

Article

Deep Learning Algorithms to Identify Autism Spectrum Disorder in Children-Based Facial Landmarks

Hasan Alkahtani ^{1,2}, Theyazn H. H. Aldhyani ^{1,3,*}  and Mohammed Y. Alzahrani ^{1,4}¹ King Salman Center for Disability Research, Riyadh 11614, Saudi Arabia; hsalkahtani@kfu.edu.sa (H.A.)² College of Computer Science and Information Technology, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia³ Applied College in Abqaiq, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia⁴ Department of Computer Sciences and Information Technology, Albaha University, Al-Baha 65799, Saudi Arabia

* Correspondence: talldhyani@kfu.edu.sa

Abstract: People with autistic spectrum disorders (ASDs) have difficulty recognizing and engaging with others. The symptoms of ASD may occur in a wide range of situations. There are numerous different types of functions for people with an ASD. Although it may be possible to reduce the symptoms of ASD and enhance the quality of life with appropriate treatment and support, there is no cure. Developing expert systems for identifying ASD based on the facial landmarks of children is the main contribution for improvements in the healthcare system in Saudi Arabia for detecting ASD at an early stage. However, deep learning algorithms have provided outstanding performances in a variety of pattern-recognition studies. The use of techniques based on convolutional neural networks (CNNs) has been proposed by several scholars to use in investigations of ASD. At present, there is no diagnostic test available for ASD, making this diagnosis challenging. Clinicians focus on a patient's behavior and developmental history. Therefore, using the facial landmarks of children has become very important for detecting ASDs as the face is thought to be a reflection of the brain; it has the potential to be used as a diagnostic biomarker, in addition to being an easy-to-use and practical tool for the early detection of ASDs. This study uses a variety of transfer learning approaches observed in deep CNNs to recognize autistic children based on facial landmark detection. An empirical study is conducted to discover the ideal settings for the optimizer and hyperparameters in the CNN model so that its prediction accuracy can be improved. A transfer learning approach, such as MobileNetV2 and hybrid VGG19, is used with different machine learning programs, such as logistic regression, a linear support vector machine (linear SVC), random forest, decision tree, gradient boosting, MLPClassifier, and K-nearest neighbors. The deep learning models are examined using a standard research dataset from Kaggle, which contains 2940 images of autistic and non-autistic children. The MobileNetV2 model achieved an accuracy of 92% on the test set. The results of the proposed research indicate that MobileNetV2 transfer learning strategies are better than those developed in existing systems. The updated version of our model has the potential to assist physicians in verifying the accuracy of their first screening for ASDs in child patients.

Keywords: deep learning model; autism; facial appearance; machine learning model; recognition system

Citation: Alkahtani, H.; Aldhyani, T.H.H.; Alzahrani, M.Y. Deep Learning Algorithms to Identify Autism Spectrum Disorder in Children-Based Facial Landmarks. *Appl. Sci.* **2023**, *13*, 4855. <https://doi.org/10.3390/app13084855>

Academic Editor: Yu-Dong Zhang

Received: 22 March 2023

Revised: 10 April 2023

Accepted: 11 April 2023

Published: 12 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Autism spectrum disorder (ASD) is a complex condition that makes it difficult to communicate on a day-to-day basis [1]. Autism is characterized by a range of symptoms, many of which are mild but might sometimes call for specialized treatment. Patients with ASD often struggle to communicate verbally, via gestures, or through facial expressions. Although patients with ASD are often identified by medical professionals based on neurophysiological signals, there is neither a definitive biosignature nor a pathological method that can readily diagnose autism [2]. An early diagnosis may offer opportunities

for beneficial lifestyle changes [3], even though no therapy cures the condition. Children exhibiting indications of ASD may benefit from an early diagnosis due to the malleability of brain development, which might help them improve their social lives. Some research has shown that children who receive medical care before the age of two years have higher IQs than those who do not receive it until later on in life [4]. According to the recent research [5], most children with ASD are not diagnosed until they are at least three years old [6,7].

Numerous studies have investigated the important characteristics of autism through a variety of lenses, such as facial-feature extractions [8] using eye-tracking strategies [9], face recognition [10–12], bio-medical image analysis [13], application creation [14], and speech recognition [15]. Among these methods, face recognition is particularly useful for determining a person’s emotional state, and it has the potential to accurately diagnose autism. It is a popular method used for analyzing human faces and extracting distinguishing characteristics between normal and abnormal faces, as well as for mining significant information to reveal behavioral patterns [16,17].

In light of the recent developments in the predictive analytics of facial-pattern recognition, several intensive initiatives are presently underway in the field of autism research to analyze the data on autistic children in an attempt to diagnose ASD at an earlier age. To automate the identification of facial expressions in diverse neurological illnesses, Yolcu et al. [18] introduced the CNN technique to detect ASD. The first CNN was taught to divide important facial components, while the second CNN was utilized to detect facial expressions. They divided the system into four convolutional neural network models by Haque and Valles [19] were able to recognize human facial emotions in 2018. They did this by modifying Kaggle’s Facial Expression Recognition 2013 (FER2013) dataset of young, autistic children with lighting effects (darker or brighter shades of contrast). Figure 1 presents an overview of the numerous diagnostic methods for autism.

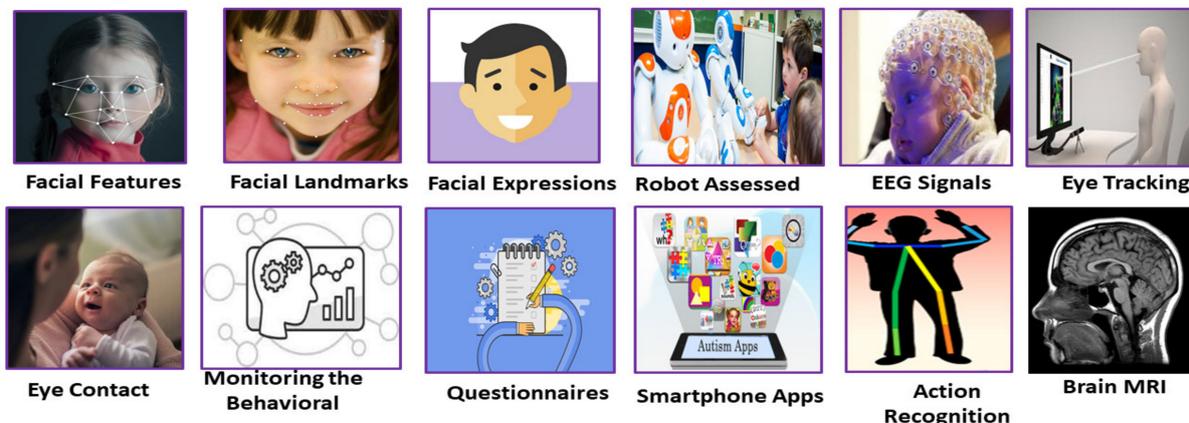


Figure 1. Different types of technology to detect autism.

We introduce a face-recognition system based on transfer learning for a more precise autism diagnosis in this study [20]. We perform this by amassing a large dataset of facial appearances from children with and without autism, which we then examine using a number of machine learning, deep learning, and other superior pre-trained models. Then, the effectiveness of these classifiers is assessed using a number of different metrics, including accuracy, area under the curve (AUC), sensitivity, specificity, fall-out rate, and miss rate, among others. Subsequently, we enhance MobileNet-V1, which presently demonstrates the best level of accuracy in this predictive analysis. Figure 2 shows the process of the transfer learning model via the computer vision method.

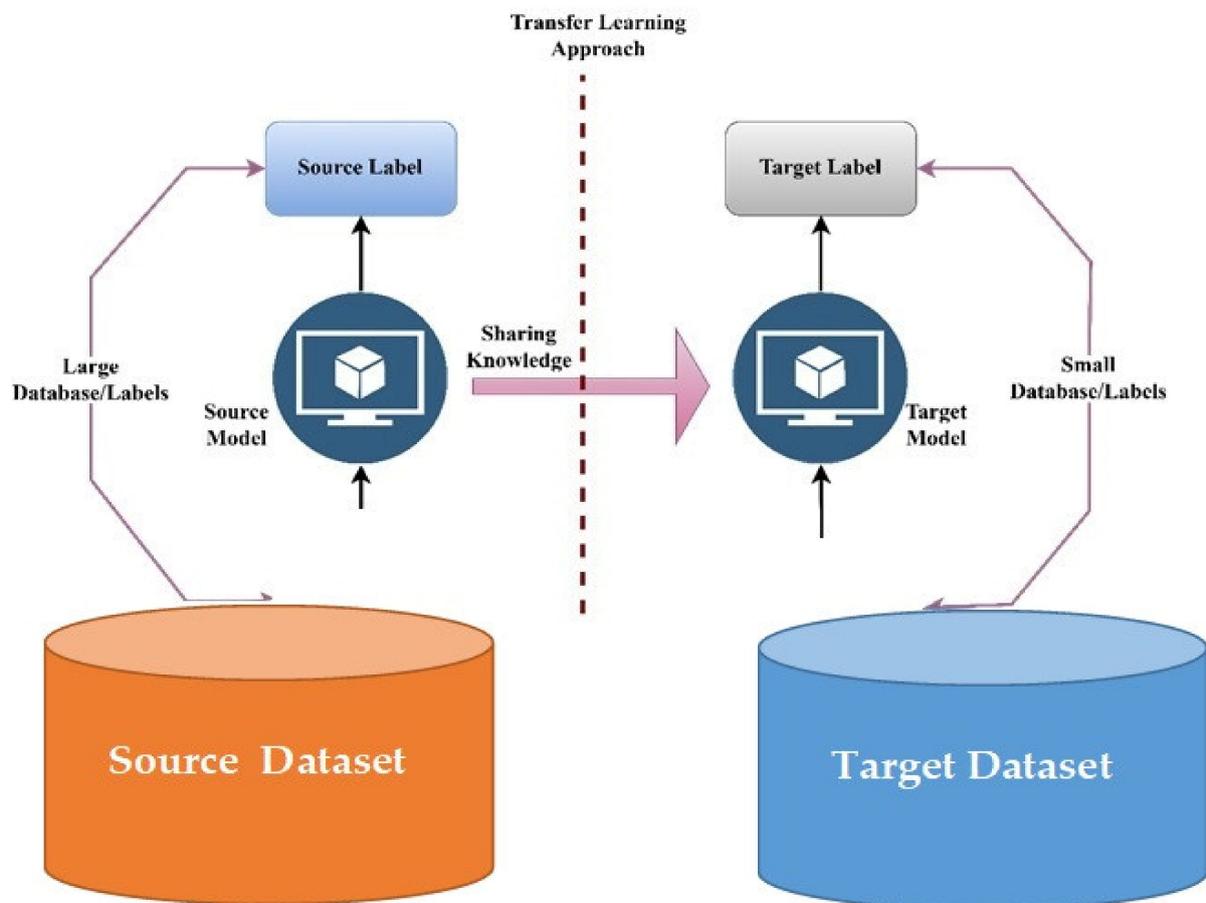


Figure 2. Transfer learning model.

When compared to children who do not have autism, autistic individuals have a specific facial feature that is not shared by children who do not have autism, as discovered by the researchers at the University of Missouri who researched the diagnosis of children based on their facial features (non-autistic children). These characteristics include the upper face that is exceptionally broad, including eyes that are positioned far apart, and the middle of the face that is unusually short, including the cheeks and nose. The proposed system makes use of a number of different transfer learning methods that are based on deep convolutional neural networks (CNNs) in order to identify autistic children based on their facial appearance.

The main contributions of the proposed system are as follows:

- We propose a more robust face-recognition framework based on transfer learning, which has the potential to provide high accuracy in identifying children with autism.
- In this study, we analyze an enhanced MobileNet model, which outperforms hybrid deep learning with deep learning methods to identify ASD.
- Various machine learning and deep learning models are developed to detect ASD.
- ASD children present characteristics that include an upper face that is exceptionally broad, including eyes that are positioned far apart, and a middle section of the face that is unusually short, including the cheeks and nose; therefore, developing an expert system for identifying ASD based on the facial landmarks of children is the main contribution for improving the healthcare system in Saudi Arabia for detecting ASD at an early stage.

2. Background of Study

Our knowledge of ASD, its categorization, and our capacity to research its essential characteristics have all advanced significantly in recent years [21–23]. ASD Test is a smartphone application developed by Thabtah et al. [24] to collect ASD data for newborns, children, and adults. Q-CHAT and AQ-10 evaluation tools were used as the foundation for the creation of this app. Data could be collected for persons with ASDs using this program, which was then placed in a repository for ML at the University of California, Irvine. Omar et al. [25] suggested a method where random forest (RF), classification and regression trees (CART), and random forest–iterative Dichotomizer 3 were all tested on the AQ-10 and 250 real-world datasets (ID3). Sharma et al. [26] investigated these datasets using a variety of methods, such as naive Bayes (NB), stochastic gradient descent (SGD), K-nearest neighbors (KNN), random tree (RT), and K-Star, with the assistance of the CFS-greedy stepwise feature selector. Using a variety of tree-based classifiers, Satu et al. [27] examined samples of children ranging in age from 16 to 30 years in order to establish what characteristics define a “normal” or “autistic” young person. Erkan et al. [28] studied comparable datasets and employed KNN, SVM, and RF to evaluate the method that was superior in terms of its ability to diagnose ASDs. In contrast, Thabtah et al. [29] employed IG and CHI to construct feature subsets of adults and adolescents; these subsets served as inputs to LR for detecting ASDs. These subsets were used to differentiate between those with and without ASDs. Akter et al. [30] collected and adjusted the datasets, including information obtained from newborns, children, adolescents, and adults, and examined them using a variety of classifiers, with SVM exhibiting the best performance for toddlers and AdaBoost delivering the best results for older children and adults; GLMBoost also performed fairly well for the dataset consisting of teenagers. Hossain et al. [31] proposed multilayer perceptron (MLP) and sequential minimal optimization (SMO) methods for detecting ASD. The SMO algorithm was shown to be the most accurate, with a success rate of 91% across all child datasets, 99.9% across all adolescent datasets, and 97.58% across all adult datasets.

Because of the societal effects of ASD on different nations, the research into the subject of diagnosing ASD based on children’s facial characteristics is rapidly expanding. This technique has the potential to be a benchmark for identifying children with ASD, as well as children without this condition. Recent research [32–42] demonstrates the potential of deep neural networks for the diagnosis of a variety of diseases. In particular, CNN models have been shown to be especially effective in this area. As CNNs are so adept at learning by automatically extracting the hidden features from a high volume of images, they are the feature extractors of choice when it comes to object identification and image categorization procedures. This ability allows CNNs to become increasingly accurate over time. Training CNN models takes a long period of time and high number of computer resources [35], despite the fact that CNNs are very effective and precise in their predictions.

3. Proposed System

Using a transfer learning individual with machine learning algorithms for autistic facial appearances and landmark detection, the purpose of this research was to determine whether children show signs of ASD at an early age. In this study, deep learning with machine learning methods were applied to automatically extract the main characteristics of facial landmark features from children for detecting ASD. These features are very difficult, if not impossible, to distinguish by eye owing to their level of complexity. Then, we ran these characteristics through a series of layers, with the diagnosis of ASD being obtained from the most densely packed layer at the top. Figure 3 shows the framework the of ASD diagnosis system.

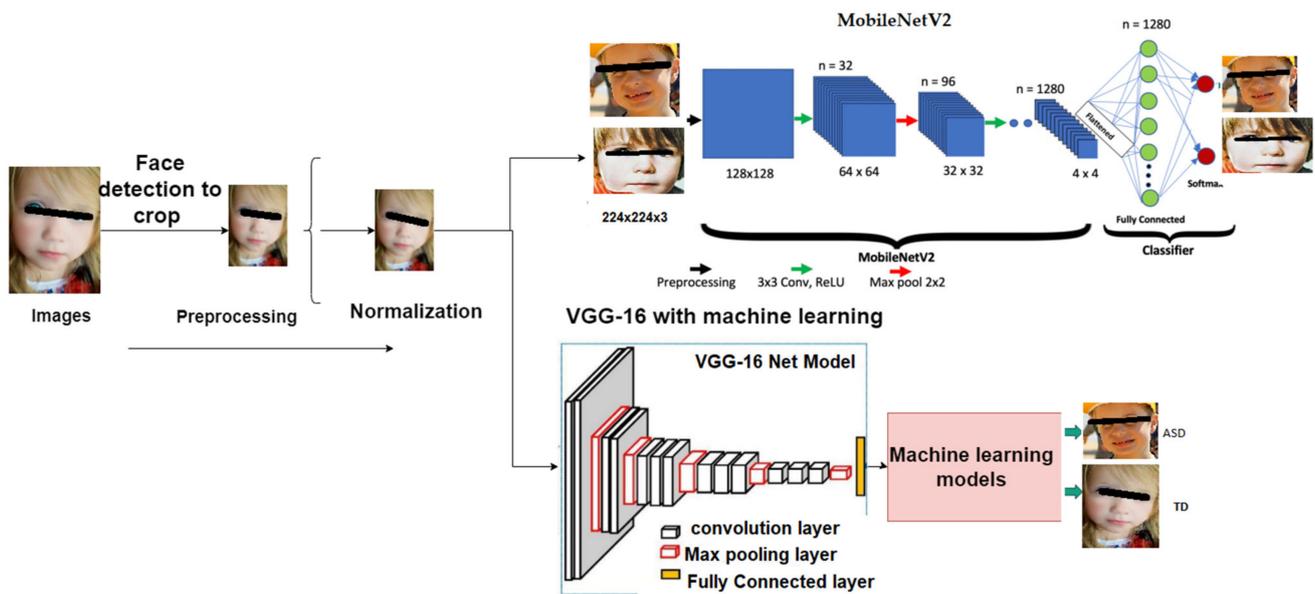


Figure 3. Framework of the proposed recognition system.

3.1. Dataset

This study used facial images gathered from the Kaggle’s autistic children dataset. This is the only publicly available dataset of its kind; thus, we used it to create the models we proposed. Children aged 2–14 years were included in the dataset, with the majority being between the ages of 2 and 8 years. All of the photos were 2D RGB jpegs. The dataset had two classes: the autistic class contained images of children with autism, while the non-autistic class contained photographs of children not diagnosed with autism. A test folder comprised the photos required to test the model after it had been trained. The test folder had two subfolders classified as autistic and non-autistic. Each subfolder contained 100 images $224 \times 224 \times 3$ in size, in a jpeg format. The autistic subfolder contained the facial images of children with autism, and the non-autistic subfolder contained facial images of children randomly collected from online searches. The dataset contained 2940 images where 1327 images were of autistic children and 1613 were of non-autistic children. Figure 4 shows a sample of the dataset.



Figure 4. Images of autistic and non-autistic children.

3.2. Deep Learning Models

CNNs place a burden on neural networks, which have been shown to be beneficial in the classification and detection of ASDs. ConvNets are reminiscent of inscriptions, which may be because their names are derived from the veiled layers that they include. A CNN is composed of three parts: convolution, pooling, and a fully connected layer [6]. Convolution is used to collect local image attributes, pooling is used to reduce dimensionality, and the fully connected layer is used to provide the output that is needed. This mode highlights local image features and uses brighter pixels to illustrate the image boundaries. It also allows for better processing [43,44].

An activation function, in addition to convolution and pooling functions, may be found embedded inside each foundational layer of the network. The output of the convolution process is the same image that is fed into it along with one filter. The performance of the neural network is affected by a variety of factors, including the image size (224 × 224 × 3), height, width, and channels. The height of the image channel was 200 by 200 pixels and the processing size was 49,152 bytes. The image channel used the RGB color model (224 × 224 × 3). For instance, if the image dimensions are 2048 by 2048 in 3 dimensions, the required weighted extent size is 12 million.

$$C = \sum_1^i \sum_1^j I_{ij}F_{ij} \tag{1}$$

F denotes a convolution kernel or filter, while rows and columns are indicated by the letters i and j . For instance, after multiplying the image by the kernel, Figure 5 displays both the raw image and filter. Figure 6 presents the newly developed output value in a two-dimensional format. The image is broken up into perceptrons using convolution, which are then planed along the (y), (x), and (z) axes, respectively. Each layer has a number of filters that may be used to locate certain attributes. The following annotations are added to the X -sized feature maps that are produced by layer L :

$$C_i^L = B_i^L + \sum_{j=1}^{x^{(L-1)}} F_{i,j}^L * C_j^{(L-1)} \tag{2}$$

where B_i^L : bias matrix.

$F_{i,j}^L$: filter linking j th based on the feature map in the layers.

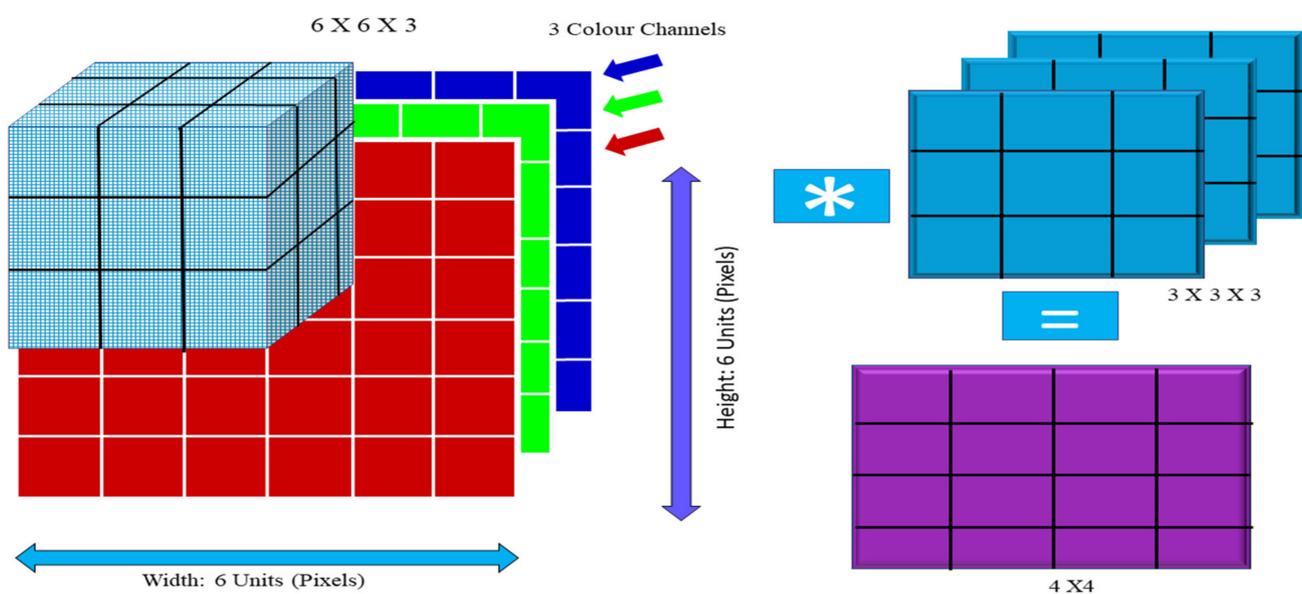


Figure 5. Filter matrix of the convolution operation.

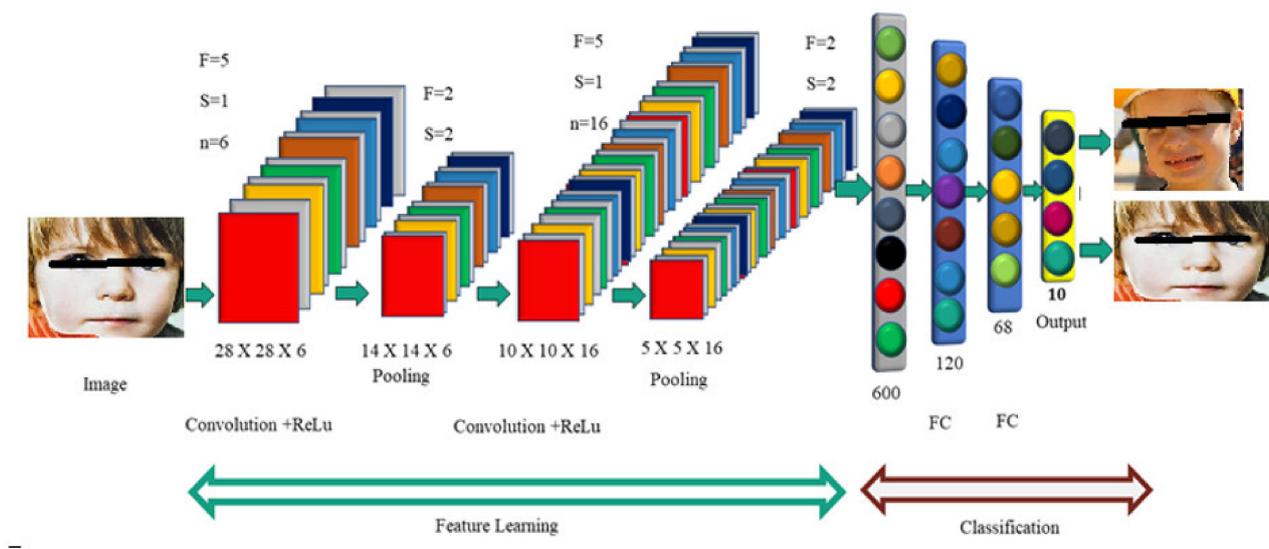


Figure 6. Convolution neural network.

For example, the blue channel may have a value of -1 , whereas the other channels might have values of $+1$ or 0 . Utilize the dot product in order to calculate the value of the convolution. Convolution is a process that warps images. In this demonstration, (I_s) equals 1. If both (I_s) and (I_s) are true, then (F_s) (F_s) describes the same image size (F_s) .

$$C_s = ((I_s - F_s)/S) + 1 \tag{3}$$

If the values of I_s , F_s , and S are 6, 3, and 1, respectively, then $C_s = (6 - 3 + 1) = 4$. The following is how the output dimension was calculated:

$$\lfloor (Width - F_s + 2P)/S \rfloor + 1 \tag{4}$$

where F_s is the filter size, padding is P , S is stride, and Equation (4) is the floor value.

3.2.1. MobileNet Architecture

MobileNet is a deep learning model developed to effectively conduct image classifications in different technology platforms, such as mobile devices, embedded systems, or low-power PCs that do not have a GPU [38]. Figure 7 provides a visual representation of the MobileNet model’s underlying architectural framework. One key feature that sets this CNN model apart from others is the presence of depth-wise separable filters, which perform both depth-wise and point-wise convolutions.

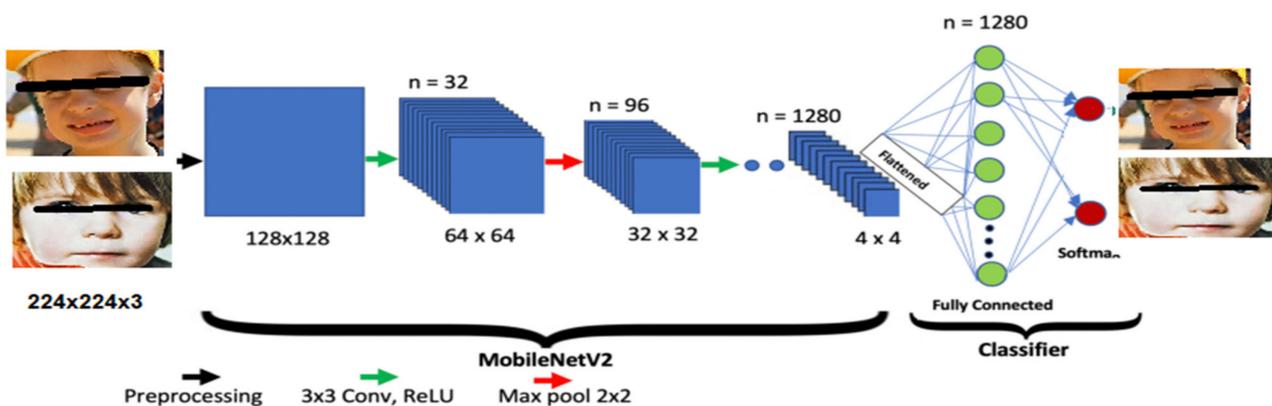


Figure 7. MobileNet’s structure.

Unlike traditional convolution procedures, depth-wise convolution makes use of each channel of the input image in its own right to extract feature maps using a variety of filtering techniques. When increasing the number of channels in the output image to the proper number, using filter masks sized (1 1) results in a significant reduction in the length of processing time required by the computer. This model is well-known as a straightforward deep neural network. MobileNet may be used for a variety of applications, some of which include object detection, facial-attribute recognition, fine-grain categorization, and geographic localization. The important parameter values of the MobileNet model are presented in Table 1.

Table 1. Parameters of the MobileNet model.

Input	Values
Image input	$224 \times 224 \times 3$
Initial learning rate	0.0001
Epochs	100
Optimizer	RMSprop
Batch sizes	80

3.2.2. Hybrid VGG-16 with Machine Learning Models

The Visual Geometry Group (VGG) at the University of Oxford is responsible for the development of a pre-trained image-recognition model known as the VGG-16 model. This model was trained using a large collection of photos known as the ImageNet dataset. This dataset includes over 14 million images and over 1000 distinct categories. During the training phase, the model identifies attributes from the image itself, which allows it to recognize and categorize items in the images it is shown. The proper operation of the deep neural network may be confirmed when the size of the network is increased. Networks A, A-LRN, B, C, D, and E were all built by the VGG. Figure 8 shows that VGG-16 networks C and D; each contain 13 convolution layers, 3 layers that are entirely linked, and 16 layers overall. Convolution in network C is conducted using a 1×1 filter size. However, in network D, this is not the case. Network D commonly employs a filter size of 3×3 for convolution operations, increasing the total number of parameters that need training to 138 million. Network E, also known as the VGG-19 network, receives its name from the fact that it has 16 convolution layers, followed by 3 layers that are totally connected over 19 levels. All VGG networks have ReLu; however, they do not really use it since executing local response normalization takes more time and memory space during training [10]. The most significant difference between AlexNet and VGG is that AlexNet has an 11×11 kernel with a stride of 4×4 , whereas VGG uses a 3×3 kernel with a stride of 1×1 . The 1×1 convolution filter that VGG provides is useful for both predictive modeling and classification work. The network employs a technique called multiscaling to boost the quality of the data, which increases the number of inputs and solves the problem of overfitting [11]. The final categorization of the items in the photographs is accomplished via the use of layers that are completely linked. In this investigation, the VGG-16 model was employed; however, the top layers were stripped away, which implies that the model does not include any of the layers that are already completely linked. This was performed in order to extract the features that were included in the pictures, and the output of the model was a feature map that contained the features that were retrieved from the image. The input shape of the model was configured to have the dimensions of $224 \times 224 \times 3$, which was the default size for the VGG-16 model. Following the extraction of the characteristics from the photos, we classified the results using a number of different machine learning models (autistic or normal).

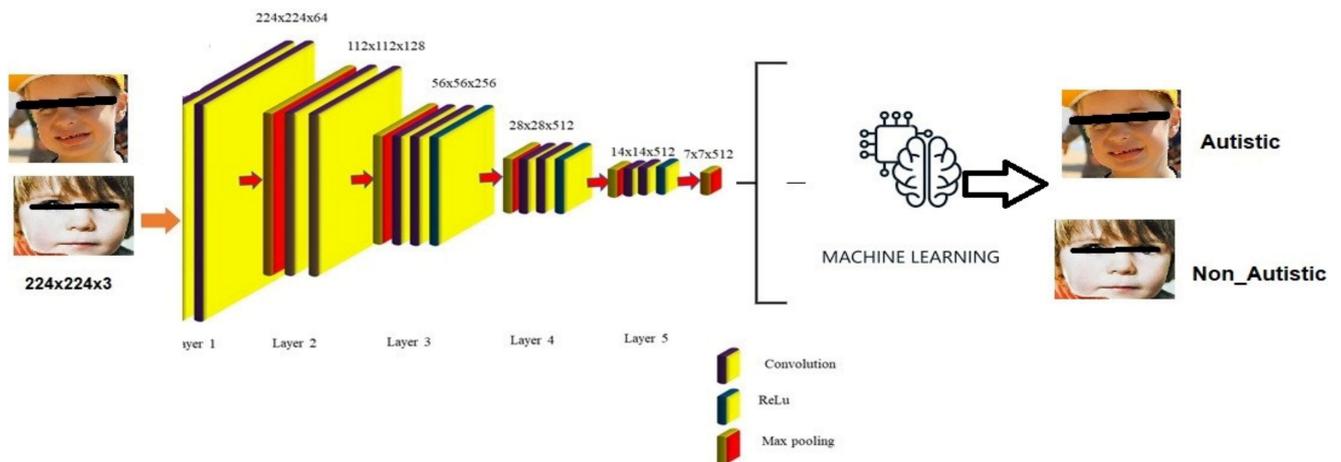


Figure 8. VGG-16 structure.

3.3. Measuring the ASD Diagnosis System's Performance

To assess the efficacy of the provided algorithms in detecting ASDs, the evaluation metrics included sensitivity, specificity, precision, and recall scores, as well as the F1-score. The equations relating to these parameters are presented below:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (5)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (7)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (8)$$

$$F1 - score = 2 \times \frac{precision \times Sensitivity}{precision + Sensitivity} \times 100 \quad (9)$$

4. Experiments

The deep learning MobileNet-V1 and VGG-16 models hybridized with various machine learning models, namely, logistic regression, LinearSVC, random forest, decision tree, gradient boosting, MLPClassifier, AdaBoost, and K-nearest neighbors, were applied to detect ASDs. As a result, Jupyter Notebook, based on Python, was responsible for all of the calculations that were altered for developing a classification of ASDs in children. Several methods may be used to detect ASDs. Individual classification models were trained using the 80% training set, and their efficacy was evaluated using the 20% test set. Both sets were utilized in conjunction with the training set. For the purpose of determining how well these classifiers functioned, a number of assessment measures, including accuracy, area under the curve (AUC), sensitivity, specificity, precision, and F1-score, were computed.

4.1. Results

4.1.1. Results of the MobileNet Model

The training of the MobileNet learning models with the assistance of the Keras API Library was applied. Matplotlib, Sklearn, and Pandas are just a few examples of the data visualization and analysis tools utilized in the process of determining the effectiveness of the models. We used a standard set of hyperparameters with the following values to compare the performance of MobileNet-based models trained with different optimizers

with a batch size of 80 and a learning rate of 0.001 across 100 epochs. The dataset comprised a total of 2940 images, of which 2540 were utilized for training, 300 were utilized for testing, and 100 were utilized for validation. As can be observed in Table 2, when we were building the data frame, we provided an image a value of 0 for children in the normal control (NC) group and a value of 1 for children who were diagnosed with ASD. It is observed that the MobileNet model achieved an accuracy of 92%.

Table 2. Results of MobileNet model.

	Precision %	Recall %	F1-Score %	Support
Autism	92	92	92	150
Normal	92	92	92	150
Accuracy		92		300
Macro avg	92	92	92	300
Weighted avg	92	92	92	300
Overall results	90.47	92	92	

Figure 9 is the MobileNet model’s confusion matrix, which includes indicators, such as the true-negative and false-positive rates, as well as valid-positive and false-negative ones. When the MobileNet model was applied to the total of 300 test images, 140 images were classified as normal (with 19 false-positives and 10 false-negatives), and 131 images were classified as autism-related (true-positive). According to the evaluation metrics, the MobileNet model achieved good accuracy results.

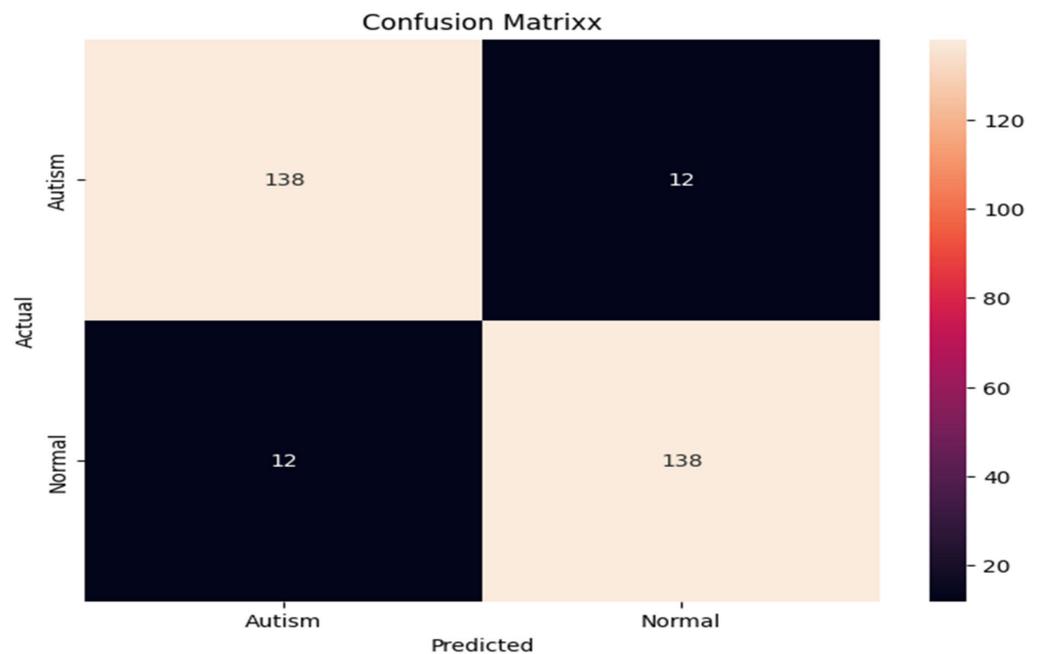


Figure 9. MobileNet model confusion matrix.

Figure 10 provides a graphical representation of the MobileNet model’s accuracy, where the percentage accuracy of the model is represented in % (Y-axis) and each iteration of 17 epochs (X-axis). In order to obtain an accurate image of how effectively the training system functioned, we examined how well the validation system operated. A break in the process of optimizing led to a significant increase in precision that was eventually brought up to 100 epochs. During the course of the validation procedure, the MobileNet model’s efficiency increased from 75% to 92% during the testing process phase. Because of this, we

decided to employ a categorical cross-entropy function for obtaining the accuracy loss of the MobileNet model. In a period of 100 epochs, the validation loss decreased from 7 to 1 during the testing phase.

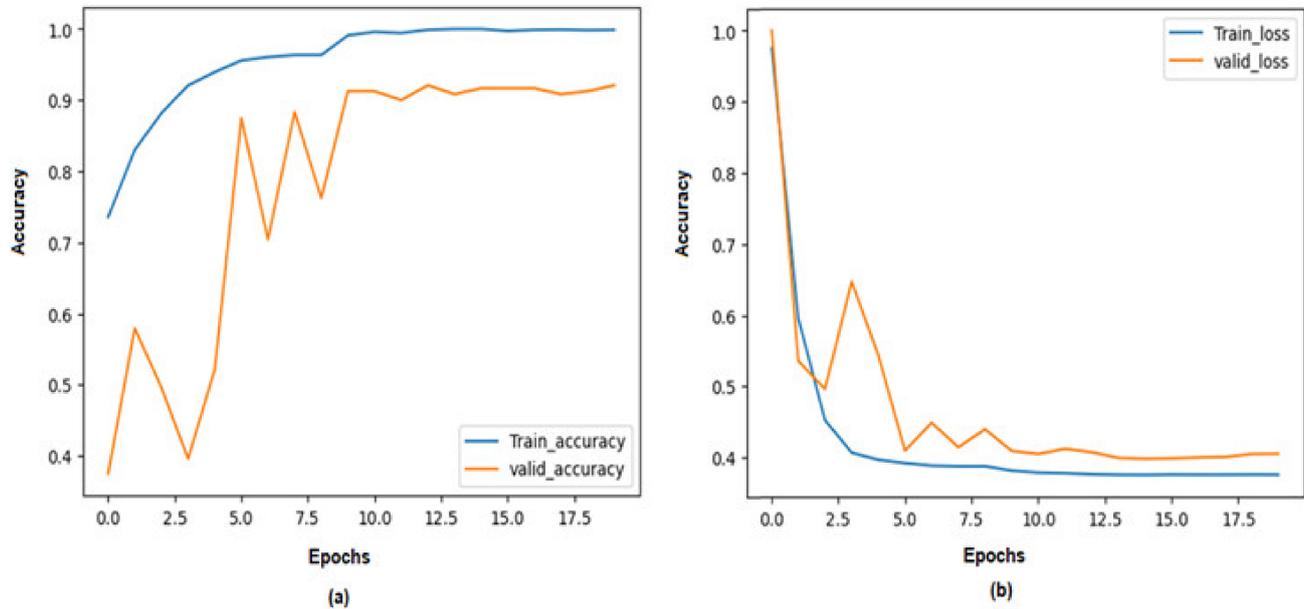


Figure 10. Performance of MobileNet. (a) Accuracy; (b) loss.

4.1.2. Results of VGG-16 with Machine Learning

The results of the VGG-16 deep learning model hybridized with various machine learning models, namely, logistic regression, LinearSVC, random forest, decision tree, gradient boosting, MLPClassifier, AdaBoost, and K-nearest neighbors, are presented in the study. In this study, we made use of the VGG-16 model without its top layers. This means that the model does not contain the layers that are totally related to one another. This facilitated the model’s ability to extract the images’ features, and the end result of this procedure was a “feature map” that included those extracted features. The dimensions that are common for a VGG-16 model, which are $224 \times 224 \times 3$, were applied to the model’s input shape. In order to classify the outcomes of feature extraction from the photographs, we made use of several different machine learning models (autistic or normal). The Kaggle ASD dataset includes a total of 2940 images; of those, 2540 were used for training, 300 were used for testing, and 100 were used for validation. The outcomes of VGG-16 using a logistic regression model are shown in Table 3. It can be observed that VGG-16 using logistic regression is 82.14 percent accurate.

Table 3. Results of VGG-16 with logistic regression.

	Precision %	Recall %	F1-Score %	Support
Autism	83	81	82	291
Normal	82	83	82	297
Accuracy		82.14		
Macro avg	82	82	82	588
Weighted avg	82	82	82	5888

Figure 11 shows the confusion matrix of VGG-16 with logistic regression. For the 300 images, the model produced a false-positive outcome for 50 images, a true-positive outcome (autism) for 247 images, and classified 236 images as normal.

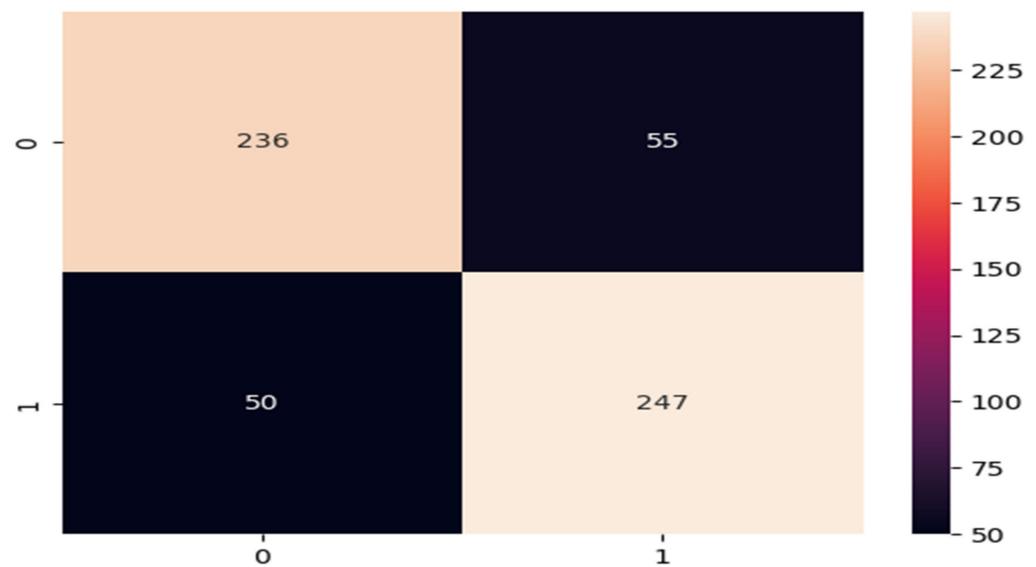


Figure 11. Confusion matrix of VGG-16 with logistic regression.

The results of the VGG-16 and linear SVC hybrid models for detecting ASD are presented in Table 4. In its predictions, the model presented an F1-score of 81%, along with a precision score of 82% and a recall score of 80%. The average accuracy of the model was 81.46%.

Table 4. Results of VGG-16 with linear SVC.

	Precision %	Recall %	F1-Score %	Support
Autism	82	80	81	291
Non-autistic	81	83	82	297
Accuracy		81.46		
Macro avg	81	81	81	588
Weighted avg	81	81	81	588

The performances of the VGG-16 and linear SVC hybrid models are shown in the confusion matrix presented in Figure 12. The confusion matrix has four indicators for the performance of the VGG-16 and linear SVC models. The model produced 50 false-positives, 247 true-positives (autism), and classified 232 images as normal out of a total of 300 test images.

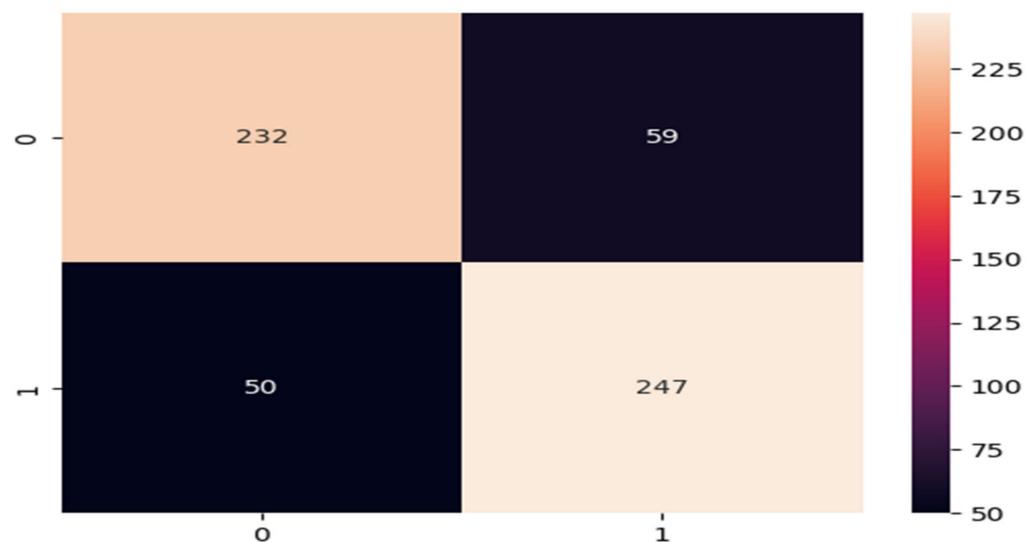


Figure 12. Confusion matrix of VGG-16 with linear SVC.

The results of VGG-16 with random forest are presented in Table 5. The results of VGG-16 with random forest are not satisfactory; the accuracy of the model was 78.06%.

Table 5. Results of VGG-16 with random forest.

	Precision %	Recall %	F1-Score %	Support
Autism	75	84	79	291
Non-autistic	82	72	77	297
Accuracy		78.06		
Macro avg	79	78	78	588
Weighted avg	79	78	78	588

The confusion matrix of VGG-16 with random forest is presented in Figure 13. There is a very high rate of false-positives (83 images). The number of true-positives (214 images) is much lower than that of the previous model.

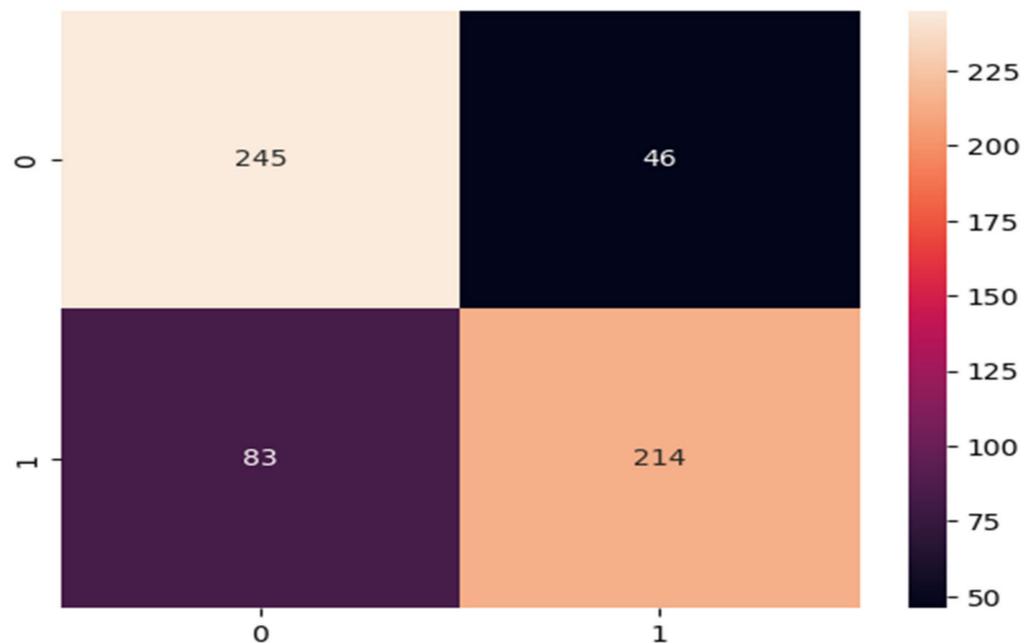


Figure 13. Confusion matrix of VGG-16 with random forest.

Table 6 shows the results of VGG-16 with a decision tree. This hybrid achieved an accuracy of 66.15%. Figure 14 displays the VGG-16 decision tree confusion matrix. We achieved a significant number of false-positives (97 pictures) and a low number of genuine negatives (189 images).

Table 6. Results of VGG-16 with decision tree.

	Precision %	Recall %	F1-Score %	Support
Autism	66	65	66	291
Non-autistic	66	67	67	297
Accuracy		66.15		
Macro avg	66	66	66	588
Weighted avg	66	66	66	588

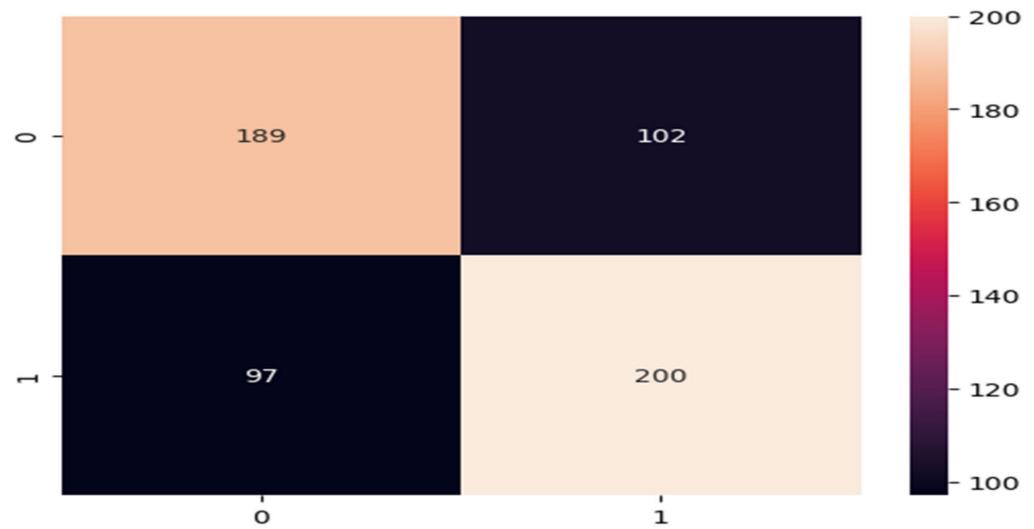


Figure 14. Confusion matrix of VGG-16 with decision tree.

The results of VGG-16 with gradient boosting are shown in Table 7. VGG-16 with gradient boosting achieved an accuracy of 75.15%, superior to that of the decision tree algorithm.

Table 7. Results of VGG-16 with gradient boosting.

	Precision %	Recall %	F1-Score %	Support
Autism	74	78	76	291
Non-autistic	77	73	75	297
Accuracy		75.51		
Macro avg	76	76	76	588
Weighted avg	76	76	76	588

The confusion matrix of VGG-16 with gradient boosting is presented in Figure 15. The number of true-positives was 218, while the number of true-negatives was 226. False-positives and false-negatives numbered 79 and 65, respectively.

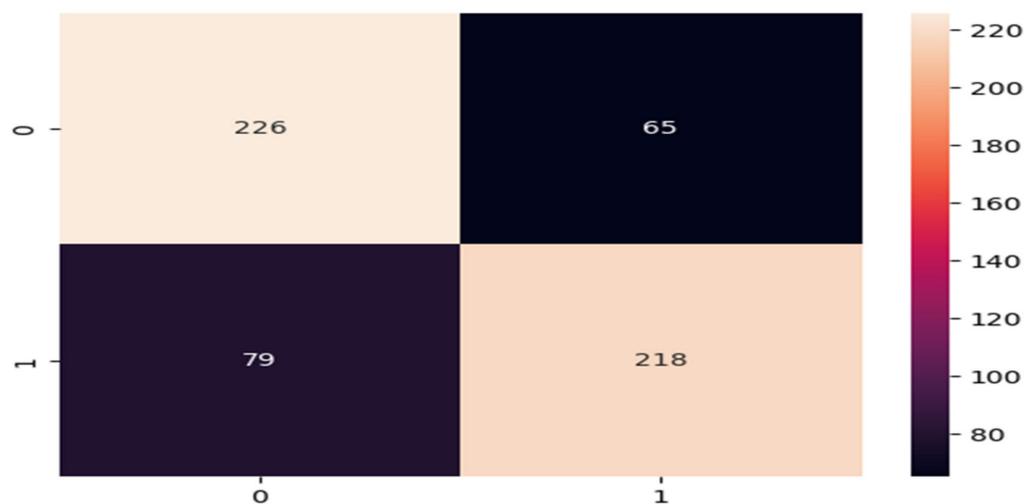


Figure 15. Confusion matrix of VGG-16 with gradient boosting.

Table 8 shows the results of VGG-16 with the MLPClassifier. This model achieved an accuracy of 77.04%. The confusion matrix of VGG-16 with the MLPClassifier is presented in Figure 16. The number of true-positives was 73, and 218 images were classified as normal.

Meanwhile, 235 images were classified as autism-related. The results of this model are not satisfactory.

Table 8. Results of VGG-16 with MLPClassifier.

	Precision %	Recall %	F1-Score %	Support
Autism	78	75	76	291
Normal	76	79	78	297
Accuracy		77.04		
Macro avg	77	77	77	588
Weighted avg	77	77	77	588

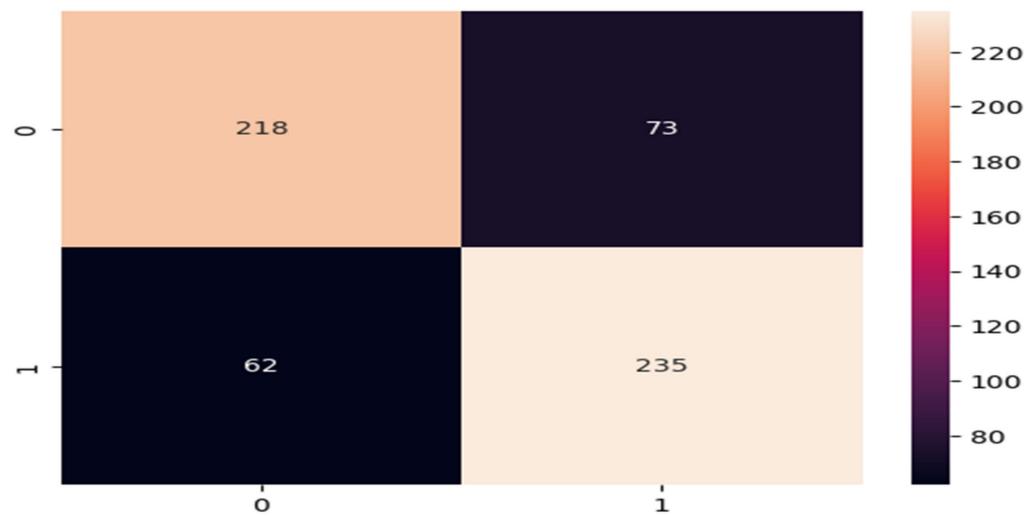


Figure 16. Confusion matrix of VGG-16 with MLPClassifier.

Finally, the results of VGG-16 with KNN are shown in Table 9. This model performed poorly, with a recall of 0.04% and an F1 score of 0.08%. The confusion matrix of this model is presented in Figure 17. It was noted that there were less than 13 false-negative and true-negative images.

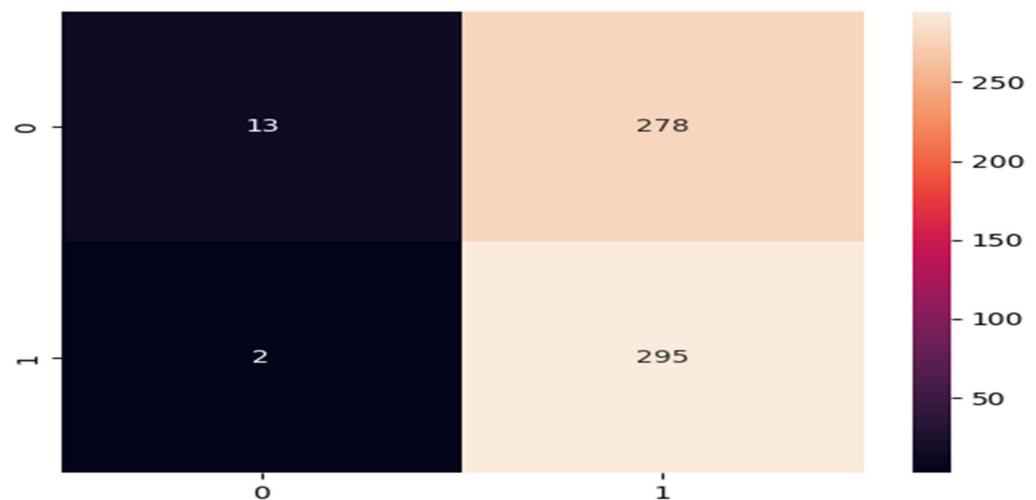


Figure 17. Confusion matrix of VGG-16 with KNN.

Table 9. Results of VGG-16 with KNN model.

	Precision %	Recall %	F1-Score %	Support
Autism	87	0.04	0.08	291
Non-autistic	0.51	0.99	0.68	297
Accuracy		52.38		
Macro avg	69	52	38	588
Weighted avg	69	52	38	588

In summary, we can observe that the MobileNet method for the diagnosis of ASD based on facial expression recognition is reliable, efficient, and easy to disseminate. It provides a one-of-a-kind approach for identifying facial expressions based on attention, which allows us to concentrate on critical facial characteristics, such as the brows and lips. We recommend this method be used to help to detect ASD in children.

5. Discussion

The use of facial landmarks has promise as a screening tool for autism spectrum disorder (ASD). Creating a concise and informative paradigm for ASD diagnosis is critical for overcoming the challenges of dealing with children. Therefore, developing such systems is a challenging task because of the intricacies of attentional behavior present in individuals with ASDs. Early diagnosis and intervention result in the most favorable outcomes for patients receiving treatment for ASDs. Hospitalization has traditionally been necessary to diagnose ASD in children; this process is not only expensive and time consuming, but it is also susceptible to expert prejudice [45]. In this work, we presented a method that is objective, usable, and effective for diagnosing children with ASD based on their facial expressions.

Our concept reduces the gap between autism categorization and facial analysis, making automated autism categorization an alternative method that is more efficient in terms of both cost and time. Our deep learning model employed MobileNet as well as VGG-16 with various machine learning models in order to complete both of these objectives. The model was trained and validated using 2940 images; 80% of the data were used for training and 20% were used for testing. The data was obtained from children with autism, as well as children who did not have autism. According to the results of our study, a single image may be all that is required to make an accurate diagnosis of autism in children. It is possible that this diagnostic method could be used for a number of other illnesses as well.

Our results reveal that the proposed model achieved an accuracy of 92%. The accuracy of MobileNet's results was evaluated using the ROC metric, with true-positives for normal and pathological classes presented on the y -axis and false-positives indicated on the x -axis. The results of the MobileNet model compared to different existing systems are presented in Table 10. Figure 18 presents the results of the measurements used to determine the ROC of the MobileNet model.

Table 10. Significant results of ASD diagnosis system against existing diagnosis systems.

References	Model	Dataset	Accuracy %
[10]	MobileNet	Kaggle dataset	87
[24]	Xception	Kaggle dataset	90
(2022) [46]	Xception	Kaggle dataset	91
Proposed		Kaggle dataset	92

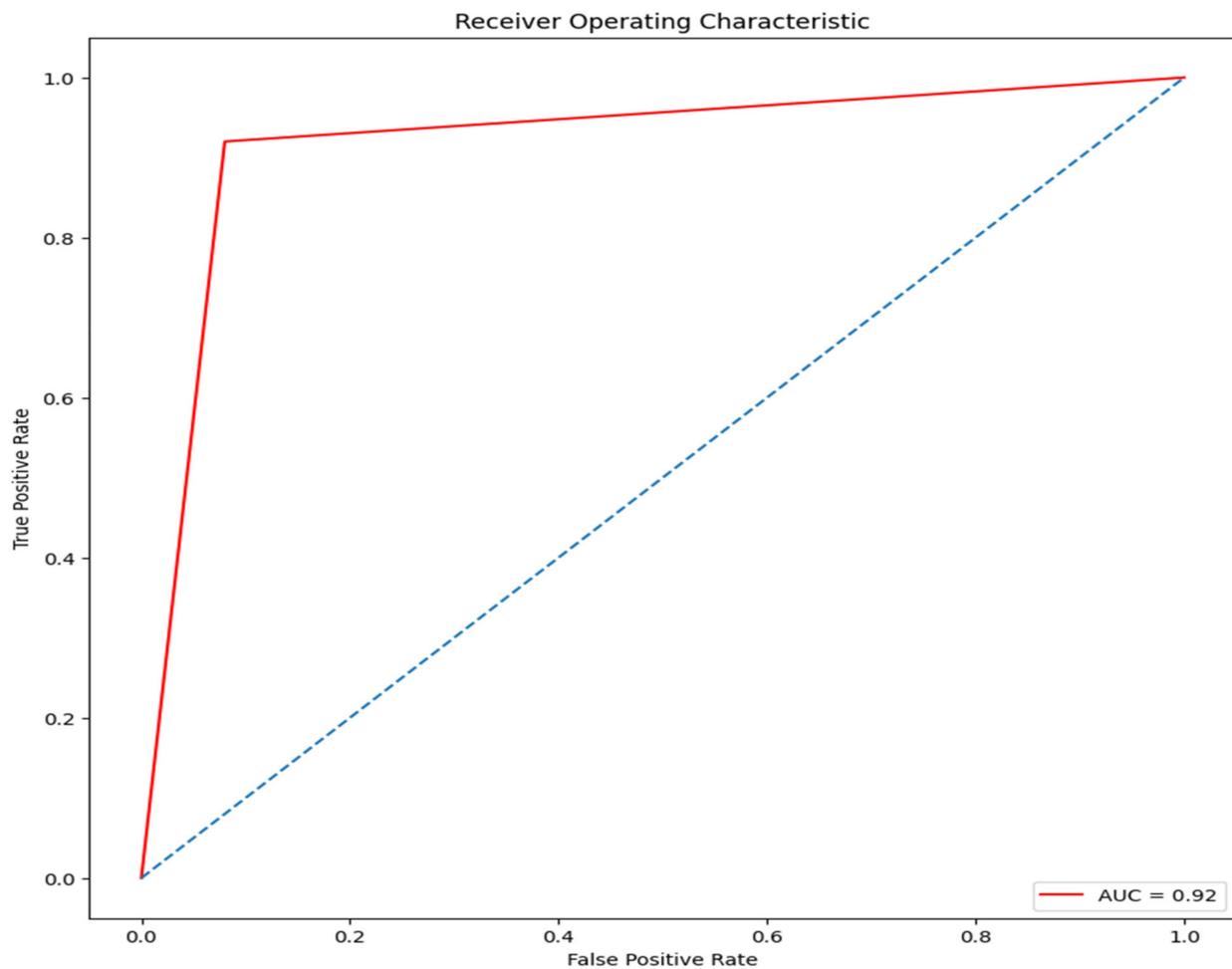


Figure 18. ROC of MobileNet model.

Khosla et al. [46] stated that they were able to achieve an accuracy of 87% using MobileNet on the new autism dataset. Rahman and Subashini used the Xception model, achieving an accuracy of 90%. A comparison of the accuracy reported by a number of studies is shown in Figure 19. Alsaade and Alzahrani developed an exception model to detect ASD, using an old dataset that was considerably enhanced to help with classifying the algorithm with high accuracy. The proposed expert system increased the accuracy by 1% compared to recent, existing studies, due to the complexity of the dataset. This is the only standard dataset that has ASD images, which were collected from different individuals with different ages. Furthermore, the capturing of these images using different devices allowed the system to show the complexity of increasing its accuracy.

If ASD is identified at an earlier age through the use of facial images, this will have a tremendous impact not only on the child, but also on their parents and the clinician. When a child is diagnosed with ASD at an early age, the benefits include the fact that it only takes the physician a few moments to determine whether or not the child is autistic or developing normally after being shown images of the child's face. This is one of the benefits of early ASD diagnoses. However, it is more challenging when an expert performs a manual diagnosis of autism or the typical development in a child based on a visual interpretation of its facial traits. The model increases in accuracy when it is trained on a larger training set. If the suggested model were to be implemented in the form of a mobile application, then the parents would be able to perform the screening on their own children, which would facilitate the process of preparing for referrals or diagnostic testing.

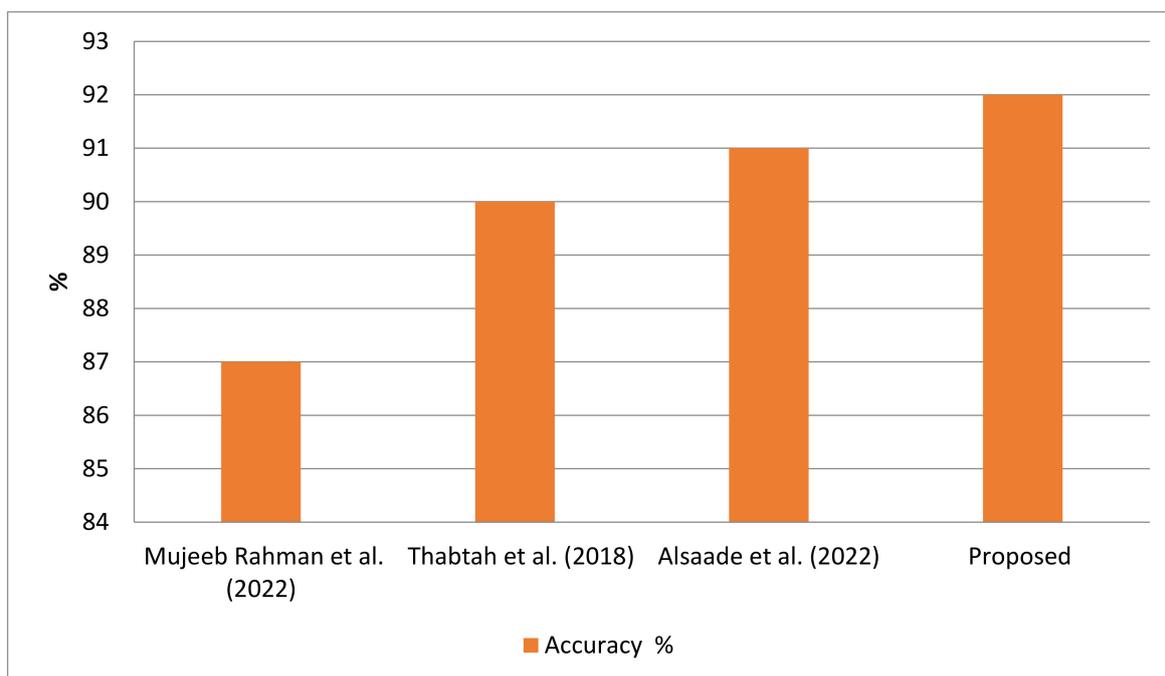


Figure 19. Comparison of the proposed and existing systems. Histograms represent references [10,24,46] (from left to right).

Our approach comes with a number of disadvantages that need to be considered. For instance, a significant proportion of both positive and negative results were erroneous. An incorrect diagnosis caused by a false-positive outcome might result in unnecessary medical procedures and increased anxiety for the patient's parents. The contrary is also true; delays in the diagnosis and treatment of a child might be caused by a false-negative result. Because of this, it is important to bear in mind that the suggested method is not capable of identifying ASD by itself; rather, the results need to be accurate from a clinical standpoint.

The recommended approach would have a favorable influence on ASD diagnoses due to its simplicity, speed, and accuracy in comparison to the parent-administered screening techniques used to date for early ASD screening. This is because the suggested method would be able to detect ASD earlier.

6. Conclusions

The early diagnosis of ASD in children has been shown to have significant positive effects on the diagnosed child's long-term health outcomes. All the detection methods used at present rely on the judgement of professionals, despite the fact that this approach is both subjective and costly. In this study, we recommended using a deep learning system that integrates several facial landmark features in order to identify children who have ASD. The utilization of this system could reduce costs and increase the effectiveness of the detection process. To start, we devised a unique approach for recognizing the properties of facial landmarks. The goal of this study was to examine whether certain facial characteristics may be utilized as biomarkers to accurately and reliably distinguish autistic children from normally developing children.

In this study, we made use of a dataset that is available to the general public and includes images of the faces of both autistic and typically developing children. Pre-trained models for binary ASD classification were developed and assessed using logistic regression, LinearSVC, random forest, decision tree, gradient boosting, MLPClassifier, and K-nearest neighbors methods. Hybrid VGG-16 models employing these and other machine learning methods were also constructed. When compared to the other models we examined, we discovered that the MobileNet-V2 model had the best accuracy (92%). The results suggest

that still images of children's faces might be used to rapidly gather diagnostic indicators of ASD, thus enabling an ASD screening approach that is both rapid and accurate.

Through the examination of images of children's faces, this study addressed a pressing issue in the field of computer vision, which is the screening of individuals for ASDs. The result of this study lend credence to the observations made by clinicians who noted differences in the facial characteristics of children diagnosed with ASD and children who are developing typically. We believe that this computer vision solution will help to address the major causes of racial disparity in ASD diagnosis and screening methods. These major causes include subjective screening or diagnosis, difficulty in accessing professional medical services, and the financial obstacles that families face in numerous regions, and especially in less-developed countries. The development of an expert system for recognizing autism spectrum disorder in children based on facial landmarks is the key contribution for improving the healthcare system in Saudi Arabia with the aim of detecting autism spectrum disorder at an earlier stage. The performance accuracy of the proposed system is in need of improvement; therefore, advanced AI algorithms will be suggested in our future research.

Author Contributions: H.A., T.H.H.A. and M.Y.A.; methodology, T.H.H.A.; software, T.H.H.A.; validation, H.A. and T.H.H.A. formal analysis, H.A. and M.Y.A.; investigation, M.Y.A., T.H.H.A. and H.A.; resources, H.A., T.H.H.A. and M.Y.A.; data curation, T.H.H.A. and H.A.; writing—original draft preparation, M.Y.A. and H.A.; writing—review and editing, H.A.; visualization, T.H.H.A. and H.A.; supervision, T.H.H.A.; project administration, T.H.H.A. and H.A.; funding acquisition, T.H.H.A. and H.A. All authors have read and agreed to the published version of the manuscript.

Funding: The author extend their appreciation to the king Salman center for disability research for funding this work through research group No: KSGR-2022-013.

Data Availability Statement: <https://www.kaggle.com/datasets/cihan063/autism-image-data> Accessed Date 10 December 2022.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Lord, C.; Elsabbagh, M.; Baird, G.; Veenstra-Vanderweele, J. Autism Spectrum Disorder. *Lancet* **2018**, *392*, 508–520. [[CrossRef](#)] [[PubMed](#)]
2. Kojovic, N.; Natraj, S.; Mohanty, S.P.; Maillart, T.; Schaer, M. Using 2D Video-Based Pose Estimation for Automated Prediction of Autism Spectrum Disorders in Young Children. *Sci. Rep.* **2021**, *11*, 1–10. [[CrossRef](#)] [[PubMed](#)]
3. Georgoula, C.; Ferrin, M.; Pietraszczyk-Kedziora, B.; Hervas, A.; Marret, S.; Oliveira, G.; Rosier, A.; Crutel, V.; Besse, E.; Severo, C.A.; et al. A Phase III Study of Bumetanide Oral Liquid Formulation for the Treatment of Children and Adolescents Aged between 7 and 17 Years with Autism Spectrum Disorder (SIGN 1 Trial): Participant Baseline Characteristics. *Child Psychiatry Hum. Dev.* **2022**, *8*, 1–13. [[CrossRef](#)] [[PubMed](#)]
4. Zuckerman, K.E.; Broder-Fingert, S.; Sheldrick, R.C. To Reduce the Average Age of Autism Diagnosis, Screen Preschoolers in Primary Care. *Autism* **2021**, *25*, 593–596. [[CrossRef](#)]
5. Goh, K.L.; Morris, S.; Rosalie, S.; Foster, C.; Falkmer, T.; Tan, T. Typically Developed Adults and Adults with Autism Spectrum Disorder Classification Using Centre of Pressure Measurements. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 844–848, ISSN 2379-190X.
6. Speaks, A. What is autism. Retrieved Novemb. **2011**, *17*, 2011. Available online: <https://www.autismspeaks.org/what-autism> (accessed on 20 January 2023).
7. Asghar, J.; Tabasam, M.; Althobaiti, M.M.; Adnan Ashour, A.; Aleid, M.A.; Ibrahim Khalaf, O.; Aldhyani, T.H.H. A Randomized Clinical Trial Comparing Two Treatment Strategies, Evaluating the Meaningfulness of HAM-D Rating Scale in Patients With Major Depressive Disorder. *Front. Psychiatry* **2022**, *13*, 873693. [[CrossRef](#)]
8. Guillon, Q.; Hadjikhani, N.; Baduel, S.; Rogé, B. Visual Social Attention in Autism Spectrum Disorder: Insights from Eye Tracking Studies. *Neurosci. Biobehav. Rev.* **2014**, *42*, 279–297. [[CrossRef](#)]
9. Kanhirakadavath, M.R.; Chandran, M.S.M. Investigation of Eye-Tracking Scan Path as a Biomarker for Autism Screening Using Machine Learning Algorithms. *Diagnostics* **2022**, *12*, 518. [[CrossRef](#)]
10. Mujeeb Rahman, K.K.; Subashini, M.M. Identification of Autism in Children Using Static Facial Features and Deep Neural Networks. *Brain Sci.* **2022**, *12*, 94. [[CrossRef](#)]
11. Aldhyani, T.H.H.; Verma, A.; Al-Adhaileh, M.H.; Koundal, D. Multi-Class Skin Lesion Classification Using a Lightweight Dynamic Kernel Deep-Learning-Based Convolutional Neural Network. *Diagnostics* **2022**, *12*, 2048. [[CrossRef](#)]

12. Akter, T.; Satu, M.S.; Barua, L.; Sathi, F.F.; Ali, M.H. Statistical Analysis of the Activation Area of Fusiform Gyrus of Human Brain to Explore Autism. *Int. J. Comput. Sci. Inf. Secur. IJCSIS* **2017**, *15*, 331–337.
13. Aldhyani, T.H.H.; Nair, R.; Alzain, E.; Alkahtani, H.; Koundal, D. Deep Learning Model for the Detection of Real Time Breast Cancer Images Using Improved Dilation-Based Method. *Diagnostics* **2022**, *12*, 2505. [[CrossRef](#)] [[PubMed](#)]
14. Schelinski, S.; Borowiak, K.; von Kriegstein, K. Temporal Voice Areas Exist in Autism Spectrum Disorder but Are Dysfunctional for Voice Identity Recognition. *Soc. Cogn. Affect. Neurosci.* **2016**, *11*, 1812–1822. [[CrossRef](#)] [[PubMed](#)]
15. Jiang, X.; Chen, Y.F. Facial Image Processing. In *Applied Pattern Recognition*; Bunke, H., Kandel, A., Last, M., Eds.; Studies in Computational Intelligence; Springer: Berlin/Heidelberg, Germany, 2008; pp. 29–48.
16. Garcia, C.; Ostermann, J.; Cootes, T. Facial Image Processing. *Eurasip J. Image Video Process.* **2008**, *2007*, 070872. [[CrossRef](#)]
17. Yolcu, G.; Oztel, I.; Kazan, S.; Oz, C.; Palaniappan, K.; Lever, T.E.; Bunyak, F. Deep Learning-Based Facial Expression Recognition for Monitoring Neurological Disorders. In Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, USA, 13–16 November 2017; pp. 1652–1657.
18. Yolcu, G.; Oztel, I.; Kazan, S.; Oz, C.; Palaniappan, K.; Lever, T.E.; Bunyak, F. Facial Expression Recognition for Monitoring Neurological Disorders Based on Convolutional Neural Network. *Multimed. Tools Appl.* **2019**, *78*, 31581–31603. [[CrossRef](#)]
19. Haque, M.I.U.; Valles, D. A Facial Expression Recognition Approach Using DCNN for Autistic Children to Identify Emotions. In Proceedings of the 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 1–3 November 2018; pp. 546–551.
20. Rudovic, O.; Utsumi, Y.; Lee, J.; Hernandez, J.; Ferrer, E.C.; Schuller, B.; Picard, R.W. CultureNet: A Deep Learning Approach for Engagement Intensity Estimation from Face Images of Children with Autism. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 339–346.
21. Satu, M.S.; Azad, M.S.; Haque, M.F.; Imtiaz, S.K.; Akter, T.; Barua, L.; Rashid, M.; Soron, T.R.; Al Mamun, K.A. Protoy: A Smart Phone Based Mobile Application to Detect Autism of Children in Bangladesh. In Proceedings of the 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 17–19 December 2019; pp. 1–6.
22. Akter, T.; Ali, M.H.; Khan, M.I.; Satu, M.S.; Moni, M.A. Machine Learning Model to Predict Autism Investigating Eye-Tracking Dataset. In Proceedings of the 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Online, 5–7 January 2021; pp. 383–387.
23. Akter, T.; Ali, M.H.; Khan, M.I.; Satu, M.S.; Uddin, M.; Alyami, S.A.; Ali, S.; Azad, A.; Moni, M.A. Improved Transfer-Learning-Based Facial Recognition Framework to Detect Autistic Children at an Early Stage. *Brain Sci.* **2021**, *11*, 734. [[CrossRef](#)]
24. Thabtah, F.; Kamalov, F.; Rajab, K. A New Computational Intelligence Approach to Detect Autistic Features for Autism Screening. *Int. J. Med. Inform.* **2018**, *117*, 112–124. [[CrossRef](#)]
25. Omar, K.S.; Mondal, P.; Khan, N.S.; Rizvi, M.R.K.; Islam, M.N. A Machine Learning Approach to Predict Autism Spectrum Disorder. In Proceedings of the 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Bazar, Bangladesh, 7–9 February 2019; pp. 1–6.
26. Sharma, M. Improved Autistic Spectrum Disorder Estimation Using Cfs Subset with Greedy Stepwise Feature Selection Technique. *Int. J. Inf. Technol.* **2019**, *14*, 1251–1261. [[CrossRef](#)]
27. Satu, M.S.; Sathi, F.F.; Arifen, M.S.; Ali, M.H.; Moni, M.A. Early Detection of Autism by Extracting Features: A Case Study in Bangladesh. In Proceedings of the 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, 10–12 January 2021; pp. 400–405.
28. Erkan, U.; Thanh, D.N. Autism Spectrum Disorder Detection with Machine Learning Methods. *Curr. Psychiatry Res. Rev. Former. Curr. Psychiatry Rev.* **2019**, *15*, 297–308.
29. Thabtah, F.; Abdelhamid, N.; Peebles, D. A Machine Learning Autism Classification Based on Logistic Regression Analysis. *Health Inf. Sci. Syst.* **2019**, *7*, 12. [[CrossRef](#)]
30. Akter, T.; Satu, M.S.; Khan, M.I.; Ali, M.H.; Uddin, S.; Lio, P.; Quinn, J.M.; Moni, M.A. Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders. *IEEE Access* **2019**, *7*, 166509–166527. [[CrossRef](#)]
31. Hossain, M.D.; Kabir, M.A.; Anwar, A.; Islam, M.Z. Detecting Autism Spectrum Disorder using Machine Learning. *arXiv* **2020**, arXiv:2009.14499.
32. Yang, M.; Cao, M.; Chen, Y.; Chen, Y.; Fan, G.; Li, C.; Wang, J.; Liu, T. Large-Scale Brain Functional Network Integration for Discrimination of Autism Using a 3-D Deep Learning Model. *Front. Hum. Neurosci.* **2021**, *15*, 277. [[CrossRef](#)] [[PubMed](#)]
33. Ahsan, M.M.; Alam, E.T.; Trafalis, T.; Huebner, P. Deep Mlp-Cnn Model Using Mixed-Data to Distinguish between COVID-19 and Non-COVID-19 Patients. *Symmetry* **2020**, *12*, 1526. [[CrossRef](#)]
34. Ahsan, M.M.; Gupta, K.D.; Islam, M.M.; Sen, S.; Rahman, M.; Shakhawat Hossain, M. COVID-19 Symptoms Detection Based on Nasnetmobile with Explainable AI Using Various Imaging Modalities. *Mach. Learn. Knowl. Extr.* **2020**, *2*, 490–504. [[CrossRef](#)]
35. Ahsan, M.M.; Ahad, M.T.; Soma, F.A.; Paul, S.; Chowdhury, A.; Luna, S.A.; Yazdan, M.M.S.; Rahman, A.; Siddique, Z.; Huebner, P. Detecting SARS-CoV-2 from Chest X-Ray Using Artificial Intelligence. *IEEE Access* **2021**, *9*, 35501–35513. [[CrossRef](#)] [[PubMed](#)]
36. Ahsan, M.M.; Mahmud, M.; Saha, P.K.; Gupta, K.D.; Siddique, Z. Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies* **2021**, *9*, 52. [[CrossRef](#)]
37. Ahsan, M.M.; Nazim, R.; Siddique, Z.; Huebner, P. Detection of COVID-19 Patients from CT Scan and Chest X-Ray Data Using Modified MobileNetV2 and LIME. *Healthcare* **2021**, *9*, 1099. [[CrossRef](#)]

38. Ahsan, M.M.; Luna, S.A.; Siddique, Z. Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. *Healthcare* **2022**, *10*, 541. [[CrossRef](#)]
39. Ahsan, M.M.; Siddique, Z. Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review. *Artif. Intell. Med.* **2022**, *128*, 102289. [[CrossRef](#)]
40. Haque, R.; Islam, N.; Islam, M.; Ahsan, M.M. A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning. *Technologies* **2022**, *10*, 57. [[CrossRef](#)]
41. Pham, H.; Jones, M.A.; Gai, T.; Islam, W.; Danala, G.; Jo, J.; Zheng, B. Identifying an optimal machine learning generated image marker to predict survival of gastric cancer patients. In Proceedings of the Medical Imaging 2022: Computer-Aided Diagnosis, San Diego, CA, USA, 4 April 2022; Volume 12033, pp. 408–414.
42. Mishra, P.; Passos, D. Realizing transfer learning for updating deep learning models of spectral data to be used in new scenarios. *Chemom. Intell. Lab. Syst.* **2021**, *212*, 104283. [[CrossRef](#)]
43. Khosla, Y.; Ramachandra, P.; Chaitra, N. Detection of autistic individuals using facial images and deep learning. In Proceedings of the 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 16–18 December 2021; pp. 1–5.
44. Krithika, L.B.; Priya, G.G.L. MAFONN-EP: A Minimal Angular Feature Oriented Neural Network based Emotion Prediction system in image processing. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 1320–1329. [[CrossRef](#)]
45. Mandell, D.S. Psychiatric Hospitalization among Children with Autism Spectrum Disorders. *J. Autism Dev. Disord.* **2008**, *38*, 1059–1065. [[CrossRef](#)]
46. Alsaade, F.W.; Alzahrani, M.S. Classification and Detection of Autism Spectrum Disorder Based on Deep Learning Algorithms. *Comput. Intell. Neurosci.* **2022**, *2022*, 8709145. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.