Autism Spectrum Disorder Detection Using Facial Images and Deep Convolutional Neural Networks

Lalitha Kumari Gaddala[1*], Koteswara Rao Kodepogu[1], Yalamanchili Surekha[1], Mangalarapu Tejaswi[1], Kethineni Ameesha[1], Lakshman Saketh Kollapalli[1], Siva Kalyan Kotha[1], Vijaya Bharathi Manjeti[2]

[1] CSE Department, PVP Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh 520007, India
[2] Department of CSE, GITAM School of Technology, Visakhapatnam 530045, India

Corresponding Author Email: glalitha@pvpsiddhartha.ac.in

ABSTRACT

Autism Spectrum Disorder (ASD) is a prevalent neurodevelopmental disorder, affecting approximately 1% of the global population. It is characterised by deficits in social communication, interaction, and a propensity for repetitive behaviours. Despite its prevalence, the diagnosis of ASD remains challenging due to the lack of conspicuous disparities between neuroimages of affected individuals and their neurotypical counterparts. This study aims to enhance the accuracy and efficiency of ASD diagnosis by integrating deep learning techniques with conventional diagnostic procedures. In this work, we present a novel approach to detect and classify ASD using facial images processed through deep Convolutional Neural Networks (CNNs). We utilised the Visual Geometry Group models (VGG16 and VGG19) to construct our deep learning models. The models were trained and validated using an extensive dataset of facial images. The proposed models have demonstrated promising results, achieving an accuracy rate of 84% in the classification of ASD individuals. This study's findings suggest the potential of deep learning applications in refining the diagnostic process of Autism Spectrum Disorder. Further research is recommended to optimise these models and validate their effectiveness on a broader scale.

1. INTRODUCTION

Patients with autism spectrum disorder (ASD) have an aberrant gaze while viewing images of the outside environment as well as an extraordinary sensitivity to social cues. However, it is unclear how people see the world in first person. One of the most fundamental cognitive processes in humans is the capacity to pay attention to what is important in the surroundings. To social signals like human faces and social contexts, however, people with ASD show substantial attentional difficulties [1]. According to earlier studies, people without ASD spend substantially longer looking at the eyes on human faces than people with ASD. When comparing social versus non-social stimuli, people with autism spectrum disorder pay less attention to human faces and other social indications like human voices and body movements and place greater focus on non-social objects.

Genetics have a significant role in autism. The concordance percentage varies between 60 and 90% for identical twins and between 5 and 10% for fraternal twins. Autism has been associated with a large number of genes and gene variants. Genes are responsible for the creation of synaptic circuits, which aid in communication between different brain areas. Since environmental factors contribute disproportionately to new gene mutations, many of them should be connected to an elevated risk of ASD. In this study, ASD images are predicted using kid pictures as the dataset. The datasets were gathered via the Kaggle service. The study's objective is to categorise facial photographs as autistic or not.

ASD is seen as a neurological disorder that affects a person's ability to engage in activities, communicate, and think. Compared to children without the disease, children with autism are less receptive to behaviour. Children with autism have a variety of challenges, including learning deficits, poor attention, and psychological problems including perceptual, locomotory, despair, anxiety, etc. A major time, financial, and value commitment is necessary for ASD diagnosis. Early intervention aids in recommending the ideal course of treatment and medication for the patient and would help stop the problem from growing worse. The main advantage is a decrease in the expense and burden of delayed diagnosis. Cases of autism spectrum disorder are increasing and multiplying rapidly all over the world. According to the World Health Organisation (WHO), autism affects one in 160 kids. Even fewer autistic persons require additional care and support than can live alone. The identification of autistic symptoms in people and the determination of whether an individual needs a complete assessment thus necessitate a model prediction that is swift, accurate, and time-efficient. The Centres for Disease Control and Prevention (CDC) describe the autism spectrum disorder (ASD) as a developmental condition that can lead to significant social, communication, and behaviour issues. It is estimated that 1 in 59 US children aged 8 or younger have ASD, and this number is rising [2]. However, there are significant and enduring racial and cultural differences in the prevalence of ASD and accessibility to therapies and treatments. Children from racial and ethnic minority groups reported lower rates of ASD diagnosis and higher rates of misdiagnosed or delayed diagnosis as compared to White

children. In spite of the combined rough ASD commonness of 16.8 per 1000 (1 out of 59) youngsters in 2018, are suffering from ASD. The sample images of the ASD can be seen in Figure 1.

(a)

(b)

Figure 1. Sample image from the autism dataset

ASD (17.2 per 1000) was essentially higher than that of non-African American kids (16.0 per 1000), Latino kids (14.0 per 1000), and Asian/Pacific Islander youngsters (14.0 per 1000), (13.5 per 1000) [2].

The following are the main causes of the gap in ASD prevalence and postponed diagnosis in the United States:

1. The diagnostic process is subjective because ASD is presently diagnosed by behavioural observation. As a result, only clinicians with extensive expertise can reliably identify ASD in children as young as 2 years old, with an average age for diagnosis of 4-5 years [3].
2. The diagnostic process is subjective due to the fact that ASD is currently diagnosed through behavioural observation. As a result, only clinicians with extensive expertise can

reliably identify ASD in early children as early as two years old, with a 4-5 year average age for diagnosis [3].
3. Many households cannot afford professionals and experts, and availability is much worse in undeveloped areas.
4. Lack of screening and knowledge is another issue, especially in rural areas.
5. Additionally, compared to White children, children from racial and ethnic minorities are less likely to get an ASD diagnosis overall and are more likely to have their diagnoses misplaced.

Neuromarkers for early diagnosis are always being sought after. Scientists have worked hard to find key indicators that can help an expert diagnose an ASD. Neurologists have long recognised the tight relationship between underlying neurological issues and atypical face traits caused by errors in the embryonic development process, known as facial dysmorphologies. The classification of photographs is one area where machine learning (ML) approaches have gained popularity recently [4]. Machine learning algorithms have the amazing ability to learn from hidden patterns found in vast volumes of data, which makes them excellent predictors. The two components of a machine learning-based image classifier are a feature extractor and an algorithm. The most popular feature extractor is the Convolutional Neural Network (CNN), and there are several machine learning techniques to choose from to determine which one best matches the data.

According to the research mentioned above, autistic children have different facial features from TDs who are of a similar age and gender. CNN's capacity to recognise images may potentially help with the early identification of ASD in children. We set out to create the best CNN-based model that can effectively diagnose autism in children with the highest degree of sensitivity and specificity using features derived from face pictures in light of the aforementioned facts. In this study, we classified face photos into two categories—Autistic and Non-Autistic—using Convolutional Neural Networks VGG16 and VGG19.

2. METHOD

2.1 Data acquisition

The only ASD face picture collection that is openly available. The 2936 face images in this sample are evenly split between children with TD and those with ASD [5]. As can be seen in the study [5], the 3014 photographs in the original dataset [6] plainly displayed problems. Since the contributor also said that he was unable to obtain any photographs of ASD from organisations or trustworthy sources, all of the images in the Kaggle dataset were found online [6]. We used the dataset [7], which had 2936 photographs after deleting photos that were obviously erroneous.

In this sample, 89% of the children are White, and 11% are Tan youngsters. The main purpose of this dataset was to show how race affects the creation of deep learning systems that use face photos. In the data set, half of the photos were categorised as being non-autistic and the other half as being autistic, the defference between Autistic and Non-Autistic sample dataset can be seen in Figure 2.

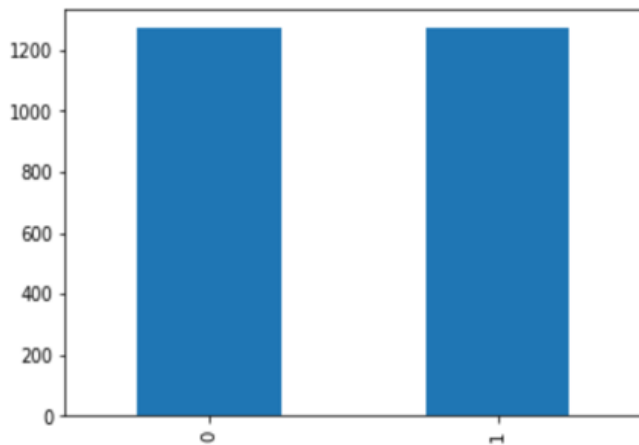The simple Autistic image from the sample dataset can be seen in Figure 3.

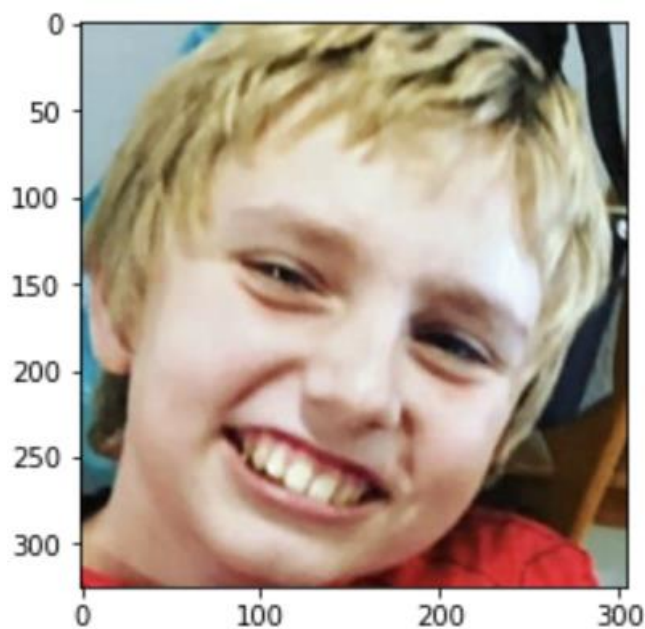**Figure 2.** Differentiation between Autistic and Non-Autistic sample dataset

Figure 3. Autistic image from the sample dataset

2.2 Methodology

After collecting the dataset from the mentioned sources, we will use our dataset to train, test and validate our model. The methodology that we use will be as follows:

1)   Load the training dataset: We will start by first training our dataset. The two files in our dataset have the names "Non_Autistic.135.jpg" and "Autistic.391.jpg," with the terms "Non_Autistic.135" and "Autistic.391" denoting the type (whether the person is autistic or not) and the second word denoting the image number, respectively.

2)   Load Test Dataset: Included in our dataset are the files "Non_Autistic.135.jpg and Autistic.391.jpg."

3)   Multilayer neural networks will be used as deep neural networks.

4)   Convolutional Neural Networks will be utilised because the dataset only contains pictures. A method to separate the data into distinct groups is the CNN classification. An picture will pass through a variety of convolutional layers and filters each time it is processed [8].

5)   Model Preparation: Next, we'll set up the model by taking the following actions:

6)   Next, we'll discuss hyperparameters.

7)   The output layer is then flattened to one dimension.

8)   A completely There will also be an additional connected layer with 512 hidden units and ReLU activation.

9)   The dropout rate will then be increased by 0.5.

10)   A final sigmoid layer will then be added for classification.

11)   Data Preparation: In order to create a matrix, we will prepare the data and extract the necessary information from the images.

12)   Getting ready the training dataset:

a) Training Generator: In this section, the train data will be prepared and generated.

b) Validation Generator: Using the verified data, we will develop a validation generator to filter the data and assure the performance quality.

c) Validation data: Information used to assess any model metrics and the loss at the conclusion of each epoch. These data won't be used to train the model.

13)   Model Fitting: At this point, we'll train the data using a predetermined batch size and number of epochs.

14)   Creating a testing generator will allow us to generate the test dataset from the test photos that have already been imported.

15)   Prediction: Using the model that was previously assessed, we will now predict the outcome for the test dataset.

16)   There will be the following results: The image's name is:- actual name (image name 0 or 1); for instance, autistic.127.jpg(1); 0/1 is a prediction in this case.

17)   Our forecasts will be kept in the submission_13010030.csv file. This file includes the anticipated outcome, and we will further determine its correctness, precision, etc.
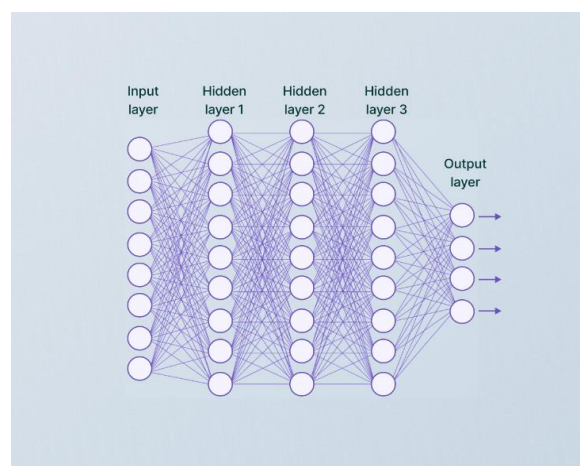
2.3 Convolution Neural Network

Figure 4. Representation of Convolution Neural Network

Over the past several years, deep (CNNs) have seen significant use in computer vision because they exhibit strong exclusionary abilities while maintaining high levels of performance [8]. Our study concentrated on a deep-learning-based method since practically all face recognition systems employ CNNs as its deep learning tool [9] and because deep learning networks have shown their feasibility as a model for facial identification. Deep (CNNs) have been widely used in

computer vision over the past few years due of their strong exclusionary capabilities and high levels of performance [9]. Our study focused on a deep-learning-based approach since Almost all face recognition algorithms employ CNNs as a deep learning tool [10] and deep learning networks have demonstrated their potential as a model for facial identification. Representation of Convolution Neural Network as seen in Figure 4.

2.4 Convolution Neural Network(VGG16)

Convolutional Neural Networks utilise deep learning techniques to prioritise several features in photos and can distinguish between one image and another. Pre-processing time for ConvNet is lower than for other classification methods. Block diagram of VGG16 can be seen as in Figure 5 [11].
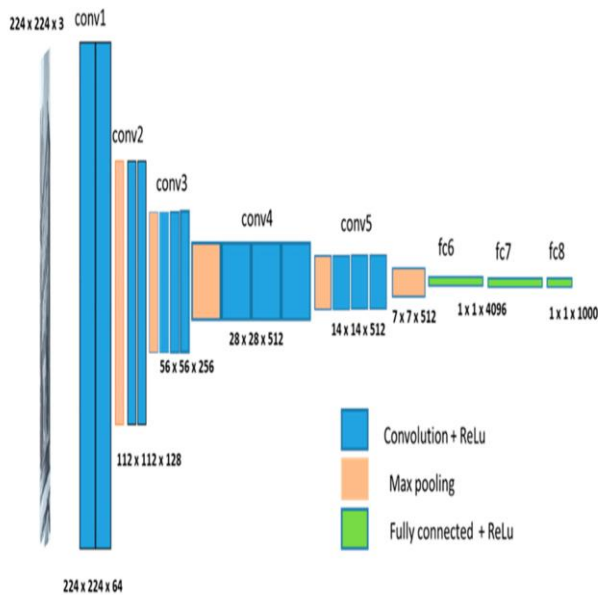

Figure 5. Block diagram of VGG16

Convolutional Neural Networks like Visual Geometry Group (VGG16) are said to be superior vision models when compared to more recent technologies like ResNet and Inception. We employ a deeper network and a 3 3 convolution kernel that is smaller. The field of view of a 5 5 convolution kernel is comparable to the stack of two 33 convolution kernels, and the field of view of a 7 7 convolution kernel to the stack of three 33 convolution kernels. This allows for the possibility of having fewer parameters (3 stacked 33 structures with only 7 structural parameters (333) / (77) = 55%); nonetheless, they contain more parameters. The feature learning capability of CNN is enhanced by the non-linear modification. The convolutional structure of the VGG Net is extended with a 1 * 1 convolution kernel. Without altering the input or output dimensions, non-linear transformation is utilised to increase the network's expressiveness and reduce processing. The last convolutional layer has been utilised among the fully connected layers. To avoid the model's input having a fixed size, the last convolution layer should be taken into account (224224). Input values are analysed by VGG, which then generates a value sequence. VGG16's top layer can recognise up to 1000 distinct image classifications. Our study's

main emphasis is the categorization of just two characteristics as autistic or not. As a result, we will reduce the output layer to a single dimension, Block diagram of VGG19 can be seen as in Figure 6 [12].
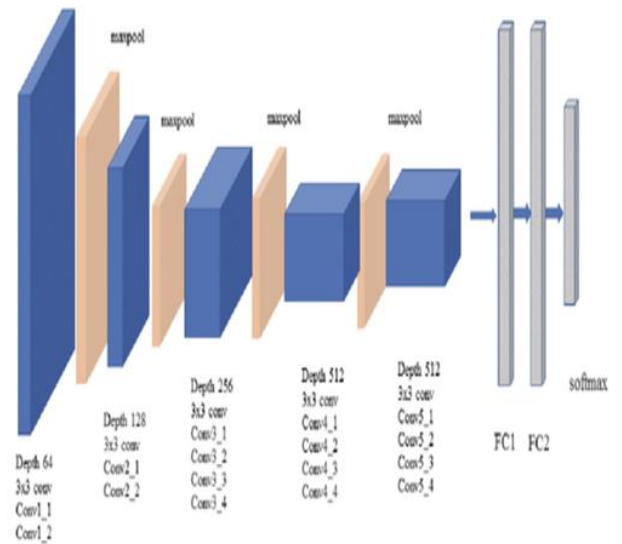

Figure 6. Block diagram of VGG19

2.5 Convulution Neural Network (VGG19)

The VGG-19 model is an improvement over the VGG-16. Convolution neural network model with 19 layers. The model is built by stacking convolutions together, however the diminishing gradient issue limits the model's depth. This issue makes deep convolution networks difficult to train. The model was trained on ImageNet to categorise 1,000 distinct types of objects, like the other models being considered. VGG analyses the input values before producing a value sequence. The top layer of VGG16 can identify up to 1000 various picture categories. Though just two variables are employed in our study to determine whether a person is autistic or not. The output layer will then be reduced to a single dimension as a result.

2.6 Training

A model is trained when it is shown how to categorise or predict something using existing preliminary data. To begin with, the divided original dataset is used to create the train dataset, which will be used to train the model. We will create a validation generator to screen out the data and utilise the verified data to ensure performance quality after preparing and creating the train data. The validation data are the data used to assess any model metrics and the loss at the conclusion of each epoch. These data won't be used to train the model. After using the validation dataset to judge the performance of our model, we will train the data for a particular number of epochs and batch size. We evaluated using a batch size of 12 with 10 epochs to train our model. We trained two models, in which one is based on VGG16 and the other is based on VGG19. Utilising the validation dataset, we will compute the evaluation metrics to assess the performance of our model at the end of each epoch [13].

2.7 Evaluation metrics

Evaluation metrics are used to assess how effectively a statistical or machine learning model is functioning. The machine learning models or algorithms employed in each project must be evaluated. A model may be assessed using a variety of different assessment metrics. These consist of measurements such as classification accuracy, logarithmic loss, confusion matrix, and others. When we talk about accuracy, we usually mean classification accuracy, which is the proportion of successful predictions to all input data. Logarithmic loss, commonly referred to as log loss, is used to penalise incorrect classifications. A confusion matrix creates a matrix for us and provides a summary of the model's overall performance. In this work, we have used evaluation metrics like Accuracy, Sensitivity, Specificity, and Precision to evaluate the performance of our model [14, 15].

Accuracy:
Accuracy is a common evaluation statistic for classification problems. It displays the percentage of all forecasts that result in accurate predictions.

$$Accuracy = \frac{True\ Positive\ +\ True\ Negative}{Total\ Predicted} \quad (1)$$

Sensitivity:
Sensitivity provides the true positive rate (TPR), which is the ratio of genuine positives to all positives.

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

Precision:
Precision (either rightly or wrongly) is the ratio of correctly classified positive samples (True Positive) to all positively classified samples.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3)$$

Specificity:
Specificity is the proportion of genuine negatives that the model correctly identifies. This means that more real negatives, often known as false positives since they were previously believed to be positive, would be recorded.

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (4)$$

The above-mentioned evaluation metrics has been used to evaluate the strength of the model compared to other models during its training on validation dataset and as well as on training dataset after its training.

3. RESULTS AND CONCLUSION

We used both VGG16 and VGG19 based models in this investigation to divide face image categories into autistic and non-autistic groups. The findings from the validation dataset for each epoch of the VGG16 training process are shown in Table 1.

Our final model based on VGG16 has achieved remarkable and promising results with an accuracy of 0.853175 and a loss

of 0.353069 during training. The following are the model's assessment metrics based on VGG16.

In our investigation, the model that is based on VGG16 greatly outperformed the model that is based on VGG19. Table 2 shows the evaluation metrics for the training dataset's validation dataset at each epoch as shown in Table 3.

Table 1. Evaluation metrics of models of model based on VGG16

Epoch	Loss	Accuracy	Validation Loss	Validation Accuracy
1	0.6572	0.6341	0.5755	0.7024
2	0.5690	0.7111	0.5220	0.7222
3	0.5223	0.7375	0.5221	0.7460
4	0.4653	0.7792	0.4794	0.7381
5	0.4713	0.7735	0.5642	0.7262
6	0.4208	0.8069	0.4388	0.7698
7	0.4119	0.8135	0.5462	0.7183
8	0.4103	0.8228	0.5140	0.7341
9	0.3754	0.8351	0.4250	0.7857
10	0.3594	0.8452	0.3755	0.7937

Table 2. Final results of model based on VGG16

Metrics	Value/Score
Accuracy	86.33333333333333%
Precision	84.66666666666667%
Sensitivity	87.58620689655172%
Specificity	85.16129032258064%

Table 3. Evaluation metrics of metrics of model based on VGG19

Epoch	Loss	Accuracy	Validation Loss	Validation Accuracy
1	0.6572	0.6341	0.5755	0.7024
2	0.5690	0.7111	0.5220	0.7222
3	0.5223	0.7375	0.5221	0.7460
4	0.4653	0.7792	0.4794	0.7381
5	0.4713	0.7735	0.5642	0.7262
6	0.4208	0.8069	0.4388	0.7698
7	0.4119	0.8135	0.5462	0.7183
8	0.4103	0.8228	0.5140	0.7341
9	0.3754	0.8351	0.4250	0.7857
10	0.3594	0.8452	0.3755	0.7937

The evaluation metrics of the model based on VGG16 are as follows and listed as in Table 4.

Table 4. Final results of model based on VGG19

Metrics	Value/Score
Accuracy	84.0%
Precision	86.66666666666667%
Sensitivity	82.27848101265823%
Specificity	85.91549295774648%

4. CONCLUSION AND FUTURE

In this study, we suggested a comprehensive computerised approach for face image-based autism identification. This study used a Convolutional Neural Network with transfer learning to create a deep learning-based online interface for identifying autism. The CNN architecture has the right models to extract the facial landmarks, which can classify faces into autistic and non-autistic types. This is done by producing sequences of facial characteristics and calculating the distance

between facial features. Parents and physicians will find it easier to recognise ASD in children with the help of this sort of software. Children with autism may gain from having a correct diagnosis of their condition by picking an appropriate therapy route. With VGG16 as its pre-trained model, the model has a precision of 90% and an accuracy of 84%. This study addresses a gap in the literature by screening young children for ASD using facial photos. Clinical observations that children with ASD and normally developing children have distinct facial characteristics are supported by the study's findings. We are confident that this computer vision method will help address the main issues that lead to racial disparity in the diagnosis or screening of ASD, such as the subjectivity of screening or diagnosis, the difficulty in obtaining access to skilled medical care, and the financial struggles families face globally, particularly in developing countries. Future work may focus on developing the technique into a straightforward mobile application that would let families take a photo with their phone and quickly get a very accurate screening result. More research should be done to combine image- and video-based approaches into one system that enables the detection of both behavioural and facial phenotypic issues in ASD in order to further minimise misclassifications.

REFERENCES

[1] Wang, S., Adolphs, R. (2017). Social saliency. In Computational and Cognitive Neuroscience of Vision, pp. 171-193.

[2] Baio, J., et al. (2018). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2014. MMWR Surveillance Summaries, 67(6), 1-23. https://doi.org/10.15585/mmwr.ss6904a1

[3] Zwaigenbaum, L., Penner, M. (2018). Autism spectrum disorder: Advances in diagnosis and evaluation. BMJ, 361: k1674. https://doi.org/10.1136/bmj.k1674

[4] Sidey-Gibbons, J., Sidey-Gibbons, C. (2019). Machine learning in medicine: A practical introduction. BMC Medical Research Methodology, 19(1): 64.

[5] Akter, T., et al. (2021). Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage. Brain Sciences, 11(6): 734. https://doi.org/10.3390/brainsci11060734

[6] Rajaram, M. (n.d.). Concerns with 'Detect Autism' dataset. Kaggle. https://www.kaggle.com/code/melissarajaram/concerns-with-detect-autism-dataset, accessed on August 6, 2021.

[7] Musser, M. (2020). Detecting Autism spectrum disorder in children with computer vision. https://towardsdatascience.com/detecting-autism-spectrum-disorder-in-children-with-computer-vision-8abd7fc9b40a, accessed on August 1, 2021.

[8] Vo, T., Nguyen, T., Le, T. (2018). Race recognition using deep convolutional neural networks. Symmetry, 10(11): 564. https://doi.org/10.3390/sym10110564

[9] Chaudhuri, A. (2020). Deep learning models for face recognition: A comparative analysis. In Deep Biometrics, pp. 1-17. Springer.

[10] Gwyn, T., Roy, K., Atay, M. (2021). Face recognition using popular deep net architectures: A brief comparative study. Future Internet, 13(6): 144.

[11] Xiao, J., Huang, J., Li, H. (2020). Application of a novel and improved VGG-19 network in the detection of workers wearing masks. Journal of Physics: Conference Series, 1518: 012041. https://doi.org/10.1088/1742-6596/1518/1/012041

[12] Simonyan, K., Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

[13] Atluri, S., Bandi, V.S.S., Chunchu, S., Kagitha, P.R.P., Donepudi, S. (2022). Segmentation of potholes from road images. 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT). http://dx.doi.org/10.1109/ICSSIT53264.2022.9716468

[14] Nitesh, B., Madhuri, A., Sai Manogna, B., Naga Jogendra Babu, K., Ishwarya, N., Mohan Trivendra, G. (2022). Brain tumor segmentation using U-Net based on Inception. 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS). Coimbatore, India.

[15] Alsaade, F.W., Alzahrani, M.S. (2022). Classification and detection of autism spectrum disorder based on deep learning algorithms. Computational Intelligence and Neuroscience, 2022: 8709145. https://doi.org/10.1155/2022/870914