ARTICLE

Facial Image-Based Autism Detection: A Comparative Study of Deep Neural Network Classifiers

Tayyaba Farhat[1,2], Sheeraz Akram[3,*], Hatoon S. AlSagri[3], Zulfiqar Ali[4], Awais Ahmad[3] and Arfan Jaffar[1,2]

[1]Faculty of Computer Science and Information Technology, The Superior University, Lahore, 54600, Pakistan

[2]Intelligent Data Visual Computing Research (IDVCR), Faculty of Computer Science and Information Technology, The Superior University, Lahore, 54600, Pakistan

[3]Information Systems Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, 12571, Saudi Arabia

[4]School of Computer Science and Electronic Engineering (CSEE), University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK

*Corresponding Author: Sheeraz Akram. Email: sAkram@imamu.edu.sa

ABSTRACT

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by significant challenges in social interaction, communication, and repetitive behaviors. Timely and precise ASD detection is crucial, particularly in regions with limited diagnostic resources like Pakistan. This study aims to conduct an extensive comparative analysis of various machine learning classifiers for ASD detection using facial images to identify an accurate and cost-effective solution tailored to the local context. The research involves experimentation with VGG16 and MobileNet models, exploring different batch sizes, optimizers, and learning rate schedulers. In addition, the "Orange" machine learning tool is employed to evaluate classifier performance and automated image processing capabilities are utilized within the tool. The findings unequivocally establish VGG16 as the most effective classifier with a 5-fold cross-validation approach. Specifically, VGG16, with a batch size of 2 and the Adam optimizer, trained for 100 epochs, achieves a remarkable validation accuracy of 99% and a testing accuracy of 87%. Furthermore, the model achieves an F1 score of 88%, precision of 85%, and recall of 90% on test images. To validate the practical applicability of the VGG16 model with 5-fold cross-validation, the study conducts further testing on a dataset sourced from autism centers in Pakistan, resulting in an accuracy rate of 85%. This reaffirms the model's suitability for real-world ASD detection. This research offers valuable insights into classifier performance, emphasizing the potential of machine learning to deliver precise and accessible ASD diagnoses via facial image analysis.

KEYWORDS

Autism; Autism Spectrum Disorder (ASD); disease segmentation; features optimization; deep learning models; facial images; classification

1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by difficulties in social interaction, communication, and repetitive behaviors. ASD is a developmental condition that manifests in early childhood, making learning, speaking, expressing emotion, and socializing difficult [1–3]. ASD has risen significantly in recent years. However, due to a lack of research, childcare facilities, trained professionals for ASD, and cultural and contextual variables, the autism ratio could be higher than previously estimated [4]. ASD is characterized by two main symptoms: difficulties with social communication and a lower IQ, causing autistic individuals to be less accepted in society. Individuals with ASD have trouble maintaining eye contact, interpreting facial expressions, using gestures effectively, and comprehending others' perspectives. They may also struggle with responding to their name, sharing interests with others, and engaging in reciprocal conversation. Furthermore, they may display restricted or repetitive behaviors, such as lining up objects, repeating words or phrases, or fixating intensely on specific topics. Common self-stimulatory behaviors include hand-flapping or rocking, and individuals with ASD may have altered sensitivity to sensory stimuli, such as light, noise, touch, taste, or smell [5–7].

Early intervention and detection are crucial in helping autistic individuals, families, and caregivers. Detecting ASD in early childhood and assisting with social interactions and communication can help individuals become independent and socially accepted. Traditionally, ASD is detected using a checklist based on questions posed to parents or caregivers about the behavior and patterns of children. This method is performed several times to ensure the presence of autism. The timely detection of ASD Autism Spectrum is paramount to prevent severe difficulties in social interaction and daily life tasks. In addition to traditional checklist-based ASD detection, analyzing behavioral, genetic, neuroimaging, facial images, or speech data can assist in identifying associated patterns or features of ASD [8–12].

Unlike traditional algorithms, machine learning enables the development of an algorithm that can learn and improve over time based on data [13–17]. Conventional diagnostic methods often rely on subjective observations, which can hinder the accuracy and timeliness of diagnoses [18]. Facial features can reflect ASD, such as the eyes, nose, mouth, or ears, due to genetic or developmental factors. Individuals with ASD may exhibit distinctive features, such as a broader upper face, a shorter middle face, or wider-set eyes [19]. Machine learning techniques, like CNNs with transfer learning, can measure facial features and classify images into ASD or non-ASD groups [20,21]. Pakistan reported over 400,000 cases of autism in April 2021. However, social stigma and lack of awareness about the disorder may have led to underestimating the actual number [22]. In Pakistan, there is a limited availability of centers, experts, and instructors for ASD. Researchers are currently focusing on using machine learning algorithms to detect autism using facial images. However, the datasets used for this purpose are limited, and experimentation has been limited to typical CNN models. There is a need to experiment with different learning rate schedulers, optimizers, and batch sizes to explore the maximum potential of deep neural network models for autism detection.

Our study used deep neural networks to identify facial images. We conducted experiments with different learning rates, batch sizes, and network architectures using n-fold cross-validation and Orange. We also collected data from autism centers in Pakistan and collaborated with local centers to introduce ml-based autism detection methodologies. This paper consists of a literature review demonstrating the research conducted using machine learning for autism detection, including the classifiers used and their respective accuracies. The research methodology section explains the flow used in this study, and the experiment section shows the experimentation setting for all the experiments.

The result section provides a detailed analysis of the results, while the discussion section explains the work and evaluates all classifiers used. Lastly, we conclude the study in the conclusion section.

2 Literature Review

ASD is a neurodevelopmental disorder causing difficulties in social skills and communication, and individuals demonstrate repetitive behaviors [23]. Autism is a developmental disorder that can be assessed by monitoring verbal communication, social interaction, and nonverbal cues in early childhood [24–27]. Research on ASD covers a wide range of symptoms and dimensions. Fig. 1 summarizes the research areas covering a wide spectrum of research for ASD. ASD detection is crucial in research as early detection can reduce the effect of this neurodevelopmental disorder.
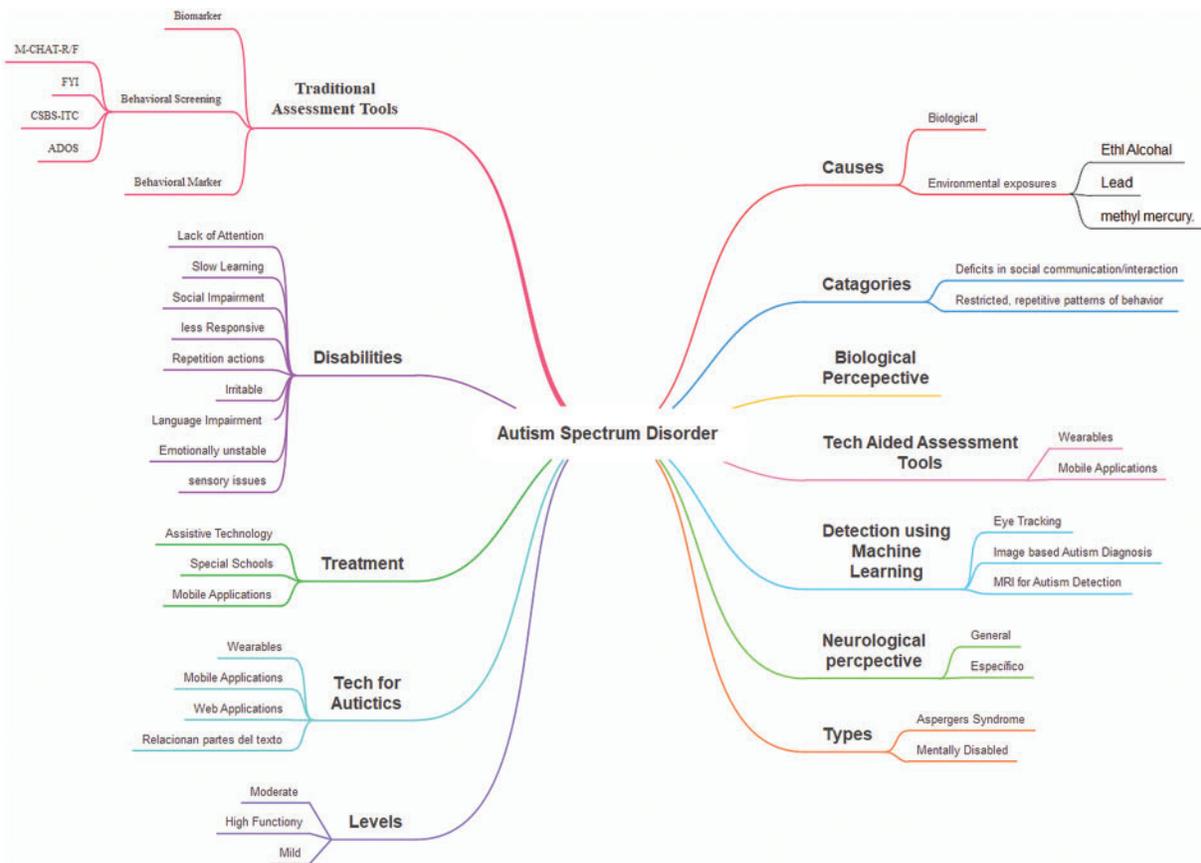

Figure 1: Research spectrum of ASD

The detection of ASD considered the most significant research dimension, can be achieved using traditional assessment tools and ML-based classifiers. Detection of Autism currently relies on interviews-based behavioral assessment tools like the Autism Diagnostic Observation Schedule (ADOS) [28] and the Autism Diagnostic Interview-Revised (ADIR) [29]. This traditional checklist-based assessment tool requires parents/caregivers to answer questions in the presence of an expert. These methods are time-consuming and costly. The assessments are conducted repeatedly with an interval of months to confirm ASD. However, these approaches do not identify the biological basis for developmental delays, as neuroanatomy is unclear [30,31]. Diagnosing autism through traditional

methods can be a lengthy and repetitive process. Traditional tools rely on behavioral assessments to analyze the child's behavior after age. Delayed diagnosis of autism can profoundly impact individuals, leading to lifelong effects. Compared to traditional ASD assessment methods, Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) approaches provide greater accuracy and earlier detection. Supervised learning is used by ML and DL classifiers to recognize patterns, features, and biomarkers associated with autism and detect them in new data. Lately, many researchers have explored deep neural networks as they can accurately identify ASD at an early age. Table 1 summarizes some of the studies on autism detection using facial images. Most researchers use facial image datasets, Magnetic Resonance Imaging (MRI), and checklist analysis for autism detection.

Table 1: Literature on machine and deep learning for autism detection

Paper	Limitations	Contribution	Classifiers	Test accuracy
[32]	–	A CNN system with 91% accuracy and 0.53.	Deep neural network (DNN) classifier	91%
[33]	Limited facial images labeled from online data.	Funding support for the research.	CNN	90%
[34]		A comprehensive review of existing supervised machine learning.	SVM (support vector machines)	80%
[35]	Small sample sizes in many of the studies	A systematic review of technological approaches for autism.	Symptoms pulled from standardized assessment protocols	78.90%
[36]	Many at-risk toddlers did not attend an evaluation even after repeated invitations.	We are assessing early ASD screening and repeat screenings from 12 months using longitudinal data.	Autism diagnostic observation schedule, toddler autism symptom inventory, autism diagnostic interview	The overall sensitivity of 71% and specificity of 95.9%
[37]	Web searching tasks are unsuitable for young children; a small number of participants.	Eye tracking for autism detection in infants, toddlers, and young.	Eye tracking data from visual processing tasks	Around 74%

The literature reviewed in the proposed study demonstrates that ML-based autism detection can detect ASD at an early age. Table 1 suggests limitations, contributions, classifiers, and accuracy achieved in the literature. We can interpret that all research accuracy achieved is insufficient, the dataset used is limited, and the methodology used is weak. Reference [38] applied ML algorithms to analyze Electroencephalogram (EEG) signals and accurately distinguish between individuals with ASD and typically developing individuals. The study demonstrated the potential of ML techniques in providing objective and efficient ASD detection based on neurophysiology.

Reference [39] applied models such as Support Vector Machines (SVM), Random Forest Classifier (RFC), Naïve Bayes (NB), Logistic Regression (LR), and K-Nearest Neighbors (KNN) to their dataset and constructed predictive models based on the outcome. The paper's main objective was determining if a child is susceptible to ASD in its nascent stages, which would help streamline the diagnosis process. Based on their results, Logistic Regression gave the highest accuracy for their selected dataset. Another study summarized recent progress in machine learning models for diagnosing ASD and Attention-Deficit/Hyperactivity Disorder (ADHD) using functional and structural MRI. They outlined and described machine learning, especially deep learning, techniques suitable for addressing research questions in this domain, pitfalls of the available methods, and future directions for the field [40].

Reference [41] proposed a web-based deep learning application to detect ASD using facial images. The authors of the study experimented using MobileNet, XceptionNet, and InceptionV3. Based on the validation data, it achieved an accuracy of 95% for MobileNet, 94% for Xception, and 0.8 for InceptionV3. The dataset contains 3,014 facial images of children, including 1,507 autistic children and 1,507 non-autistic children.

During this study, the literature review revealed a few gaps that must be addressed. These gaps include lower accuracy, limited and insufficient datasets, experimentation on deep learning classifiers without changing hyperparameters, and testing on datasets that lack diversity. To overcome these gaps, the study conducted experiments using various classifiers and collected a local dataset to verify the best-performing classifier's authenticity.

3 Proposed Methodology

This section will discuss the approach used to develop a model that deals with the challenges of detecting ASD. In the proposed study, we preprocessed the collected data and experimented with VGG16 using multiple optimizers, learning rate schedulers, and batch sizes. Finally, we performed testing using the best-performing model on a locally collected dataset of facial images. Fig. 2 visually represents our proposed DL-based ASD detection system. The acquired data undergoes an initial preprocessing phase, followed by division into training, validation, and test sets for the subsequent classification. In the experimentation section, we discussed the parameters investigated to optimize the choice of optimizer, learning rate scheduler, and batch size. Our analysis encompasses VGG16 MobileNet and employs automated ASD detection with the Orange tool. This study performs a comparative analysis of VGG16 and MobileNet by changing hyperparameters. Detailed information regarding the dataset, preprocessing steps, and experimentation is presented in Sections 3.1, 3.2, and 3.3, respectively.

3.1 Facial Image Dataset

We collected the facial image dataset from the public library Kaggle, which includes data for individuals with autism and those without autism, with ages ranging from 2 to 14. Table 2 describes the dataset; Fig. 3 presents images of autistic and non-autistic, respectively.

Our research analyzed the accuracy of facial recognition in identifying ASD. With a dataset of 2500 images lacking crucial information, we obtained 50 confidential images from nearby ASD centers to fill the gaps. We ensured the confidentiality of these images by not sharing them on social media or documenting them.
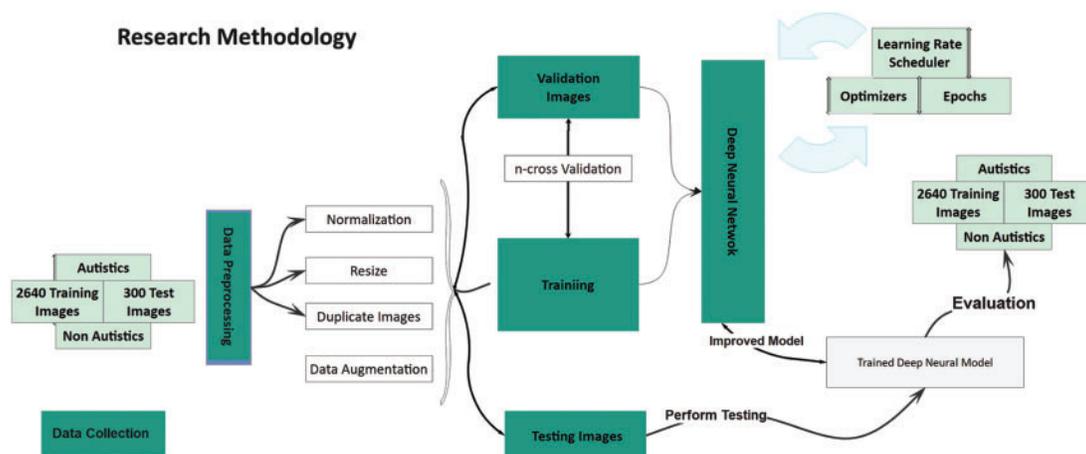
Figure 2: Research methodology

Table 2: Facial dataset description

	Autistic	Non-autistic	Total	Image dimension	Image type
Training	1250	1250	2500	2D image	RGB image
Validation	50	50	100	2D image	RGB image
Testing	150	150	300	2D image	RGB image
Local dataset for testing	25	25	50	2D image	RGB image
Image size			240×240		
File type			JPG		
Male-Female ratio			3-1		
Autistic to non-autistic ratio			1-1		
Age range			2–14 years		

(a)　　　　(b)
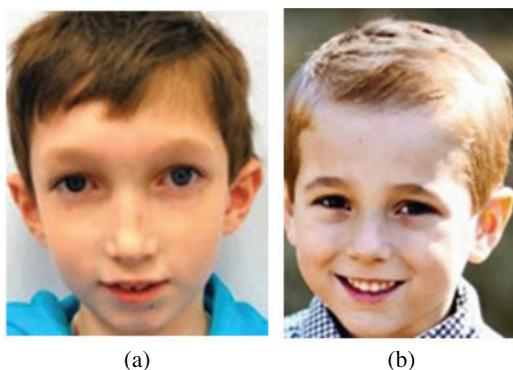
Figure 3: Images from dataset (a) autistics (b) non-autistic

3.2 Data Preprocessing

To effectively train a Convolutional Neural Network (CNN) model, a set of facial images was first cropped and divided into three distinct groups: training (consisting of 2,500 images), validation (comprised of 100 images), and testing (consisting of 150 images). We alternated between class images to ensure a balanced class representation. To effectively prepare the images for feature extraction, we utilized the Keras preprocessing Image Data Generator to normalize the pixel values by scaling them from a range of [0,255] to [0,1]. Feature extraction involves using deep neural networks that utilize input, hidden, and output layers, with weights and biases transformed by hidden layers and passed through activation functions to produce predictions. We used the Stochastic Gradient Descent (SGD) optimizer to optimize the algorithm and minimize error [42].

3.3 Experimentation

In this research, the experimentation phase fine-tunes our machine learning models by exploring optimizers, learning rate schedulers, and batch sizes. This ensures that our models deliver the best performance in autism detection within the given constraints. Optimizers offer different mechanisms for weight updates, learning rate schedulers control the evolution of the learning rate during training, and batch sizes affect both training efficiency and generalization capacity. Our approach strikes the right balance between accuracy, convergence speed, and computational efficiency.

This study utilized the computational capabilities of Google Colab Premium, including its powerful GPU and RAM, to conduct experiments on preprocessed datasets. We explored the performance of deep learning models, including VGG16 and MobileNet. The aim of experimenting with these deep learning models was to test the effectiveness of these models in classifying images of individuals with autism. The experiment analyzed classifiers' training and validation accuracy by varying the optimizer, learning rate scheduler, and batch size. The reason for choosing VGG16 and MobileNet is their high accuracy and precision in capturing facial features.

The VGG16 model has 16 layers, including 13 convolutional and three fully connected layers. It processes 224×224 pixel images and extracts features using convolutional filters and max pooling. The model minimizes cross-entropy loss using SGD during training to update weights and biases and produce predictions [43]. MobileNet is a deep learning architecture that relies on depth wise separable convolutions to reduce the number of parameters required compared to other neural networks. This approach results in more lightweight models that maintain high accuracy, making it suitable for mobile and embedded devices with limited computational resources [44]. According to previous research, to enhance the precision of the classifiers, we can conduct further experimentation and change optimizers and learning rate schedulers [45]. We aim to improve classifiers' efficiency by utilizing various experimentation methods, such as modifying the number of epochs, batch size, and learning rate scheduler and incorporating 5-fold cross-validation.

After preprocessing the images in the initial phase, this study conducted experiments with learning rate schedulers to fine-tune the hyperparameters in deep-learning models. A learning rate scheduler adjusts the learning rate during training based on a predetermined schedule to strike a balance between the speed of convergence and the final performance of the model. Common approaches include constant, step-based, exponential decay, time-based decay, and cyclical learning rates [46]. Eqs. (1) and (2) represent the learning rate of exponential decay and time-based decay, respectively. This study used exponential and time-based learning schedules with VGG16.

$$learning_{rate} = initial_{learning_{rate}} * \left(e^{\frac{-iteration}{decay_{rate}}} \right) \tag{1}$$

$$learning_{rate} = initial_{learning_{rate}} * \left(e^{\frac{-time}{decay_{rate}}} \right) \tag{2}$$

After experimenting with both learning rate schedulers on VGG16 and MobileNet using batch sizes 2, we found that exponential decay yielded good results on VGG16 and MobileNet with batch size 2 on 100 epochs. When training a deep learning model, optimizers adjust the weights and minimize the loss function. In this study, we utilized SGD, Adam, and Adagrad optimizers while training the VGG16 model with 100 epochs and a batch size of 2.

$$w = w - learning_{rate} * gradient$$

$$v = (momentum * v) - (learning_{rate} * gradient)$$

$$w = w + v$$

$$accumulated_{squared_{gradients}} = accumulated_{squared_{gradients}} + gradient^2$$

$$weights = W - \frac{learning_{rate} * gradient}{\sqrt{accumulated_{squared_{gradients}} + epsilon}} \tag{3}$$

Adagrad adjusts the learning rate for each parameter based on the sum of squares of past gradients but may become slow over time. Eq. (3) mathematically represents Adagrad in detail. Adam combines SGD and momentum to provide an efficient optimization method for deep learning models [45]. The Adam update rule can be expressed mathematically as in Eq. (4). Through experimentation; we discovered that using the Adam optimizer with a batch size of 2 over 100 epochs yielded remarkable performance on VGG16. Once we selected the optimal learning rate scheduler and optimizer, we focused on determining the best batch size for VGG16 and MobileNet. The results of our experiments, which identified classifiers that achieved notable accuracy, are presented in the results section.

$$m = \beta_1 * m + (1 - \beta_1) * gradient$$

$$v = \beta_2 * v + (1 - \beta_2) * gradient^2$$

$$weights = weights - \frac{learning_{rate} * m}{\sqrt{v} + epsilon} \tag{4}$$

As a free and open-source tool, Orange offers a visual programming interface that facilitates building ML workflows through blocks representing various tasks [46]. This study used the Orange tool to evaluate its built-in features. Visual Geometry Group 19 (VGG19) loaded and read the dataset, and then performed feature extraction. We trained three classifiers (KNN, RF, and Gradient Boost), and the evaluation included generating a Receiver Operating Characteristics (ROC) curve and accuracy matrix. Training data compromised 80% of the data for VGG19 with 200 epochs, and the testing used the remaining 20%. Orange's visualization module visualized the accuracy metric and ROC curve. Fig. 4 shows the architecture of the Orange tool for Autism detection.
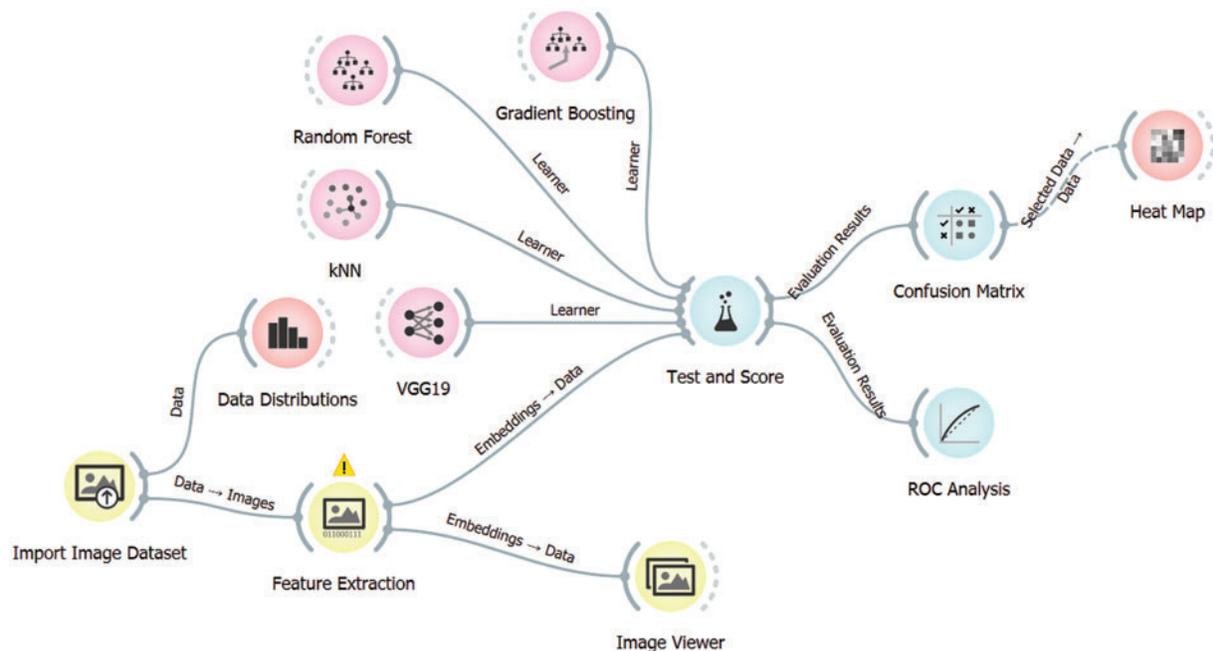
Figure 4: Architecture diagram for autism detection using orange

4 Results

This section will analyze the performance of all the classifiers used in our experiments. We will discuss the results of our experimentation and provide a detailed evaluation of the classifiers. We conducted experiments with various batch sizes, epochs, and techniques, including 5-fold cross-validation to enhance performance in facial image analysis for individuals with autism. Image preprocessing was made easier using data generators. The VGG16 architecture attained a validation accuracy of 98% and a test accuracy of 88% after training for 100 epochs with a batch size of 2. Fig. 5 displays the training and validation accuracy curves for VGG16. As depicted in Fig. 6, the accompanying confusion matrix reveals that the trained model exhibited 16 false positives and 19 false negatives out of 150 images of individuals with and without autism. ROC curve, Fig. 7 demonstrates that it has a sharp angle, which indicates a high True Positive Rate (TPR) for a relatively low False Positive Rate (FPR). The classifier seems to detect positive instances and avoid false positives accurately.

The VGG16 model was trained using 5-fold cross-validation with a batch size of 2 for 100 epochs. The model achieved an outstanding validation accuracy of 99%, while the testing accuracy was 87%. However, the validation accuracy curves (presented in Fig. 8) fluctuate, indicating a possible issue with the dataset. The confusion matrix (Fig. 9) and the ROC curve (Fig. 10) further confirm that the VGG16 model with 5-fold cross-validation obtained 99% and 87% test accuracy rates. The model's FPR was 15, and the FNR was 23 out of 150 images of individuals with and without autism.
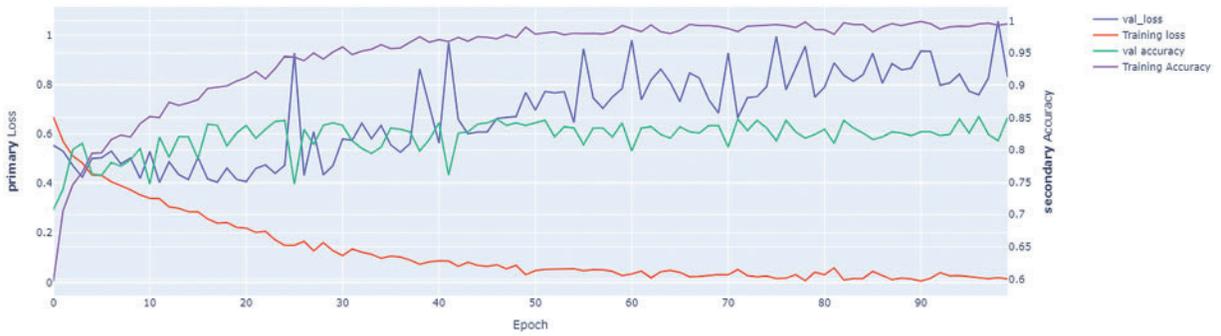
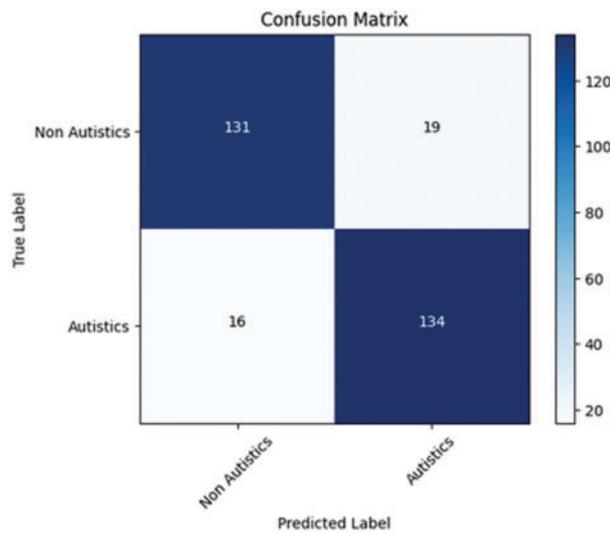**Figure 5:** Vgg-16 training/validation accuracy and training/validation loss
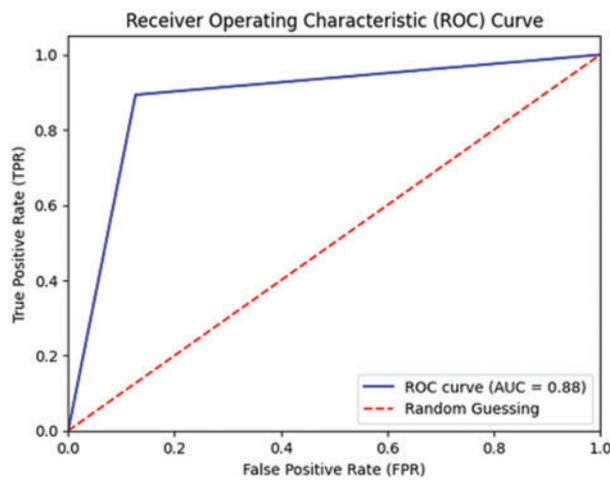

Figure 6: Vgg-16 confusion matrix
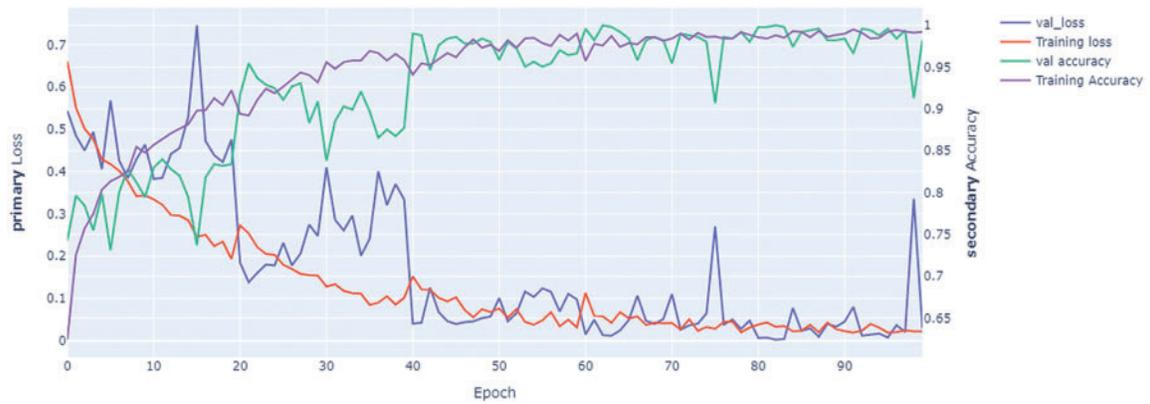

Figure 7: Vgg-16 roc curve

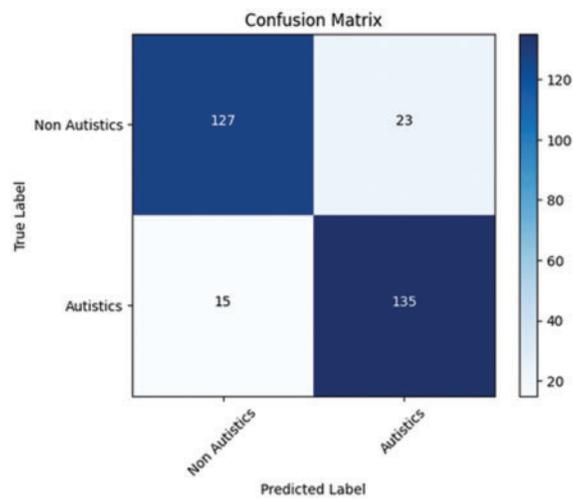**Figure 8:** Vgg-16 5-fold cross-validation complete training report

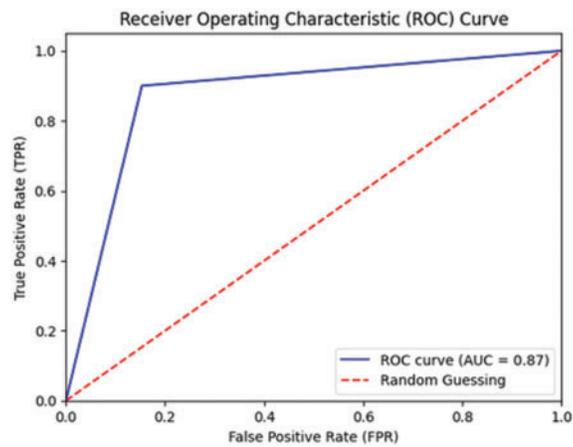**Figure 9:** Vgg-16 5-fold confusion matrix

Figure 10: Vgg-16 5-fold roc curve

Fig. 11 displays the training and validation accuracy of VGG16, which was trained with a batch size of 16. The model achieved a validation accuracy of 93% and a testing accuracy of 86%. The confusion matrix, shown in Fig. 12, reveals that the trained model had 13 false positives and 29 false negatives out of 150 images of individuals with and without autism. Furthermore, Fig. 13 presents the ROC curve for VGG16 with a batch size 16.
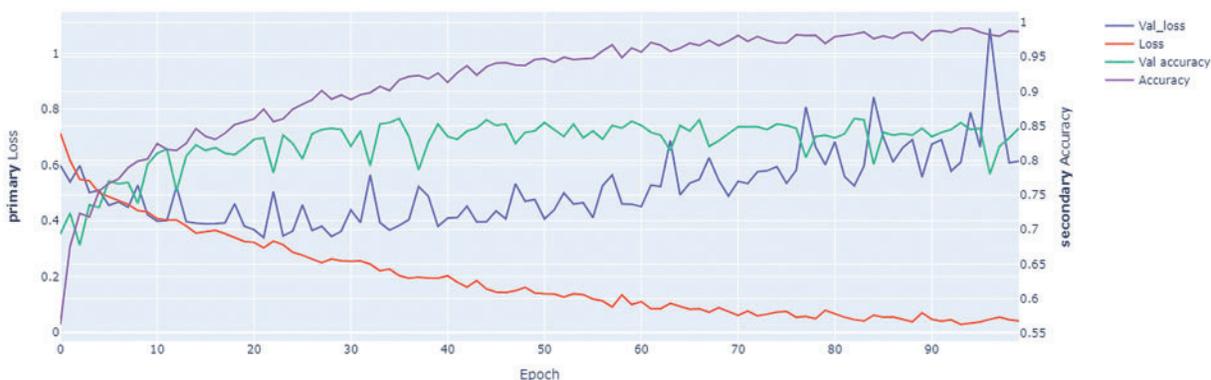

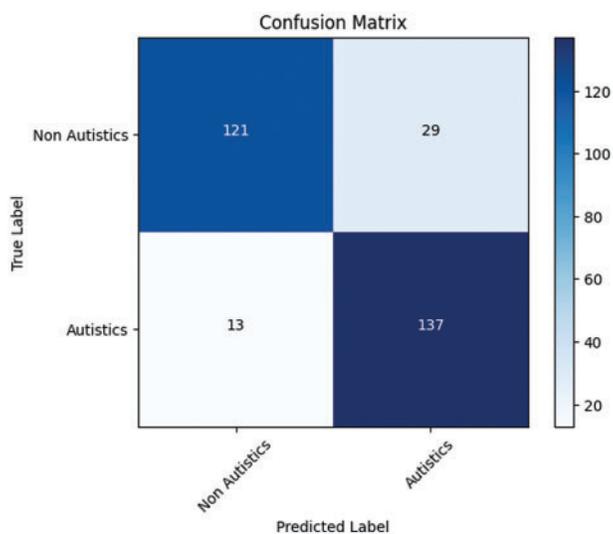**Figure 11:** Vgg-16 batch size 16 cross-validation complete training report

Figure 12: Vgg-16 batch size 16 confusion matrix

We trained VGG16 using a batch size of 32, and its training and Fig. 14 presents validation accuracy equal 98% and test accuracy equivalent 87%. Additionally, Fig. 15 shows the confusion matrix, which indicates that the model, out of 150 images of individuals with and without autism, produced 15 false positives and 26 false negatives. The ROC curve in Fig. 16 illustrates the performance of the trained model with a batch size of 32. From these plots, we analyzed that although VGG16 with 32 batch size provided an accuracy of 88%, FNR is 26, which is a bit higher.
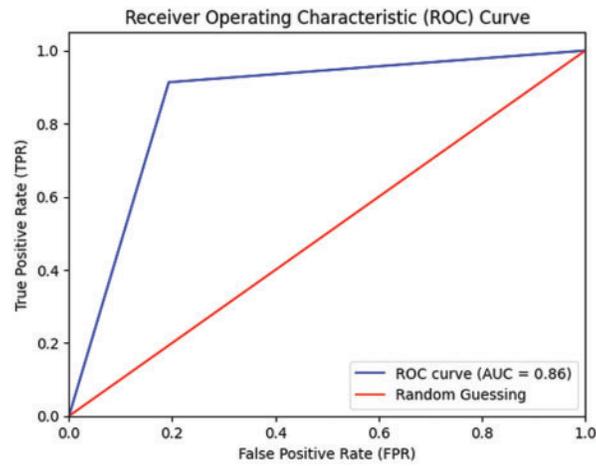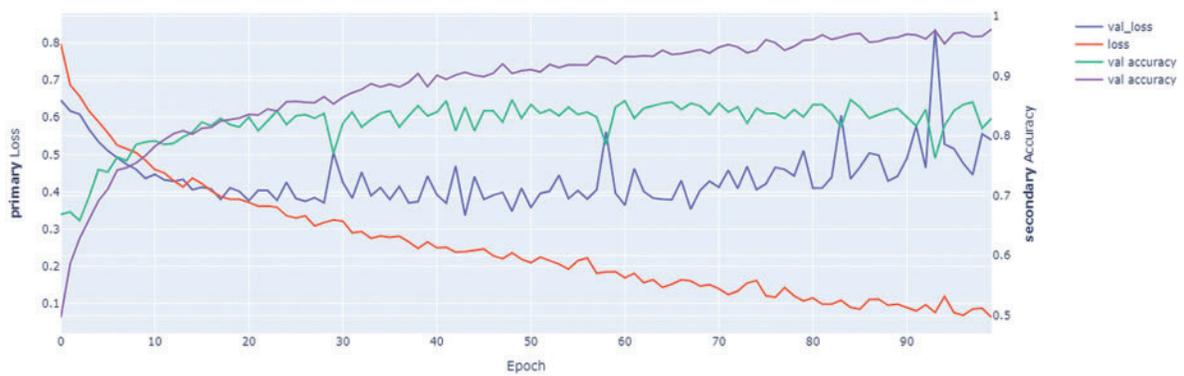
Figure 13: Vgg-16 batch size 16 roc curve

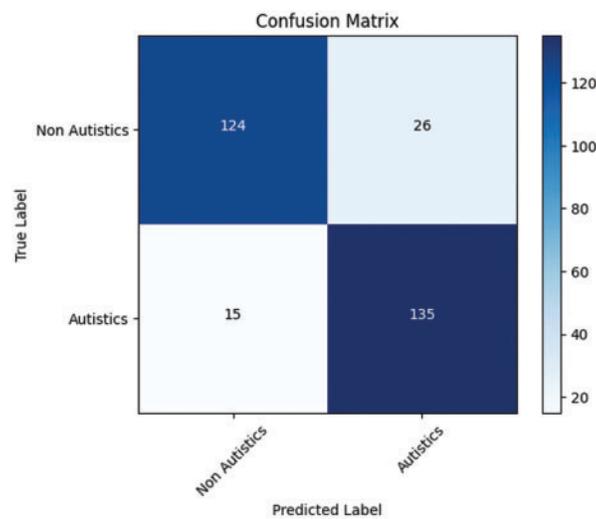**Figure 14:** Vgg-16 experimentation results with 32 batch size

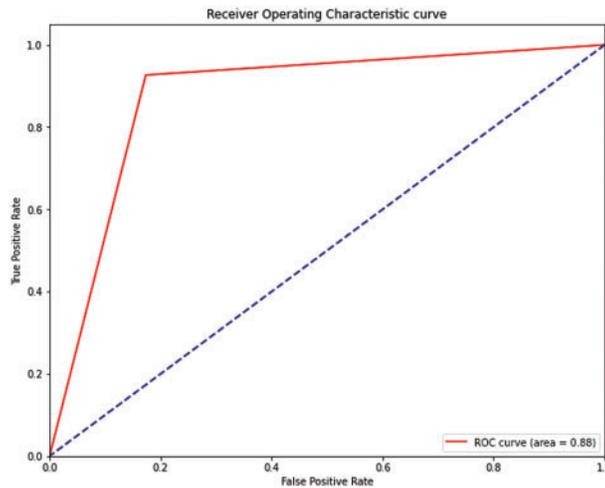**Figure 15:** Vgg-16 batch size 32 confusion matrix

Figure 16: Vgg-16 batch size 32 roc curve

We trained VGG16 with a batch size of 64. It exhibited a validation accuracy of 94% and a test accuracy of 88%. The training and validation accuracy curves for VGG16 are shown in Fig. 17. The confusion matrix in Fig. 18 reveals that the model produced 19 false positives and 16 false negatives out of 150 images of individuals with and without autism. The ROC curve in Fig. 19 also exhibits a high TPR and a relatively low FPR.
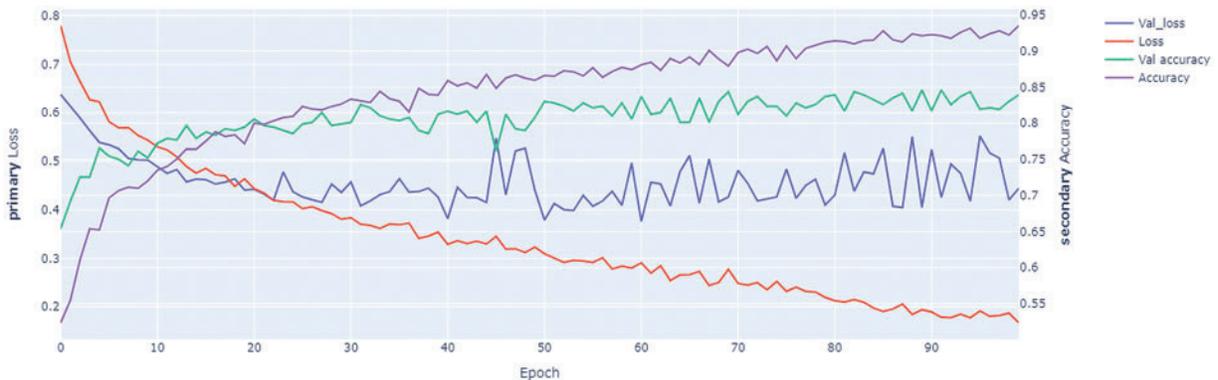

Figure 17: Vgg-16 experimentation results with 64 batch size

We trained MobileNet with a batch size of 64. Fig. 20 visualizes the training and validation accuracy curves for MobileNet, showcasing the accuracy progression throughout the training process. The confusion matrix, presented in Fig. 21, provides insights into the model's performance. Out of 150 images of individuals with and without autism, the trained MobileNet model exhibited 15 false positives and 23 false negatives. The model incorrectly classified 15 instances as positive (individuals with autism) when negative (individuals without autism) and 23 as negative when, they were positive. To assess the model's ability to balance TPR and FPR, the ROC curve is displayed in Fig. 22. The sharp angle in the ROC curve suggests a high TPR for a relatively low FPR. This behavior indicates that the MobileNet classifier is effective in achieving a high sensitivity by correctly identifying positive instances (individuals with autism) while simultaneously maintaining a low rate of false positives.
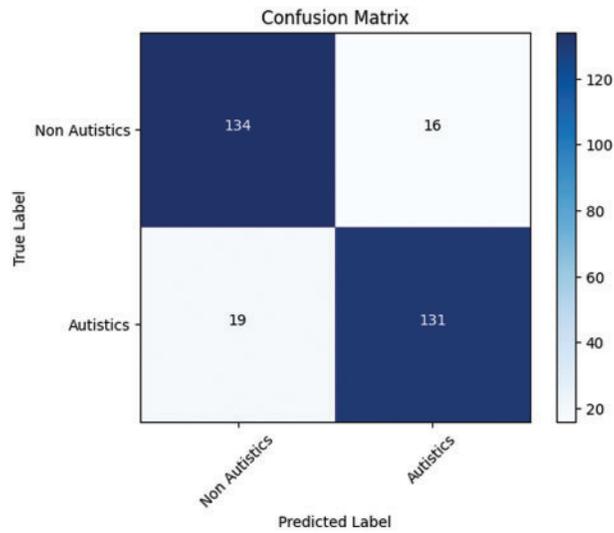
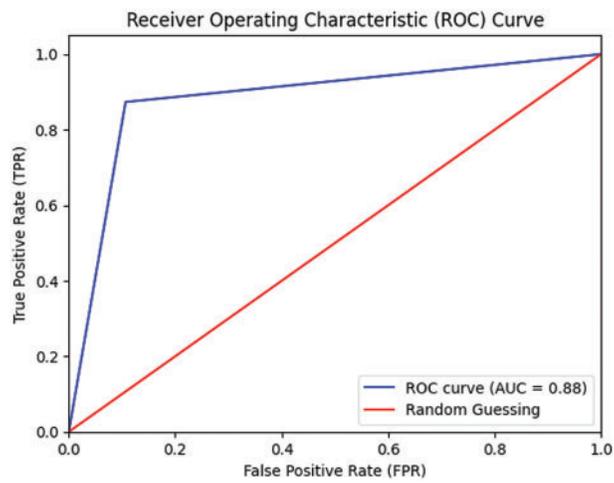**Figure 18:** Vgg-16 batch size 64 confusion matrix
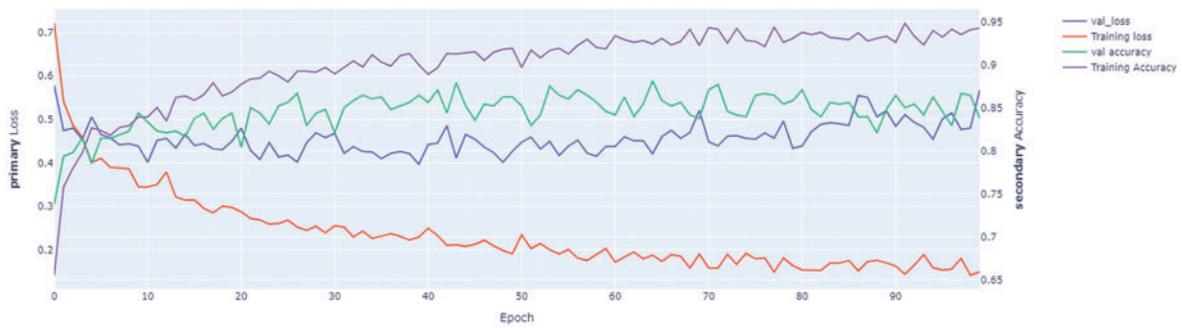

Figure 19: Vgg-16 batch size 64 roc curve

Figure 20: Mobilenet accuracy/loss experiment result
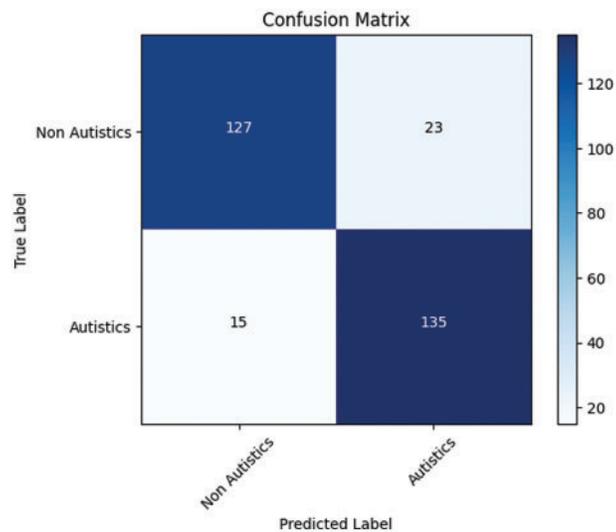
Figure 21: Mobilenet confusion matrix
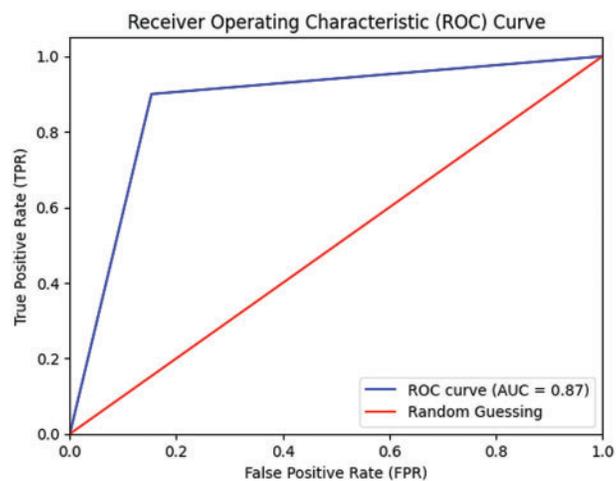
Figure 22: Mobilenet roc curve

4.1 Using Orange for Machine Learning

Utilizing the Orange Tool, we employed machine learning techniques to analyze an image dataset comprising autistic and non-autistic pictures. We utilized the VGG16 neural network to extract features from the dataset. Later, the VGG19 is trained for classification. The training process involved 200 epochs with Relu activation and Adam optimizer and was supplemented by the application of Gradient Boost, KNN, and SVM for a more comprehensive analysis. All applied classifiers' performances are presented in Figs. 23–26.

The KNN classifier achieved a validation accuracy of 78% and a test accuracy of 77.4%, with precision, recall, and F1 score values ranging from 72% to 70.9%. The Random Forest classifier showed similar validation accuracy (78%), but a lower test accuracy of 71.1%, with precision, recall, and F1 score values around 71.2% and 71%. The Gradient Boost classifier demonstrated a higher validation

accuracy of 83.9% but a lower test accuracy of 75.2%. However, its precision, recall, and F1 score values remained at 75.2%. Lastly, the VGG19 classifier achieved the highest validation accuracy of 85.5% and test accuracy of 77.8%, with precision, recall, and F1 score values all at 77.8%.

		Predicted		Σ
		Autistic	Non_Autistic	
Actual	Autistic	1132	334	1466
	Non_Autistic	318	1147	1465
	Σ	1450	1481	2931

Figure 23: Orange tool vgg-19

		Predicted		Σ
		Autistic	Non_Autistic	
Actual	Autistic	1097	369	1466
	Non_Autistic	479	986	1465
	Σ	1576	1355	2931

Figure 24: Orange tool random forest

		Predicted		Σ
		Autistic	Non_Autistic	
Actual	Autistic	891	575	1466
	Non_Autistic	268	1197	1465
	Σ	1159	1772	2931

Figure 25: Orange tool kNN

		Predicted		Σ
		Autistic	Non_Autistic	
Actual	Autistic	1099	367	1466
	Non_Autistic	359	1106	1465
	Σ	1458	1473	2931

Figure 26: Orange tool gradient boosting

4.2 Additional Test Data

During our visit to local autism centers in Pakistan, we identified the limited availability of experts and resources for autism detection. We began collaborating on developing ML solutions to address the challenge. We collected a dataset of 50 images each from autistic and typically developed children, ensuring strict confidentiality. Using the VGG16 model with 5-fold cross-validation, we tested the

locally collected dataset and achieved an impressive prediction accuracy of 85%. This performance remained remarkable despite race, facial features, and ethnicity variations between the training and testing datasets.

5 Discussion

Based on the provided results, a comparative evaluation of different classifiers applied to the autistic dataset reveals variations in their performance. We considered the following evaluation metrics: validation accuracy, test accuracy, precision, recall, and F1 score. The VGG16 classifier exhibited a validation accuracy of 97% and a test accuracy of 88%, accompanied by precision, recall, and F1 score values of 87%, 88%, and 88%, respectively. In contrast, the VGG16 model with 5 folds achieved a higher validation accuracy of 99% but a slightly lower test accuracy of 87%. Its precision and recall were 88% and 90%, resulting in an F1 score of 85%.

Furthermore, the VGG16 classifier with a batch size of 16 displayed a validation accuracy of 93%, a test accuracy of 86%, and a precision-recall balance with values of 93% and 91%, respectively, leading to an F1 score of 87%. Similarly, the VGG16 models with batch sizes of 32 and 64 achieved validation accuracies of 98% and 94.5%, test accuracies of 87% and 88%, and F1 scores of 88%, respectively. These models' precision, recall, and F1 score values varied between 84%-89% and 87%-93%. Lastly, the MobileNet classifier exhibited a validation accuracy of 92%, test accuracy of 87, and precision, recall, and F1 score values of 85%, 90%, and 88%, respectively. In conducting machine learning analysis using the Orange Tool, we observed that the KNN classifier demonstrated a validation accuracy of 78% and a test accuracy of 77.4%. Its precision, recall, and F1 score values were 72% to 70.9%.

Similarly, the Random Forest classifier displayed a validation accuracy of 78%, though with a slightly lower test accuracy of 71.1%. Its precision, recall, and F1 score values were approximately 71.2% and 71%. The Gradient Boost classifier exhibited a higher validation accuracy of 83.9% but a lower test accuracy of 75.2%, with consistent precision, recall, and F1 score values of 75.2%. Finally, the VGG19 classifier demonstrated the highest validation accuracy of 85.5% and test accuracy of 77.8%, with all precision, recall, and F1 score values at 77.8%. The VGG16 classifier with 5 folds demonstrated superior performance to other classifiers based on the evaluation metrics. It achieved a higher validation accuracy of 99%, indicating its ability to generalize well to unseen data. Although its test accuracy was slightly lower at 87%, the precision (88%) and recall (90%) values highlight its balanced performance in correctly identifying positive cases (autistics) and minimizing false negatives. The resulting F1 score of 85% indicates a favorable trade-off between precision and recall. The utilization of 5-fold cross-validation further enhances the reliability of the VGG16 model by effectively utilizing the available data for training and evaluation. This approach helps mitigate overfitting and accurately represents the model's performance on unseen data. Furthermore, test results on locally collected data highlight the potential of machine learning, specifically the VGG16 model, as a reliable tool for autism detection in the local context. Overcoming variations in facial characteristics and cultural attributes demonstrates its applicability to diverse populations. Table 3 presents the performance of the tested classifiers for autism detection using facial images.

We compared the VGG16 and MobileNet classifiers in detecting autism. VGG16 achieved exceptional potential with 99% validation accuracy using a 5-fold validation strategy. However, test accuracy was slightly lower at 87%. MobileNet showed strong performance with 92% validation and 87% test accuracy. Both models used the Adam optimizer, but VGG16 outperformed MobileNet in accuracy. MobileNet used a batch size of 64, while VGG16 was trained using 2 batch size.

Table 3: Comparing the performance of applied classifiers

Classifier	Validation accuracy	Test accuracy	Precision	Recall	F1 score
VGG16	0.97	0.88	0.87	0.88	0.88
VGG16 5 folds	0.99	0.87	0.88	0.90	0.85
VGG16 16 batch size	0.93	0.86	0.93	0.91	0.87
VGG16 32 batch size	0.98	0.87	0.84	0.93	0.88
VGG16 64 batch size	0.945	0.88	0.89	0.87	0.88
MobileNet	0.92	0.88	0.85	0.90	0.88
KNN	0.78	0.774	0.72	0.72	0.709
Random forest	0.78	0.711	0.712	0.711	0.71
Gradient boost	0.839	0.752	0.752	0.752	0.752
VGG19	0.855	0.778	0.778	0.778	0.778

6 Conclusion

This study employed various deep learning classifiers, including VGG16 and MobileNet, to classify autism spectrum disorder (ASD) through facial images. Additionally, the machine learning tool Orange was utilized for this purpose. The researchers tested VGG16 with 5-fold cross-validation and different batch sizes. Among the classifiers applied, VGG16, with 5-fold cross-validation, demonstrated the highest performance, achieving an impressive test accuracy of 87%. This result highlights the effectiveness of the cross-validation technique in improving the robustness and generalizability of the model. Furthermore, collecting locally gathered data from Pakistan was crucial in raising awareness about autism within the local community. By introducing a deep learning-based approach to autism detection, this study has the potential to significantly reduce the cost and enhance the efficiency of autism diagnosis. Early detection of autism is vital for timely intervention and support, and using facial images in combination with deep learning techniques offers a promising avenue for accurate and accessible diagnosis. The findings of this research contribute to the field of autism detection, providing valuable insights into the performance of different classifiers and the potential of machine learning in ASD diagnosis. Integrating locally collected data underscores the importance of context-specific research and highlights the significance of addressing specific regions' unique challenges and requirements.

Author Contributions: The authors confirm contribution to the paper as follows: study concept and design: T. Farhat, S. Akram, H. S. AlSagri, Z. Ali, A. Ahmad and A. Jaffar; data collection: T. Farhat, S. Akram and Z. Ali; methodology: T. Farhat, S. Akram, H. S. AlSagri, Z. Ali and A. Ahmad; analysis and interpretation of results: T. Farhat, S. Akram, Z. Ali, A. Ahmad and A. Jaffar; draft manuscript

preparation: T. Farhat, S. Akram, Z. Ali, A. Ahmad and A. Jaffar. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data is openly available in a public repository. The data supporting this study's findings are openly available in Kaggle at https://www.kaggle.com/datasets/imrankhan77/autistic-children-facial-data-set.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   B. Zablotsky, L. I. Black, M. J. Maenner, L. A. Schieve and S. J. Blumberg, "Estimated prevalence of autism and other developmental disabilities following questionnaire changes in the 2014 National Health Interview Survey," *National Vital Statistics Reports*, vol. 14, no. 87, pp. 1–21, 2015.

[2]   D. Christensen and J. Zubler, "From the CDC: Understanding autism spectrum disorder: An evidence-based review of ASD risk factors, evaluation, and diagnosis," *The American Journal of Nursing*, vol. 120, no. 10, pp. 30–40, 2020.

[3]   J. E. Robison, "Autism prevalence and outcomes in older adults," *Autism Research*, vol. 12, no. 3, pp. 370–374, 2019.

[4]   A. de Leeuw, F. Happé and R. A. Hoekstra, "A conceptual framework for understanding the cultural and contextual factors on autism across the globe," *Autism Research*, vol. 13, no. 7, pp. 1029–1050, 2020.

[5]   M. Hoogman, D. van Rooij, M. Klein, P. Boedhoe, I. Ilioska *et al.,* "Consortium neuroscience of attention deficit/hyperactivity disorder and autism spectrum disorder: The ENIGMA adventure," *Human Brain Mapping*, vol. 43, no. 1, pp. 37–55, 2022.

[6]   L. Chetcuti, M. Uljarević and K. Hudry, "Editorial perspective: Furthering research on temperament in autism spectrum disorder," *Journal of Child Psychology and Psychiatry*, vol. 60, no. 2, pp. 225–228, 2019.

[7]   D. J. Morris-Rosendahl and M. A. Crocq, "Neurodevelopmental disorders—The history and future of a diagnostic concept," *Dialogues in Clinical Neuroscience*, vol. 1, no. 22, pp. 65–72, 2022.

[8]   P. Kalaivaani and R. Thangarajan, "Enhancing the classification accuracy in sentiment analysis with computational intelligence using joint sentiment topic detection with MEDLDA," *Intelligent Automation & Soft Computing*, vol. 26, no. 1, pp. 71–79, 2020.

[9]   E. C. Bacon, A. Moore, Q. Lee, C. Carter Barnes, E. Courchesne *et al.,* "Identifying prognostic markers in autism spectrum disorder using eye tracking," *Autism*, vol. 24, no. 3, pp. 658–669, 2020.

[10]   B. Trost, W. Engchuan, C. M. Nguyen, B. Thiruvahindrapuram, E. Dolzhenko *et al.,* "Genome-wide detection of tandem DNA repeats that are expanded in autism," *Nature*, vol. 586, no. 7827, pp. 80–86, 2020.

[11]   M. B. Nejad and M. E. Shiri, "A new enhanced learning approach to automatic image classification based on Salp Swarm Algorithm," *Computer Systems Science and Engineering*, vol. 34, no. 2, pp. 91–100, 2019.

[12]   K. Jha, A. Gupta, A. Alabdulatif, S. Tanwar, C. O. Safirescu *et al.,* "CSVAG: Optimizing vertical handoff using hybrid cuckoo search and genetic algorithm-based approaches," *Sustainability*, vol. 14, no. 14, pp. 8547–8557, 2022.

[13]   D. Bzdok and A. Meyer-Lindenberg, "Machine learning for precision psychiatry: Opportunities and challenges," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 3, no. 3, pp. 223–230, 2018.

[14]   G. Choy, O. Khalilzadeh, M. Michalski, S. Do, A. E. Samir *et al.,* "Current applications and future impact of machine learning in radiology," *Radiology*, vol. 288, no. 2, pp. 318–328, 2018.

[15]   F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," *Informatics for Health and Social Care*, vol. 44, no. 3, pp. 278–297, 2019.

[16] P. Kassraian-Fard, C. Matthis, J. H. Balsters, M. H. Maathuis and N. Wenderoth, "Promises, pitfalls, and basic guidelines for applying machine learning classifiers to psychiatric imaging data, with autism as an example," *Frontiers in Psychiatry*, vol. 7, pp. 177–187, 2016.

[17] S. Kalikar, A. Sinha, S. Srivastava and G. Aggarwal, "Early detection of autism spectrum disorder (ASD) using machine learning techniques: A review," in *Proc. of Third Int. Conf. on Communication, Computing and Electronics Systems*, Coimbatore, India, vol. 1, pp. 1015–1027, 2022.

[18] T. Falkmer, K. Anderson, M. Falkmer and C. Horlin, "Diagnostic procedures in autism spectrum disorders: A systematic literature review," *European Child & Adolescent Psychiatry*, vol. 22, no. 6, pp. 329–340, 2013.

[19] E. Bal, E. Harden, D. Lamb, A. V. van Hecke, J. W. Denver *et al.,* "Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state," *Journal of Autism and Developmental Disorders*, vol. 40, pp. 358–370, 2010.

[20] K. Aldridge, I. D. George, K. K. Cole, J. R. Austin, T. N. Takahashi *et al.,* "Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes," *Molecular Autism*, vol. 2, pp. 1–12, 2011.

[21] Z. A. Ahmed, T. H. Aldhyani, M. E. Jadhav, M. Y. Alzahrani, M. E. Alzahrani *et al.,* "Facial features detection system to identify children with autism spectrum disorder: Deep learning models," *Computational and Mathematical Methods in Medicine*, vol. 22, no. 1, pp. 1–9, 2022.

[22] "400k Pakistani children suffer from autism," [Online]. Available: https://tribune.com.pk/story/2292799/400k-pakistani-children-suffer-from-autism (accessed on 08/09/2023)

[23] "ASHA," [Online]. Available: https://www.asha.org/practice-portal/clinical-topics/autism (accessed on 08/09/2023)

[24] J. Dubreucq, F. Haesebaert, J. Plasse, M. Dubreucq and N. Franck, "A systematic review and meta-analysis of social skills training for adults with autism spectrum disorder," *Journal of Autism and Developmental Disorders*, vol. 52, no. 4, pp. 1598–1609, 2022.

[25] A. Pickles, A. Le Couteur, K. Leadbitter, E. Salomone, R. Cole-Fletcher *et al.,* "Parent-mediated social communication therapy for young children with autism (PACT): Long-term follow-up of a randomised controlled trial," *The Lancet*, vol. 388, no. 10059, pp. 2501–2509, 2016.

[26] L. A. Vismara and S. J. Rogers, "Behavioral treatments in autism spectrum disorder: What do we know?," *Annual Review of Clinical Psychology*, vol. 6, pp. 447–468, 2010.

[27] T. M. Dawson, H. S. Ko and V. L. Dawson, "Genetic animal models of Parkinson's disease," *Neuron*, vol. 66, no. 5, pp. 646–661, 2010.

[28] C. Lord, M. Rutter, S. Goode, J. Heemsbergen, H. Jordan *et al.,* "Austism diagnostic observation schedule: A standardized observation of communicative and social behavior," *Journal of Autism and Developmental Disorders*, vol. 19, no. 2, pp. 185–212, 1989.

[29] C. Lord, M. Rutter and A. Le Couteur, "Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders," *Journal of Autism and Developmental Disorders*, vol. 24, no. 5, pp. 659–685, 1994.

[30] K. Riddle, C. J. Cascio and N. D. Woodward, "Brain structure in autism: A voxel-based morphometry analysis of the autism brain imaging database exchange (ABIDE)," *Brain Imaging and Behavior*, vol. 11, pp. 541–551, 2017.

[31] V. Subbaraju, M. B. Suresh, S. Sundaram and S. Narasimhan, "Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging: A spatial filtering approach," *Medical Image Analysis*, vol. 35, pp. 375–389, 2017.

[32] S. R. Arumugam, S. G. Karuppasamy, S. Gowr, O. Manoj and K. Kalaivani, "A deep convolutional neural network based detection system for autism spectrum disorder in facial images," in *Proc. of 2021 Fifth Int. Conf. on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, Tamilnadu, India, IEEE, vol. 5, pp. 1255–1259, 2021.

[33] Y. Yang, "A preliminary evaluation of still face images by deep learning: A potential screening test for childhood developmental disabilities," *Medical Hypotheses*, vol. 144, pp. 109978–109988, 2020.

[34] K. K. Hyde, M. N. Novack, N. LaHaye, C. Parlett-Pelleriti, R. Anden *et al.,* "Applications of supervised machine learning in autism spectrum disorder research: A review," *Review Journal of Autism and Developmental Disorders*, vol. 6, pp. 128–146, 2019.

[35] A. V. Dahiya, E. DeLucia, C. G. McDonnell and A. Scarpa, "A systematic review of technological approaches for autism spectrum disorder assessment in children: Implications for the COVID-19 pandemic," *Research in Developmental Disabilities*, vol. 109, pp. 103852–103862, 2021.

[36] A. T. Wieckowski, T. Hamner, S. Nanovic, K. S. Porto, K. L. Coulter *et al.,* "Early and repeated screening detects autism spectrum disorder," *The Journal of Pediatrics*, vol. 234, pp. 227–235, 2021.

[37] V. Yaneva, S. Eraslan, Y. Yesilada and R. Mitkov, "Detecting high-functioning autism in adults using eye tracking and machine learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 6, pp. 1254–1261, 2020.

[38] S. Alhassan, A. Soudani and M. Almusallam, "Energy-efficient EEG-based scheme for autism spectrum disorder detection using wearable sensors," *Sensors*, vol. 23, no. 4, pp. 2228–2238, 2023.

[39] K. Vakadkar, D. Purkayastha and D. Krishnan, "Detection of autism spectrum disorder in children using machine learning techniques," *SN Computer Science*, vol. 2, pp. 1–9, 2021.

[40] T. Eslami, F. Almuqhim, J. S. Raiker and F. Saeed, "Machine learning methods for diagnosing autism spectrum disorder and attention-deficit/hyperactivity disorder using functional and structural MRI: A survey," *Frontiers in Neuroinformatics*, vol. 1, pp. 62–72, 2021.

[41] Z. A. Ahmed, T. H. Aldhyani, M. E. Jadhav, M. Y. Alzahrani, M. E. Alzahrani *et al.,* "Facial features detection system to identify children with autism spectrum disorder: Deep learning models," *Computational and Mathematical Methods in Medicine*, vol. 2022, no. 1, pp. 1–9, 2022.

[42] M. Jogin, M. Madhulika, G. Divya, R. Meghana and S. Apoorva, "Feature extraction using convolution neural networks (CNN) and deep learning," in *Proc. of 2018 3rd IEEE Int. Conf. on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, India, IEEE, vol. 3, pp. 2319–2323, 2018.

[43] M. Hosseinzadeh, J. Khoramdel, Y. Borhani and E. Najafi, "A new fuzzy logic based learning rate scheduling method for crop classification with convolutional neural network," in *Proc. of 2022 8th Int. Conf. on Control, Instrumentation and Automation (ICCIA)*, Tehran, Islamic Republic of Iran, IEEE, vol. 8, pp. 1–5, 2022.

[44] A. Pujara, "Image classification with MobileNet," BuildIn. [Online]. Available: https://builtin.com/machine-learning/mobilenet (accessed on 10/22/2023)

[45] P. Verma, V. Tripathi and B. Pant, "Comparison of different optimizers implemented on the deep learning architectures for COVID-19 classification," in *Proc. of Materials Today*, Uttarakhand, India, vol. 46, pp. 11098–11102, 2021.

[46] D. Vaishnav and B. R. Rao, "Comparison of machine learning algorithms and fruit classification using orange data mining tool," in *Proc. of 2018 3rd Int. Conf. on Inventive Computation Technologies (ICICT)*, Coimbatore, India, IEEE, vol. 3, pp. 603–607, 2018.