from numerous Kagle databases, including the All-Age-Face dataset, the FacialAge dataset, and the FGNET dataset. The three datasets are pooled and divided into three categories: child, adolescent, and adult.

The child class comprises children aged 0 to 12 years, the adolescent class of children aged 12 to 17 years, and the adult class of children aged 18 to 77 years. The data for children is 1500 photographs, the data for adolescents is 1500 images, and the data for adults is 1500 images. The image from the three datasets is a cropped portrait of a person with a focus on the face. The images were then identified according to their age group. The majority of the images in this dataset are in.jpeg and.png format. It too has a resolution of 224 × 224. Figure 8 depicts partial images from the dataset for each class.

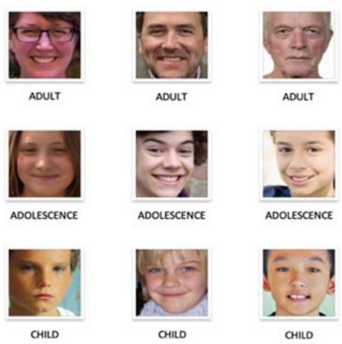

Figure 8. Sample of facial image for each class

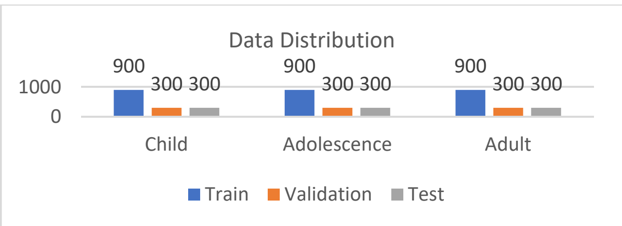

Figure 9. Data Distribution

The total number of images collected is 4500, which are evenly distributed among the three classifications mentioned above. Each class comprises 1500 photos, with 60% being training data, 20% being validation data, and 20% being testing data. The graph in Figure 9 depicts the distribution of picture data into three categories.

B. Finding Hyperparameter

This procedure yields the optimal hyperparameter value for the learning rate as well as the two best bottleneck layers to include in the training. The MobileNet model is utilized in the process of discovering hyperparameters since it is the fastest runtime model. The learning rates at 0.001, 0.0001, and 0.00001 are compared. The results reveal that a learning rate of 0.00001 produces the best results in terms of testing and validation accuracy. Various numbers of bottleneck layers can be incorporated in the re-training process, which is also tested. The first experiment employs 0 layer, implying that all layers from the original model design are frozen. The second training begins with the last one block, the third with the last two blocks, and the fourth with the last three blocks. The results reveal that training from the last one and two blocks produces the best results. After 99 epochs, it gives the highest testing and validation accuracy.

C. Deep Learning Model Implementation

Table 1 shows summary validation and testing of twelve models in deep learning. The results are analyzed in each section. Confusion matrix of each model is not displayed, except model "Xception 1 Bottleneck Layer" which is shown in Figure 10, to give the illustration of whole confusion matrix model.

Table 1. Summary validation and Testing for Some Models

Model	Validation (%)	Testing (%)
Xception 1 Bottleneck Layer	82	82.66
Xception 2 Bottleneck Layer	84	83
VGG16 1 Bottleneck Layer	82	82.33
VGG16 2 Bottleneck Layer	85	85
VGG19 1 Bottleneck Layer	81.67	81.78
VGG19 2 Bottleneck Layer	82	82.22
MoblieNet 1 Bottleneck Layer	83	83.89
MoblieNet 2 Bottleneck Layer	85	85.44
MobileNetV2 1 Bottleneck Layer	82	81.67
MobileNetV2 2 Bottleneck Layer	82	81.99
ResNet101V2 1 Bottleneck Layer	82	83.56
ResNet101V2 2 Bottleneck Layer	81	81.67

1) Xception 1 Bottleneck Layer

The validation accuracy of the Xception 1 bottleneck layer is 82%, and the testing accuracy is 82.66%. The confusion matrix of the testing findings for the Xception 1 bottleneck layer is shown in Figure 10.

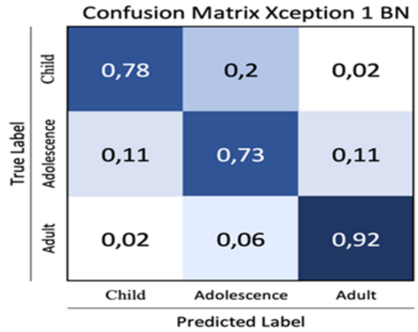

Figure 10. Confusion Matrix Xception 1 bottleneck layer

Table 2 summarizes the classification report on the testing dataset for the Xception 1 bottleneck layer. It has an average accuracy of 83 percent, an average recall of 83 percent, and an average f1-score of 83 percent, as shown in Table 2.

Table 2. Precision, recall, and f1-score for Xception 1 bottleneck layer

Class	Precision	Recall	F1-score	Support
Child	0,85	0,78	0,81	300
Adolescence	0,75	0,78	0,77	300
Adult	0,88	0,92	0,9	300
accuracy			**0,83**	900
average	**0,83**	**0,83**	**0,83**	900

2) Xception 2 Bottleneck Layer

The validation accuracy of Xception 2 bottleneck layer is 84%, while the testing accuracy is 83%.

Table 3. Precision, recall, and f1-score for Xception 2 bottleneck layer

Class	Precision	Recall	F1-score	Support
Child	0,82	0,82	0,82	300
Adolescence	0,79	0,73	0,76	300
Adult	0,87	0,94	0,91	300
accuracy			**0,83**	900
average	**0,83**	**0,83**	**0,83**	900

The classification report on the testing dataset for the Xception 2 bottleneck layer is summarized in Table 3. As demonstrated in Table 3, it has an average precision of 83%, an average recall of 83%, and an average f1-score of 83%.

3) VGG16 1 Bottleneck Layer

The validation accuracy of the VGG16 1 bottleneck layer is 82 %, while the testing accuracy is 83.33 %. Table 4 provides the classification report for the VGG16 1 bottleneck layer testing dataset. As demonstrated in Table 4, it has an average precision of 83%, an average recall of 83%, and an average f1-score of 83%.

Table 4. Precision, Recall, And F1-Score For Vgg16 1 Bottleneck Layer

Class	Precision	Recall	F1-score	Support
Child	0,89	0,77	0,82	300
Adolescence	0,75	0,8	0,78	300
Adult	0,87	0,93	0,9	300
accuracy			**0,83**	900
average	**0,84**	**0,83**	**0,83**	900

4) VGG16 2 Bottleneck Layer

The VGG16 2 bottleneck layer has a validation accuracy of 85% and a testing accuracy of 85%. Table 5 provides the classification report for the VGG16 2 bottleneck layer testing dataset. As demonstrated in Table 5, it has an average precision of 85%, an average recall of 85%, and an average f1-score of 85%.

Table 5. Precision, Recall, And F1-Score For VGG16 2 Bottleneck Layer

Class	Precision	Recall	F1-score	Support
Child	0,87	0,82	0,85	300
Adolescence	0,78	0,81	0,8	300
Adult	0,9	0,92	0,91	300
accuracy			**0,85**	900
average	**0,85**	**0,85**	**0,85**	900

5) VGG19 1 Bottleneck Layer

VGG19 1 bottleneck layer has a validation accuracy of 81,67% and a testing accuracy of 81,78%. Table 6 provides the classification report for the vgg19 1 bottleneck layer testing dataset. As indicated in Table 6, it has an average precision of 83 %, recall of 82 %, and f1-score of 82 %.

Table 6. Precision, Recall, And F1-Score For VGG19 1 Bottleneck Layer

Class	Precision	Recall	F1-score	Support
Child	0,94	0,69	0,8	300
Adolescence	0,71	0,83	0,76	300
Adult	0,86	0,93	0,89	300
accuracy			**0,82**	900
average	**0,83**	**0,82**	**0,82**	900

6) VGG19 2 Bottleneck Layer

The validation accuracy of the VGG19 2 bottleneck layer is 82 %, while the testing accuracy is 82,22 %. Table 7 provides the classification report for the VGG19 1 bottleneck layer testing dataset. It has an average precision of 84 %, an average recall of 82 %, and an average f1-score of 82 %, as shown in Table 7.

Table 7. Precision, recall, and f1-score for VGG19 2 bottleneck layer

Class	Precision	Recall	F1-score	Support
Child	0,97	0,66	0,78	300
Adolescence	0,73	0,84	0,78	300
Adult	0,83	0,97	0,9	300
accuracy			**0,82**	900
average	**0,84**	**0,82**	**0,82**	900

7) MoblieNet 1 Bottleneck Layer

The validation accuracy of MobileNet 1 bottleneck layer is 83 %, while the testing accuracy is 83,89 %. Table 8 highlights the categorization report for the MobileNet 1 bottleneck layer testing dataset. As indicated in Table 8, it has an average precision of 84%, an average recall of 84%, and an average f1-score of 84%.

Table 8. Precision, recall, and f1-score for Mobilenet 1 bottleneck layer

Class	Precision	Recall	F1-score	Support
Child	0,87	0,82	0,84	300
Adolescence	0,8	0,77	0,78	300
Adult	0,85	0,93	0,89	300
accuracy			**0,84**	900
average	**0,84**	**0,84**	**0,84**	900

8) MobileNet 2 Bottleneck Layer

The bottleneck layer of MobileNet 2 has a validation accuracy of 85 % and a testing accuracy of 85,44 %. The classification report on the testing dataset for the MobileNet 2 bottleneck layer is summarized in Table 9. As demonstrated in Table 9, it has an average precision of 85%, an average recall of 85%, and an average f1-score of 85%

Table 9. Precision, recall, and f1-score for Mobilenet 2 bottleneck layer

Class	Precision	Recall	F1-score	Support
Child	0,88	0,82	0,85	300
Adolescence	0,80	0,80	0,80	300
Adult	0,89	0,94	0,91	300
accuracy			**0,85**	900
average	**0,85**	**0,85**	**0,85**	900

9) MobileNetV2 1 Bottleneck Layer

MobileNetV2 1 bottleneck layer has a validation accuracy of 82% and a testing accuracy of 81.67%. Table 10 highlights the categorization report for the MobileNetV2 1 bottleneck layer testing dataset. As demonstrated in Table 10, it has an average precision of 82%, an average recall of 82%, and an average f1-score of 82%.

Table 10. Precision, recall, and f1-score for MobileNetV2 1 bottleneck layer

Class	Precision	Recall	F1-score	Support
Child	0,84	0,76	0,8	300
Adolescence	0,74	0,8	0,77	300
Adult	0,88	0,89	0,88	300
accuracy			**0,82**	900
average	**0,82**	**0,82**	**0,82**	900

10) MobileNetV2 2 Bottleneck Layer

MobileNetV2 2 bottleneck layer has a validation accuracy of 82% and a testing accuracy of 81.99%. Table 11 highlights the categorization report for the MobileNetV2 2 bottleneck layer testing dataset. As indicated in Table 11, it has an average precision of 82%, an average recall of 82%, and an average f1-score of 82%.

Table 11. Precision, recall, and f1-score for MobileNetV2 2 bottleneck layer

Class	Precision	Recall	F1-score	Support
Child	0,83	0,8	0,81	300
Adolescence	0,75	0,76	0,76	300
Adult	0,87	0,9	0,89	300
accuracy			**0,82**	900
average	**0,82**	**0,82**	**0,82**	900

11) ResNet101V2 1 Bottleneck Layer

ResNet101V2 1 bottleneck layer has a validation accuracy of 82% and a testing accuracy of 83.56%. Table 12 highlights the classification report for the ResNet101V2 1 bottleneck layer testing dataset. As indicated in Table 12, it has an average precision of 84%, an average recall of 84%, and an average f1-score of 84%.

Table 12. Precision, recall, and f1-score for ResNet101V2 1 bottleneck layer

Class	Precision	Recall	F1-score	Support
Child	0,87	0,78	0,82	300
Adolescence	0,76	0,82	0,79	300
Adult	0,89	0,91	0,9	300
accuracy			**0,84**	900
average	**0,84**	**0,84**	**0,84**	900

12) ResNet101V2 2 Bottleneck Layer

The validation accuracy of ResNet101V2 2 bottleneck layer is 81%, while the testing accuracy is 81.67%. Table 13 highlights the classification report for the ResNet101V2 2 bottleneck layer testing dataset. As indicated in Table 13, it has an average precision of 82%, an average recall of 82%, and an average f1-score of 82%.

D. Evaluation

Table 14 shows the f1-score values for each class for each model. As a result, the adolescent class has the lowest performance, followed by a kid class and an adult class with the highest performance. On average, 12 trials with varied topologies result in a juvenile class f1-score of 0.7766, a child class f1-score of 0.8158, and an adult class f1-score of 0.8966.

Table 13. Precision, recall, and f1-score for ResNet101V2 2 bottleneck layer

Class	Precision	Recall	F1-score	Support
Child	0,87	0,73	0,79	300
Adolescence	0,73	0,82	0,77	300
Adult	0,87	0,9	0,88	300
accuracy			**0,82**	900
average	**0,82**	**0,82**	**0,82**	900

Table 14. Class ranking based on f1-score

Model		MobileNet 2 Block Layer	VGG16 2 Block Layer	MobileNet 1 Block Layer	ResNet101V2 1 Block Layer	VGG16 1 Block Layer	Xception 2 Block Layer	Xception 1 Block Layer	VGG 19 2 Block Layer	MobileNetV2 2 Block Layer	MobileNetV2 1 Block Layer	VGG 19 1 Block Layer	ResNet101V2 2 Block Layer
F1 Score	Child	0,91	0,91	0,89	0,9	0,9	0,91	0,9	0,9	0,89	0,88	0,89	0,88
	Adolescence	0,85	0,85	0,84	0,82	0,82	0,82	0,81	0,78	0,81	0,8	0,8	0,79
	Adult	0,8	0,8	0,78	0,79	0,78	0,76	0,77	0,78	0,76	0,77	0,76	0,77
Average		85%	85%	84%	84%	83%	83%	83%	82%	82%	82%	82%	81%

Teenage class performance is lower than the other two classes because early teen characteristics with ages 12–13 are frequently discovered as children, whereas adolescents with ages 17 are sometimes detected as adult classes.

Table 15. Percentage of in-class misconception

Child predicted as Adolescence				Child predicted as Adult			
No	**Model**	**1 block**	**2 block**	**No**	**Model**	**1 block**	**2 block**
1	Xception	11%	15%	1	Xception	2%	2%
2	VGG16	9%	10%	2	VGG16	1%	2%
3	VGG19	4%	2%	3	VGG19	0%	0%
4	MobileNetV1	10%	10%	4	Mobile NetV1	3%	1%
5	MobileNetV2	10%	13%	5	Mobile NetV2	5%	3%
6	ResNetV2	9%	9%	6	ResNet V2	3%	2%
average		9,33%		average		2%	

Table 15, Table 16, and Table 17 show the proportion of the model's fallacy when predicting classes based on 12 trials with data testing up to 300 facial pictures. The error rate for forecasting an adolescent class when the model predicts the class entered a kid class is 20.08 percent on average. Adult predicted adolescent classes had a 5.6% average. Later classes of children were anticipated to be adolescents by an average of 9.33%, whereas adult courses were predicted to be adolescents by an average of 11%. The misconception is thought to have accumulated as a result of each individual's varied growth during adolescence.

Table 16. Percentage of in-class misconception for adolescents

Adolescence predicted as Child				Adolescence predicted as Adult			
No	**Model**	**1 block**	**2 block**	**No**	**Model**	**1 block**	**2 block**
1	Xception	20%	16%	1	Xception	6%	4%
2	VGG16	20%	16%	2	VGG16	6%	6%
3	VGG19	28%	28%	3	VGG19	7%	3%
4	MobileNetV1	16%	15%	4	MobileNetV1	4%	5%
5	MobileNetV2	22%	18%	5	MobileNetV2	6%	7%
6	ResNetV2	20%	22%	6	ResNetV2	6%	8%
average		20,08%		average		5,67%	

Table 17. Percentage of errors in adult class

Adult predicted as Child				Adult predicted as Adolescence			
No	**Model**	**1 block**	**2 block**	**No**	**Model**	**1 block**	**2 block**
1	Xception	2%	2%	1	Xception	11%	12%
2	VGG16	3%	1%	2	VGG16	11%	9%
3	VGG19	3%	6%	3	VGG19	13%	13%
4	Mobile NetV1	2%	2%	4	MobileNetV1	14%	10%
5	Mobile NetV2	2%	2%	5	MobileNetV2	10%	11%
6	ResNet V2	2%	5%	6	ResNetV2	9%	9%
average		2,67%		average		11,00%	

Teenagers are a transitional stage from childhood to adulthood. Some people grow quickly, while others are reliant on both internal and external forces. Prediction errors in the other class are very tiny in the 2% range, owing to less clear pictures or faces that do not occur in the entire dataset.

Table 18. Model ranking based on lowest Testing Loss

Model / Result	1 Block			2 Block		
	Accuracy	Loss	Runtime	Accuracy	Loss	Runtime
MobileNet	83,89%	0,43	4s	85,44%	0,44	6s
Xception	82,56%	0,56	10s	83,00%	0,67	16s
MobileNetV2	81,67%	0,68	9s	81,89%	0,81	9s
ResNet101V2	83,56%	0,77	15s	81,67%	1,08	26s
VGG16	83,33%	1,04	5s	85,00%	0,92	6s
VGG19	81,78%	1,24	6s	82.22%	1,34	8s

In this experiment, each model has unique features in terms of performance testing accuracy, testing loss, and testing runtime. It can be shown that the more blocks in a model, the longer the testing time. The MobileNet model has the highest accuracy, the lowest loss, and the shortest runtime in this scenario. Table 18 shows the results of testing on each model.

6. Conclusion

Based on the results of 12 trials against 6 such models, it is clear that VGG16 and MobileNet have the highest accuracy testing values, but VGG16 also has a high loss value, implying that it has an unfavorable performance. So, based on 6 models tested in this experiment, the model with the lowest loss testing value was MobileNet, which had a loss value of 0.43 at 1 block layer and 0.47 at 2 block layer. MobileNet has an accuracy testing value of 85.44 % with two block layers. As a result, the study concludes that MobileNet is the best model for age classification instances based on facial photos. The ANN+Gabor Filter approach was utilized in the study, and the datasets were FG-NET, yielding a test accuracy of 83%.

References

[1]. Agbo-Ajala, O., & Viriri, S. (2019, November). Face-based age and gender classification using deep learning model. In *Pacific-Rim Symposium on Image and Video Technology* (pp. 125-137). Springer, Cham.

[2]. Melangi, S. (2020). Klasifikasi Usia Berdasarkan Citra Wajah Menggunakan Algoritma Artificial Neural Network dan Gabor Filter. *Jambura Journal of Electrical and Electronics Engineering*, *2*(2), 60-67.

[3]. Kusuma, G. P. (2019). Classification of C2C e-Commerce Product Images using Deep Learning Algorithm. *International Journal of Advanced Computer Science and Applications*, *10*(9).

[4]. Putra, M. T. D., and Kusuma, G. P. (2019). Batik Classification using Deep Learning. *International Journal of Recent Technology and Engineering (IJRTE),* 8(4), 11416–11421. http://www.doi.org/10.35940/ijrte.D9039.118419

[5]. Ramesha, K. B. R. K., Raja, K. B., Venugopal, K. R., & Patnaik, L. M. (2010). Feature extraction based face recognition, gender and age classification. *Int. J. Comput. Sci. Eng.*, *1*(1S), 14-23.

[6]. Qawaqneh, Z., Mallouh, A. A., & Barkana, B. D. (2017). Deep convolutional neural network for age estimation based on VGG-face model. *arXiv preprint arXiv:1709.01664*.

[7]. Anand, A., Labati, R. D., Genovese, A., Munoz, E., Piuri, V., & Scotti, F. (2017, November). Age estimation based on face images and pre-trained convolutional neural networks. In *2017 IEEE symposium series on computational intelligence (SSCI)* (pp. 1-7). IEEE.

[8]. Fei-Fei L, Karpathy A, Johnson J (2016) CS231n:Convolutional neural networks for visual recognition. Stanford.

[9]. Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798-1828.

[10]. Shao, S., McAleer, S., Yan, R., & Baldi, P. (2018). Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics*, *15*(4), 2446-2455.

[11]. Soekhoe, D., Putten, P. V. D., & Plaat, A. (2016, October). On the impact of data set size in transfer learning using deep neural networks. In *International symposium on intelligent data analysis* (pp. 50-60). Springer, Cham.

[12]. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018, October). A survey on deep transfer learning. In *International conference on artificial neural networks* (pp. 270-279). Springer, Cham.

[13]. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[14]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

[15]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[16]. Agbo-Ajala, O., & Viriri, S. (2021). Deep learning approach for facial age classification: a survey of the state-of-the-art. *Artificial Intelligence Review*, *54*(1), 179-213.