2024 IEEE International Conference on Computing, Applications and Systems (COMPAS) | 979-8-3315-2976-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/COMPAS60761.2024.10796046

Computer Vision-based Interactive Autism Detection System using Deep Learning

Badhon Parvej, Sikder Md Mahbub Alam, Faizul Islam Fahim,
Md. Naimul Pathan, Muhammad Aminur Rahaman* {SMIEEE}
Dept of Computer Science and Engineering, Green University of Bangladesh, Bangladesh
Email: badhonparvej481@gmail.com, iimahbubsikder@gmail.com faizulislam837@gmail.com,
naimulpathan99@gmail.com, aminur@cse.green.edu.bd

***Abstract*—Autism Spectrum Disorder (ASD) is a multifaceted neurodevelopmental disorder that influences communication, language abilities, and social interactions, affecting both verbal and nonverbal exchanges. Identifying and diagnosing ASD at an early stage is essential for creating timely educational strategies, providing family assistance, and ensuring suitable medical care is available. Early screening can facilitate the development of optimal therapeutic strategies at the right time. By analyzing facial features, significant markers encoded in human faces can be studied to identify ASD. In this research, we gathered facial images of children with ASD from publicly available sources and employed deep-learning classifiers to sort them. Furthermore, the classifier was utilized to differentiate various ASD subgroups, relying exclusively on the autistic image data. The need for automated diagnosis techniques for ASD has become increasingly important.**

To address this, we made use of a dataset containing 2,940 facial images of children with autism and those who are typically developing. We evaluated the outcomes of existing methods and enhanced the VGG16 model to improve its training and fitting processes. Through substantial modifications to variables such as epoch number and batch size, we increased the accuracy from 77% to 86.3%, while also calculating metrics like precision, specificity, and sensitivity. Moreover, we optimized the training and testing procedures compared to the base code. To enhance the evaluation of the results, we calculated several metrics, such as true positive, true negative, false positive, and false negative rates, in addition to accuracy, precision, sensitivity, and specificity. These results were documented in the submissions.csv file. Additionally, we implemented the VGG19 CNN model and compared its performance with VGG16. The deep learning approach achieved the highest accuracy of 88% with VGG19, followed by 86.3% with VGG16.

***Index Terms*—Autism Spectrum Disorder, Computer Vision, Deep Learning, Max pooling, VGG16, VGG19.**

I. Introduction

ASD is a neurodevelopmental disorder that disrupts social interactions by influencing both verbal and nonverbal communication. The early identification and diagnosis of ASD are essential for timely educational planning, appropriate treatment, family support, and ensuring the child receives necessary medical care without delay. Social communication is the main emphasis of the revised ASD diagnosis criteria [1] [2], mostly because they have communication issues. The child may begin speaking later than their peers and might avoid making eye contact. Their body language often mirrors their behavior and emotions, and they tend to display similar facial expressions as well. Children with autism were compared to typically developing kids who have autistic relatives in the study. Children with autism were shown to have more facial asymmetries than typically developing kids. Facial asymmetry is another indicator of how severe autism symptoms [3]. ASD knows no racial, ethnic, or socioeconomic boundaries, although boys are diagnosed more frequently than girls. [4] Boys are four to five times more likely to have ASD than girls. Over the past two decades, the prevalence of ASD has consistently risen, with current estimates indicating that 1 in 36 children are impacted. The rising incidence of ASD has been linked to genetic factors, a family history of mental health issues, premature birth, and prenatal exposure to psychotropic drugs or pesticides. [5] [6] The correct therapy strategy might be developed and implemented at the right time with the aid of early ASD screening. DL approaches have been extensively employed throughout the past century to evaluate and categorize medical pictures for the diagnosis of illnesses. These methods aid medical professionals in the detection and identification of illnesses. By examining facial characteristics, eye contact, and head movement, it is possible to recognize ASD in individuals whose faces encode significant signals. This study presents an upgraded transfer-learning-based autism face recognition framework aimed at improving early diagnosis accuracy of children with ASD.

Rest of the paper is organized as the the proposed methodology is described in Section II, Section III describes the result analysis with comparisons and the paper is concluded in Section IV

II. Proposed Methodology for Autism Detection

Using VGG16 and VGG19 deep learning models, this research aims to differentiate between autistic and typically developing children by analyzing their facial characteristics. A child's facial features can help indicate whether they have autism or are typically developing. [7]. Key facial characteristics were extracted from images using the VGG16 and VGG19 models. A major advantage of deep learning algorithms is their ability to capture subtle details in images that are imperceptible to the human eye. [8] [9].

979-8-3315-2976-5/24/$31.00

A. Dataset

Our study compared the facial images of autistic and typically developing kids that were taken from the open source platform, which is accessible to everyone online [10]. In this Table I demonstrates that the dataset consists of 2,940 facial images, with half representing children with autism and the other half representing typically developing children. This data was gathered from online sources, such as autism-related websites and Facebook pages.

TABLE I: DIVIDING THE DATASET FOR TRAINING, TESTING, AND VALIDATION.

Total face images	Training dataset	Validation dataset	Testing dataset
2940	2540	100	300

B. Preprocessing

As part of data preprocessing, the images were cleaned and cropped. Piosenkaâs [11] [12] dataset, which was sourced from online collections, required preprocessing prior to the deep learning model's training. The faces in the original photos were automatically cropped by the dataset creator. After that, the dataset was divided into 2,540 pictures for training, 100 for validation, and 300 for testing. The dataset was scaled using a normalization technique, which changed all picture parameters from the range of [0, 255] to [0, 1].

C. Models of Convolutional Neural Networks

Computer vision, a field of artificial intelligence, has seen significant advancements aimed at aiding humans in various facets of everyday life, including medical applications t12. Consequently, the CNN algorithm has played a crucial role in research related to behavior and psychology, as well as in the diagnosis of diseases. Figure 1 illustrates the architecture of the proposed system for our model, which focuses on detecting autism by distinguishing between Autism Spectrum Disorder (ASD) and Typically Developing (TD) children. Our objective is to identify autistic children with the highest accuracy possible. Therefore, we employed a CNN model to enhance performance and utilized a confusion matrix for evaluation.

1) Fundamental Elements of the CNN Model: Convolutional neural networks (CNNs) are unquestionably among the most well-known methods in deep learning [13]. In order to categorize an input image, it applies weights and biases that can be learned to ascertain the importance of certain features [14]. A neuron can be considered a simulation of the connection patterns found in human brain neurons, as it connects and interacts with other neurons. In this section of the article, The input layer, convolutional layer, pooling layer, fully connected layer, activation function, and output prediction will all be covered.

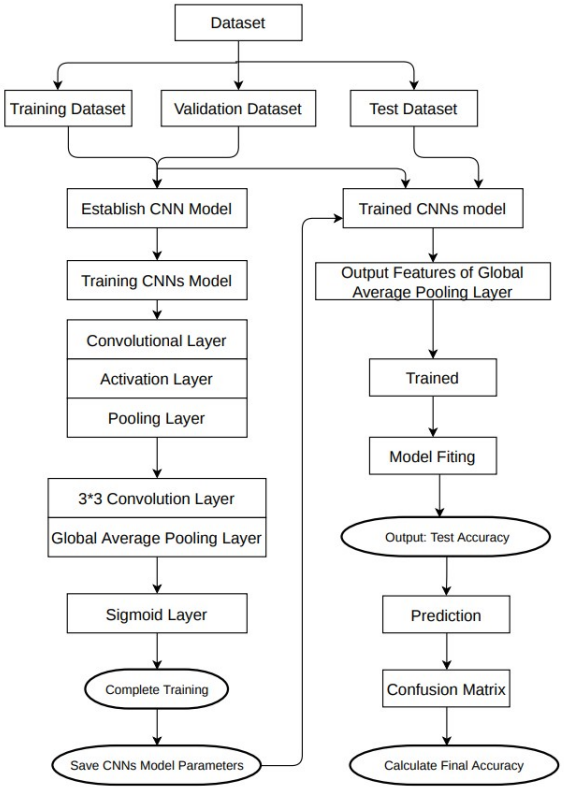

Fig. 1: System architecture for our proposed model

2) Convolutional Layer with a Pooling Layer: An image input is supplied to the convoluteional layer in the form of a matrix of pixel values. Its main goal is to simplify the images while retaining crucial features that aid in autism detection [15] [16]. The first layer of the Convolutional Neural Network (CNN) model extracts fundamental features, such as edges and colors. The CNN architecture supports the addition of more layers, allowing for the extraction of higher-level features that improve visual comprehension. To reduce the computational load from the output of the convolutional layer, which generates a significant number of parameters that may complicate matrix operations, either max pooling or average pooling techniques are used [17] [18]. Max pooling involves selecting the maximum values within each stride window, while average pooling calculates the mean value within the same window. In this study, the model specifically implemented max pooling.

3) Activation Function and Fully Connected Layer.: The fully connected (FC) layer produces outputs by combining high-level characteristics in a nonlinear manner after receiving input from the hidden layers. The input image is represented as a column vector in the FC layer. The forward neural network in the model has a flattened output layer, and it also includes backpropagation pathways. During backpropagation, the neural network reduces loss errors by utilizing the number of training iterations, which allows it to identify additional features. By increasing the number of hidden layers and training iterations, deep learning models frequently enhance

performance and help neural networks recognize basic input attributes more successfully. After obtaining parameters from the fully connected layer, the softmax classifier assesses these features to predict the output, as illustrated. A photograph falls under class 0 if its Softmax output is zero, and class 1 if it has a Softmax output of one. In this study, class 0 denotes an autistic person, whereas class 1 denotes a healthy person.

D. Deep Learning Models

VGG16 and VGG19 are two face feature-based models that have been pre-trained for use in our paper's autism identification algorithm.

1) VGG16: A Convolutional Neural Network (CNN), or ConvNet, is a unique type of artificial neural network. It is made up of an input layer, an output layer, and several hidden layers. VGG16, illustrated in Figure 2, is a notable CNN model recognized as one of the most effective computer vision frameworks to date. The developers of this model assessed existing networks and enhanced their depth by employing an architecture featuring very small (3 Ã 3) convolutional filters, resulting in significant improvements over previous configurations.

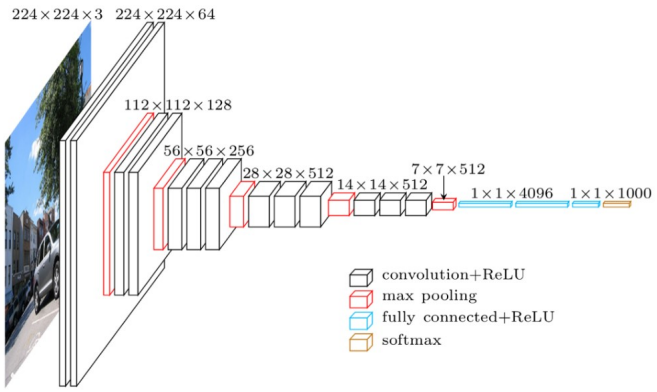

Fig. 2: VGG16 model architecture

2) VGG19: Figure 3 depicts the 19-layer deep artificial neural network model known as the Visual Geometry Group network or VGG19 for short. VGG19 is frequently used with the ImageNet dataset and is based on the Convolutional Neural Network (CNN) architecture. Its importance stems from its simple implementation, marked by the arrangement of 3 Ã 3 convolutional layers, which increases its depth.

- Input layer: The purpose of the input layer is to receive an input image with dimensions of 224 x 224 x 3. The middle 224Ã224Ã3 area from each image was chosen by the model's designers in order to maintain a uniform input size for the ImageNet competition.
- Convolutional layers: With the convolution stride set to 1 pixel, the convolutional layers of VGG utilize a small receptive field of 3Ã3, representing the smallest practical size that allows for the preservation of spatial resolution in both horizontal and vertical directions. The stride refers to the number of pixel shifts across the input matrix.

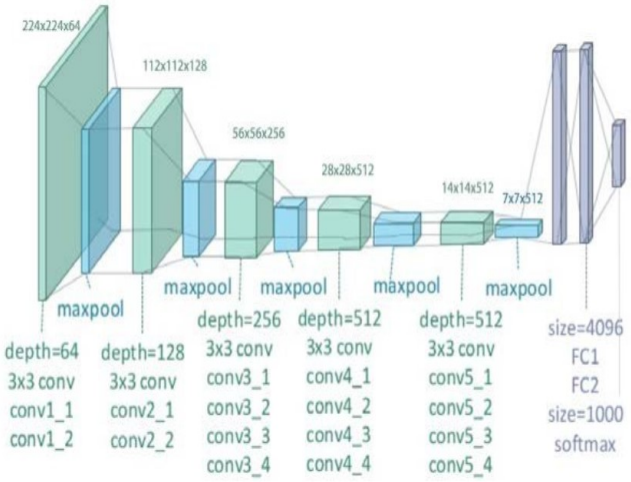

Fig. 3: VGG19 Model Architecture

- Hidden layers: The ReLU activation function is used in all hidden layers of the VGG network. VGG seldom utilizes Local Response Normalization (LRN) due to its longer training duration and higher memory demands. Furthermore, LRN does not affect the model's overall accuracy.
- Fully connected layers: There were 4,096 channels total in the first two FC layers and 1,000 channels in the third tier.

III. Result Analysis and Discussion

The main results of the system development are described in this study together with the conclusions drawn from the deep learning models.

A. Experimental Setup

A test was conducted on various Python libraries and hardware tools that are being used to create sophisticated autism diagnosis systems (ASD). Hardware point of view use at least processor core I5 and 8 GB RAM. For Software and libraries use Keras library, TensorFlow library, Matplotlib, and Numpy.

B. Evaluation Metrics.

For the two pre-trained models, this study employs a variety of performance assessment criteria, including a confusion matrix, accuracy, sensitivity, precision, and specificity as shown in Eq.1, Eq.2, and Eq.3. The confusion matrix serves as a table that shows the true and false outcomes of test findings, demonstrating the characteristics related to detection performance. In the confusion matrix for the VGG16 model, out of the 150 autistic children, 134 were True Positives and 25 were False Negatives. Furthermore, True Negatives represented 125 out of 150 typically developing children, and there were 16 False Positives. On the other hand, the confusion matrix for the VGG19 model showed that there were 16 False Negatives and

130 True Positives out of 150 autistic children. True Negatives amounted to 134 out of 150 normally developing children, and False Positives reached 20. The following are the formulae for these measures:
Accuracy

$$\frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

Specificity

$$\frac{TN}{TN+FP} \times 100\% \quad (2)$$

Sensitivity

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

In this context, TN denotes True Negative, FP signifies False Positive, and TP represents True Positive. Specificity refers to the model's capacity to identify typically developing children, while sensitivity pertains to its ability to recognize autistic children. Table III-B presents the number of layers utilized, along with the batch size and the number of epochs. Additionally, it details the optimizer and the activation function employed in the model.

TABLE II: The parameters utilized in our system's pre-trained deep learning models.

Name of parameter:	Value
Max pooling layer size globally:	3 * 3
The batch size:	12
The Number of epochs:	100
Layer of density:	128, 64
Classification layer of output:	Softmax
The Optimizer:	ADAM
The Activation function:	Rule

TABLE III: Evaluating the results of our system's pre-trained deep learning models.

Model name	Precision	Specificity	Sensitivity	Accuracy
VGG16	89.33%	88.65%	84.27%	86.33%
VGG19	87.01%	86.66%	89.09%	88%

In this Table III we are showing the comparison of Precision Specificity Sensitivity and final Accuracy. In this System, we train different types of autistic children and non-autistic children on their facial features. VGG19 achieved the highest Accuracy 88.0%.

C. Performance Measurement and Comparison Analysis

This section provides the test results from the studies carried out to detect Autism Spectrum Disorder (ASD). The findings for the employed deep learning models are summarized in Table III. To identify ASD, this study utilized two different pre-trained deep learning models, VGG16 and VGG19. In order to identify the facial characteristics that set autistic children apart from children with typical development, both models underwent testing and training. Figures 4 and 5 show the confusion metrics for the two deep learning models.

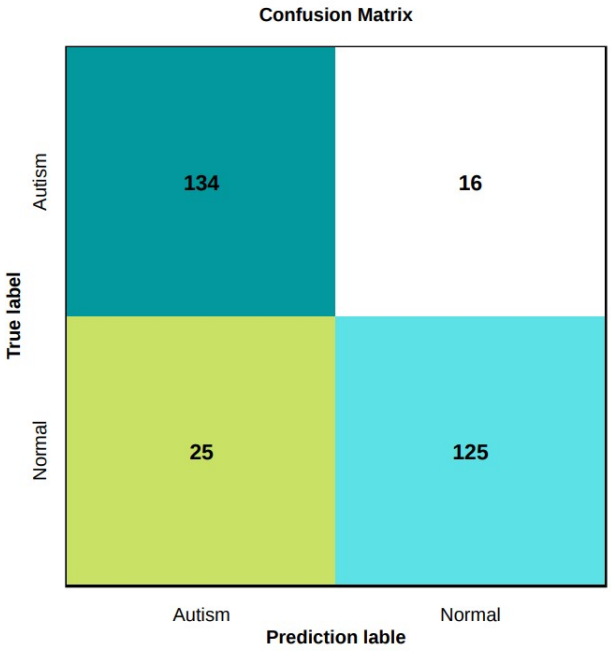

Fig. 4: Confusion Matrix of VGG16.

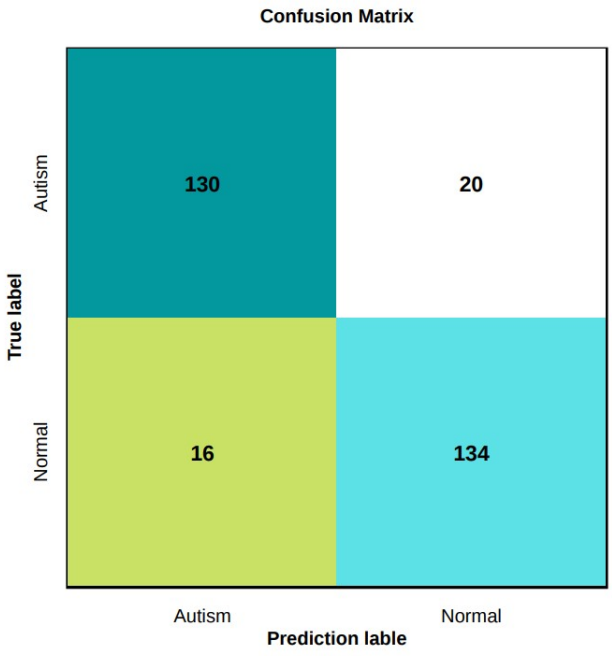

Fig. 5: Confusion Matrix of VGG19

According to the results, the VGG19 model outperformed the other model in terms of testing accuracy (88%), whereas the VGG16 model performed worse (86.3%). Even though the data generator assembled the information from internet sources, which revealed differences in age and image quality, the VGG19 model performed better with a lower number of mistakes. Figures 7 and 9 demonstrate the efficacy of the VGG19 model in training and confirming data for ASD detection. The y-axis displays the accuracy percentage, while the x-axis displays the number of epochs. During the training phase, the VGG19 model's accuracy rose from 56% to 98% over 100 iterations; however, during the validation phase, it fell to 8%. The model's training and validation losses are shown in Figure 9.

On the other hand, Figures 6 and 8 show how well the VGG16 model detects ASD. The model's performance was subpar, as the visual representation shows. Figure 6 shows the accuracy percentages for both training and validation, with the score percentage plotted on the y-axis and the number of epochs on the x-axis. While it performed reasonably well, it was less effective than the other deep learning models. Figure 8 illustrates the training and validation losses for the VGG16 model. 6 displays how it performs.

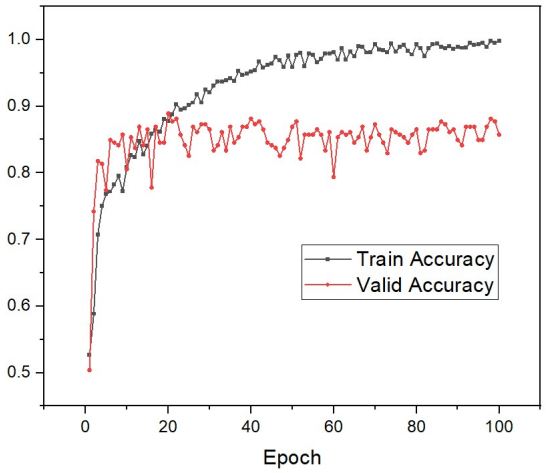

Fig. 6: VGG16 Train and validation accuracy

The status of the comparison between VGG16 and VGG19 is finally shown in figure 10. The VGG19 model was shown to be the best deep-learning model for diagnosing ASD. The VGG19 achieved a height accuracy that is 88%.

IV. Results and Discussion

Children with ASD frequently struggle with social interaction and day-to-day activities because they are unable to express their thoughts, feelings, likes, dislikes, and pains. In some cases, they become afraid that they are the only ones in a crowded room who don't understand them, which leads them to withdraw from society. This study examined the use of two cutting-edge deep learning models, VGG19 and VGG16, in

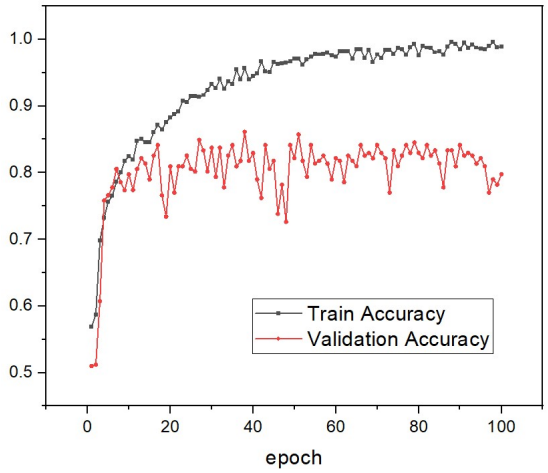

Fig. 7: VGG19 Train and validation accuracy

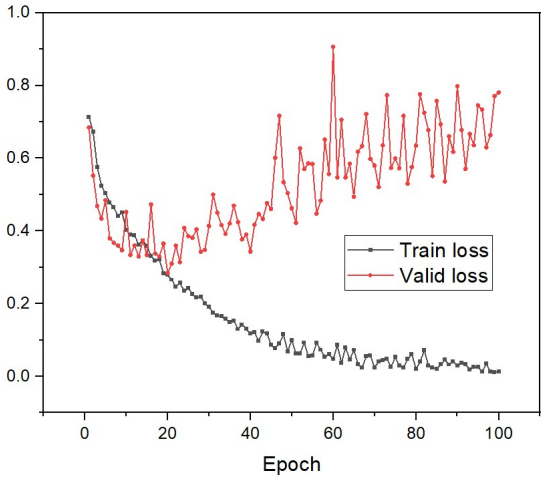

Fig. 8: VGG16 Train and validation loss

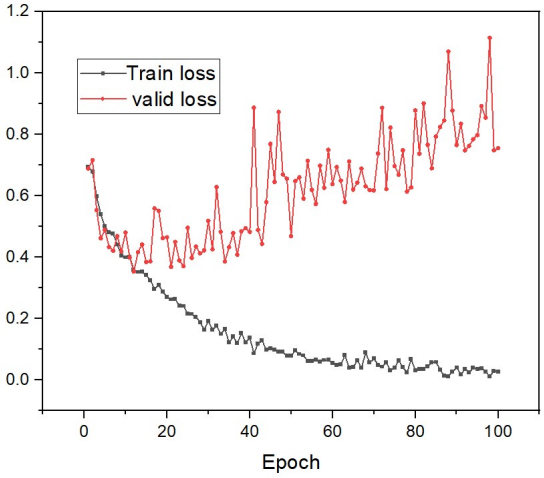

Fig. 9: VGG19 Train and validation loss

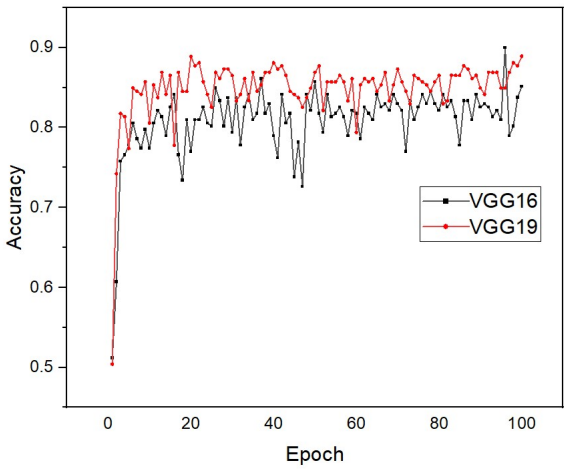

Fig. 10: Comparison between VGG16 and VGG19

the diagnosis of autism. Based on an analysis of the empirical data, the VGG19 model achieved a maximum accuracy of 88 percent. The comparison prediction research findings between the VGG19 model and the current system are shown in the table. Autism must be diagnosed as soon as possible, and creating an intelligent system using artificial intelligence (AI) can help. Images of the face are used to detect autism using the VGG19 and VGG16. Two deep learning algorithms were utilized in this work to identify autism, with the VGG19 model achieving 88% accuracy and the VGG16 model achieving 86.4%.

V. Conclusion

For the purpose of identifying and categorizing ASD patients, we suggested a CNN (Convolutional Neural Network) architecture in this study. The training period will be shortened since our proposed CNN architecture can achieve better detection performance with fewer parameters. As a result, when compared to earlier models of a similar kind, our proposed model is faster and simpler. All models were trained using a publicly available dataset from the internet, with VGG19 achieving the highest accuracy. Computer vision serves as an automated tool for specialists and families, enabling quicker and more accurate diagnoses of autism. These computational methods effectively implement intricate behavioral and psychological assessments necessary for diagnosing autism, which usually require significant time and resources. In general, our developed system surpassed all existing systems.

Acknowledgments

This work was partly supported by the Center for Research, Innovation, and Transformation (CRIT) of the Green University of Bangladesh (GUB).

References

[1] F. Tamilarasi and J. Shanmugam. Convolutional neural network based autism classification. pages 1208–1212, 06 2020.

[2] Md Tahmid Zoayed, Sayma Arshe, and Plabon Banik. Autism detecting model using image. 06 2022.

[3] Luciano Nunes, PlÃ¡cido Pinheiro, Mirian Pinheiro, Monica Pompeu, Marum SimÃ£o Filho, Rafael Comin-Nunes, and Pedro Pinheiro. *A Hybrid Model to Guide the Consultation of Children with Autism Spectrum Disorder*, pages 419–431. 10 2019.

[4] Yi L. Liu W, Li M. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. pages 888–98, 4 2016.

[5] Peebles D. Thabtah F. A new machine learning model based on induction of rules for autism detection. pages 264–286., 3 2016.

[6] Seyed Reza Shahamiri and Fadi Thabtah. Autism ai: a new autism screening system based on artificial intelligence. *Cognitive Computation*, 12, 07 2020.

[7] Jordan Hashemi, Geraldine Dawson, Kimberly L. H. Carpenter, Kathleen Campbell, Qiang Qiu, Steven Espinosa, Samuel Marsan, Jeffrey P. Baker, Helen L. Egger, and Guillermo Sapiro. Computer vision analysis for quantification of autism risk behaviors. *IEEE Transactions on Affective Computing*, 12(1):215–226, 2021.

[8] James M. Rehg, Gregory D. Abowd, Agata Rozga, Mario Romero, Mark A. Clements, Stan Sclaroff, Irfan Essa, Opal Y. Ousley, Yin Li, Chanho Kim, Hrishikesh Rao, Jonathan C. Kim, Liliana Lo Presti, Jianming Zhang, Denis Lantsman, Jonathan Bidwell, and Zhefan Ye. Decoding children's social behavior. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3414–3421, 2013.

[9] Marco Leo, Marco Del Coco, Pierluigi Carcagni, Cosimo Distante, Massimo Bernava, Giovanni Pioggia, and Giuseppe Palestra. Automatic emotion recognition in robot-children interaction for asd treatment. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 537–545, 2015.

[10] Mohammed Saeed Alzahrani Fawaz Waselallah Alsaade. "classification and detection of autism spectrum disorder based on deep learning algorithms", computational intelligence and neuroscience. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2022.

[11] Fernando De la Torre, Wen-Sheng Chu, Xuehan Xiong, Francisco Vicente, Xiaoyu Ding, and Jeffrey Cohn. Intraface. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 1, pages 1–8, 2015.

[12] Yin Li, Alireza Fathi, and James M. Rehg. Learning to predict gaze in egocentric video. In *2013 IEEE International Conference on Computer Vision*, pages 3216–3223, 2013.

[13] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[14] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pages 94–101, 2010.

[15] Tanaya Guha, Zhaojun Yang, Ruth B. Grossman, and Shrikanth S. Narayanan. A computational study of expressive facial dynamics in children with autism. *IEEE Transactions on Affective Computing*, 9(1):14–20, 2018.

[16] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4645–4653, 2017.

[17] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for everyone. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2176–2184, 2016.

[18] Ming Jiang and Qi Zhao. Learning visual attention to identify people with autism spectrum disorder. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3287–3296, 2017.