Computer Vision Based Autism Classification System using Deep Learning

1st Badhon Parvej
Dept. of CSE
Green University of Bangladesh
Dhaka, Bangladesh
badhonkhan481@gmail.com

2nd Afrin Sufian
Dept. of CSE
Green University of Bangladesh
Dhaka, Bangladesh
afsufian17@gmail.com

3rd Sikder Md Mahbub Alam
Dept. of CSE
Green University of Bangladesh
Dhaka, Bangladesh
iimahbubsikder@gmail.com

4th Dr. Muhammad Aminur Rahman
Associate Professor, Department of CSE
Green University of Bangladesh
Dhaka, Bangladesh
aminur@cse.green.edu.bd

***Abstract*—A complicated neuro-developmental illness called autism spectrum disorder (ASD) impacts communication, language, and social skills. In social interactions, it influences both spoken and nonverbal communication. Early ASD detection and diagnosis are crucial for early educational planning and treatment, family support, and timely delivery of appropriate medical care for the child. It may be possible to develop the best therapeutic strategy at the correct moment with the aid of early ASD screening. By examining facial features, it is possible to recognize ASD by studying the significant markers encoded in human faces. In order to categorize them, we have collected facial images of children with ASD from the public domain and employed deep learning classifiers. Additionally, using only autistic picture data, this classifier is used to differentiate between different ASD groups. Thus, it is becoming increasingly important to develop automated diagnosis techniques for ASD. This was accomplished using a dataset of autistic children that contains 2940 facial images of children with autism and typical children. We examined the outcomes of the existing methods and updated the VGG 16 for training and fitting the model. Subsequently, by making major adjustments to the epoch number batch size and other variables, we increased the accuracy from 77% to 86.3% and also calculate precision, specificity, and sensitivity.Additionally, we have enhanced training and testing in comparison to base code. In addition, we have calculated a number of factors to help us evaluate the outcomes more effectively, including true positive, true negative, false positive, and false negative results, as well as accuracy, precision, sensitivity, and specificity of the prediction and result in the submissions.csv file. We have also implemented the vgg19 model of cnn and compared the results of both the models .We obtained that deep learning achieved the highest accuracy result of 88% of VGG 19, followed by VGG16 86.3%.**

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

I. Introduction

ASD is a neurodevelopmental disease that impacts social relationships by affecting both verbal and nonverbal communication. Early ASD detection and diagnosis are crucial for early educational planning and treatment, family support, and timely delivery of appropriate medical care for the child. Social communication is the main emphasis of the revised ASD diagnosis criteria [1] [2]. mostly because they have communication issues. He or she might not talk as early as other kids. He or she might not want to look people in the eye. Their conduct and emotions are reflected in their body language. They have the same facial pattern 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 often affects people of all racial, ethnic, and socioeconomic backgrounds. Boys experience the disease far more frequently than girls do. [4] Boys are four to five times as likely than females to have ASD. Over the past 20 years, there has been a steady rise in the frequency of ASD; current estimates put the number of affected children at 1 in 36.A increased incidence of ASD has been associated with hereditary variables, parental history of mental problems, preterm births, and prenatal exposure to psychotropic medicines 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. In order to better accurately identify children with ASD in the early stages, an enhanced transfer-learning-based autism face recognition framework is presented in this study.

II. Materials and Methods

The facial characteristics of autistic and typically developing children are used in this work to identify autism using a deep learning model based on the VGG16 and VGG19. If a youngster has autism or is healthy, it may be told by their facial traits [7]. Important face traits were retrieved from the photos using the VGG16 and VGG19 models. One benefit of deep learning algorithms is their capacity to extract tiny features from images that a human eye cannot see [8] [9].

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 show that the number of photos 2940 face made up the dataset, half of which were of children with autism and the other half of which were of typically developing kids. This information was compiled using online resources, including Facebook pages and websites dedicated to autism.

TABLE I
Splitting of the dataset for training, testing, and validation.

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

B. Preprocessing

The photos were cleaned up and cropped as part of the data preprocessing. Piosenka's [11] [12] collection of data from online sources necessitated preprocessing before the deep learning model could be trained on them. The face in the original image was automatically cropped by the -e dataset builder. The dataset was divided into 2,540 training photos, 100 validation images, and 300 testing images. The normalization approach was used to scale; the dataset was rescaling all of the picture parameters from [0, 255] to [0, 1].

C. Convolutional Neural Network Models

A branch of artificial intelligence known as "computer vision" has been extensively developed to help humans in their daily lives, such as through medical applications [13]. As a result, the CNN algorithm has contributed to study on behavior and psychology as well as the diagnosis of diseases. This figure 1 shows the architecture of the proposed system for our model. Our proposed system working on autism classification Autism spectrum disorder (ASD) vs Typically developing (TD). We are finding out the autistic child using the best accuracy result. So we used CNN model for getting better performance and used the Confusion matrix.

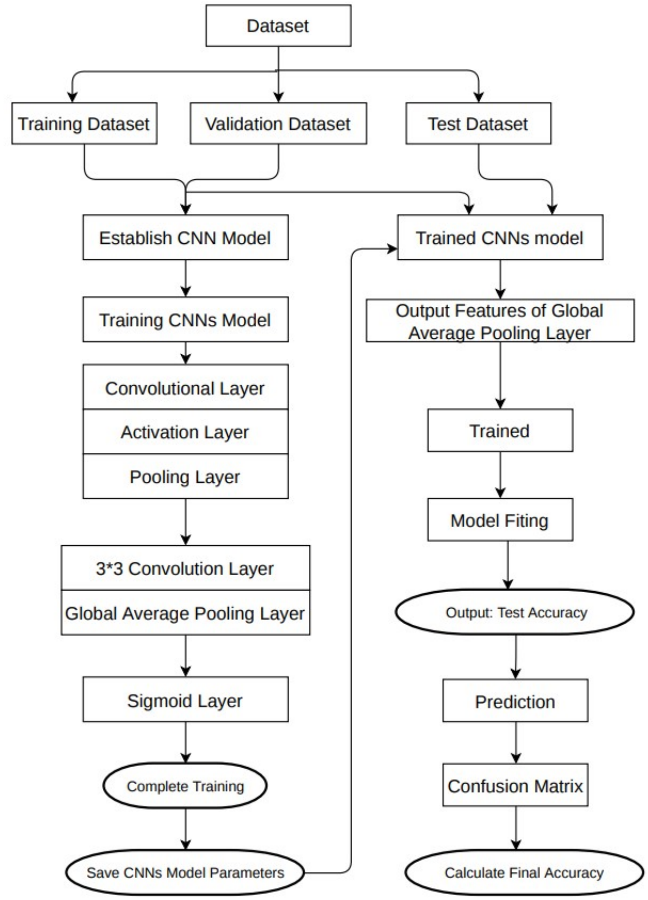

Fig. 1. System architecture for our proposed model

1) Basic Components of the CNN Model: Undoubtedly one of the most well-known deep learning techniques is the convolutional neural network (CNN) [14] It accepts the input image and assigns weights and biases that can be learned importance in order to determine the image's class [15]. By connecting and communicating with other cells, a neuron can be thought of as a mimic of the contact pattern of the neurons in the human brain. We shall discuss the input layer, convolutional layer, activation function, pooling layer, fully connected layer, and output prediction in this section of the article.

2) Convolutional Layer with a Pooling Layer: The convolutional layer receives a picture as a matrix of pixel values as input. The convolutional layer's goal is to simplify the images without sacrificing any of the key characteristics that will aid in the detection of autism [16] [17]. Low level features like edges and color are extracted by the CNN model's first layer. We were able to add more layers to the CNN model because of how it was built, allowing it to extract high-level features that will aid with visual comprehension. The number of weights was decreased by utilizing either the max pooling or average pooling approaches because the convolution layer outputs a large number of parameters that could significantly slow down the matrices' arithmetic operations [18] [19]. The maximum values in each window of the stride are the basis for max pooling, whereas the mean value of each window of the stride is the basis for average pooling. The model used in this investigation was based on maximum pooling.

3) Fully Connected Layer and Activation Function.: The fully connected (FC) layer is a nonlinear collection of high-level features that took input from the hidden layers and were shown as outputs. In the FC layer, the input image is shown as a column vector. The model is composed of the back-propagation pathways and the forward neural network. The forward neural network's output layer is flattened. The neural network in backpropagation undervalues the loss mistakes by using the number of training iterations, and it discovers more features. Most deep learning models perform very well as the number of hidden layers and training iterations are increased, enabling the neural network to effectively extract low-level input . After obtaining the parameters from the FC layer, the softmax classifier calculates the characteristics to anticipate 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 is also comprehended as a ConvNet, which is a type of artificial neural network. A convolutional neural network has an input layer, an output layer, and various hidden layers. VGG16 is a type of CNN (Convolutional Neural Network) show the figure 2 that is considered to be one of the best computer vision models to date. The creators of this model evaluated the networks and increased the depth using an architecture with very small (3 × 3) convolution filters, which showed a significant improvement on the prior-art configurations.

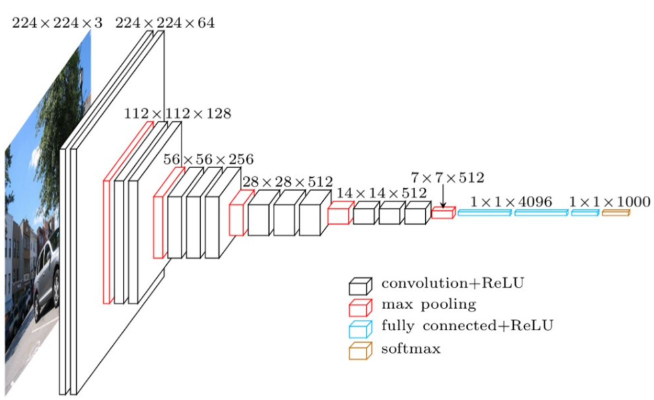

Fig. 2. VGG16 model architecture

2) VGG19: VGG19 stands for the visual geometry group network (VGG19), in this figure 3 is a in-depth artificial neural network model with a 19 artificial multilayers process. VGG19 is based on the CNN technique, is normally implemented on the ImageNet dataset, and is valuable because of its easy implemenatation, as 3 × 3 convolutional layers are connected to its upper side to increase with the deep level.

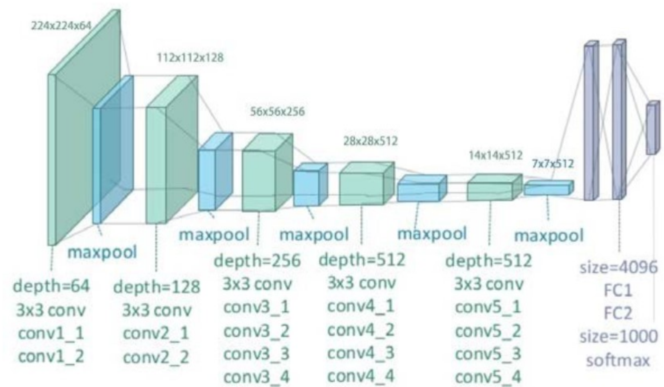

Fig. 3. VGG19 Model Architecture

- Input layer: This layer's purpose is to receive an image input with the dimensions 224*224*3. To keep the input size of the image consistent for the ImageNet race, the model's creators chose the central 224*224*3 patch in each image.
- Convolutional layers: After the convolution stride is held at 1 pixels, the convolutional layers of the VGG use a minimal receptive field of 3*3, which is the smallest practical size that still releases left and right as well as up and down to preserve the spatial resolution. The number of pixel changes across the input matrix is known as the stride.
- Hidden layers: All the hidden layers of the VGG network use ReLU. VGG rarely affects Local Response Normalization due to its longer training time and memory requirements (LRN). It also has no effect on 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. Experiments

This article presents the deep learning models' findings. part, and the key outcomes of system development are declared.

A. Experimental Setup

A test was conducted on various Python libraries and hardware tools are being used to create sophisticated autism diagnosis systems (ASD). In this table II shows the requirements of our system.

TABLE II
EXPERIMENTAL SETUP FOR OUR SYSTEM

Hardware	Software/libraries
Processor core I5	Keras library
	TensorFlow library
8 GB RAM	Matplotlib
	Numpy

B. Evaluation Metrics.

For the two pretrained models, this study employs a variety of performance assessment criteria, including a confusion matrix, accuracy, sensitivity, precision, and specificity. An example of a measure of classification performance is a confusion matrix, which is a table containing the true and false values of the test results. In the confusion matrix of the VGG16 model, the True Positives represented 134 of the 150 autistic children, the False Negatives represented 25, the True Negatives represented 125 of the 150 normally developing children, and the False Positives represented 16.In the confusion matrix of the VGG19 model, the True Positives represented 130 of the 150 autistic children, the False Negatives represented 16, the True Negatives represented 134 of the 150 normally developing children, and the False Positives represented 20. The following are the formulae for these measures:
Accuracy= (TP + TN)/(FP+FN+TP+FN)*100%
Specificity= TN/(TN+FN)*100%
Sensitivity= TP/(TP+FP)*100%
Where TN stands for True Negative, FP for False Positive, TP for True Positive, and FP for False Positive. Specificity is the model's ability to recognize normal children, whereas sensitivity is the model's ability to recognize autistic children. In this table III-B shows us how many layers we used and the batch size and number of epoch , we also used one optimizer and a activation function.

TABLE III
PARAMETERS USED IN THE PRETRAINED DEEP LEARNING MODELS FOR OUR SYSTEM.

Parameter name	Value
Global max pooling layer size	3 * 3
Batch size	12
Number of epochs	100
Dense layer	128, 64
Output classification layer	Softmax
Optimizer	ADAM
Activation function	Rule

TABLE IV
TESTING RESULTS OF THE PRETRAINED DEEP LEARNING MODELS FOR OUR SYSTEM.

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 IV we are showing the comparison Precision Specificity Sensitivity and final Accuracy. In this System we are train different type autistic child and non-autistic child on their facial features. VGG19 achieved highest Accuracy 88.0%.

IV. RESULTS

The test results from the investigations done to look for ASD are shown in this section. The testing outcomes for the used deep learning models are presented in IV. Two separate pretrained deep learning models, namely VGG16 and VGG19, were used in these tests to identify ASD. Each model underwent training and testing to identify the characteristics that, based on face features, classify youngsters as autism or normal. The confusion metrics for the two deep learning models are shown in Figures 4 and 5.

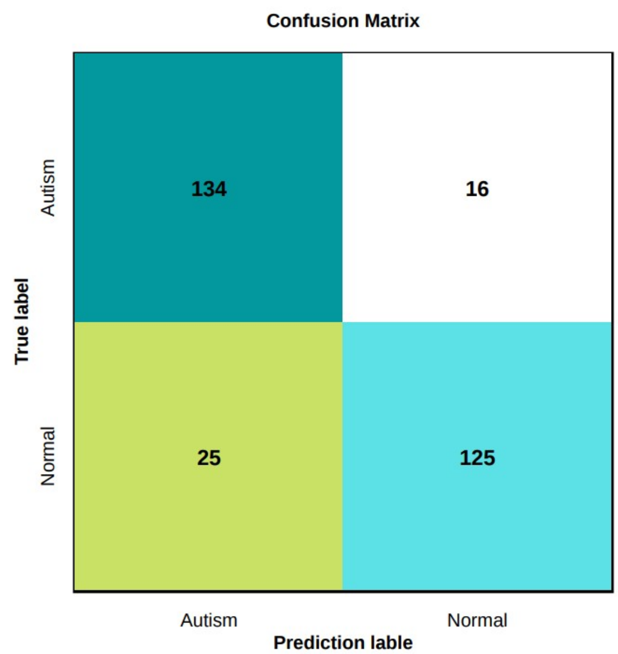

Fig. 4. Confusion Matrix of VGG16.

They demonstrate that among the two models, the VGG19 model had the greatest testing accuracy (88%) while the VGG16 model had the lowest performance level (86.3%). The VGG19 model demonstrated the superior accuracy, with just a tiny proportion of mistakes, despite the dataset being compiled from Internet sources by the data generator, which clearly indicated the difference in ages and the caliber of the photos. Figures 7 and 9, where the y-axis represents the score percentage and the x-axis the number of epochs, show the performance of the VGG19 model for the training and validation of the data for ASD detection. The VGG19 model's accuracy increased from 56 to 98% during the training phase after 100 iterations, then it decreased to 82% during the validation phase. Figure 9 shows its training and validation losses.Figures 6 and 8 depict how well the VGG16 model performed in identifying ASD. The visual plot demonstrates that the model's performance was subpar. Figure 6, with the

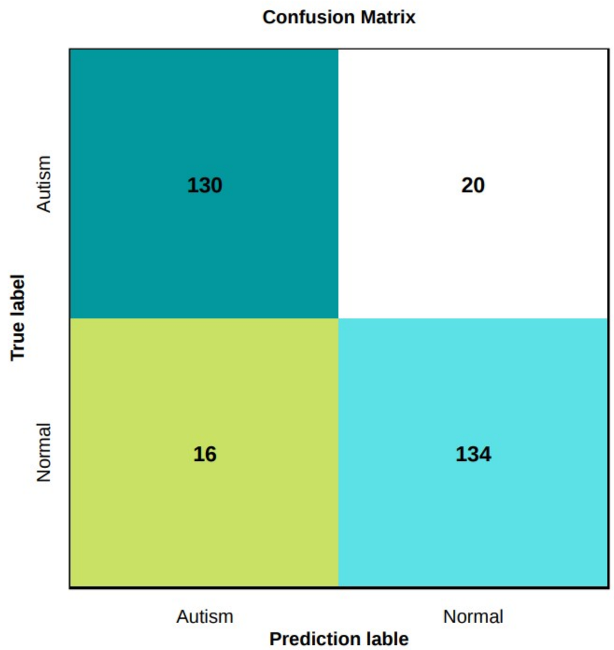

Fig. 5. Confusion Matrix of VGG19

score% on the y-axis and the number of epochs on the x-axis, displays the accuracy percentages for both training and validation. It performed well, however in comparison to other deep learning models, it performed worse. Figure 8 displays its entropy training and validation losses. Figure 6 displays how it performs.

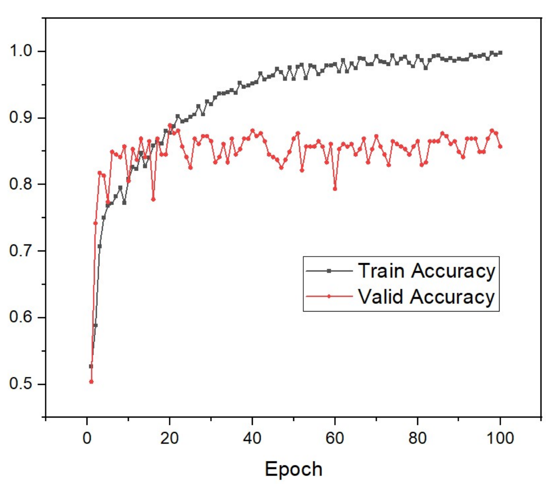

Fig. 6. VGG16 Train and validation accuracy

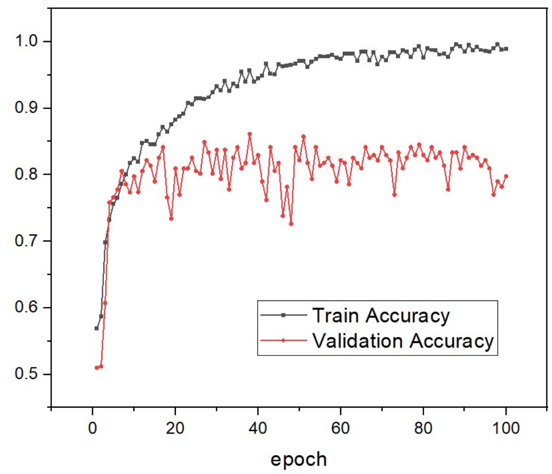

Fig. 7. VGG19 Train and validation accuracy

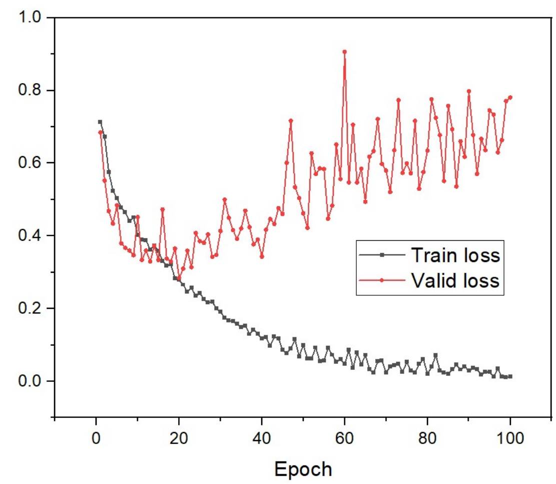

Fig. 8. VGG16 Train and validation loss

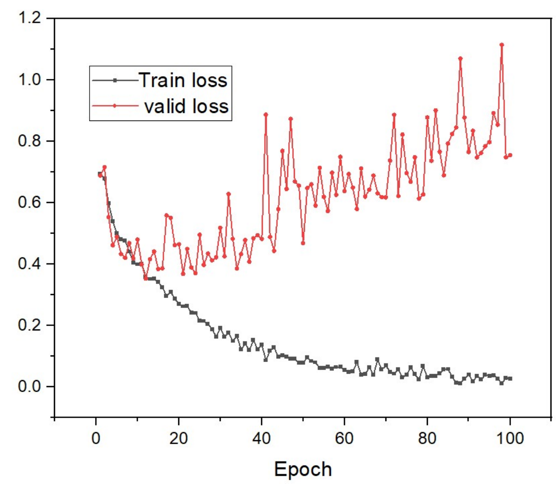

Fig. 9. VGG19 Train and validation loss

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 heights accuracy that is 88%.

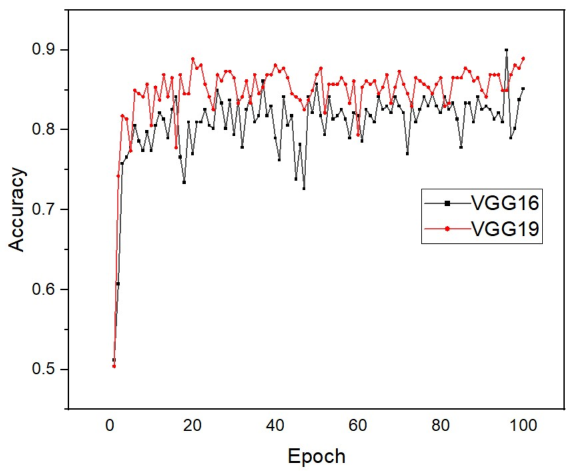

Fig. 10. Comparison between VGG16 and VGG19

V. 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 one in a crowded room who doesn't understand them, which leads them to withdraw from society. Two cutting-edge deep learning models, VGG19 and VGG16, were taken into consideration in this study for application in autism diagnosis. When these models' empirical data were given, it was found that the VGG19 had the highest accuracy, at 88 percent. The table shows the findings of the comparative prediction study of the VGG19 and the current system.Early diagnosing autism is important. Developing an intelligence system based on AI can help identify autism early.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%.

VI. 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 suggested CNN architecture can achieve better classification performance with less parameters. As a result, when compared to previous models of a similar nature, our suggested model is quicker and less complicated.Each model was trained on a publicly available data set on the internet, and VGG19 gives the best result when accuracy was achieved. Computer vision as automatic tools for specialists and families to accurately and more quickly diagnose autism computer techniques contribute to the successful conduct of complex behavioral and psychological analyses for autism diagnosis, which require a longer time and great effort. In general, our developing system surpassed all the existing systems.

References

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

[2] M. T. Zoayed, S. Arshe, and P. Banik, "Autism detecting model using image," 06 2022.

[3] L. Nunes, P. Pinheiro, M. Pinheiro, M. Pompeu, M. Simão Filho, R. Comin-Nunes, and P. Pinheiro, *A Hybrid Model to Guide the Consultation of Children with Autism Spectrum Disorder*, 10 2019, pp. 419–431.

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

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

[6] S. R. Shahamiri and F. Thabtah, "Autism ai: a new autism screening system based on artificial intelligence," *Cognitive Computation*, vol. 12, 07 2020.

[7] J. Hashemi, G. Dawson, K. L. H. Carpenter, K. Campbell, Q. Qiu, S. Espinosa, S. Marsan, J. P. Baker, H. L. Egger, and G. Sapiro, "Computer vision analysis for quantification of autism risk behaviors," *IEEE Transactions on Affective Computing*, vol. 12, no. 1, pp. 215–226, 2021.

[8] J. M. Rehg, G. D. Abowd, A. Rozga, M. Romero, M. A. Clements, S. Sclaroff, I. Essa, O. Y. Ousley, Y. Li, C. Kim, H. Rao, J. C. Kim, L. L. Presti, J. Zhang, D. Lantsman, J. Bidwell, and Z. Ye, "Decoding children's social behavior," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3414–3421.

[9] M. Leo, M. Del Coco, P. Carcagni, C. Distante, M. Bernava, G. Pioggia, and G. Palestra, "Automatic emotion recognition in robot-children interaction for asd treatment," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, 2015, pp. 537–545.

[10] M. S. A. 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] F. De la Torre, W.-S. Chu, X. Xiong, F. Vicente, X. Ding, and J. Cohn, "Intraface," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1, 2015, pp. 1–8.

[12] Y. Li, A. Fathi, and J. M. Rehg, "Learning to predict gaze in egocentric video," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 3216–3223.

[13] J. Hashemi, Q. Qiu, and G. Sapiro, "Cross-modality pose-invariant facial expression," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 4007–4011.

[14] 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*, vol. 24, no. 7, pp. 971–987, 2002.

[15] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. 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*, 2010, pp. 94–101.

[16] T. Guha, Z. Yang, R. B. Grossman, and S. S. Narayanan, "A computational study of expressive facial dynamics in children with autism," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 14–20, 2018.

[17] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4645–4653.

[18] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye tracking for everyone," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2176–2184.

[19] M. Jiang and Q. Zhao, "Learning visual attention to identify people with autism spectrum disorder," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3287–3296.