


Article

Empirical Study of Autism Spectrum Disorder Diagnosis Using Facial Images by Improved Transfer Learning Approach

Md Shafiu Alam¹, Muhammad Mahbubur Rashid¹, Rupal Roy¹, Ahmed Rimaz Faizabadi¹ ,
Kishor Datta Gupta²  and Md Manjurul Ahsan^{3,*} 

¹ Department of Mechatronics Engineering, International Islamic University Malaysia, Kula Lumpur 43200, Malaysia

² Computer and Information Science, Clark Atlanta University, Atlanta, GA 30314, USA

³ School of Industrial and Systems Engineering, University of Oklahoma, Norman, OK 73019, USA

* Correspondence: ahsan@ou.edu

Abstract: Autism spectrum disorder (ASD) is a neurological illness characterized by deficits in cognition, physical activities, and social skills. There is no specific medication to treat this illness; only early intervention can improve brain functionality. Since there is no medical test to identify ASD, a diagnosis might be challenging. In order to determine a diagnosis, doctors consider the child's behavior and developmental history. The human face can be used as a biomarker as it is one of the potential reflections of the brain and thus can be used as a simple and handy tool for early diagnosis. This study uses several deep convolutional neural network (CNN)-based transfer learning approaches to detect autistic children using the facial image. An empirical study is conducted to select the best optimizer and set of hyperparameters to achieve better prediction accuracy using the CNN model. After training and validating with the optimized setting, the modified Xception model demonstrates the best performance by achieving an accuracy of 95% on the test set, whereas the VGG19, ResNet50V2, MobileNetV2, and EfficientNetB0 achieved 86.5%, 94%, 92%, and 85.8% accuracy, respectively. Our preliminary computational results demonstrate that our transfer learning approaches outperformed existing methods. Our modified model can be employed to assist doctors and practitioners in validating their initial screening to detect children with ASD disease.

Keywords: deep learning; convolutional neural network (CNN); ASD diagnosis; facial image; transfer learning



Citation: Alam, M.S.; Rashid, M.M.; Roy, R.; Faizabadi, A.R.; Gupta, K.D.; Ahsan, M.M. Empirical Study of Autism Spectrum Disorder Diagnosis Using Facial Images by Improved Transfer Learning Approach. *Bioengineering* **2022**, *9*, 710. <https://doi.org/10.3390/bioengineering9110710>

Academic Editor: Luca Mesin

Received: 21 September 2022

Accepted: 14 November 2022

Published: 18 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 complicated condition that interferes with a person's day-to-day communication [1]. The autistic person mostly experiences minor disabilities but sometimes requires special care. ASD patients mostly have communication issues; thus, they cannot express themselves through words, gestures, or facial expressions while interacting with others. Although medical experts often detect ASD patients based on the neurophysiological signs caused by ASD, there is no certain biosignature or pathological procedure that can identify autism at any time [2]. Despite a lack of proper treatment, an early diagnosis might provide some opportunity to improve the individual's lifestyle [3]. Due to the flexibility in brain development, an early diagnosis might help children with ASD symptoms to improve their social life. There is also research supporting that the children who were intervened before the age of two achieved better IQ scores than those who got their medical attention after four years of age [4]. A recent study shows that no more than 30% of ASD children are detected while they are over the age of three [5].

ASD is an ailment that affects various parts of our brain. ASD results from polymorphism, which is the genetic influence caused by human gene interaction [6]. According to the World Health Organization (WHO) report, about 1 in 100 children have ASD. The percentage is the highest in the USA, where approximately 1 in 44 children has ASD, and the

ratio is 4 times higher in boys than girls, according to the Centers for Disease Control and Prevention (CDC) in 2021 [7].

There is no specific treatment for autism spectrum disorder. However, to lessen the symptoms, improve cognitive capacity, improve daily living skills, and boost the capabilities of ASD patients, different intervention techniques have been well-thought-out by the experts [8]. By applying these intervention methods, a proper diagnosis of ASD should be made as early as possible. There are some known procedures to diagnose this autism spectrum disorder. The primary and conventional method used by experts is interview-based, where the condition of the patients is assessed by the different questionnaire protocols such as ADOS-2, ADI-R, CARS, Q-CHAT, and AQ-10 [9]. These methods are easy, effective, and lead to an accurate diagnosis. The main flaw of these methods is biasness, such as the physician's competence, skill, and timetable. In addition, the patient's parents or attendant cannot always give accurate data or fill out the questionnaire forms correctly. All these factors can influence the accuracy of interview-based ASD diagnosis.

Another method of diagnosis of ASD is from different modalities of neuroimaging data such as Magnetic resonance imaging (MRI), Electroencephalography (EEG), Electro-corticography (ECoG), Functional near-infrared spectroscopy (fNIRS), and Magnetoencephalography (MEG) [10]. However, such techniques are often not affordable to the people who live in economically depressed areas.

Although there are enough tools for the diagnosis of ASD patients, below are a few primary reasons for late detection:

- ASD is diagnosed mainly by interactive sessions, so it requires clinical experts to diagnose children near two years of age [11];
- It is difficult for the parents to visit the specialists, and the availability of such physicians is much lower in rural communities or underdeveloped countries [12];
- Parents who are not familiar with and aware of ASD do not often consider the growth issues as their children's disease;
- In addition, children from racial and ethnic minority backgrounds who receive a primary screening are less likely to have further medical exams due to the high costs associated with the expensive equipment and skilled personnel required for these tests [13].

As a result, a simple-to-use tool is required for rapid primary screening without the involvement of experts or costly pathological tests. The method should be cost-effective, dependable, and time and resource-efficient. In this regard, detecting ASD from static facial images of the children via a user interface—such as a website or a mobile application—could be highly convenient. This procedure avoids unnecessary harm to little infants due to lengthy medical protocols and is free from human biasness and high costs. Thus, this study aims to demonstrate the method's feasibility and precision through appropriate dataset and accuracy scores. The face is an important human biomarker because the central nervous system receives and processes information from facial components directly. The ability to distinguish between different facial expressions is a fundamental feature that can lead to identifying brain asymmetry or neurodevelopmental disorders [14].

Detecting ASD based on facial expressions is a fairly new area of study, and researchers are currently conducting feasibility studies and developing the relevant algorithms. Due to the unique characteristics of each patient, facial recognition can be the most accurate method of diagnosis. A group of scientists from the University of Missouri found that children diagnosed with autism have some facial markers, such as a wide upper face, including wide-set eyes. Their faces are often seen with a shorter middle region, including the cheeks and nose, which differ from those of children without the disorder [15].

Diagnosing ASD using facial features is a rapidly growing field of research, owing to the social impact on developing countries. To ease the early detection of ASD, this method can be a milestone for the primary screening of the ASD or normal child. Recent studies demonstrate the potentiality of the deep neural network, particularly the application of CNN models in various disease diagnosis [16–25]. Due to its remarkable ability to learn by

automatically extracting the hidden features from a large volume of images, convolutional neural networks (CNNs) are the widely used feature extractors for object detection or image classification work. Although CNNs are incredibly efficient and accurate, training the models requires a significant amount of time and computational resources [18]. Thus, instead of beginning from scratch, it is more convenient to employ pre-trained models that have previously been developed using supercomputers and massive datasets. Transfer learning is a concept that involves using the weights and parameters of these pre-trained models to modify the final output according to the application of the desired tasks, which results in better classification or prediction accuracy [26].

A systematic literature review is conducted to identify the potential research in ASD domains that consider facial images. Figure 1 illustrates our systematic approach, which has been used to determine the referenced literature. The initial search is carried out using three popular datasets: Scopus, Web of Science, and IEEE Xplore. From the figure, it can be observed that minimal research has been conducted on ASD diagnosis using deep learning approaches. To identify additional potential research papers, we have also randomly searched and selected some of the recent studies from Google Scholar.

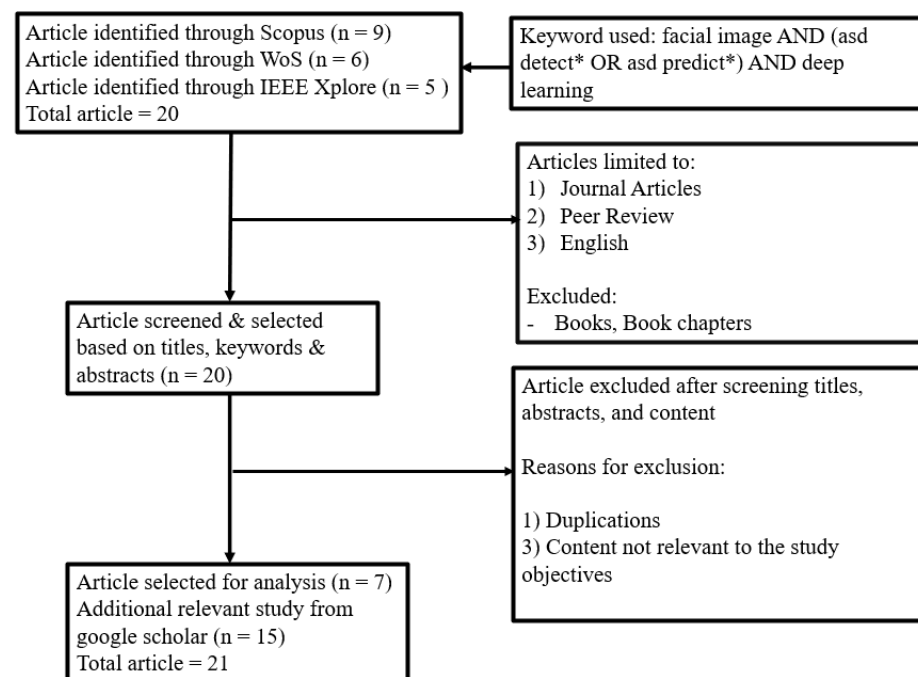


Figure 1. Flow diagram of the ASD article selection procedure for ablation study.

Some excellent progress has been made in screening ASD from facial images in recent times. Akter et al. (2021) introduced transfer learning models to identify the ASD faces with a 2D image dataset adopted from the Kaggle website. The authors considered both shallow and deep models for diagnosing Autism in young children ages 2 to 14 and achieved the highest accuracy with improved MobileNet-V1, applying the k-means clustering algorithm [27]. Hosseini et al. (2022) also used the MobileNet model to improve autism detection significantly [28]. The image features were extracted from the pre-trained deep learning models where three fully connected layers topped by a dense layer were used to predict ASD. However, to achieve higher accuracy, the author ignores the picture of young children from the datasets. As an effect, they were able to reduce the false positive and false negative rates, ultimately leading to an accuracy of around 95%. Rahman and Subashini (2021) later used the same dataset and focused their research on higher-layer deep learning models such as Xception and EfficientNetB, with a particular emphasis on the area under the curve (AUC) [29]. Yukti et al. (2021) used the MobileNet, InceptionV3, and InceptionResNetV2 models and cleaned the duplicates from the dataset using the

MD5 hash algorithm, although they have reported lower accuracy compared to earlier research [30]. Alsaade and Alzahrani (2022) trained the CNN-based models, Xception, VGG19, and NASNETMobile, using the same dataset and got the highest accuracy of 91% for Xception [31]. All of these CNN-based models used to extract characteristics from the images in the Kaggle autistic image dataset, which were trained extensively on the ImageNet dataset, contain 14 million images categorized into 1000 categories.

However, most of the proposed CNN models deal with a higher amount of hyperparameters, ultimately leading to higher computational time and are not often plausible for different sizes of datasets. In addition, concerns were raised about the performance of the existing models, as their performance validity on noisy datasets is frequently questioned, and their results are typically provided without sufficient statistical measurements. Therefore, there is a need to develop a CNN architecture that can be used to detect ASD with minimum hyperparameters, ultimately allowing the development of an efficient CNN-based ASD diagnosis model. Considering this opportunity, In this work, we use the 2D facial image and pre-trained deep learning models to diagnose ASD early. The transfer learning approach is used to extract the feature from the images, and we use the publicly available Kaggle dataset. The significant contributions of our work can be summarized as follows:

1. Pre-process the dataset for training after organizing and resizing the images;
2. Conduct the ablation study by tuning hyperparameters during training and validating the models' performance after each iteration; As a result, a comprehensive empirical study was introduced;
3. After determining the optimal set of hyperparameters, the optimizer for model training explains the facts behind the low accuracy with prediction probabilities;
4. Analyzing model performance to establish the research's future direction in terms of dataset pre-processing and imposing feature maps.

The rest of this paper is structured as follows: Section 2 discusses the facial image dataset and pre-trained deep learning models to detect ASD. In Section 3, an ablation study has been done for hyperparameter optimization to find the best ASD detection CNN model. In Section 4, the performance of the different models is compared with the findings of contemporary research. Finally, Section 5 concludes the paper with the contributions of this research and future work.

2. Materials and Methods

This study aims to predict Autism Spectrum Disorder (ASD) in children at an early age utilizing a transfer learning-based paradigm for autistic facial recognition. In this research, we used pre-trained deep learning models to extract robust features automatically, which were practically impossible to recognize by visual inspection due to their intricacy. We then fed these features through several layers, where the topmost dense layer resulted in the diagnosis of ASD.

2.1. Dataset

A large dataset is required to train deep learning models for optimal performance [17]. While training for all potential circumstances, the model will acquire significantly higher accuracy. We used the Kaggle repository's autistic children dataset [32] to build our suggested models, as it is the only free available dataset of this kind online. This dataset contains 2D RGB images of children aged from 2 to 14 years, where most fall between the ages of 2 and 8. The male-to-female ratio in the dataset was approximately 3:1, whereas the autistic class and normal control class had a nearly 1:1 ratio. The images are divided into 3 groups training set, testing set, and validation set consisting of 2536 (86.38%), 300 (10.22%), and 100 (3.41%) images, respectively, and each group has a ratio of 1:1 for ASD and NC class. The contributor, Gerry Piosenka, collected the images from the internet source, and there is no clinical history, ASD severity, ethnicity, or socioeconomic status of the children in the dataset. Many of the images are not of the best quality in terms of

facial alignment, brightness, or image size. To get an accurate and consistent outcome, the training set for the ML model should comprise a particularly comprehensive collection of photos that covers the entire spectrum of ASD, covering all these labelings of ASD severity, ethnicity, or socioeconomic status. Note that EEG or MEG data can also be utilized to determine the ASD diagnosis and comprehend the brain's functions [33,34]. However, gathering this information is hard and time-consuming, whereas obtaining facial images is relatively easier [35].

2.2. Transfer Learning for ASD Diagnosis

Recent research demonstrates the exceptional capabilities of deep convolutional neural networks (CNNs) for the classification of image data with an extremely high accuracy rating [19]. With the development of deep CNN-based models that have established themselves as a promising tool for facial recognition, computer vision is gaining popularity day by day [36,37]. Face recognition or object classification models work by extracting features from a specific object, making it possible to tell the difference between an autistic and non-autistic face by learning from a huge set of images. Transfer learning is a machine learning technique that applies a model developed for one task to another. The internal structure of the model remains the same; however, in this study, the models are utilized to extract distinct features from autistic and normal faces, and then the topmost layers are modified for classification purposes. The motivation of this research is to find a suitable framework to predict autism in children from facial images. To achieve this goal, we have designed and implemented the research as per the following flow

1. **Data attainment:** We adopted the image dataset from Kaggle data repository containing 2940 images of children aged 2–11 years. Then we divided the dataset into three subsets, train, test, and validation, having 2540, 300, and 100 images, respectively. The raw data were cleaned, labeled, and resized to give a reasonable shape as the input for the deep learning models;
2. **Select the transfer learning models:** The CNN-based models are chosen based on their demonstrated performance and accuracy in earlier studies. Additionally, we keep in mind that the model should be lightweight in terms of layers and parameters and their high accuracy;
3. **Ablation study for training parameters:** The models are trained multiple times using the same set of data, and their performance was evaluated using a variety of hyperparameters. Additionally, we used several optimizers and split ratios of the test-train data to obtain the high accuracy setting necessary for analyzing and evaluating the performance of various deep learning models in detail;
4. **Evaluation parameters:** Multiple metrics such as accuracy, the area under the curve (AUC), precision, and recall are used to verify the performance of various transfer learning approaches. We plot the ROC curve and confusion matrix for each transfer learning model at the optimal hyperparameter and optimizer settings;
5. **Choose the best performing model:** After evaluating the taken matrices and comparing the performance of different transfer learning approaches in terms of various statistical measures, the optimized model was selected for screening ASD among children;
6. **Analyze the model's performance:** To determine the effect of the prediction results on the test set, they were subsequently evaluated on various aspects containing a variety of different scenarios.

Figure 2 illustrates an overview of the proposed framework that was used to conduct the ablation study in this research work.

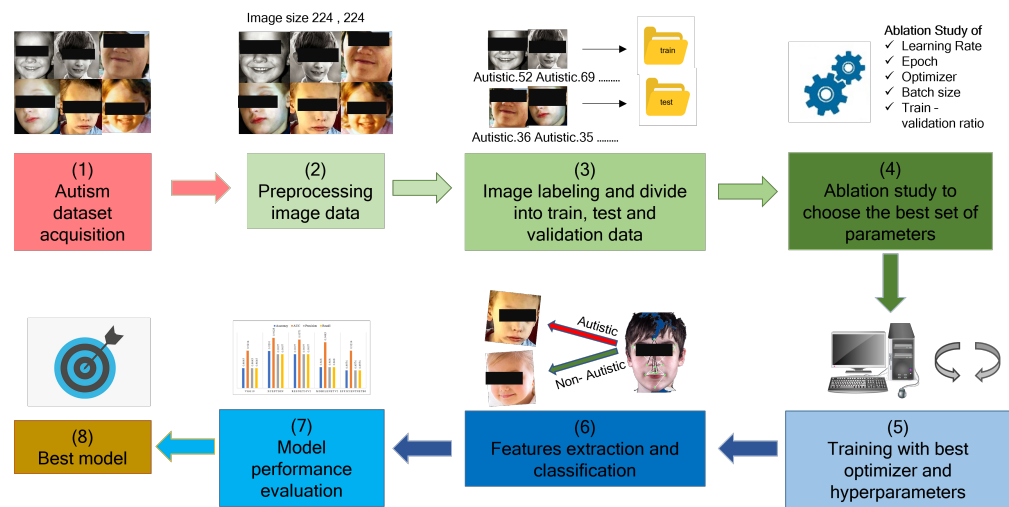


Figure 2. Proposed framework of the transfer learning approaches used in this study.

2.3. Transfer Learning Models for Feature Extraction

One of the key goals of this ablation study is to provide an overview of the algorithms used in previous research and compare their overall efficacy to detect differences or build superior approaches. This research relies on the existing peer-reviewed literature because there is no standard procedure or general criteria for selecting the best pre-trained algorithms. Therefore, the study is based on five pre-trained deep learning models based on CNNs: VGG19 [38], Xception [39], ResNet50V2 [40], MobileNetV2 [41], and EfficientNetB0 [42]. These models were selected due to their promising performance introduced by several referenced literature. Additionally, during the existing model modification, the layer of the CNN models was kept minimum in order to make it more feasible for mobile-based apps. Table 1 below states the accuracy of different transfer learning models’ size, accuracy, and depth. From the table, it can be observed that the Xception model demonstrates the highest accuracy in terms of selecting the top 1 to top 5 objects compared to the existing referenced models [43].

Table 1. Model accuracy, parameters, and depth.

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
VGG19	549	0.713	0.9	143.7 M	19
Xception	88	0.79	0.945	22.9 M	81
ResNet50V2	98	0.76	0.93	25.6 M	103
MobileNetV2	14	0.713	0.901	3.5 M	105
EfficientNetB0	29	0.771	0.933	5.3 M	132

2.3.1. VGG19 Model

VGG19 is a CNN-based image recognition architecture with extremely small (3 × 3) convolution filters, demonstrating that increasing the depth to 19 weighted layers improves prior art design. This model was developed for the 2014 ImageNet Challenge, where the developer Karen Simonyan and Andrew Zisserman won the contest due to localizing and classifying tasks, attaining state-of-the-art performance. They have made VGG19 models publicly available to allow further study on the use of deep visual representations in computer vision. The VGG19 requires a 224 × 224 RGB picture as input. The only pre-processing is performed here by subtracting the mean RGB value derived from the training set from each pixel. The image is processed through a stack of convolutional (Conv.) layers, each with a very narrow receptive field: 3 × 3 (the smallest size that adequately captures the concepts of left/right, up/down, and center). Five max-pooling layers follow several

of the conventional layers to perform spatial pooling. After stacking convolutional layers, whose depths can vary considering on the architecture, three fully connected (FC) layers are utilized: the first 2 layers contain 4096 channels each, and the 3rd layer comprises 1000 channels incorporated with a soft-max function that can be used to make the final prediction [38].

2.3.2. MobileNetV2 Model

MobileNetV2 is a cutting-edge module that utilizes inverted residuals and a linear bottleneck. This simple model can be extended for use in mobile phone applications. The model may accept a low-dimensional input, reduce the number of operations, and consume less memory while maintaining a greater level of accuracy. This model uses depth-wise separable convolution, which divides the convolution operation into two distinct layers. The first layer is depth-wise convolution, which is highly efficient because it performs just one filtering operation. The second layer acquires additional characteristics due to the linear calculation of the inputs. MobileNetV2 significantly lowered the computational cost of standard model layers by a factor of k^2 , thus often saving 8 or 9 times the computational resources required for 3×3 depth-wise separable convolution than other conventional models [41].

2.3.3. EfficientNetB0 Model

EfficientNet is a scaled-up model that optimizes both efficiency and accuracy. The EfficientNetB0's fundamental layer is a mobile inverted bottleneck MBConv with compound scaling of all three components—depth, width, and resolution—where α , β , and γ are the constants associated with depth, width, and resolution, respectively. These constants should be derived for the best outcome; larger models take more system resources. EfficientNetB0 may search for these scaling coefficients with less system capacity being a smaller model and then use the same for other heavy models [42].

2.3.4. ResNet50V2 Model

This model comprises several residual units that propagate both forward and backward utilizing identity mapping. Propagation can occur between blocks with a high degree of accuracy in terms of classification performance. Training will be significantly more accessible and more generalized with the assistance of these residual mappings. ResNet models are often over 100 layers deep and exhibit exceptional accuracy in ImageNet or COC competitions [40].

2.3.5. Xception Model

This model is based on Google's Inception model and has a straightforward modular structure. The model is based on three major blocks—entry, center, and exit—using a separable convolutional layer with Relu activation functions. The output of the convolution layer has been max-pooled incorporated with residual networks at the end of each compartment. To get the output, the input picture of size $299 \times 299 \times 3$ is passed through the entry flow, and then the output will consist of feature maps of $19 \times 19 \times 728$ at the end. Even after nine passes through the segment, the image's feature size remains constant at the middle flow's end. For a standard-sized input image, the output of the last component contains 2048 features. Lastly, the prediction layer gets the features through a fully connected (FC) layer. The final layers will be modified for binary classification to train and test the models on the ASD dataset. [39]. Figure 3 illustrates the architecture of modified Xception regarding the number of layers, weight, and height.

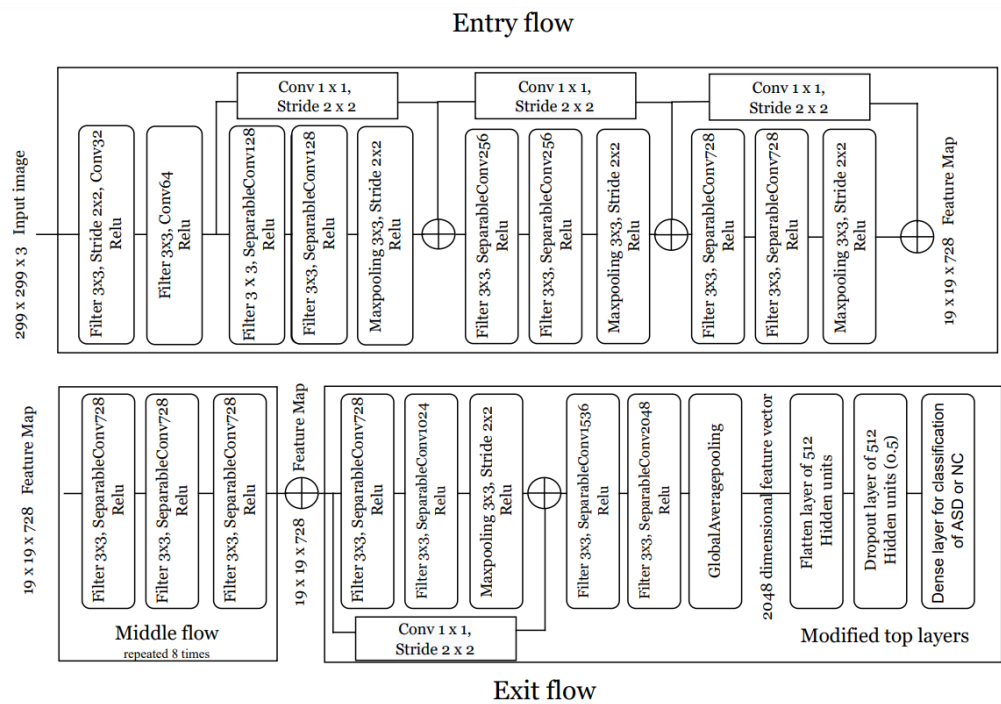


Figure 3. Modified Xception model structure (number of layers height, width, and filter size).

2.3.6. Classification Layer

After completing the final level of deep CNN models, facial features were extracted. The features were max-pooled and fed to the FC layer having 512 hidden units. In order to avoid overfitting, a drop-out layer weighted 0.5 was introduced at this stage. Finally, a dense layer is used to forecast the ASD classes after the investigation (refer to Figure 4).

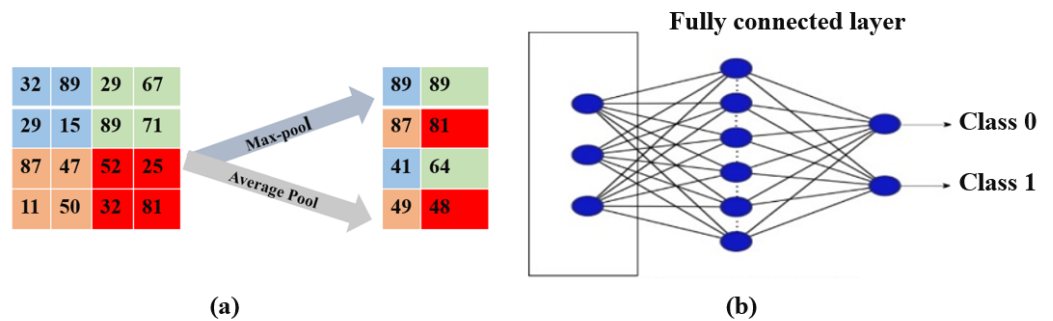


Figure 4. Demonstration of (a) max-pooling and average-pooling, and (b) classification layers.

2.4. Evaluation Matrices

We used some of the most widely used statistical measures such as accuracy, precision, and recall to represent our study findings and compare them with others. The mathematical equation of Accuracy, Precision, and Recall can be calculated using the following formula [18]:

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{1}$$

$$Precision = \frac{T_p}{T_p + F_p} \tag{2}$$

$$Recall = \frac{T_p}{T_n + F_p} \tag{3}$$

where,

- True Positive (T_p) = ASD children identified as ASD children;
- True Negative (T_n) = Healthy children identified as healthy;
- False Positive (F_p) = Healthy children identified as ASD children;
- False Negative (F_n) = ASD children identified as healthy.

3. Result

This section details the various stages of the ablation study we conducted to optimize the performance of ASD detection from static RGB facial image data. The codes were executed using the Kaggle platform and were written in the Python programming language. Kaggle is an incredible platform that enables a user to run the code on a dedicated GPU, provides the opportunity to share the user’s dataset, and publish the results. We conducted a series of ablation studies evaluating the accuracy and AUC to determine the optimal hyperparameter and optimizer combination. We completed the training using deep transfer learning models using the Keras API Library. Several data processing libraries such as matplotlib, sklearn, and pandas are used to analyze and visualize the models performance. To observe the CNN-based models’ performance with different optimizers, we have taken a fixed set of hyperparameters-epochs of 50, a learning rate of 0.001, and a batch size of 32. To select the best optimizer for these 5 models, we split the train image set into a 90–10 percent ratio for validation; such a split is more common in machine learning domains [17]. The Kaggle ASD dataset contains 2940 images, where 2540 were used for training, 300 for testing, and 100 for validation purposes. The image was labeled as ‘0’ for Normal control (NC) children and ‘1’ for ASD children while generating a data frame, as shown in Table 2.

Table 2. Characteristics of the Kaggle image datasets used during the study.

Dataset	Number	Class	Label
Training set	2540	Normal Control (NC)	NC-0
Testing set	300	Autistic (ASD)	ASD-1
Validation set	100		

Table 3 summarizes the comparative testing accuracy of the deep learning models following training and validation with a 90–10 split ratio for the different optimizers. The optimizer Adagrade, Adam, and Adamax were selected as many referenced literature exhibits better model performance using these three optimizers [17]. According to Table 3, the highest accuracy and AUC values are 86.61% and 91.74%, respectively. We obtained the best result with the Adagrade optimizer and an initial accumulator value of 0.01.

Table 3. Model performance for different optimizers.

Model	Adagrade		Adam		Adamax	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
VGG19	0.8169	0.8927	0.7857	0.8621	0.7991	0.8922
Xception	0.8571	0.9174	0.8080	0.8894	0.8303	0.8887
ResNet50V2	0.8661	0.8966	0.8169	0.8858	0.7991	0.8682
MobileNetV2	0.7991	0.8842	0.6875	0.7321	0.8258	0.8727
EfficientNetB0	0.7053	0.8143	0.6607	0.7465	0.4821	0.5075

Therefore, Adagrade was selected as the optimizer to train the model in subsequent experiments. During this time, we arbitrarily used the learning rate of 0.01, 0.001, and 0.0001 and obtained different performance results. The test results of the various learning rates are shown in Table 4. The learning rate of a model indicates how quickly it can learn features from a given dataset and is directly related to other hyperparameters such as epoch and

batch size. As illustrated in Table 4, accuracy and AUC are increased when the learning rate is set to 0.001.

Table 4. Model performance for different learning rates.

Model	Learning Rate					
	0.01		0.001		0.0001	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
VGG19	0.6875	0.7234	0.8169	0.8927	0.8438	0.8910
Xception	0.7901	0.8724	0.8571	0.9174	0.7991	0.8747
ResNet50V2	0.7857	0.8674	0.8661	0.8966	0.8259	0.8859
MobileNetV2	0.8571	0.8844	0.7991	0.8842	0.7813	0.8552
EfficientNetB0	0.6250	0.6712	0.7053	0.8143	0.7009	0.7824

The learning rate was set to 0.001 for the subsequent experiments based on the performance evaluation. We used Adagrade as an optimizer and trained the models for 50 epochs while splitting the training set into different percent ratios for validation. The ratio of 100% in Table 5 indicates that the entire training set of 2540 images is used for training, while the validation set of 100 images is used for validation. Following that, experiments were performed to validate the performance after segmenting the training dataset by percentage. Previously, we used a split of 90–10 percent, but as illustrated in Table 5, the model performs optimally when we use a training and validation set of 2540 and 100 images, respectively, for training and validation. The reason for this result is that by using a larger number of images for model training, the learning process is improved, resulting in higher accuracy and AUC values.

Table 5. Model performance for different train test split ratios.

Model Ratio	VGG19		Xception		ResNet50V2		MobileNetV2		EfficientNetB0	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
100%	0.8645	0.9214	0.9201	0.9625	0.9097	0.9571	0.8368	0.9230	0.8264	0.8990
90–10%	0.8169	0.8927	0.8571	0.9174	0.8661	0.8966	0.7991	0.8842	0.7054	0.8143
80–20%	0.8125	0.8604	0.8521	0.9161	0.8375	0.9143	0.8500	0.9132	0.7875	0.8686
70–30%	0.8021	0.8913	0.8247	0.8951	0.8207	0.8871	0.8057	0.8844	0.6957	0.7683
60–40%	0.8191	0.9229	0.8523	0.9414	0.8934	0.9083	0.8523	0.9146	0.8590	0.9253

The following two ablation studies determine the optimal batch size and epoch as described in Tables 6 and 7. While conducting the training, we encountered a few issues. Larger batch sizes require more system resources; during training for the Xception model, the system collapsed several times with a batch size of 64 images. The run time for a larger batch size is long when training the model. According to the Table 6, the optimal batch size is 32.

Table 6. Accuracy and AUC values for different batch sizes.

Batch	VGG19		Xception		ResNet50V2		MobileNetV2		EfficientNetB0	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
16	0.8446	0.8988	0.9132	0.9610	0.8750	0.9374	0.8819	0.9399	0.8541	0.9229
32	0.8645	0.9214	0.9201	0.9625	0.9097	0.9571	0.8368	0.923	0.8264	0.899
64	0.8672	0.9258	0.8945	0.9489	0.8789	0.9418	0.8281	0.9005	0.793	0.9075

As a result of the previous experiments, we now have all the parameters necessary to train the models and perform detailed metrics evaluations. The final component of our ablation is the epoch size, which is specified in Table 7. Initially, we employed a number of epochs and discovered that fewer than 50 and more than 100 led to overfitting and underfitting of models, respectively. For instance, the performance of all models is not identical, as when the iteration size is increased, some models, such as MobileNetV2 and

EfficientNetB0, perform better in terms of accuracy. While other models become overfitted after 50 epochs. Thus, we attempted another training for EfficientNetB0 that is more than 100 epochs, but it quickly overfitted after 100 epochs. If we want to consider the highest accuracy and optimize the system’s resources, 50 epochs are the optimal number.

Table 7. Model performance for epochs 50 and 100 (accuracy).

Epoch	VGG19	Xception	ResNet50V2	MobileNetV2	EfficientNetB0
50	0.8645	0.9201	0.9097	0.8368	0.8264
100	0.8646	0.8819	0.8750	0.8819	0.8646

We obtained the optimal set of model training parameters: the batch size of 32, the learning rate of 0.001, the optimizer set to Adagrade, and finally, categorical cross-entropy is considered as a loss function. For 50 epochs, we trained and validated the model, and the graphs for model accuracy are displayed in Figure 5.

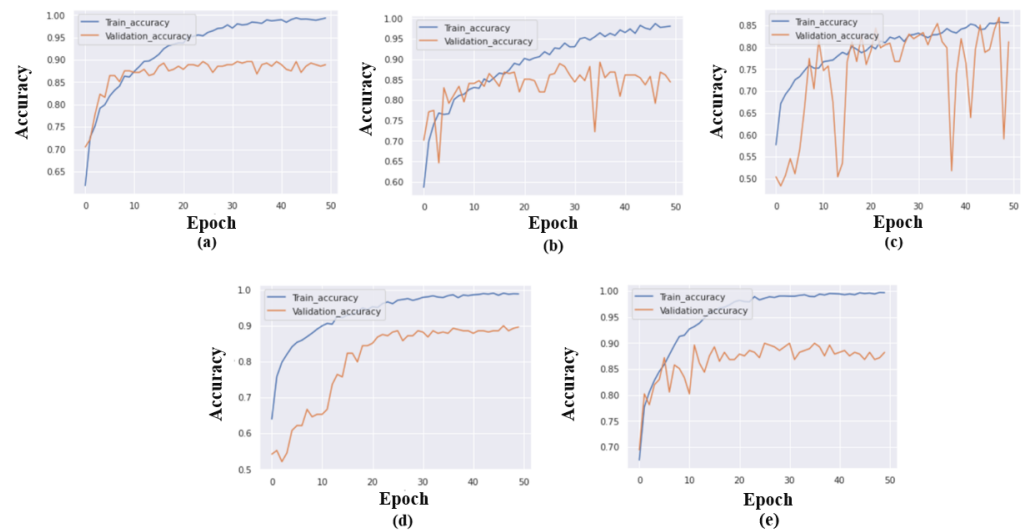


Figure 5. Plots of model accuracy of (a) Xception, (b) VGG19, (c) EfficientNetB0, (d) MobileNetV2, and (e) Resnet50V2 following each epoch applied to both training and validation set.

Figure 6 displays the models loss during each epochs.

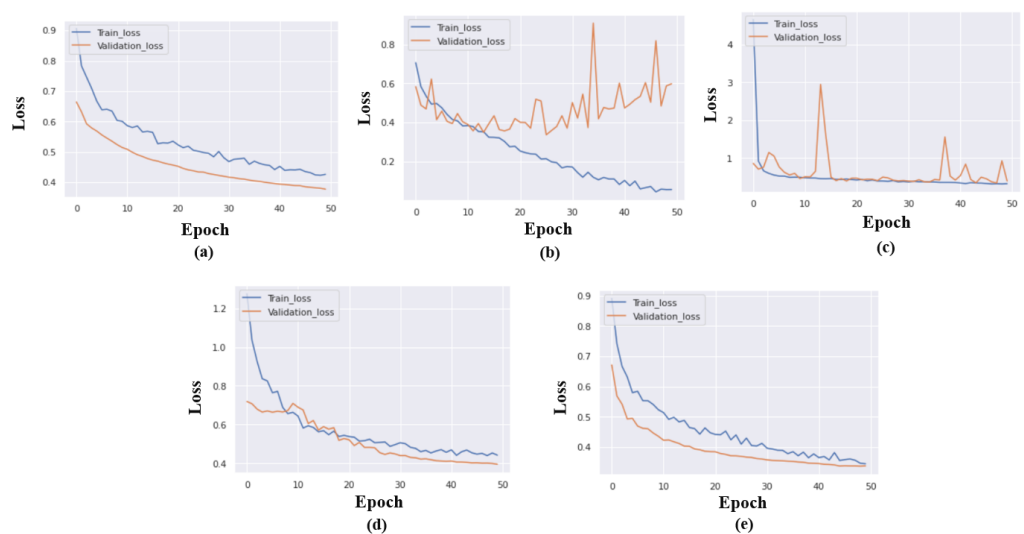


Figure 6. Plots of model loss of (a) Xception, (b) VGG19, (c) EfficientNetB0, (d) MobileNetV2, and (e) Resnet50V2 following each epoch applied to both the training and validation set.

The various metrics for each of the five models are listed in Table 8. This evaluation is based on the 300 test samples where the best value of accuracy is 92.01%. The highest AUC is 0.9625, which correctly describes the very high probability of detecting an autistic or normal child.

Table 8. Model performance for the best setting achieved from the ablation study.

Model	Accuracy	AUC	Precision	Recall
VGG19	0.8645	0.9214	0.8645	0.8645
Xception	0.9201	0.9625	0.9097	0.9097
ResNet50V2	0.9097	0.9571	0.9097	0.9097
MobileNetV2	0.868	0.9483	0.868	0.868
EfficientNetB0	0.8576	0.9214	0.8576	0.8576

The Xception model performs the best in terms of accuracy and AUC compared to the other four referenced transfer learning approaches. The ROC plot in Figure 7 clearly shows that the area under the curve is larger, implying that the prediction rate for various test samples is higher in the real-world scenario.

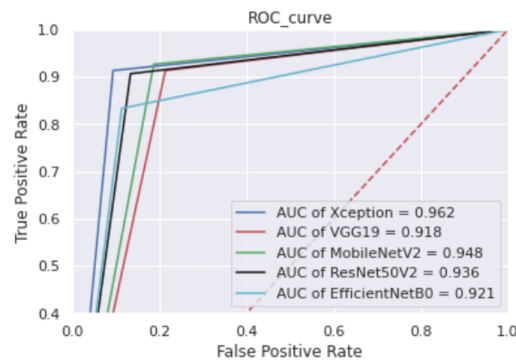


Figure 7. ROC curve of the models.

The confusion matrix in Figure 8 graphically depicts these models’ prediction performance following training and validation. Each model is evaluated based on a 300-image test set. The number in the brown box represents the images that were incorrectly predicted for each class. The overall performance of Xception is evident by the fact that the number of incorrect predictions is lower than that of other models. Only 12 images from each class are incorrectly predicted, which is a small number in comparison to other models. Overall, 24 images from both classes are incorrectly classified as the opposite class, as illustrated in Figure 8. The wrongly predicted images are shown in Figure 9.

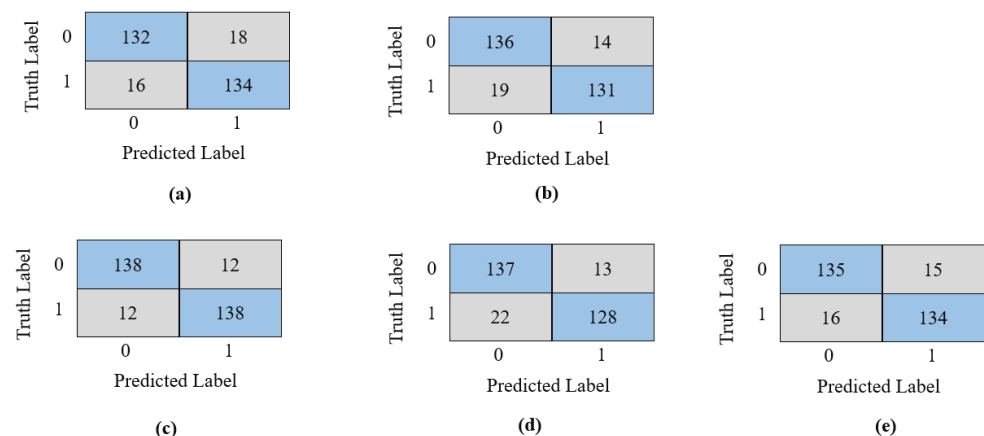


Figure 8. Confusion matrix of the five models (a) VGG19, (b) EfficientNetB0, (c) Xception, (d) MobileNetV2, and (e) ResNet50V2.



Figure 9. The 12 images in the top row are of (a) autistic children, and those misclassified as normal and vice-versa for the images are shown in (b) the lower row, which belong to normal control child (prediction of Xception).

The most prevalent causes of this misprediction occasionally include poor image quality, the existence of strong facial expressions, and in the majority of cases, alignment issues. If the image quality is inadequate for RGB photos due to low brightness, small size, or blurriness, most of the elements that distinguish a picture from NC to ASD will be lacking. Similarly, when an image of an NC child has an intense facial expression, it tends to have the same facial traits as an ASD child, resulting in a misinterpretation. Finally, if the photo is not correctly aligned or comprises just one side of the face, essential features will be difficult to extract from the photos, resulting in incorrect prediction and decreased detection accuracy.

There is another version of the dataset containing 3014 images where Hosseini et al. (2021) considered only faces that are properly aligned and resized [28]. The details of the dataset are stated in Table 9. On the test data, the Xception model demonstrated the best performance by achieving the highest accuracy of 95%, while ResNet50V2 and MobileNetV2 obtained 94% and 92% accuracy, respectively.

Table 9. Assignment of the second dataset used for the ablation study.

Dataset	Number	Class	Label
Train	2654	Normal Control (NC)	NC-0
Test	280	Autistic (ASD)	ASD-1
Valid	160		

Table 10 shows the preliminary computational results on the cleaned dataset introduced by Hosseini et al. (2021) [28]. As can be seen from the table, it is clear that, the improved Xception model outperforms all other models across all measures.

Table 10. Model performance on the second test dataset, obtained from [28].

Model	Accuracy	AUC	Precision	Recall
Xception	0.95	0.98	0.95	0.95
ResNet50V2	0.94	0.96	0.94	0.94
MobileNetV2	0.92	0.96	0.92	0.92

The ROC plot in Figure 10 clearly shows that the area under the curve is larger, implying that the prediction rate for various test samples is higher in the real-world scenario.

In Figure 11 confusion matrices were shown to understand the overall performance of the prediction better. The figure shows that our modified Xception models showed the best performance by misclassifying only 14 images, and MobileNetV2 demonstrated the worst performance by misclassifying 23 images.

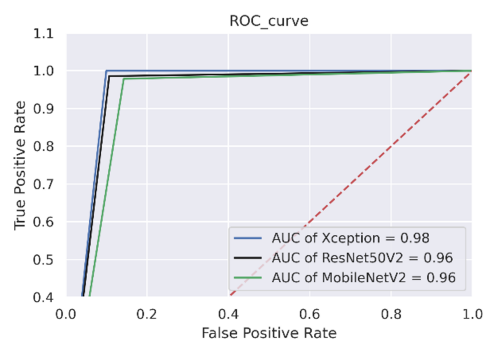


Figure 10. ROC curves for all models using the dataset introduced by Ref. [28].

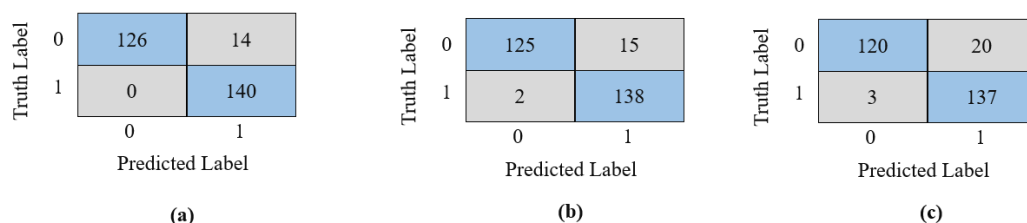


Figure 11. Confusion matrix of model (a) Xception (b) ResNet50V2 (c) MobileNetV2 on second test dataset.

4. Discussion

This study aims to conduct empirical research to detect ASD using an improved deep learning-based diagnosis tool from the facial images of the children. To detect autistic children, there are quite a few methods already that are serving in the current diagnosis process. The oldest and most accurate way to evaluate is the interview-based approach; however, the average detection trend in children takes more than three years to manifest. The importance of early detection stems from the fact that early intervention offers the best chance for ASD children to reclaim their regular lives. Thus, the justification for this type of research is obvious: to develop a simple and accurate detection approach that may be used at an early age. The most current development in new-age research is image processing, pattern recognition, and face recognition.

Additionally, because facial characteristics reflect the psychological qualities of the human brain, the facial image is an excellent candidate for ASD prediction. There are extremely few studies explicitly conducted in this field of study. To our knowledge, this is the first study to conduct a systematic ablation of various parameters and settings to achieve the highest accuracy of ASD detection from facial images. To improve the model’s predictive abilities, the bias in the training dataset is a significant bottleneck. While the dataset should contain all conceivable variations, the referenced Kaggle dataset is overpopulated with white children, posing identification challenges for black and other ethnicities.

Additionally, the quantity of photos is relatively small compared to any conventional dataset, while the visual quality is subpar in some cases. Additionally, we learned from the results that for accurate recognition, the facial expression should be neutral; otherwise, it creates confusion during prediction. The backgrounds of photos should be identical, and their alignment and brightness should be precise. Additionally, the RGB image is insufficient to extract facial features fully; instead, an image or video dataset of different modalities can significantly increase the accuracy.

Table 11 compares the findings and performance of some recent research. Table 11 exhibits the achieved accuracy from the recent research. Although computationally quite expansive, the Xception model performed the best in most research. MobileNet-V1 is another efficient model that demonstrated promising results. For instance, Tania et al. (2021) achieved a lower AUC value despite achieving the same 92% accuracy as our suggested model after adopting K means clustering using MobileNet-V1. Our Xception model in-

icates a prediction AUC of 96.26%, close to Rahman and Subashini’s (2022), but their reported accuracy was 90% only. Apart from these, our suggested model for ASD diagnosis outperforms the other referenced literature listed in Table 11.

Table 11. Performance comparison of the recent research.

Author	Dataset Used	Suggested Model	Accuracy
Khosla et al. (2021) [30]	Kaggle [32]	MobileNet	87%
Rahman and Subashini (2022) [29]	Kaggle [32]	Xception	90%
Akter et al. (2021) [27]	Kaggle [32]	MobileNet-V1	92%
Alsaade and Alzahrani (2022) [32]	Kaggle [32]	Xception	91%
Our study	Kaggle [32]	Xception	92%

For another version of the dataset containing 3014 images, Hosseini et al. (2022) asserted that they achieved a 94.6% accuracy using MobileNet [28]. Additionally, Rabbi et al. (2021) demonstrated an accuracy of 92.31% using their CNN-based models [44]; however, the process and supporting evidence are lacking. Additionally, Ahmed et al. (2020) have shown an accuracy of 95% on the same test dataset with the MobileNet model [45]. However, our proposed model outperformed the study conducted by Ref. [45] by achieving an accuracy of 95% accuracy and an AUC value of 0.98, which is higher than what they reported in the literature. Figure 12 shows the comparison of the different works plotting the accuracy performance.

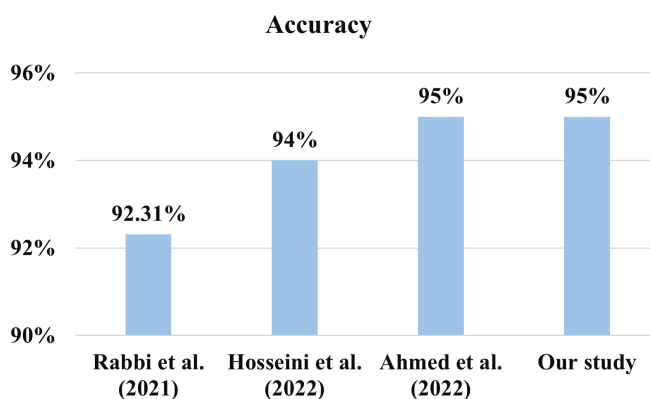


Figure 12. Graph illustrating the best accuracy on second dataset and compared with referenced literature [28,44,45].

Other researchers such Saranya and Anandan (2021) attained 92% accuracy using their CNN-based models. The work was on recognizing facial emotions in ASD patients for the aid of caregivers. They used the AFFECTNET [46] dataset to train their models, with Kaggle [31] serving only as a validation purpose. Angelina and Perkowski (2021) achieved a classification accuracy of 95% using the VGG16 model [47]. However, the accuracy was demonstrated in the East Asian dataset, which has 1154 photos and only 32 images from the Kaggle dataset. They also established that ethnic bias plays a crucial part in this diagnosis process and that a different ethnic dataset should be created for each demographic race. This idea is directly contradicted by the robustness of a particular model and could be troublesome during the model implementation.

5. Conclusions

This study aims to find the best transfer learning model for ASD classification. As an effect, We have conducted an empirical study to tune hyperparameters and optimizers for model training considering five existing and widely used CNN based models: VGG19, EfficientNetB0, Xception, MobileNetV2, and ResNet50V2. Our study reveals that VGG19 performs with 86.5% accuracy, ResNet50V2 with 94% accuracy, while MobileNetV2 and

EfficientNetB0 gives 92% and 85.76% prediction accuracy respectively. The modified Xception model demonstrates the best performance, with an accuracy of 95%, AUC of 98%, a precision of 95%, and recall values of 95%. The modified Xception models performance was further explored by visualizing using ROC curve, where more area coverage for Xception indicates a better level of prediction likelihood.

We also used a confusion matrix to evaluate each model's performance for both the positive and negative classes. Based on our observation, we found that poor image quality, the presence of extreme facial emotions, and, in the majority of cases, alignment problems are the most common causes of lower prediction rates. In the future, image augmentation may be used to help mitigate these issues. Obtaining ASD features from different modalities of data, such as thermal or 3D images, can shed new light on how to improve accuracy. The features collected by the models were not forced but instead chosen spontaneously by the model, allowing us to place attention blocks on certain regions containing discernible elements in the future. At the next level, we should also focus on the distinct actions and behavioral patterns of autistic children that have been medically demonstrated by experts using videos and can ensemble results from various modality data. The proposed approach will provide insights to future researchers and practitioners who want to make ASD screening easier, faster, and less expensive. In addition, implementing the proposed model on mobile devices as one of the feasible solutions will be one of our primary concerns in future research.

Author Contributions: Conceptualization—M.S.A. and M.M.R.; methodology—M.S.A., M.M.R. and M.M.A.; software—M.S.A.; validation—R.R., A.R.F. and K.D.G.; formal analysis—M.M.A., R.R., A.R.F. and K.D.G.; investigation—M.M.A.; writing—original draft preparation—M.S.A., M.M.R. and M.M.A.; writing—review and editing—M.S.A., M.M.A., R.R. and K.D.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lord, C.; Elsabbagh, M.; Baird, G.; Veenstra-Vanderweele, J. Autism spectrum disorder. *Lancet* **2018**, *392*, 508–520. [[CrossRef](#)]
2. Petrina, N.; Carter, M.; Stephenson, J. Recent developments in understanding friendship of children and adolescents with autism spectrum disorders. In *Encyclopedia of Autism Spectrum Disorders*; Springer: Cham, Switzerland, 2021; pp. 3875–3880.
3. 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](#)]
4. 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**, 1–13. [[CrossRef](#)]
5. 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](#)] [[PubMed](#)]
6. Hettinger, J.A.; Liu, X.; Hudson, M.L.; Lee, A.; Cohen, I.L.; Michaelis, R.C.; Schwartz, C.E.; Lewis, S.M.; Holden, J.J. DRD2 and PPP1R1B (DARPP-32) polymorphisms independently confer increased risk for autism spectrum disorders and additively predict affected status in male-only affected sib-pair families. *Behav. Brain Funct.* **2012**, *8*, 19. [[CrossRef](#)]
7. Cardinal, D.N.; Griffiths, A.J.; Maupin, Z.D.; Fraumeni-McBride, J. An investigation of increased rates of autism in US public schools. *Psychol. Sch.* **2021**, *58*, 124–140. [[CrossRef](#)]
8. Coury, D. Medical treatment of autism spectrum disorders. *Curr. Opin. Neurol.* **2010**, *23*, 131–136. [[CrossRef](#)]
9. Khodatars, M.; Shoeibi, A.; Sadeghi, D.; Ghaasemi, N.; Jafari, M.; Moridian, P.; Khadem, A.; Alizadehsani, R.; Zare, A.; Kong, Y.; et al. Deep learning for neuroimaging-based diagnosis and rehabilitation of autism spectrum disorder: A review. *Comput. Biol. Med.* **2021**, *139*, 104949. [[CrossRef](#)]

10. Ghosh, T.; Al Banna, M.H.; Rahman, M.S.; Kaiser, M.S.; Mahmud, M.; Hosen, A.S.; Cho, G.H. Artificial intelligence and internet of things in screening and management of autism spectrum disorder. *Sustain. Cities Soc.* **2021**, *74*, 103189. [CrossRef]
11. Oosterling, I.J.; Wensing, M.; Swinkels, S.H.; Van Der Gaag, R.J.; Visser, J.C.; Woudenberg, T.; Minderaa, R.; Steenhuis, M.P.; Buitelaar, J.K. Advancing early detection of autism spectrum disorder by applying an integrated two-stage screening approach. *J. Child Psychol. Psychiatry* **2010**, *51*, 250–258. [CrossRef]
12. Rossignol, D.A.; Frye, R.E. Melatonin in autism spectrum disorders: A systematic review and meta-analysis. *Dev. Med. Child Neurol.* **2011**, *53*, 783–792. [CrossRef] [PubMed]
13. Tait, K.; Fung, F.; Hu, A.; Sweller, N.; Wang, W. Understanding Hong Kong Chinese families' experiences of an autism/ASD diagnosis. *J. Autism Dev. Disord.* **2016**, *46*, 1164–1183. [CrossRef] [PubMed]
14. LaMantia, A.S. Why does the face predict the brain? Neural crest induction, craniofacial morphogenesis, and neural circuit development. *Front. Physiol.* **2020**, *11*, 610970. [CrossRef]
15. Aldrige, D.K. Is It Autism? Facial Features That Show Disorder. Available online: <https://www.cbsnews.com/pictures/is-it-autism-facial-features-that-show-disorder/> (accessed on 3 March 2022).
16. 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]
17. Ahsan, M.M.; E Alam, 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]
18. 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]
19. 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]
20. 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]
21. 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]
22. Ahsan, M.M.; Luna, S.A.; Siddique, Z. Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. *Healthcare* **2022**, *10*, 541. [CrossRef]
23. Ahsan, M.M.; Siddique, Z. Machine learning-based heart disease diagnosis: A systematic literature review. *Artif. Intell. Med.* **2022**, *128*, 102289. [CrossRef] [PubMed]
24. 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]
25. 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.
26. 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]
27. Akter, T.; Ali, M.H.; Khan, M.; Satu, M.; 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]
28. Hosseini, M.P.; Beary, M.; Hadsell, A.; Messersmith, R.; Soltanian-Zadeh, H. Deep Learning for Autism Diagnosis and Facial Analysis in Children. *Front. Comput. Neurosci.* **2021**, *15*, 789998. [CrossRef]
29. Mujeeb Rahman, K.; Subashini, M.M. Identification of Autism in Children Using Static Facial Features and Deep Neural Networks. *Brain Sci.* **2022**, *12*, 94. [CrossRef]
30. 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.
31. 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]
32. Piosenka, G. Detect Autism from a Facial Image. Available online: <https://cutt.ly/ibIXt5a> (accessed on 20 September 2022).
33. Brihadiswaran, G.; Haputhanthri, D.; Gunathilaka, S.; Meedeniya, D.; Jayarathna, S. EEG-based processing and classification methodologies for autism spectrum disorder: A review. *J. Comput. Sci.* **2019**, *15*, 1161–1183. [CrossRef]
34. Roberts, T.P.; Kuschner, E.S.; Edgar, J.C. Biomarkers for autism spectrum disorder: Opportunities for magnetoencephalography (MEG). *J. Neurodev. Disord.* **2021**, *13*, 34. [CrossRef]
35. Ewald, A.; Marzetti, L.; Zappasodi, F.; Meinecke, F.C.; Nolte, G. Estimating true brain connectivity from EEG/MEG data invariant to linear and static transformations in sensor space. *NeuroImage* **2012**, *60*, 476–488. [CrossRef]
36. Ahsan, M.M.; Li, Y.; Zhang, J.; Ahad, M.T.; Yazdan, M.M. Face recognition in an unconstrained and real-time environment using novel BMC-LBPH methods incorporates with DJI vision sensor. *J. Sens. Actuator Netw.* **2020**, *9*, 54. [CrossRef]
37. Ahsan, M.M.; Li, Y.; Zhang, J.; Ahad, M.T.; Gupta, K.D. Evaluating the Performance of Eigenface, Fisherface, and Local Binary Pattern Histogram-Based Facial Recognition Methods under Various Weather Conditions. *Technologies* **2021**, *9*, 31. [CrossRef]
38. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

39. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In *Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016*; Springer: Cham, Switzerland, 2016, pp. 630–645.
41. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
42. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning (PMLR), Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
43. Keras Application. Available online: <https://keras.io/api/applications/> (accessed on 7 March 2022).
44. Rabbi, M.F.; Hasan, S.M.; Champa, A.I.; Zaman, M.A. A Convolutional Neural Network Model for Early-Stage Detection of Autism Spectrum Disorder. In Proceedings of the 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021; pp. 110–114.
45. Ahmed, Z.A.; Aldhyani, T.H.; Jadhav, M.E.; Alzahrani, M.Y.; Alzahrani, M.E.; Althobaiti, M.M.; Alassery, F.; Alshaflut, A.; Alzahrani, N.M.; Al-Madani, A.M. Facial Features Detection System to Identify Children with Autism Spectrum Disorder: Deep Learning Models. *Comput. Math. Methods Med.* **2022**, *2022*, 3941049. [[CrossRef](#)]
46. Mollahosseini, A.; Hasani, B.; Mahoor, M.H. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Trans. Affect. Comput.* **2017**, *10*, 18–31. [[CrossRef](#)]
47. Lu, A.; Perkowski, M. Deep Learning Approach for Screening Autism Spectrum Disorder in Children with Facial Images and Analysis of Ethnoracial Factors in Model Development and Application. *Brain Sci.* **2021**, *11*, 1446. [[CrossRef](#)] [[PubMed](#)]