International Journal on Advanced Science Engineering Information Technology

Vol.13 (2023) No. 2
ISSN: 2088-5334

Performance of ShuffleNet and VGG-19 Architectural Classification Models for Face Recognition in Autistic Children

Melinda Melinda [a,*], Maulisa Oktiana [a], Yudha Nurdin [a], Indah Pujiati [a], Muhammad Irhamsyah [a], Nurlida Basir [b]

[a] *Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia*
[b] *Fakulti Sains dan Teknologi, Universiti Sains Islam Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia*
Corresponding author: [*]melinda@usk.ac.id

***Abstract*— This study discusses the face recognition of children with special needs, especially those with autism. Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder that affects social skills, ways of interacting, and communication disorders. Facial recognition in autistic children is needed to help detect autism quickly to minimize the risk of further complications. There is extraordinarily little research on facial recognition of autistic children, and the resulting system is not fully accurate. This research proposes using the Convolution Neural Network (CNN) model using two architectures: ShuffleNet, which uses randomization channels, and Visual Geometry Group (VGG)-19, which has 19 layers for the classification process. The research object used in the face recognition system is secondary data obtained through the Kaggle site with a total of 2,940 image data consisting of images of autism and non-autism. The faces of autistic children are visually difficult to distinguish from those of normal children. Therefore, this system was built to recognize the faces of people with autism. The method used in this research is applying the CNN model to autism face recognition through images by comparing two architectures to see their best performance. Autism and non-autism data are grouped into training data, 2,540, and test data, as much as 300. In the training stage, the data was validated using validation data consisting of 50 autism image data and 50 non-autism image data. The experimental results show that the VGG-19 has high accuracy at 98%, while ShuffleNet is 88%.**

***Keywords*— Face recognition system; autism; Convolutional Neural Network (CNN); ShuffleNet; VGG-19.**

Manuscript received 12 Aug. 2022; revised 6 Oct. 2022; accepted 7 Nov. 2022. Date of publication 30 Apr. 2023.
IJASEIT is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

I. INTRODUCTION

Various research on facial recognition has been conducted [1]–[5]. However, there is still limited research on the facial recognition of autistic children. Currently, the detection of autism in children from a medical point of view has been carried out with evaluations related to children's growth and development through speech, behavior, hearing tests, and genetic tests. This evaluation requires a complex process and is ineffective in getting the results of the autism diagnosis. In contrast, people with autism must receive treatment as early as possible to minimize the risk of more serious complications. This study proposes a facial recognition system for autistic children using the Deep Learning method to recognize the characteristics of autism through facial images. Visually, the faces of children with autism and non-autism are almost identical. However, a study found that the facial features of children diagnosed with autism are located on the upper face (forehead), the distance between the eyes, the middle of the face and lips, and the philtrum (the area between the nose and lips) which looks wider than normal children [6]–[11].

Research by Robson [12] explored the Convolutional Neural Network (CNN) method with the CVGG-19 Transfer Learning technique to analyze data on children with autism with an accuracy rate of 96.10%. Then research by Jahanara [13] used the MobileNet algorithm to diagnose autistic people with an accuracy of 94.6%. Meanwhile, Akter et al. [14] research explored MobileNet-VI for early detection of autism symptoms with an accuracy of 90.67%. As well as a study was conducted by Kalantarian et al. [15], which developed a crowdsource facial-data emotion labeling method in children with ASD that yielded an accuracy rate of 94% for disgust, 81% for normal expression, 92% for surprise, and 50% for fear. Banire et al. [16] proposed a method to distinguish attention from inattention in autistic children through facial features from the study obtained an accuracy using the Support Vector Machine (SVM) classification of 91% and the CNN method of 76.31%.

This study explored a CNN architecture using two architectures, namely ShuffleNet and VGG-19, for the facial recognition of autistic children. This architecture is intended to produce models with good accuracy in the recognition and classification processes and the performance of the two architectures. Based on previous research, the ShuffleNet architecture performs better with a fast computational process [21]. In contrast, using the VGG-19 method, Jahanara [13], had fairly good accuracy. Thus, this study adopted the ShuffleNet and the VGG-19 architecture for the facial recognition of children with autism.

The main contribution of this paper is as follows:

- We significantly improve the performance of autistic children's facial recognition using ShuffleNet and VGG-19 architectures.
- To the best of our knowledge, ShuffleNet has never been used for facial recognition cases for autistic children, but based on other references, ShuffleNet is quite fast in computing processes, so this study adopted the ShuffleNet architecture for computational efficiency.

The remainder of this paper is organized as follows:

- Section II describes the material and method.
- Results and discussion are presented in Section III.
- Section IV concludes the paper.

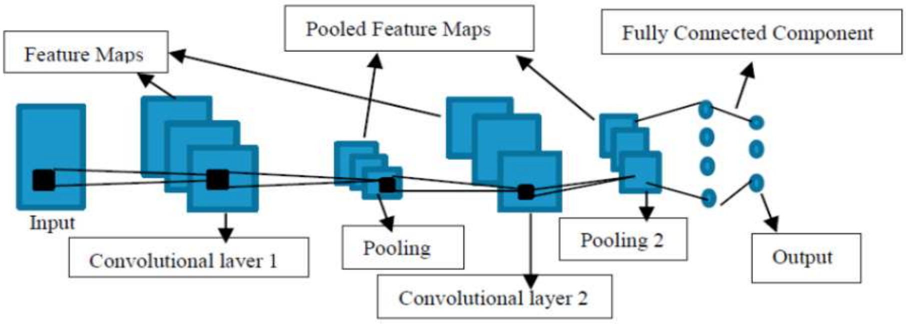


Fig. 1 The CNN algorithm structure

A. Facial Recognition System

Face recognition systems use technology to identify faces in the input image. This system will recognize and match the images contained in the dataset to the input images [22]. The face recognition process is carried out using a computerized method, including detecting and verifying a person through an image [23], [24]. Although facial recognition systems have been widely studied, several challenges, such as misalignment, illumination variations, and expression variations, require approaches and tests to improve face recognition's accuracy and precision level.

B. Convolutional Neural Network (CNN)

CNN or CovNet is one of the deep learning algorithms widely used to process data in the form of multiple arrays. The CNN algorithm will process two-dimensional data. The *CNN* algorithm is very popular in the problem of recognizing two-dimensional images, namely images or audio spectrograms, and three-dimensional images, such as video [25]. CNN consists of two layers, namely Convolution and Pooling [26]. CNN architecture in Figure 1 has several layers; the input image will go through several layers, namely the convolution, activation, pooling, and fully connected layers [27-28].

C. Shufflenet Architecture

This study uses the ShuffleNet architectural model, which is an alternative model to save computational time; the ShuffleNet architecture is quite efficient because it allows many channels or features that can encode more information, especially in exceedingly small networks [21].

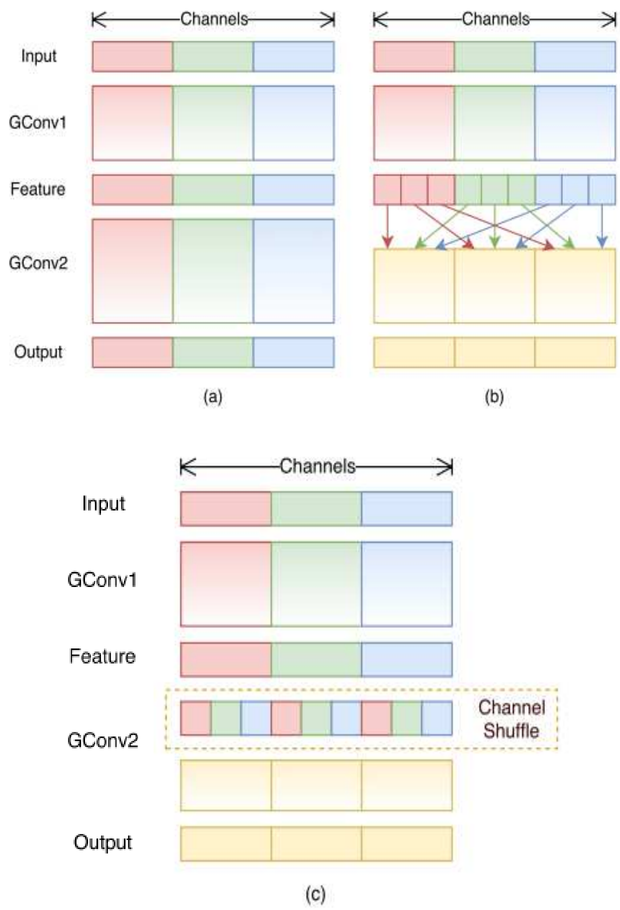


Fig. 2 ShuffleNet network convolution

In convolution stages using the ShuffleNet architectural network in Figure 2, in channel (a), There is no channel randomization; each output channel only corresponds to the input channel in the group. This property blocks the flow of information between channel groups and weakens representation. In channel (b), if group convolution allows input data from different groups, the input and output channels will be fully linked. Channel (c), almost similar to channel (b) operations, can be implemented efficiently and elegantly with channel randomization operations. Suppose a convolution layer with group g whose output has g×n channels; it reshapes the output channel dimensions to (g,n), transposes, and re-aligns it as input from the next layer.

D. Architecture Visual Geometry Group (VGG)-19

Visual Geometry Group (VGG) 19 is a variant of the VGG model with a multilayer neural network consisting of 19 layers, including 16 convolutional layers, 3 fully connected layers, 5 maxpool layers, and 1 SoftMax layer. There are several other types of VGG, such as VGG-11, VGG-16, and others, in which VGG-19 has 19.6 billion FLOPs [29–30]. VGG 19 trained on ImageNet contains millions of images with 1000 categories, a common method for image classification as it uses multiple 3x3 filters in each convolution layer. VGG has six main sizes, each consisting of several connected layers with convolution and full-connected layers with an input of 224×224×3 [31].

This VGG-19 architecture uses an alternating structure of several non-linear activation layers, uses max-pooling for downsampling, and modifies the Rectified Linear Unit (ReLU) as an activation function choosing the largest value in the area of an image. The down-sampling layer is used to increase the anti-distortion capability of an image, maintain the sample's main features, and reduce the number of parameters [32]–[34].

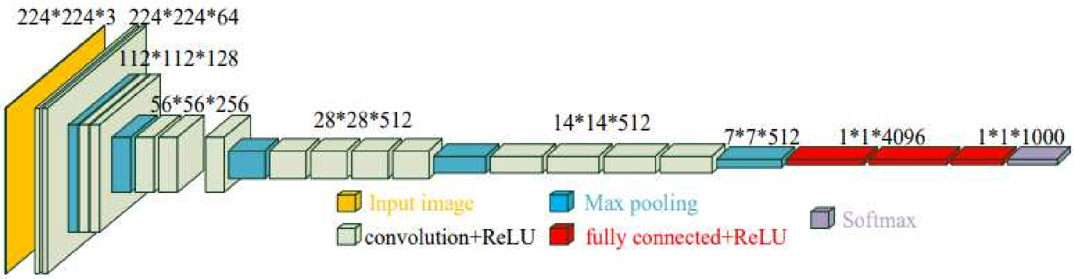
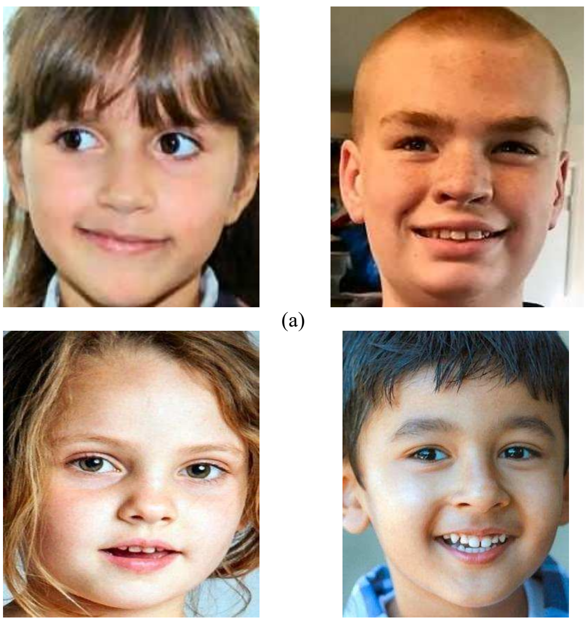


Fig. 3 The VGG-19 architecture [22]

II. Materials and Method

The experiment was conducted using facial images of children with autism and non-autism obtained from [35]. Figure 4 shows the sample of autism and non-autism facial images. A total of 2,540 images are divided into 1,270 autistic facial images and 1,270 non-autistic images with dimensions of 224×224×3.



(b)

Fig. 4 The sample of: (a) autism facial images, (b) non-autism facial images

Figure 5 demonstrates the proposed face recognition system for autistic children framework. At the dataset input stage, the data would be adjusted to the size of the filter according to the data entered. The filter is a two-dimensional matrix that extracts information for each pixel from the image and the number of channels in the image. Then the dataset is followed by a convolution layer for the entered data to produce a featured layer with values in the form of a different matrix from the input data. It then runs the Re-Lu activation function on the featured layer, which its function is to change the negative value to 0 and pass negative numbers so that it will produce a new matrix. The resulting matrix will be downsampled through a pooling layer, namely max pooling. The matrix resulting from the down-sampling process will be reprocessed to obtain the desired results. The results of the pooling layer will be flattened, which will produce a one-dimensional matrix. The one-dimensional matrix will be connected to a fully connected layer. At this layer, several hidden layers will be formed based on the defined parameter. After the hidden layer is formed, it will enter the activation function layer, i.e., the SoftMax activation function. After the activation function is used, the network layers that have been formed is optimized using Adam and SGD optimizer. This optimizer will set the learning rate value, this value will affect the epoch during the training process. After the architecture is determined, the data training stage will be based on the predetermined epoch value. Then the network architecture is validated during the training process and then the architecture will be exported into the model. The resulting model will be used in the testing phase.

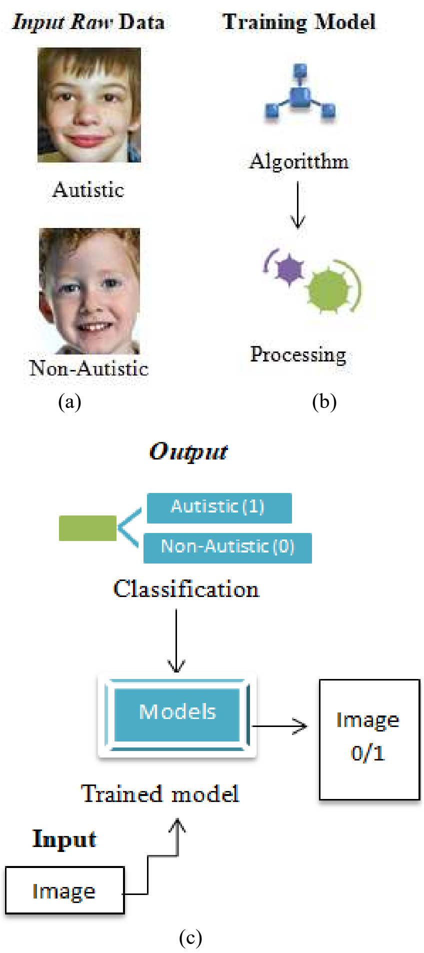


Fig. 5 The proposed framework: (a) raw data input (b) Design model (c) Model output

The results of the implementation of the architecture in this study are classification using two Convolutional Neural Network (CNN) architectures, namely ShuffleNet and Visual Geometry Group (VGG)-19, which have previously been modified. At the testing stage, the model is carried out to test the accuracy of the CNN model that has been made. Test data in the form of images of autism and non-autism that have been entered from a system that has been built using the CNN architecture will be processed to produce output in the form of an introduction to the two types of images that are represented in the form of a table with CSV format. The model output will be manually compared with the original test data. The level of accuracy will be calculated and analyzed using a confusion matrix.

Furthermore, the training phase's output is a model using two CNN architectures: ShuffleNet and VGG-19. The model will be used to recognize two types of data input at the test stage to recognize and classify facial images of people with autism and non-autism. Next, we will compare the results of classifying facial images of people with autism and non-autism through the curves generated from the process.

A. ShuffleNet Architecture Implementation

The ShuffleNet architecture used in this simulation is the second version of the ShuffleNet architecture. The characteristics of the ShuffleNet architecture are described in Table 1. Table 1 shows the ShuffleNet architecture parameters, where the data will be convoluted using Re-Lu activation. Then in the pooling layer, the Maxpooling technique will be used, namely taking the maximum value from a convolution process to determine the value of the new matrix for the next layer. In the next layer, the data will be processed in a fully connected layer where a multidimensional matrix measuring 4×4×64 must first be converted into a one-dimensional matrix using a flatten technique with a size of 1024. Then a matrix measuring 1024 will be converted into a matrix with a size of 512 using Re-Lu activation. In the next stage, the 512-sized matrix will be converted into a 2-size matrix using the SoftMax activation function based on this system, also using a dropout technique to select and remove neurons randomly. It is used to prevent overfitting and is the fully connected layer, which works by deactivating some unnecessary neurons to speed up the training process. During the learning process, the total parameters or weights generated are 581,938 parameters.

TABLE I
SHUFFLENETV2 ARCHITECTURE

Layer (type)	Output Shape	Activation	Parameter
Conv2D	None, 127,127,16	Re-Lu	208
MaxPooling2D	None, 64,64,16	Re-Lu	0
Conv2D	None, 62,62,32	Re-Lu	4640
MaxPooling2D	None, 21, 21, 32	Re-Lu	0
Conv2D	None, 17, 17, 64	Re-Lu	51264
MaxPooling2D	None, 4, 4, 64	Re-Lu	0
Flatten	None, 1024	-	0
Dense	None, 512	Re-Lu	524800
Dropout	None, 512	-	0
Dense	None, 2	Softmax	1026
Total			**581,938**

B. VGG-19 Architecture Implementation

The implementation of the VGG-19 architecture in this study can be seen in Table 2.

TABLE II
VGG-19 ARCHITECTURE

Layer (type)	Output Shape	Activation	Parameter
Conv2D	None, 224, 224, 64	Re-Lu	1792
Conv2D	None, 224, 224, 64	Re-Lu	36928
MaxPooling2D	None, 112, 112, 64	Re-Lu	0
Dropout	None, 63,63,32	-	0
Conv2D	None, 112, 112, 128	Re-Lu	73856
Conv2D	None, 112, 112, 128	Re-Lu	147584
MaxPooling2D	None, 56, 56, 128	Re-Lu	0
Conv2D	None, 56, 56, 256	Re-Lu	295168
Conv2D	None, 56, 56, 256	Re-Lu	590080
Conv2D	None, 56, 56, 256	Re-Lu	590080
Conv2D	None, 56, 56, 256	Re-Lu	590080
MaxPooling2D	None, 28, 28, 256	Re-Lu	0
Conv2D	None, 28, 28, 512	Re-Lu	1180160
Conv2D	None, 28, 28, 513	Re-Lu	2359808
Conv2D	None, 28, 28, 514	Re-Lu	2359808
Conv2D	None, 28, 28, 515	Re-Lu	2359808
MaxPooling2D	None, 14, 14, 512)	Re-Lu	0
Conv2D	None, 14, 14, 512	Re-Lu	2359808
Conv2D	None, 14, 14, 513	Re-Lu	2359808
Conv2D	None, 14, 14, 514	Re-Lu	2359808
Conv2D	None, 14, 14, 515	Re-Lu	2359808
MaxPooling2D	None, 7, 7, 512	Re-Lu	0
Flatten	None, 512	-	0
Dense	None, 512	Re-Lu	262656
Dropout	None, 512	Re-Lu	0
Dense	None, 2	SoftMax	1026
	Total		**20,288,066**

It can be observed that using the VGG-19 architecture, the data will be convoluted using Re-Lu activation. Next, the Maxpooling technique will be used in the pooling layer, namely, taking the maximum value from a convolution process to determine the value of the new matrix for the next layer. Then the data will be processed in a fully connected layer where a multi-dimensional matrix measuring 7×7×512 must first be converted into a one-dimensional matrix using the flattening technique with a size of 512. In the next step, the 512-sized matrix will be converted into a matrix of size 2 using the activation function SoftMax. This architecture also uses a dropout technique to select and eliminate neurons randomly, mainly used to prevent overfitting and speed up the training process. During the learning process, the total parameters or weights generated are 20,288,066 parameters.

E. Accuracy

Accuracy represents the ability of the model to run a system properly, and the trained model will resemble the actual system performance. The system performance can be evaluated using the Accuracy, Recall, Precision, and F-Score metrics [36].

- Accuracy:

$$\frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

- Recall:

$$\frac{TP}{TP+FN} \tag{2}$$

- Precision:

$$\frac{TP}{TP+FP} \tag{3}$$

- F-Score:

$$2 \times \frac{Recall\ x\ Precision}{Recall + Precision} \tag{4}$$

Four combinations of the actual value and the predicted value are used to see the performance of the classification system with the output of two classes. TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

III. Result and Discussions

A. Training and Validation Results using ShuffleNet

The model was developed using the ShuffleNet architecture in the training and validation stages. The dataset is grouped into training and validation data, each of which amounts to 2,540 training data consisting of images of autism and non-autism. At the training and validation stage, where the comparison between training data and validation data is 80:20, there are 2032 images used to be trained. As many as 508 images are used for validation to see the relationship between epochs with the level of accuracy and loss from the training stage. Their respective validations can be observed on the curves shown in Figure 6 and Figure 7.

Figure 6 shows that the blue curve is the accuracy curve at the training stage, and the red curve is the validation accuracy curve. On the training accuracy curve (blue color), it can be observed that at epoch 28, the level of accuracy during the training process is quite high, with a maximum accuracy value of 84%. While on the validation curve (red color), it can be seen that the number of epochs generated to get 83% accuracy is 28 epochs as well. We can conclude that as the number of epochs increases, it will affect the level of accuracy of the training and validation stages.

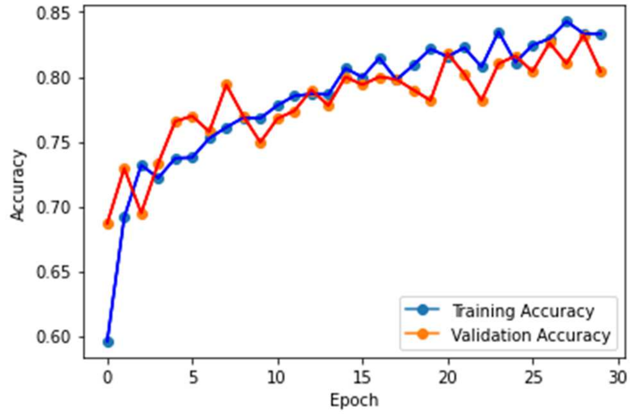


Fig. 6 The Relationship of epoch and accuracy using ShuffleNet

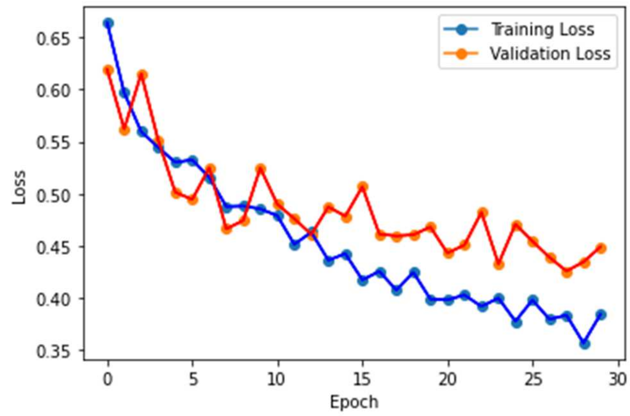


Fig. 7 The Relationship of epoch and loss using ShuffleNet

Based on Figure 7, the results curve for the training stage shows the relationship between epochs and losses using the ShuffleNet architecture. It can be observed that the blue curve is the training stage loss curve, and the red curve is the validation loss curve. The training loss curve (blue color) shows that when the number of epochs is 28, the resulting loss rate is almost close to 30%. While on the loss validation curve (red color), it can be seen that the number of epochs needed to get a loss value close to 40% is when the epochs are 27. Based on the curve obtained in Figure 7, it can be seen that the loss resulting from the training and validation stages is still quite large for the implementation of the ShuffleNet architecture.

B. Training and Validation Results using VGG-19

The model was built using the VGG-19 architecture at the training and validation stages. The dataset is grouped into training and validation data, each of which amounts to 2,540 training data consisting of images of autism and non-autism. At the training and validation stages, where the comparison between training data and validation data is 90:10, it means that there are 2286 images used to be trained, and as many as 254 images are used for validation to see the relationship between epochs with the level of accuracy and loss from the training stages. Moreover, the respective validation can be observed on the curves shown in Figure 8 and Figure 9.

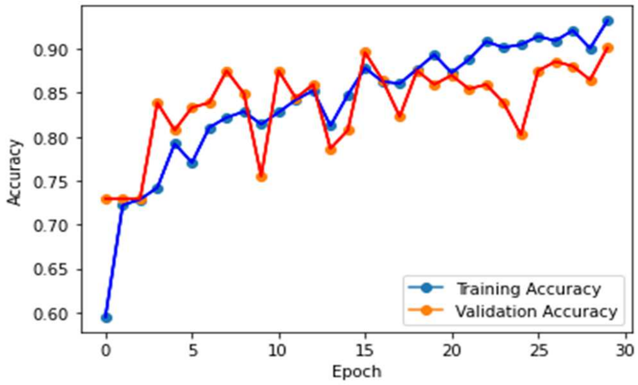


Fig. 8 The Relationship of epoch and accuracy using VGG-19

Figure 8 shows the epoch relationship curve and accuracy using the VGG-19 architecture. Based on these curves, it can be observed that the blue curve is the accuracy curve at the training stage and the red curve is the accuracy curve at the validation stage. On the training accuracy curve (blue color), it can be seen that at epoch 29, the accuracy rate during the training process was quite high, with the accuracy value reaching more than 93%. While on the validation curve (red color), it can be seen that the number of epochs produced to get 90% accuracy is 29 epochs. Based on Figure 8, it can be observed that the level of accuracy on the resulting curve from the training and validation stages will continue to change as the number of epochs increases so that the number of epochs used during training will be tested to achieve the optimum accuracy value.

Based on Figure 9, the epoch-loss relationship curve uses the VGG-19 architecture. Based on the curve, it can be observed that the blue curve is the loss curve for the training stage, and the red one is the loss curve for the validation stage. The training loss curve (blue color) shows that when the number of epochs is 29, the resulting loss rate is 10.9%. While on the loss validation curve (red color), it can be seen that the number of epochs needed to get a loss value close to 30% is when there are 29 epochs. Based on Figure 9, it can be observed that the resulting curve of the training and validation stages has a decrease in loss as the number of epochs increases.

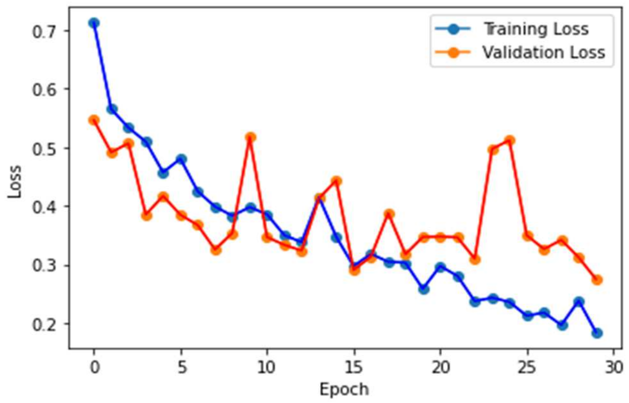


Fig. 9 The Relationship of epoch and loss using VGG-19

C. Comparison of Accuracy Results

Based on the test results in Table 1 and Table 2, the accuracy of the Shufflenet and VGG-19 architectures is obtained. To measure the performance of the classification system, the confusion matrix table can be used, as shown in the table. By using equations (1), (2), (3), and (4), the accuracy results of the two architectures can be observed in Table 3.

TABLE III
PERFORMANCE COMPARISON

Architecture CNN	***Accuracy* %**	***Recall* %**	***Precision* %**	***F-Score* %**
ShuffleNet	88%	85%	90%	87%
VGG-19	98%	97%	99%	98%

Based on Table 3, the comparison shows the best accuracy achievement using VGG-19, where the accuracy value achieved is up to 98%. This study's accuracy level is better than previous study by Jahanara [13], that obtained an accuracy of 96.10%. While ShuffleNet has an accuracy rate of 88%, the model built using this architecture is quite fast, as research by Zhang et al. [21] has done in the classification process. This is quite efficient in saving time during the computational process.

TABLE IV
PERFORMANCE COMPARISON OF SEVERAL METHOD OF FACE RECOGNITION IN AUTISTIC CHILDREN

Method	**Accuracy**	**Ref**
Convolution Neural Network (CNN) by *Transfer Learning* (CVGG-19)	96.10%	[13]
Deep Learning by MobileNet algorithm	94.6%	[6]
Proposed method	98%	

Table IV shows several performance comparisons of several face recognition methods in autistic children. Based on our studies, the results show that the method we proposed has a better accuracy value than the previous method, which is 98%.

IV. CONCLUSION

This study evaluated the performance of CNN architectures for autism facial recognition, namely VGG-19 and ShuffleNet architectures. Autism and Non-Autism look identical, but in this study, they can be distinguished by a facial recognition system using the Convolution Neural Network (CNN) algorithm. The facial recognition system using VGG-19 is much more accurate, with an accuracy rate of 98%, while ShuffleNet results in 88% accuracy. In contrast, ShuffleNet is quite fast in computing processes; therefore, ShuffleNet architecture is efficient for computational.

ACKNOWLEDGMENT

The authors express their gratitude to the Universitas Syiah Kuala for supporting this study.

REFERENCES

[1] S. Singh, D. Singh, and V. Yadav, "Face Recognition Using HOG Feature Extraction and Svm Classifier," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 9, pp. 6437–6440, 2020, doi: 10.30534/ijeter/2020/244892020.

[2] A. P. Ismail and N. M. Tahir, "Human Gait Silhouettes Extraction Using Haar Cascade Classifier on OpenCV," *Proc. - 2017 UKSim-AMSS 19th Int. Conf. Model. Simulation, UKSim 2017*, pp. 105–110, 2018, doi: 10.1109/UKSim.2017.25.

[3] A. Almansour, G. Alsaeedi, H. Almazroui, and H. Almuflehi, "I-Privacy Photo: Face Recognition and Filtering," *ACM Int. Conf. Proceeding Ser.*, pp. 131–141, 2020, doi: 10.1145/3388142.3388161.

[4] E. Setiawan and A. Muttaqin, "Implementation of K-Nearest Neightbors Face Recognition on Low-power Processor," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 13, no. 3, p. 949, 2015, doi: 10.12928/telkomnika.v13i3.713.

[5] M. M. Ghazi and H. K. Ekenel, "A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 102–109, 2016, doi: 10.1109/CVPRW.2016.20.

[6] M. Beary, A. Hadsell, R. Messersmith, and M. P. Hosseini, "Diagnosis of autism in children using facial analysis and deep learning," *arXiv*, 2020.

[7] P. Catherine Lord, P. Susan Risi, P. Pamela S. DiLavore, P. Cory Shulman, P. Audrey Thurm and P. Andrew Pickles, "Autism From 2 to 9 Years of Age," American Medical Association. All rights reserved, vol. 63, no. 6, pp.694-701, 2006.

[8] V. Anagnostopoulou, "The Effectiveness of Social StoriesTM on children with Autism Spectrum Disorder," School Teacher & Special Education Teacher, MA, National and Kapodistrian University of Athens, University of Nottingham, vol. ix, pp. 16, 2020.

[9] L. Mapeli, T. Soda, E. D'angelo and F. Prestori, "The Cerebellar Involvement in Autism Spectrum Disorder from the Social Brain to Mouse Models,"International Journal of Molecular Sciences, 2022

[10] M. P. Kazunari Yoshida, P. Emiko Koyama, P. Clement C. Zai, M. Joseph H. Beitchman, M. James L. Kennedy, P. C. P. Yona Lunsky, "Pharmacogenomic Studies in Intellectual Disabilities and Autism Spectrum Disorder: A Systematic Review", The Canadian Journal of Psychiatry, vol 66, issue 12, 2020, https://doi.org/10.1177/0706743720971950.

[11] M. Pushpal Desarkar and M. P. Daniel J. Muller, "Pharmacogenomic Studies in Intellectual Disabilities and Autism Spectrum Disorder: A Systematic Review," The Canadian Journal of Psychiatry La Revue Canadienne de Psychiatrie, pp. 1-23, 2020.

[12] B. Robson, "Autism spectrum disorder: A review of the current understanding of pathophysiology and complementary therapies in children,"Australian Journal of Herbal Medicine, pp. 128-151, 2013.

[13] S. Jahanara, "Detecting autism from facial image," *Int. J. Adv. Res. Ideas Innov. Technol.*, no. March, pp. 1–8, 2021, doi: 10.13140/RG.2.2.35268.35202.

[14] T. Akter *et al.*, "Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage," *Brain Sci.*, vol. 11, no. 6, 2021, doi: 10.3390/brainsci11060734.

[15] H. Kalantarian *et al.*, "Labeling images with facial emotion and the potential for pediatric healthcare," *Artif. Intell. Med.*, vol. 98, no. April 2018, pp. 77–86, 2019, doi: 10.1016/j.artmed.2019.06.004.

[16] B. Banire, D. Al Thani, M. Qaraqe, and B. Mansoor, "Face-Based Attention Recognition Model for Children with Autism Spectrum Disorder," *J. Healthc. Informatics Res.*, vol. 5, no. 4, pp. 420–445, 2021, doi: 10.1007/s41666-021-00101-y.

[17] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, and L. Beslay, "FaceQnet: Quality Assessment for Face Recognition based on Deep Learning," *2019 Int. Conf. Biometrics, ICB 2019*, 2019, doi: 10.1109/ICB45273.2019.8987255.

[18] S. Albawi, O. Bayat, S. Al-Azawi, and O. N. Ucan, "Social touch gesture recognition using convolutional neural network," *Comput. Intell. Neurosci.*, vol. 2018, 2018, doi: 10.1155/2018/6973103.

[19] J. Alamri, R. Harrabi, and S. Ben Chaabane, "Face Recognition based on Convolution Neural Network and Scale Invariant Feature Transform," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, pp. 644–654, 2021, doi: 10.14569/IJACSA.2021.0120281.

[20] B. Banire, D. Al Thani, M. Qaraqe, and B. Mansoor, "Face-Based Attention Recognition Model for Children with Autism Spectrum Disorder," *J. Healthc. Informatics Res.*, 2021, doi: 10.1007/s41666-021-00101-y.

[21] Zhang, X., Zhou, X., Lin, M., and Sun, J. "ShuffleNet : An extremely efficeint convolutional neural network for mobile devices," Proceeedings of the IEEE conference on computer vision and pattern recognition., pp. 6848-6856, 2018.

[22] L. Tidmarsh and F. R. Volkmar, "Diagnosis and Epidemiology of Autism Spectrum Disorders," *Can. J. Psychiatry*, vol. 48, no. 8, pp. 517–525, 2003, doi: 10.1177/070674370304800803.

[23] A. K. Jain, "Handbook of Face Recognition," pp. 1–40, 2014, [Online]. Available: papers3://publication/uuid/17FEF3DC-D6F9-408B-96E5-871ED133A4F1.

[24] K. H. Teoh, R. C. Ismail, S. Z. M. Naziri, R. Hussin, M. N. M. Isa, and M. S. S. M. Basir, "Face Recognition and Identification using Deep Learning Approach," *J. Phys. Conf. Ser.*, vol. 1755, no. 1, 2021, doi: 10.1088/1742-6596/1755/1/012006.

[25] A. K. Dubey and V. Jain, "Automatic Facial Recognition Using VGG16 based Transfer Learning Model," Journal of Information and Optimization Sciences, vol. 41, no. 7, pp. 1589-1596, 2020.

[26] M. A. Aghdam, A. Sharifi and M. M. Pedram, "Diagnosis of Autism Spectrum Disorders in Young Children Based on Resting-State Functional Magnetic Resonance Imaging Data Using Convolutional Neural Networks,"Journal of Digital Imaging, vol. 32, no. 6, pp. 899-918, 2019.

[27] M. P. Hosseini, M. Beary, A. Hadsell, R. Messermith and H. Soltanian Zadeh, "Deep Learning for Autism Diagnosis and Facial Analysis in Children," Frontiers in Computational Neuroscience, 2022.

[28] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[29] A. A. Cruz-Roa, J. E. Arevalo Ovalle, A. Madabhushi, and F. A. González Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8150 LNCS, no. PART 2, pp. 403–410, 2013, doi: 10.1007/978-3-642-40763-5_50.

[30] J. S. Asri and G. Firmansyah, "Implementasi Objek Detection Dan Tracking Menggunakan Deep Learning Untuk Pengolahan Citra Digital," *Knsi 2018*, pp. 717–723, 2018.

[31] S. J. Rashid, A. I. Abdullah and M. A Shihab, "Face Recognition System Based on Gabor Wavelets Transform, Principal Component Analysis and Support Vector Machine," International Journal on Advanced Science, Engineering and Information Technology, vol.10,no. 3. pp. 959—963, 2020.

[32] B. Achmad and K. Firdausy, "Neural Network-based Face Pose Tracking for Interactive Face Recognition System," International Journal on Advanced Science, Engineering and Information Technology, 1 vol.2, no. 1. pp.105—108, 2012.

[33] F. H. K. Zaman, A. A Sulaiman, I. M. Yassin, N. M. Tahir and Z. I. Rizman, "Development of Mobile Face Verification Based on Locally Normalized Gabor Wavelets," International Journal on Advanced Science, Engineering and Information Technology, vol.7, no. 4. pp. 1198--1205, 2017.

[34] J. Xiao, J. Wang, S. Cao, and B. Li, "Application of a Novel and Improved VGG-19 Network in the Detection of Workers Wearing Masks," in *Journal of Physics: Conference Series*, 2020, vol. 1518, no. 1, doi: 10.1088/1742-6596/1518/1/012041.

[35] "Autism_Image_Data_Kaggle." [Online]. Available: https://www.kaggle.com/cihan063/autism-image-data.

[36] C. He, M. Ma, and P. Wang, "Extract interpretability-accuracy balanced rules from artificial neural networks: A review," *Neurocomputing*, vol. 387, pp. 346–358, 2020, doi: 10.1016/j.neucom.2020.01.036.