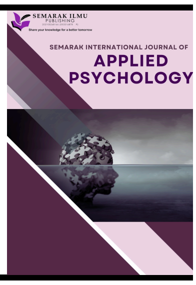
SEMARAK ILMU PUBLISHING
202103268166(003316878-P)

Semarak International Journal of Applied Psychology

Journal homepage:
https://semarakilmu.my/index.php/sijap/index
ISSN: 3030-525X



Autism Spectrum Disorder Identification from Facial Images using Fine Tuned Pre-trained Deep Learning Models and Explainable AI Techniques

Sheikh Shihab Hossain[1], Ferdib-Al-Islam[1,*], Md. Rahatul Islam[2], Shahid Rahman[3], Md. Shamim Parvej[1]

[1] Department of Computer Science and Engineering, Northern University of Business and Technology Khulna, Khulna, 9100, Bangladesh
[2] Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Japan
[3] Department of Electrical and Electronic Engineering, Canadian University of Bangladesh, Dhaka, Bangladesh

ARTICLE INFO

Article history:
Received 30 November 2024
Received in revised form 20 December 2024
Accepted 31 December 2024
Available online 20 January 2025

Keywords:
Autism Spectrum Disorder; Deep Learning; Explainable AI; LIME; Grad-CAM

ABSTRACT

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that requires early diagnosis for effective intervention. Traditional diagnostic tools, such as MRI and CT scans, are often expensive, time-consuming, and inaccessible in underprivileged regions. To address this challenge, this study leverages facial images as a cost-effective and non-invasive alternative for ASD identification. A comprehensive evaluation of twelve pre-trained deep learning models—including ResNet-50, ResNet-101, ResNet-152, MobileNetV2, MobileNetV3, AlexNet, InceptionV1 (GoogleNet), SqueezeNet, EfficientNetB0, DenseNet121, DenseNet201, and VGG16—was conducted. Among these, DenseNet121 emerged as the top-performing model, achieving an accuracy of 90.33%, precision of 92.00%, recall of 92.00%, and an F1-score of 90.00%. Explainable AI techniques, including Local Interpretable Model-Agnostic Explanations (LIME) and Gradient-weighted Class Activation Mapping (Grad-Cam), were applied to highlight facial regions crucial for the model's predictions, enhancing transparency and trust. The proposed DenseNet121 model outperformed previous works. The results demonstrate the efficacy of this approach, offering a reliable, interpretable, and accessible solution for ASD identification, particularly in resource-constrained settings.

1. Introduction

Autism spectrum disorder (ASD), is a neurological developmental disorder. It affects how people communicate and interact with others, as well as how they behave and learn [1]. Symptoms and signs of ASD appear when a child is very young. It is a chronic illness which is why there are no full treatments. A case study found that 33% of children with difficulties other than ASD have some ASD symptoms while not meeting the full classification criteria [2]. In the Southeast Asia region, it is estimated that every 1 in 160 children has ASD. Recently, the Bangabandhu Sheikh Mujib Medical University (BSMMU) in Bangladesh confirmed that almost 2 in 1000 children have been suffering

* *Corresponding author.*
E-mail address: ferdib.bsmrstu@gmail.com

https://doi.org/10.37934/sijap.5.1.2953b

from ASD in Bangladesh. Wherein, the urban prevalence is higher than the rural areas. Usually, all symptoms of ASD appear between 18 to 36 months of age. In this case, the awareness and skills of the primary healthcare service provider play a vital role in ensuring appropriate referral systems for exact intervention in the healthcare services delivery system for children with autism. Because early intervention can change the course of life of an autism-affected child [3]. Over the years, advancements in technology and the growing availability of vast datasets have spurred the exploration of innovative approaches to enhance the accuracy and efficiency of ASD diagnosis.

ASD affects a significant percentage of the pediatric population. In most cases, it can usually be identified in its early stages, but the major problem lies in existing diagnosis procedures. As a result, there is a waiting time of at least 13 months from the initial suspicion to the actual diagnosis. The diagnosis takes many hours [4], and the continuously growing demand for appointments is much greater than the peak capacity of the country's pediatric clinics [5]. ASD can be detected in many ways. Facial Expression is one of them. Using deep learning or machine learning with facial datasets can be an effective and faster way to detect ASD-affected cases.

ASD-related studies such as using brain imaging techniques like MRI and PET have helped us understand how the brain develops differently in people with autism spectrum disorder (ASD) [6-8]. A study by Daniel *et al.,* demonstrates the efficacy of functional connectome patterns and SVM-based models, achieving ASD detection accuracies ranging from 70% to 80%, with notable contributions from deep learning approaches and feature selection methods [9].

Using facial datasets for Autism Spectrum Disorder (ASD) detection offers a non-invasive, cost-effective, and accessible alternative to traditional modalities like MRI or EEG, which are expensive and require specialized equipment. Facial images can be collected easily using standard cameras or smartphones, making the approach scalable and suitable for large-scale screening in diverse settings. This method is also more child-friendly, as it avoids stressful clinical procedures and enables detection in familiar environments. Additionally, facial image analysis can reveal subtle phenotypic features associated with ASD, providing novel insights that complement traditional diagnostic techniques. In this research, we have applied twelve different pre-trained deep learning models using transfer learning on children's facial dataset for the identification of ASD. The images have gone under preprocessing steps. Then, they are fed into the models to accomplish the classification task. Explainable Artificial Intelligence (XAI) has been used in the model's output for model explanation. In this paper, Grad-CAM and LIME techniques have been applied to understand the CNN decision. These strategies aid in comprehending the specific portions of an input image that hold significance for the CNNs's prediction of a given class.

We have structured our paper as follows: The "Related Works" section summarizes the literature review performed. The "Methodology" section explains the working and methodology of the system we have proposed and its implementation. The "Result and Discussion" section portrays the inferences and results obtained. Finally, the "Conclusions" section highlights our conclusions.

2. Related Works

A study presents a novel deep-learning approach for automatic facial expression recognition [28]. The proposed architecture first segments the facial components known to be important for facial expression recognition and forms an iconized image. It then performs facial expression classification using the obtained iconized facial component image combined with the raw facial images. This approach integrates local, part-based features with holistic facial information for robust facial expression recognition. The preliminary experimental results using the proposed system achieved 93.43% facial expression recognition accuracy, which is more than a 6% improvement compared to facial expression recognition from raw input images. The proposed cascaded CNN architecture achieved a facial expression recognition accuracy of 93.43% in the preliminary experimental results. This accuracy is more than a 6% improvement compared to facial expression recognition from raw input images. The proposed system outperformed a single CNN architecture that directly recognizes facial expressions from raw facial images. The cascaded CNN architecture benefits from guided image analysis, the fusion of part-based and holistic information, and patient de-identification and privacy. The iconized images produced by the first CNN architecture facilitate the use, archiving, and communication of critical facial features while protecting patient privacy.

Another study introduces a novel approach for detecting ASD using deep learning techniques applied to facial images [32]. ASD is a developmental disability characterized by challenges in social, communication, and behavioral abilities. While individuals with ASD may not differ significantly from others in appearance, their interactions may be distinct, with some requiring assistance for basic needs. Early identification of ASD is crucial for providing timely therapy to enhance skill development. Given the neurological nature of the disorder, researchers have explored various image processing techniques, primarily based on MRI images, to predict ASD in advance. This study focuses on developing a prediction system utilizing Convolutional Neural Networks (CNNs) trained on facial photos.The dataset utilized for model training and testing is sourced from Kaggle and split into an 80:20 ratio for training and testing purposes. Results indicate that the proposed model achieves an impressive accuracy rate of 91%, with an overall loss of 0.53. This highlights the potential of deep learning-based approaches for ASD detection, offering a non-invasive and efficient means of early diagnosis.

Another work helped to establish that the findings of this study have important implications for understanding emotional processing in individuals with ASD [29]. By using physiological signals, researchers were able to objectively classify affective states in children with ASD. This approach has the potential to provide a language-free, non-invasive, and economically feasible means of recognizing and communicating internal emotional states in individuals with ASD. This study also highlighted the importance of individual-specific approaches in detecting and classifying affective states in ASD. The variability in optimal features selected across participants suggests that each individual may have unique physiological patterns associated with their emotional states. Therefore, future research should consider individualized approaches to better understand and classify emotional responses in individuals with ASD. While the study provides valuable insights into the use of physiological signals for detecting emotional states in individuals with ASD, some limitations should be addressed in future research. Firstly, the sample size was relatively small, which may limit the generalizability of the findings. Future studies should aim to include a larger number of participants to ensure the robustness of the classification results.

Another research in ASD investigates the use of machine learning techniques to identify neural connectivity patterns that can accurately classify individuals with autism from typically developing individuals [30]. The study focuses on the causal influence of brain areas during a Theory-of-Mind (ToM) task and examines the discriminative power of effective connectivity measures in predicting group membership. The study involved 15 high-functioning adolescents and adults with autism and 15 typically developing control participants. Participants were asked to view a series of comic strip vignettes in an MRI scanner and choose the most logical end to the story from three alternatives. The mean time series from 18 activated regions of interest were processed using a multivariate autoregressive model (MVAR) to obtain the causality matrices for each participant. These causal connectivity weights, along with assessment scores, functional connectivity values, and fractional anisotropy obtained from DTI data, were submitted to a recursive cluster elimination-based support vector machine (SVM) classifier to determine the accuracy of predicting a participant's group membership.

A study on an innovative approach for identifying biotypes across psychiatric disorders using neuroimaging data [31]. Schizophrenia (SZ) and autism spectrum disorder (ASD) have been perceived as distinct disorders, yet they overlap in clinical symptoms. Conventional diagnostic methods reliant on clinical manifestations often gives inaccurate results, highlighting the need for alternative biotypes using neuroimaging measures, particularly brain functional connectivity (FC). Previous studies have not effectively utilized FC in detecting biotypes, necessitating the development of innovative methodologies. Leveraging insights provided by graph theory in elucidating topological information within FC, the proposed method employs a graph kernel-based clustering technique. This involves identifying frequent subnetworks within the whole-brain FCs of all subjects, followed by computing graph kernel similarity to measure relationships between subjects for clustering purposes. This study applied that approach to functional magnetic resonance imaging (fMRI) data obtained from 137 SZ and 150 ASD subjects. Through the proposed method, researchers successfully identified meaningful biotypes demonstrating significant differences in FC profiles. The graph kernel-based clustering method presents a promising avenue for transdiagnostic biotype detection, offering potential insights into the underlying neural mechanisms associated with SZ and ASD. Mainly this research presents a pioneering methodology addressing challenges associated with conventional diagnostic frameworks, offering a robust approach towards uncovering transdiagnostic biotypes across psychiatric disorders.

Another study introduces a novel approach for detecting Autism Spectrum Disorder (ASD) using deep learning techniques applied to facial images [32]. ASD is a developmental disability characterized by challenges in social, communication, and behavioral abilities. While individuals with ASD may not differ significantly from others in appearance, their interactions may be distinct, with some requiring assistance for basic needs. Early identification of ASD is crucial for providing timely therapy to enhance skill development. Given the neurological nature of the disorder, researchers have explored various image processing techniques, primarily based on MRI images, to predict ASD in advance. This study focuses on developing a prediction system utilizing Convolutional Neural Networks (CNNs) trained on facial photos. The dataset utilized for model training and testing is sourced from Kaggle and split into an 80:20 ratio for training and testing purposes. Results indicate that the proposed model achieves an impressive accuracy rate of 91%, with an overall loss of 0.53. This highlights the potential of deep learning-based approaches for ASD detection, offering a non-invasive and efficient means of early diagnosis.

3. Methodology

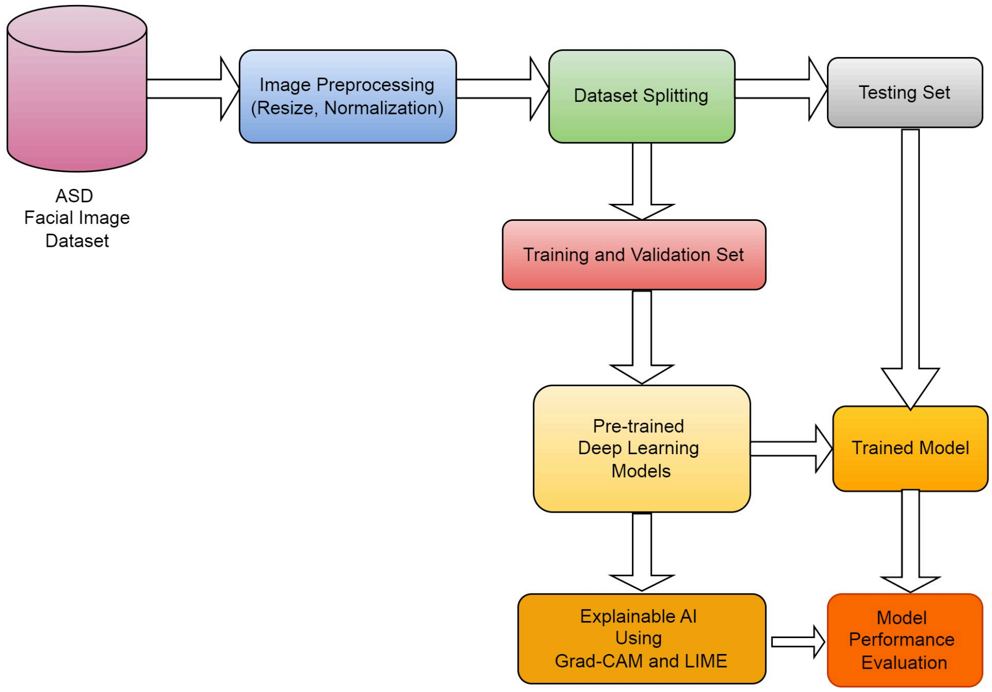


Fig. 1. Flow chart of the proposed methodology

The proposed methodology of this research has been illustrated in Figure 1. Pre-trained deep learning models are fine-tuned using the training and validation sets for Autism Spectrum Disorder (ASD) identification. The performance of the trained models is evaluated on the test set using metrics such as accuracy, precision, recall, F1-score, and a confusion matrix. To enhance the interpretability of model predictions, Explainable AI techniques like Grad-CAM and LIME are employed, providing visual explanations and insights into the facial regions contributing to the classification decisions. This approach ensures a robust evaluation of both model performance and interpretability, supporting reliable and transparent ASD identification using facial images.

3.1 Dataset Description

We have used a high quality image dataset containing facial images of children in this research [26,27]. The dataset comprises 2950 images, featuring autistic and non-autistic children, meticulously organized into training, testing, and validation sets. The dataset is structured into three main subsets - train, test, and valid have been shown in Figure 2. The dataset has two main classes - Autistic and Non_Autistic.

3.2 Data Pre-processing

The whole data set has been resized so that all images are equal in size. Using this approach, system input data can be altered to prevent errors caused by data size imbalance. We changed the images to 224×224 pixels. Then, normalization has been performed.

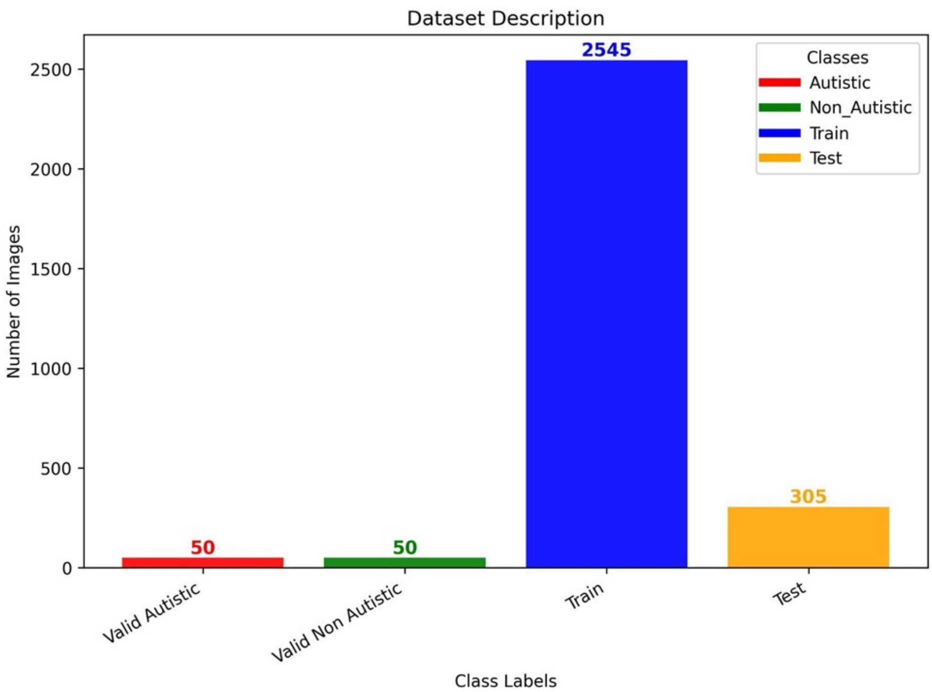


Fig. 2. Dataset description

3.3 Use of Transfer Learning-based Pre-trained Models

Transfer learning applies a variety of pre-trained models to train another desired classification issue. To train the system, twelve pre-trained models have been chosen. The loss function employed by this work was binary cross entropy. Binary cross entropy (BCE) is a good fit for the suggested system because it uses binary classifiers. In order to prevent any further updates, the weights of the pre-trained models were immobilized.

The training processes for ResNet-50 [10], ResNet-101 [11], ResNet-152 [12], MobileNetV2 [13], MobileNetV3 [14,15], AlexNet [15-17], InceptionV1 [18], SqueezeNet [19], EfficientNetB0 [20,21], DenseNet121 [22], DenseNet201 [23,24], and VGG16 [25,33] were designed to optimize their performance in identifying ASD, leveraging the strengths of these pre-trained architectures. Each model was initialized with weights pre-trained on ImageNet, ensuring a robust starting point for transfer learning. Input images were uniformly resized to 224×224 pixels with three color channels (RGB), providing consistency across the dataset and aligning with the input requirements of the models.

For all models, a global average pooling layer was added to the architecture to reduce the spatial dimensions of feature maps while retaining their semantic richness. This was followed by a fully connected dense layer with 512 units, using the ReLU activation function to learn complex patterns from the extracted features. A final dense layer with two units was employed to classify the images, using a sigmoid activation function for binary classification.

Training was conducted using the Adam optimizer, initialized with a learning rate of 0.001 to ensure efficient convergence. A learning rate scheduler was implemented to dynamically reduce the learning rate by 5% starting from the fifth epoch, facilitating stability during training and improving performance. Each model was trained over 30 epochs with a batch size of 64, balancing computational efficiency with the ability to process large amounts of data per iteration.

The training process included splitting the dataset into training, validation, and test sets to evaluate model performance at each stage and ensure generalization. An early stopping mechanism with a patience of five epochs was implemented, halting training if no improvement was observed in validation performance, thereby preventing overfitting. Models were compiled with the binary cross entropy loss function and accuracy as the primary evaluation metric, aligning with the binary-class nature of the problem.

3.4 Model Explanation Using Grad-Cam and LIME

Grad-CAM and LIME were incorporated to enhance interpretability for each model. Grad-CAM was used to generate heatmaps highlighting the regions of the input images that contributed most to the predictions, while LIME provided localized explanations by approximating the model's behavior around specific predictions [34-36]. These techniques added a critical layer of transparency, enabling medical professionals to understand and trust the models' decision-making processes.

4. Result and Discussion

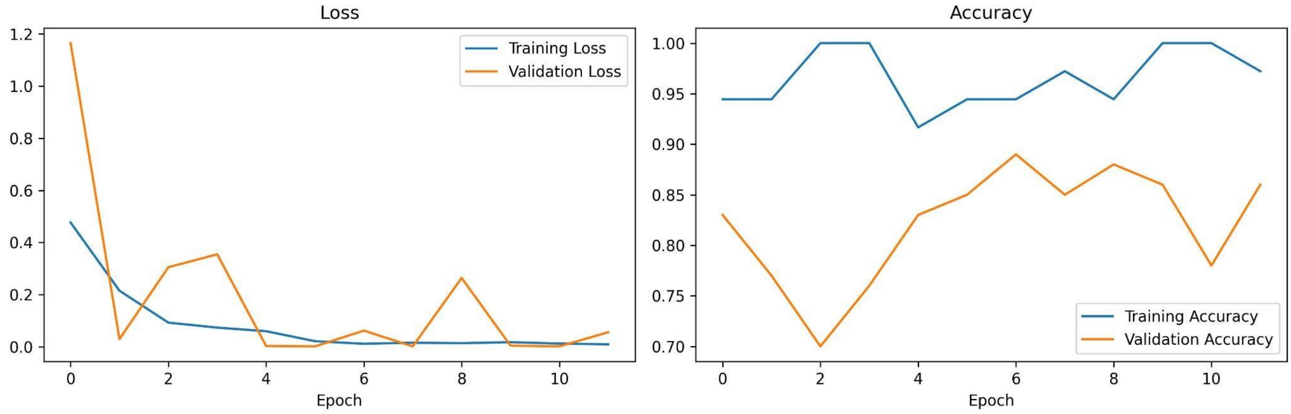
4.1 Experimental Setup

The research has been implemented on Google Colab, which is an open cloud-based notebook environment. Python has been chosen for its conciseness and ease of use. The models have been created and trained with TensorFlow and Keras version 2.15.0. Google Colab hosts considerable resources such as 12.50 GB RAM and 78 GB disc space. In order to perform model training and evaluation without any intervening steps, we have used Google Drive to load data directly into Google Colab.

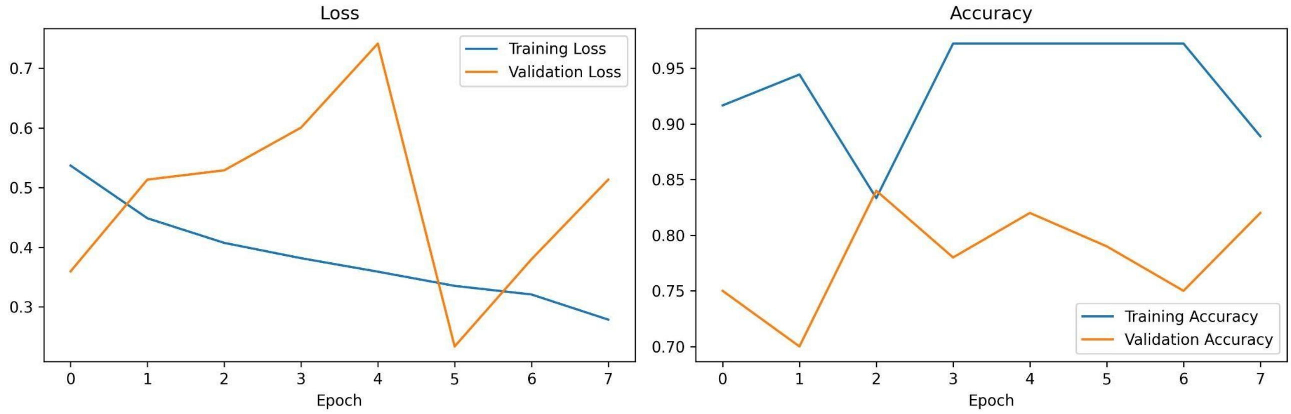
4.2 Result Analysis

Figure 3 shows Loss vs. Accuracy curve of all twelve algorithms whose was implemented in this research.

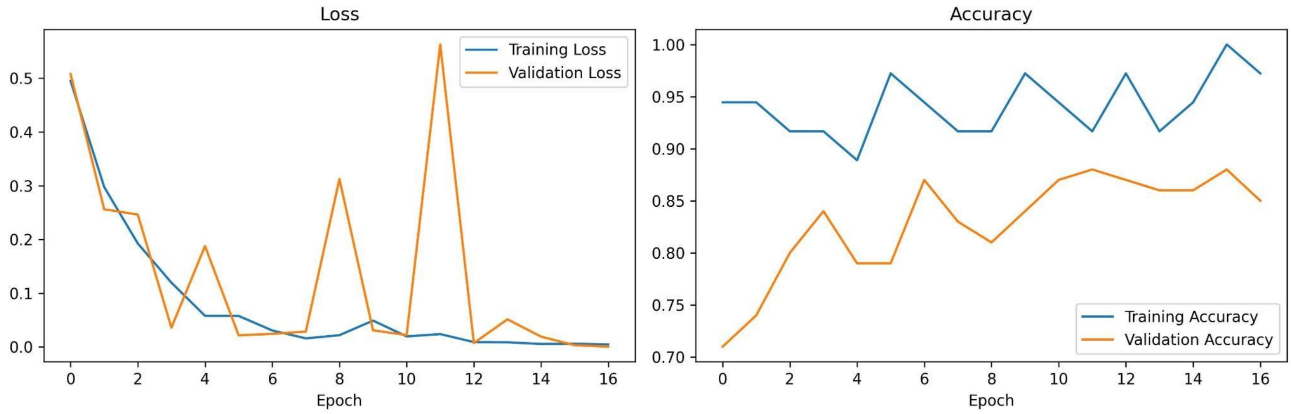
(a) ResNet-50



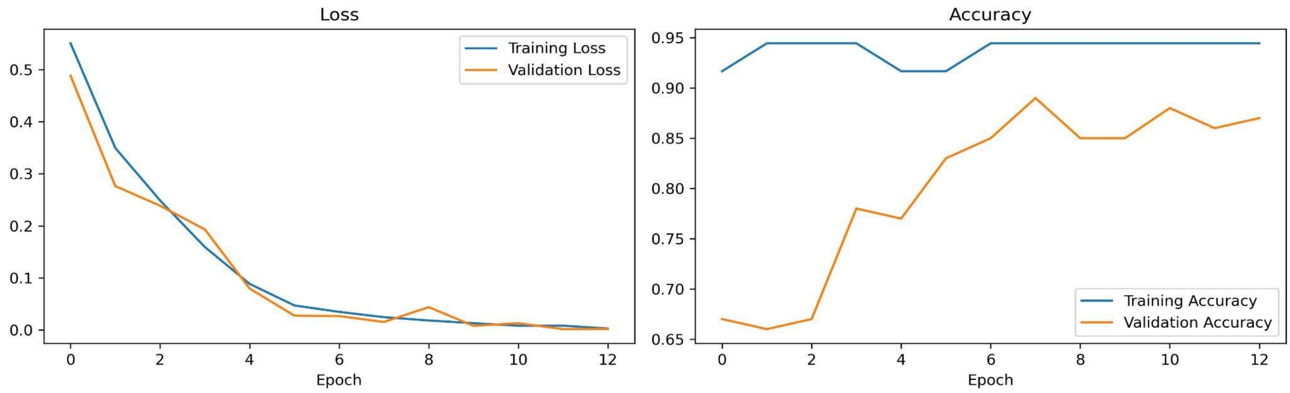
(b) ResNet-101



(c) ResNet-152

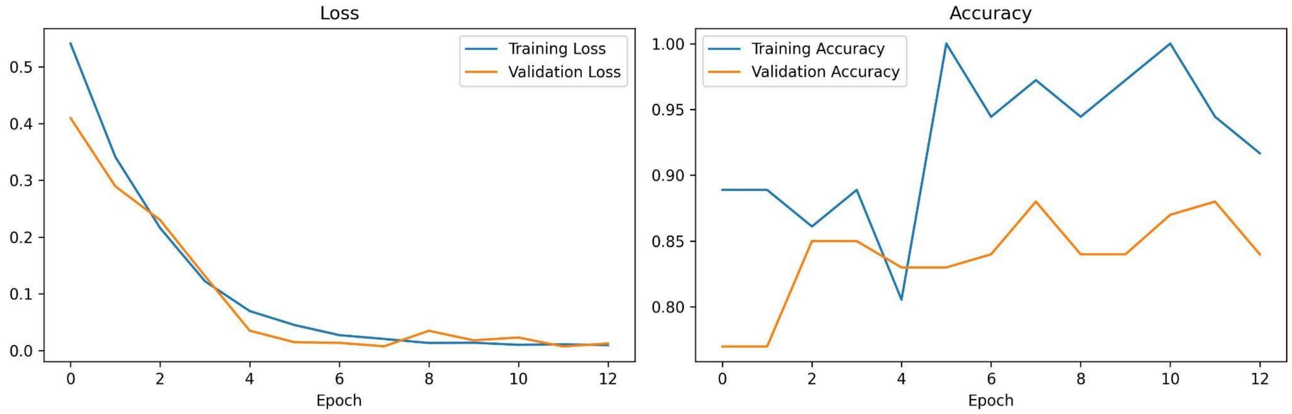


(d) MobileNetV2

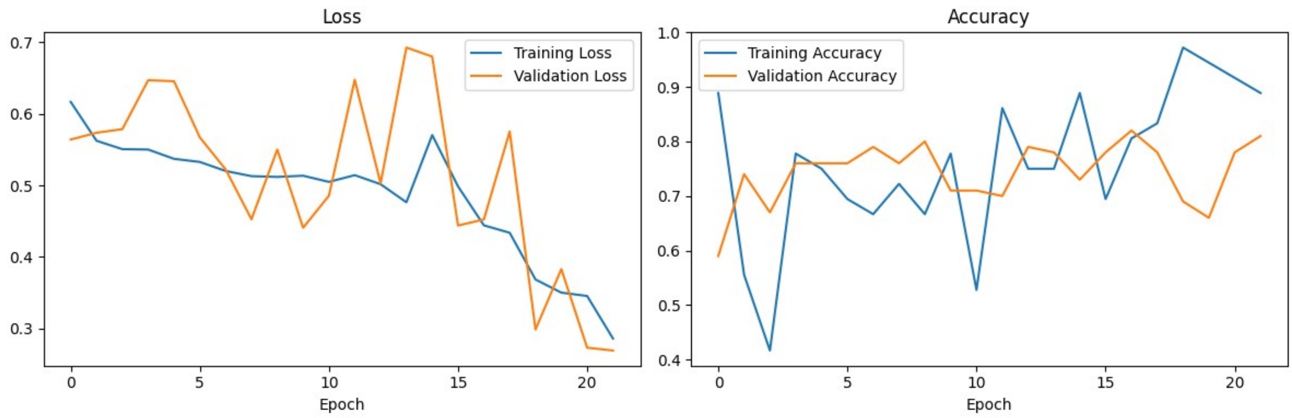


(e) MobileNetV3

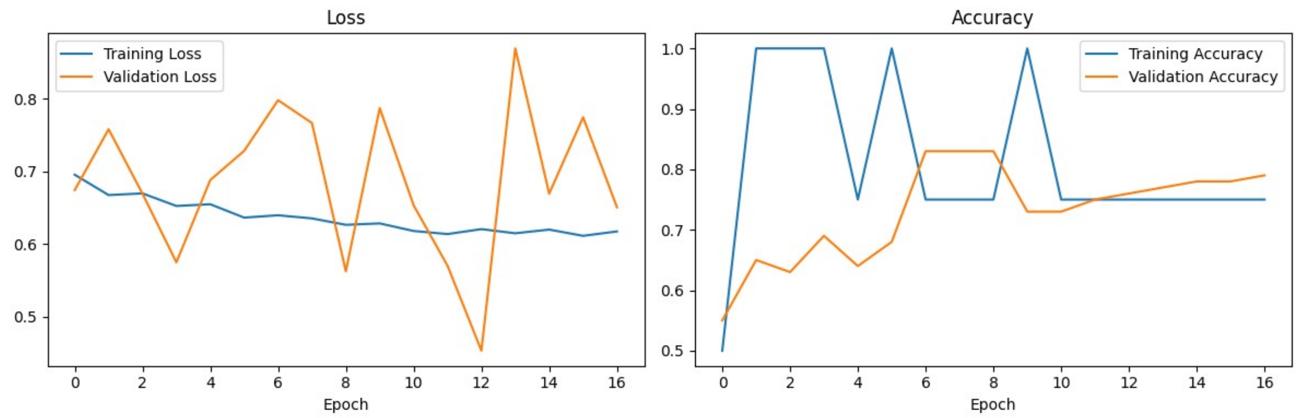
(f) AlexNet



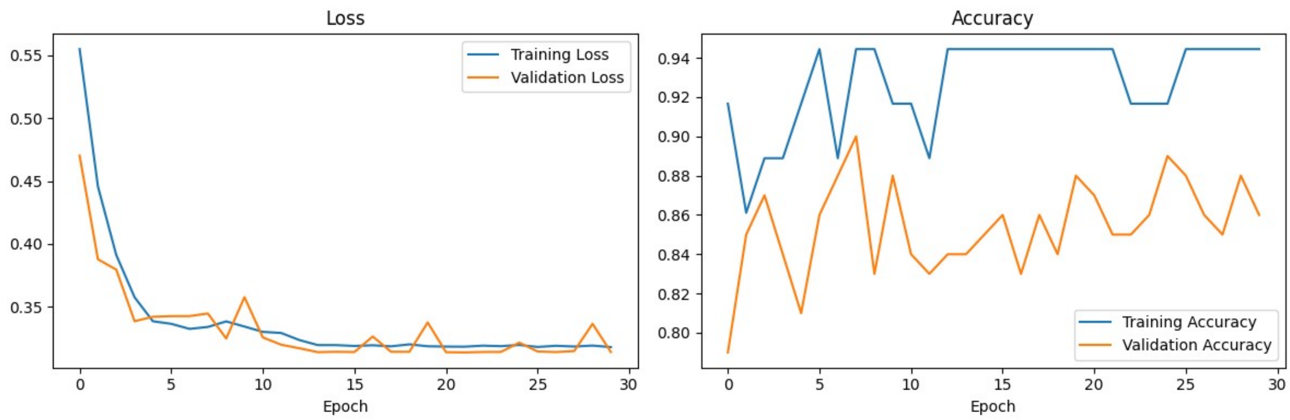
(g) InceptionV1



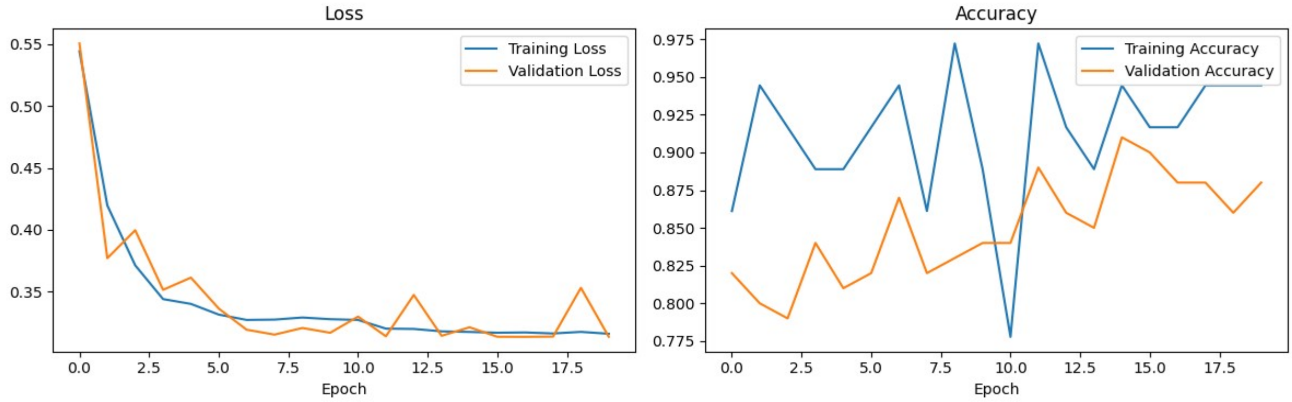
(h) SqueezeNet



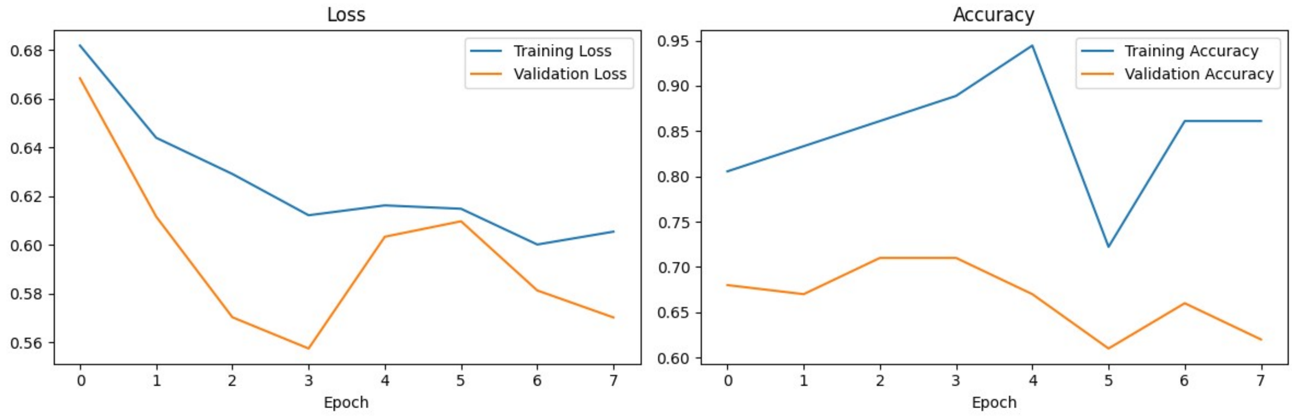
(i) EfficientNetB0



(j) DenseNet121



(k) DenseNet201



(l) VGG16

Fig. 3. Loss vs. Accuracy Graph for the Models Used in this Research

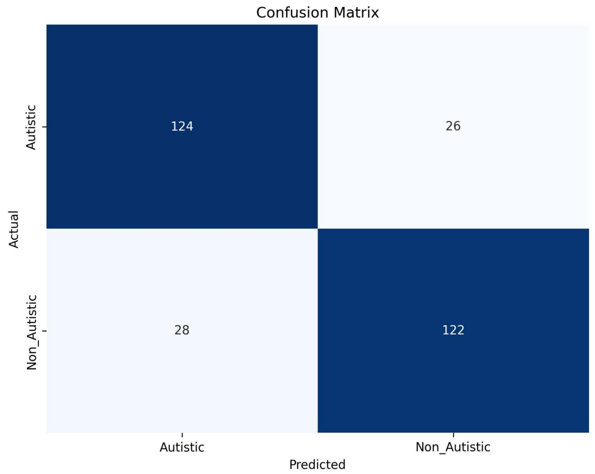
This section presents the test results of the experiments conducted to detect ASD. Table 1 summarizes the testing results of the used deep learning models. The performance results for various pre-trained deep learning models applied to the classification of Autism Spectrum Disorder (ASD) using facial images reveal significant differences in accuracy, precision, recall, and F1 score. DenseNet121 stands out as the most effective model, achieving the highest accuracy (90.33%), precision (92.00%), recall (92.00%), and F1 score (90.00%), making it the most suitable choice for this task. Among the DenseNet architectures, DenseNet121 outperformed DenseNet201, which achieved an accuracy of 89.00% and a slightly lower recall (88.00%). This indicates that the smaller architecture of DenseNet121 might generalize better for the given dataset compared to its larger counterpart. MobileNetV2 and ResNet-101 also demonstrated strong performance, with MobileNetV2 achieving

89.67% accuracy and 90.00% precision, and ResNet-101 closely following with 88.67% accuracy and 89.00% precision and recall. These results suggest that lightweight models like MobileNetV2 can balance efficiency and performance effectively, making them potential alternatives when computational resources are limited. While AlexNet, Inception V1, and MobileNetV3 achieved comparable accuracies (88.00%, 88.00%, and 87.67% respectively), they exhibited variability in precision and recall. For instance, InceptionV1 achieved a higher recall (90.00%) but lower precision (87.00%) than AlexNet, indicating a trade-off between correctly identifying true positives and avoiding false positives. ResNet-152 showed a noticeable drop in performance, with an accuracy of 82.00%, likely due to overfitting on the dataset. Similarly, SqueezeNet and EfficientNet underperformed compared to the top models, with accuracies of 86.00% and 85.33%, respectively. These results highlight that not all architectures are equally effective for ASD classification, particularly when the dataset characteristics do not align with the model's strengths. VGG16 demonstrated the lowest performance across all metrics, with an accuracy of 75.33% and a recall of only 71.00%, reflecting challenges in effectively extracting features from facial images for ASD classification. The results affirm that DenseNet121, with its ability to extract meaningful features and maintain a balanced trade-off among evaluation metrics, is the optimal choice for this application.

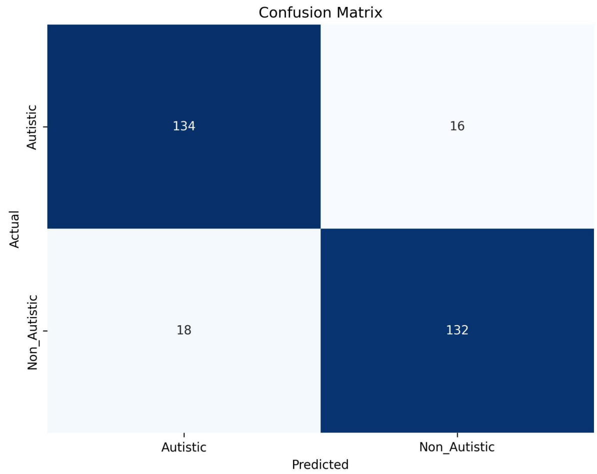
Table 1
Performances of the pre-trained deep learning models

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet-50	86.00	86.00	86.00	86.00
ResNet-101	88.67	89.00	89.00	89.00
ResNet-152	82.00	82.00	83.00	82.00
MobileNetV2	89.67	90.00	89.00	90.00
MobileNetV3	87.67	86.00	89.00	88.00
AlexNet	88.00	88.00	88.00	88.00
Inception V1	88.00	87.00	90.00	88.00
SqueezeNet	86.00	86.00	85.00	86.00
EffecientNet	85.33	86.00	85.00	85.00
DenseNet121	90.33	92.00	92.00	90.00
DenseNet201	89.00	90.00	88.00	89.00
VGG16	75.33	78.00	71.00	74.00

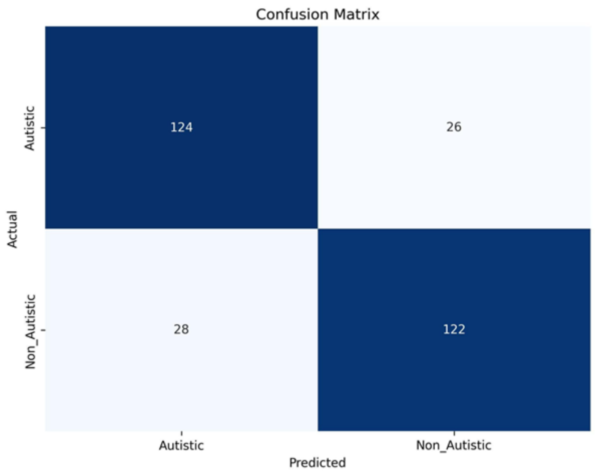
Figure 4 shows the confusion matrix of all implemented pre-trained deep learning models.



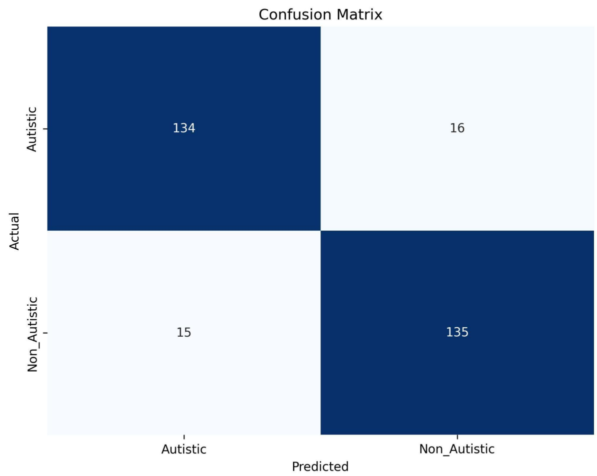
(a) ResNet-50



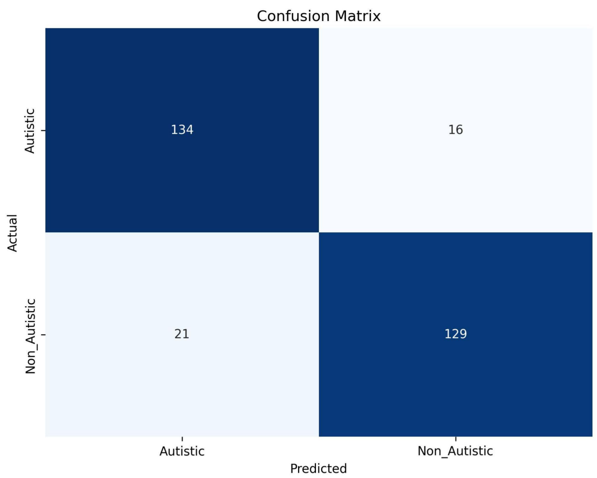
(b) ResNet-101



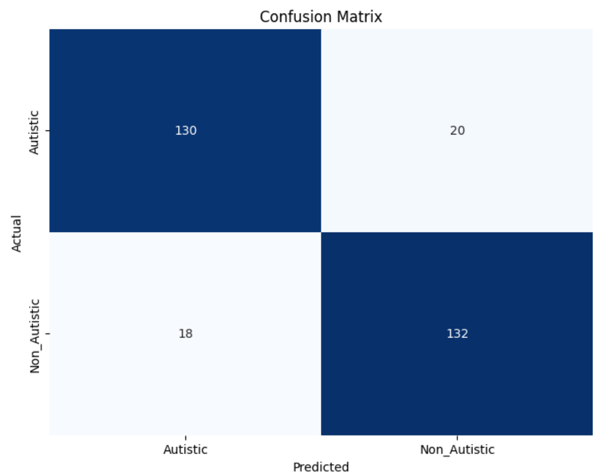
(c) ResNet-152



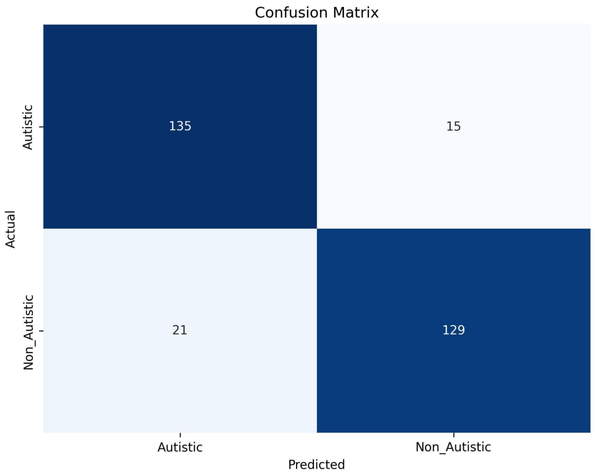
(d) MobileNetV2



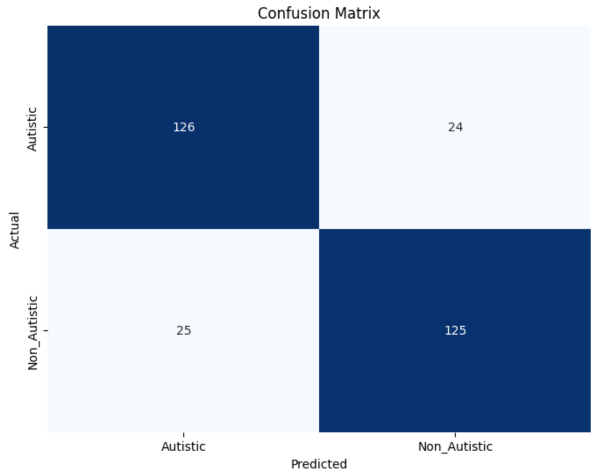
(e) MobileNetV3



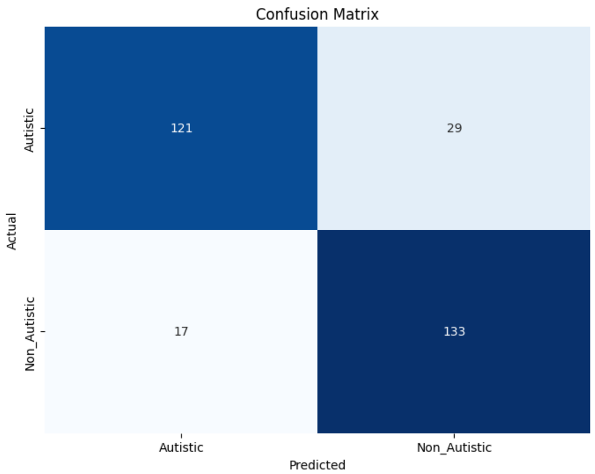
(f) AlexNet



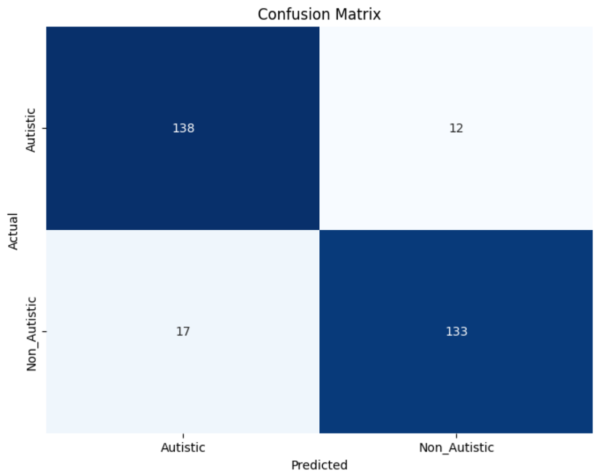
(g) InceptionV1



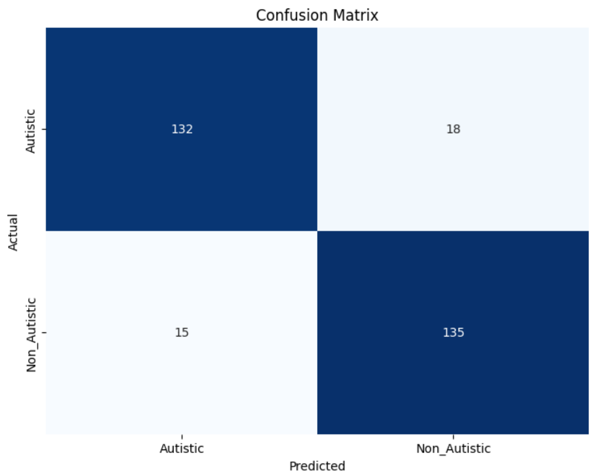
(h) SqueezeNet



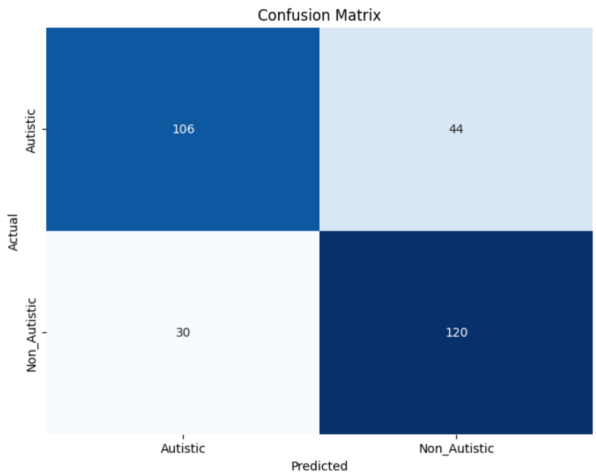
(i) EffecientNetB0



(j) DenseNet121



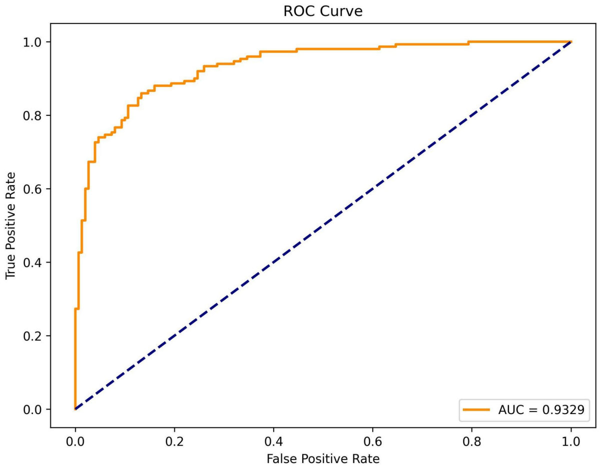
(k) DenseNet201



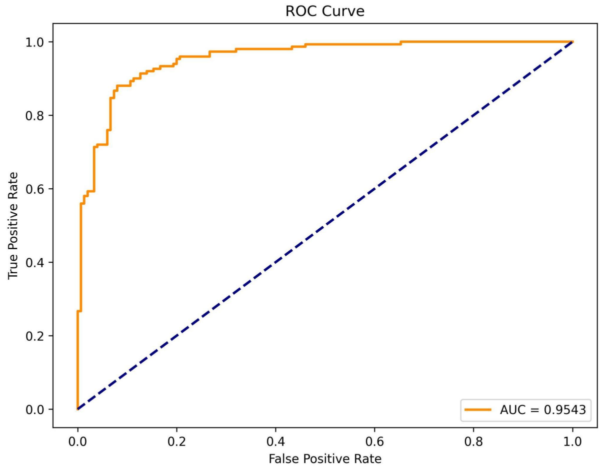
(l) VGG16

Fig. 4. Confusion matrix of the deep learning models

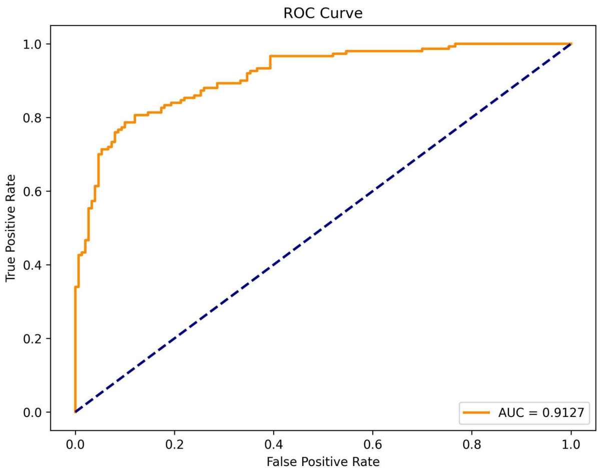
Figure 5 shows ROC and AUC of all implemented models.



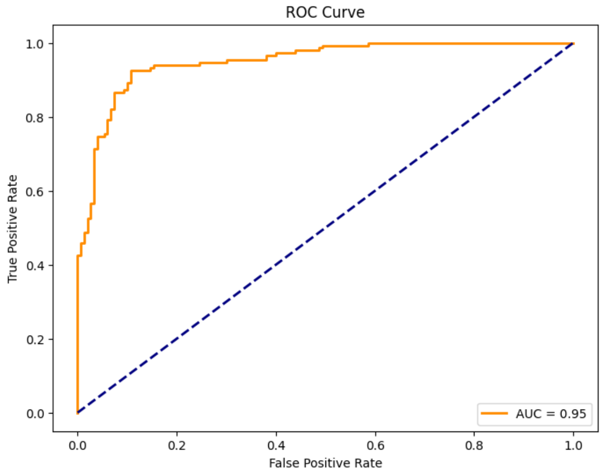
(a) ResNet-50



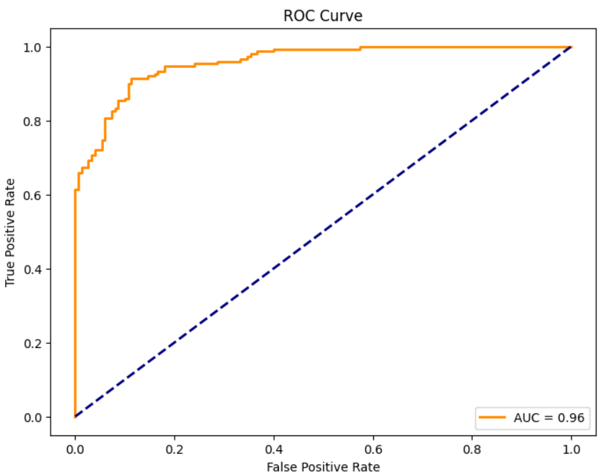
(b) ResNet-101



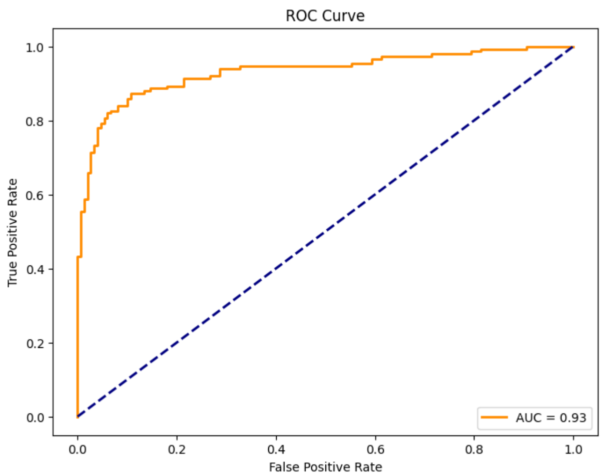
(c) ResNet-152



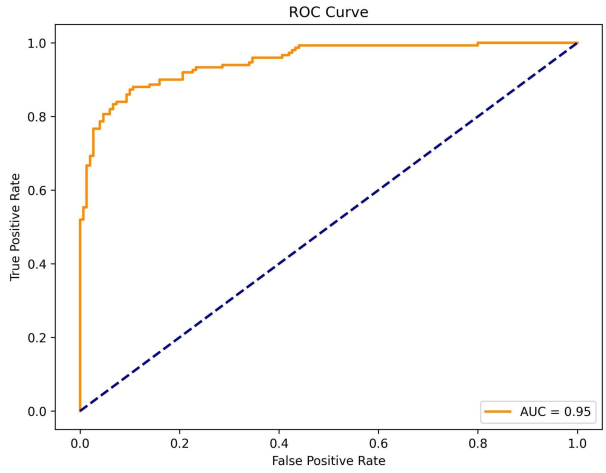
(d) MobileNetV2



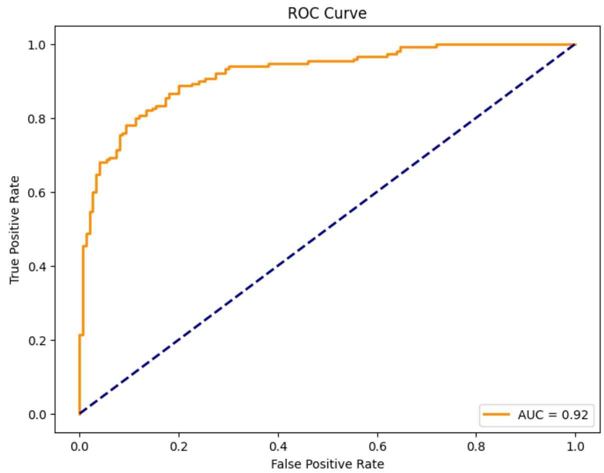
(e) MobileNetV3



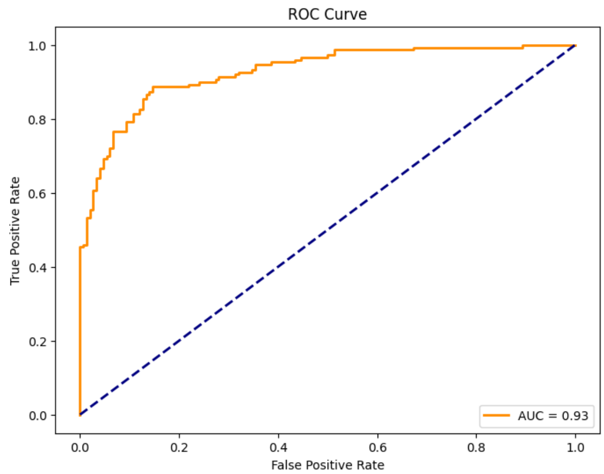
(f) AlexNet



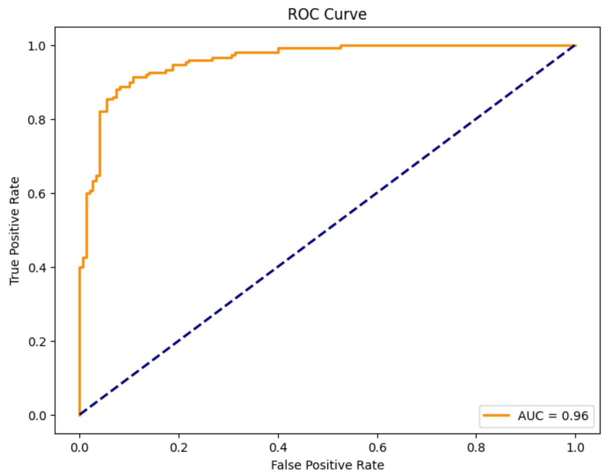
(g) InceptionV1



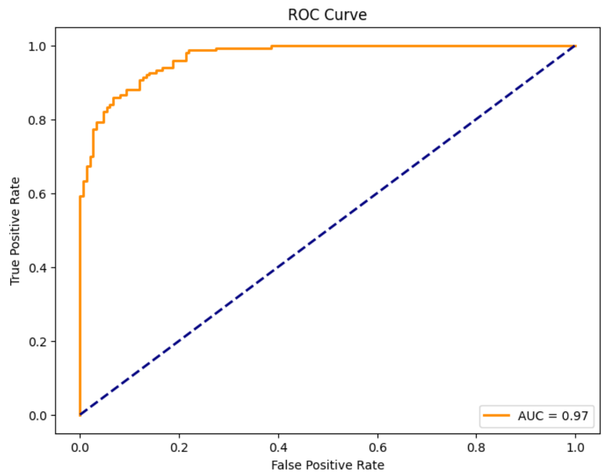
(h) SqueezeNet



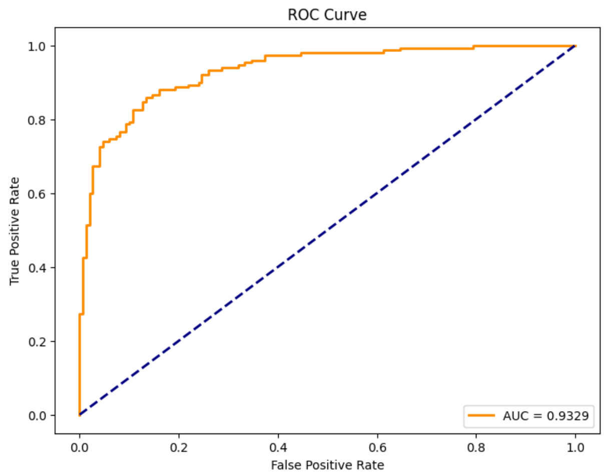
(i) EffecientNetB0



(j) DenseNet121



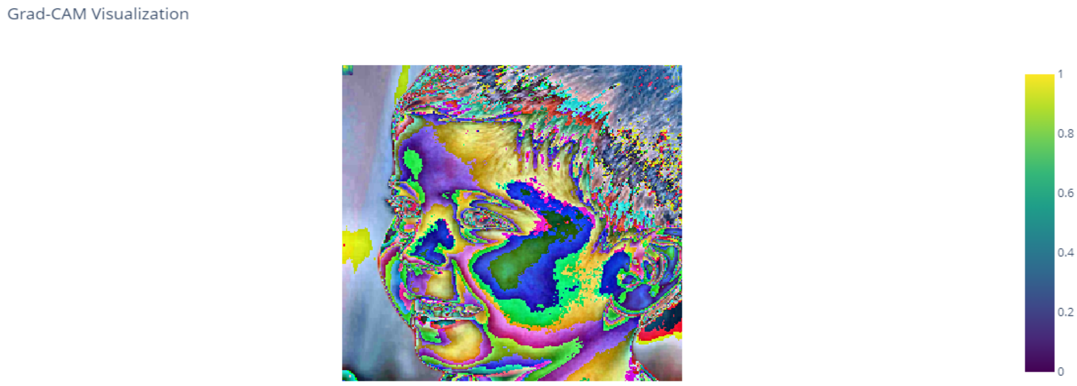
(k) DenseNet201



(l) VGG16

Fig. 5. ROC curve of the deep learning models

In Figure 6 the Grad-CAM output of the models has been shown.



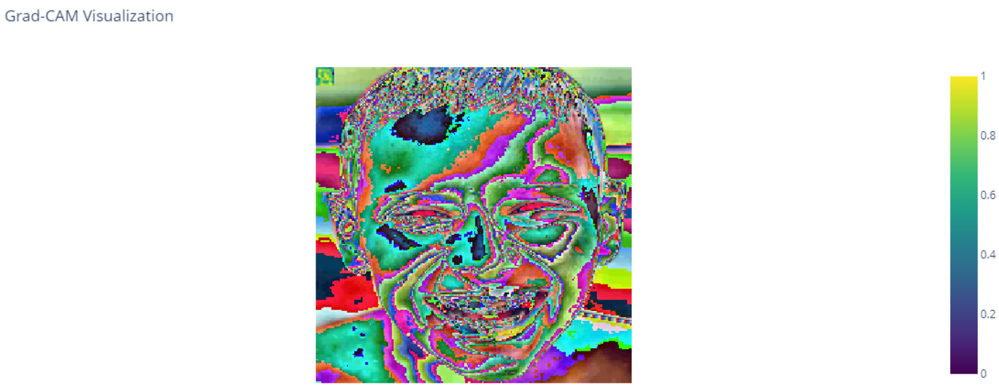
(a) ResNet-50

(b) ResNet-101

(c) ResNet-152

(d) MobileNetV2

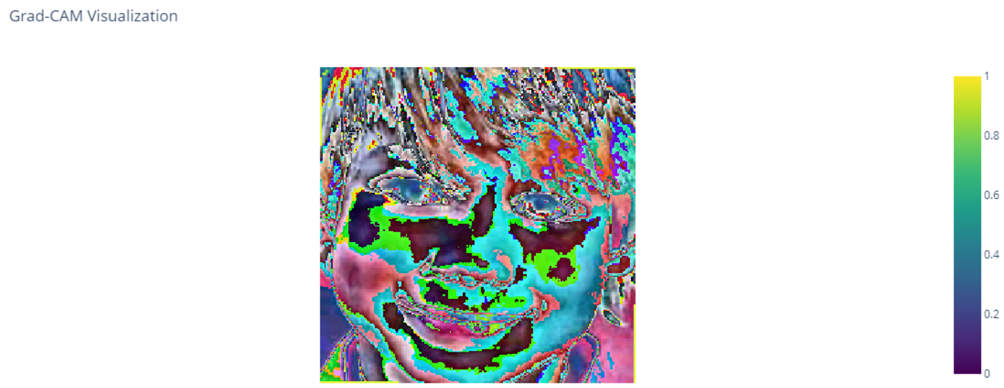
(e) MobileNetV3



(f) AlexNet

(g) Inception V1

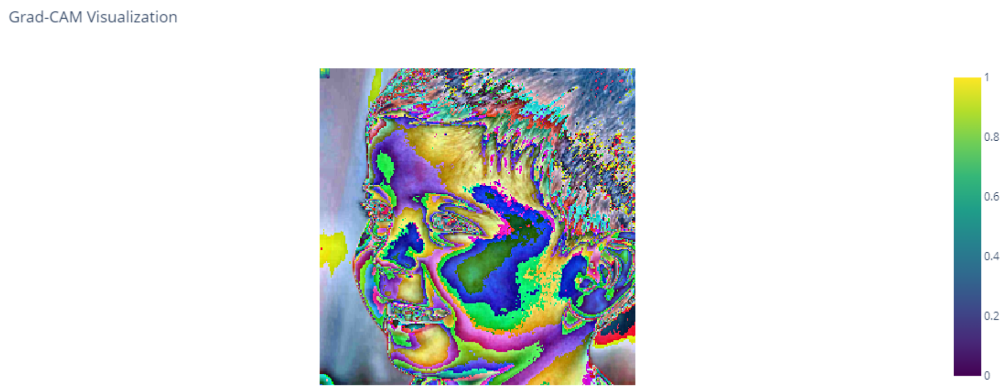
(h) SqueezeNet



(i) EffecientNetB0

(j) DenseNet121

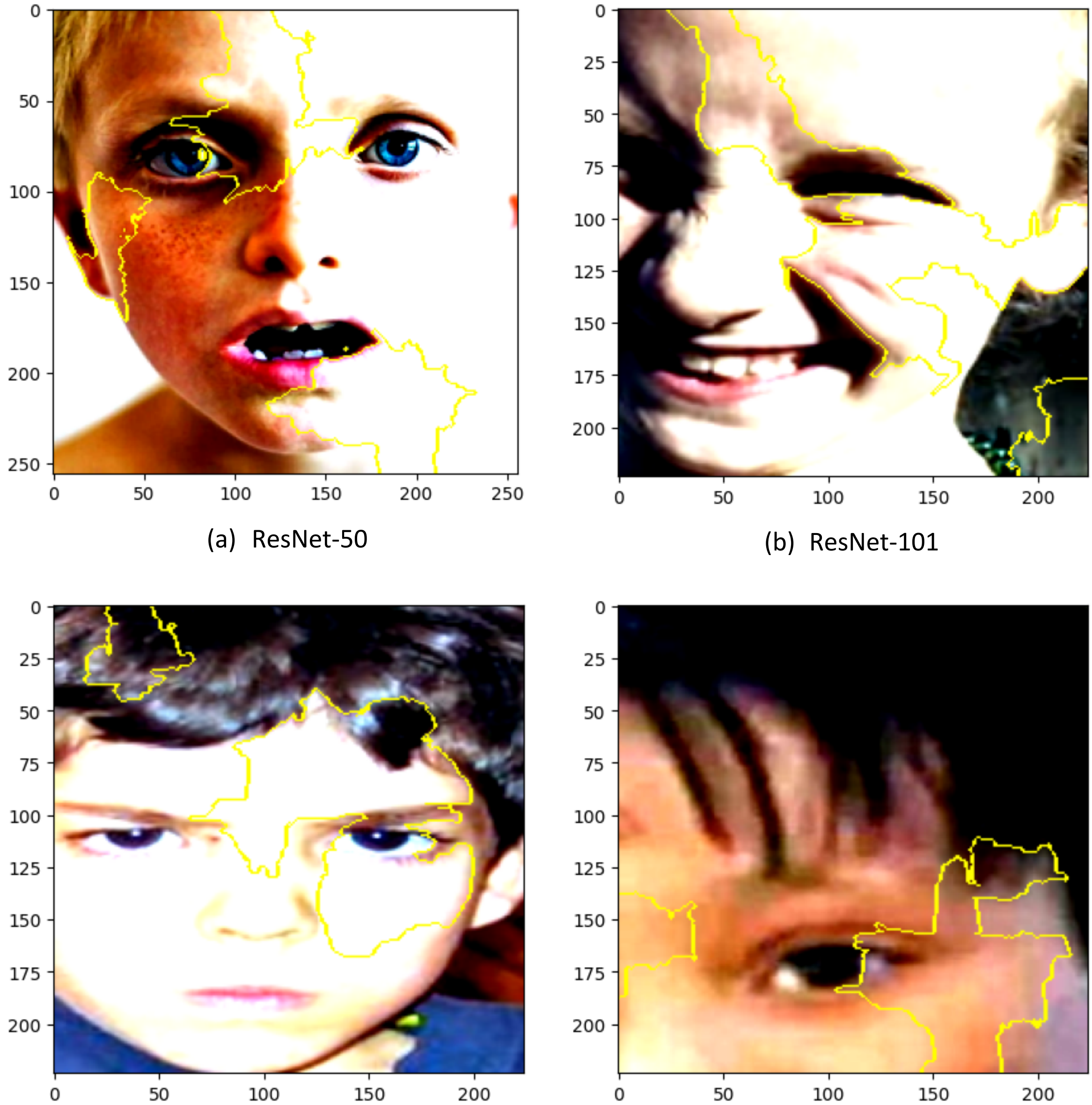
(k) DenseNet201



(l) VGG16

Fig. 6. Grad-Cam visualization of the deep learning models

Another Explainable AI – LIME output has been shown in Figure 7.

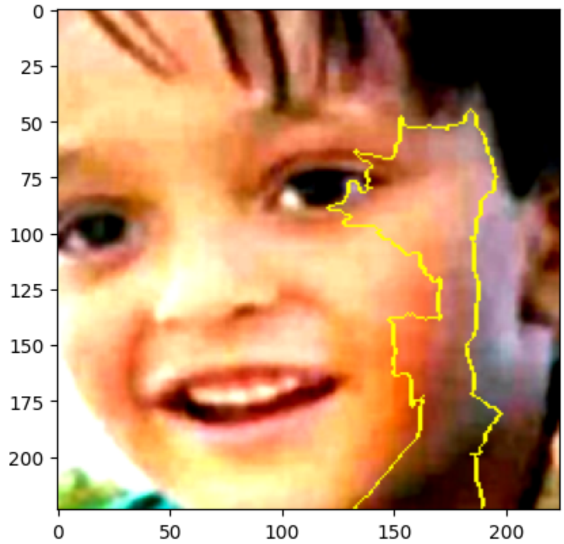


(a) ResNet-50

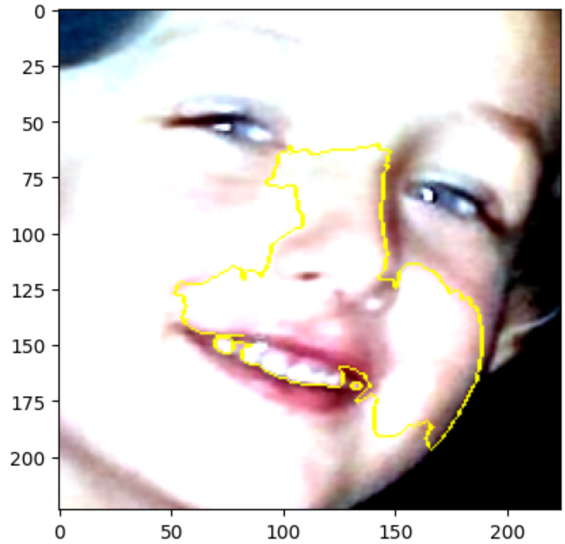
(b) ResNet-101

(c) ResNet-152

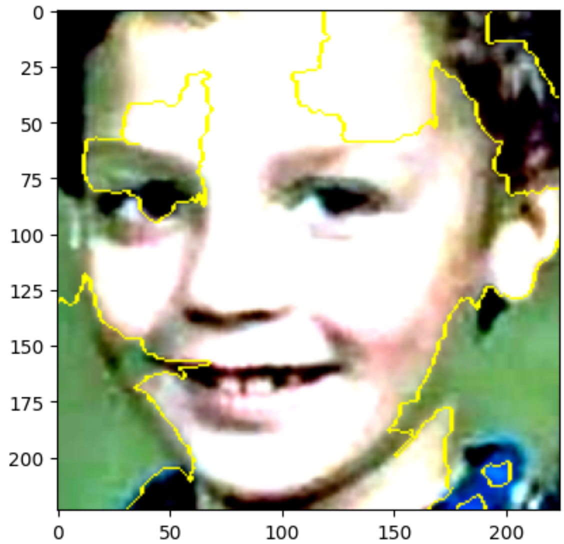
(d) MobileNetV2



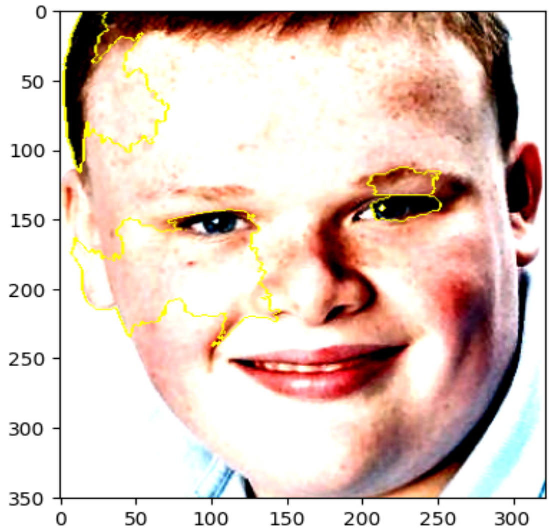
(e) MobileNetV3



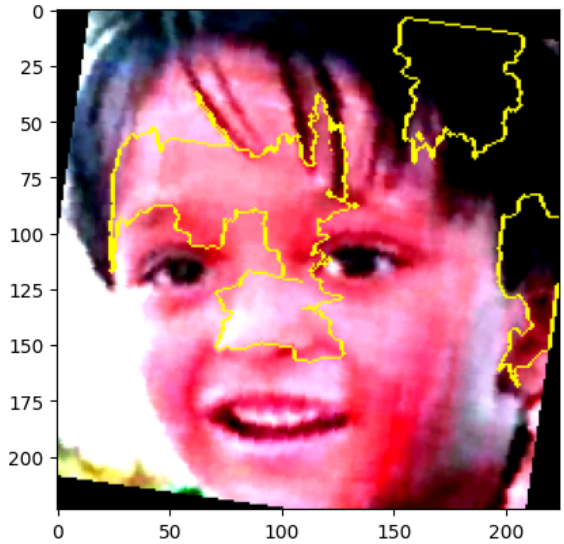
(f) AlexNet



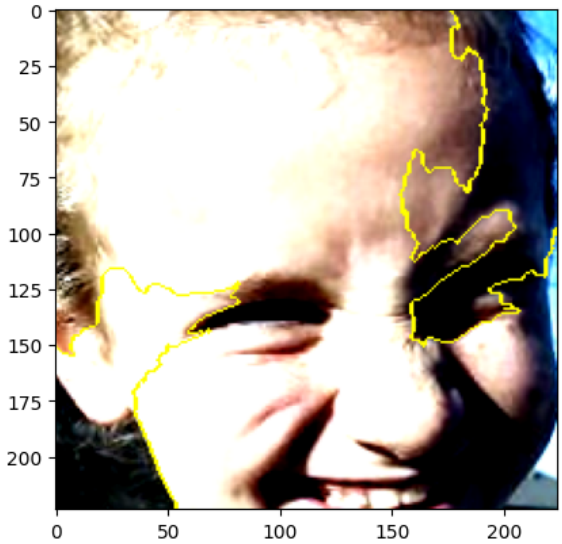
(g) InceptionV1



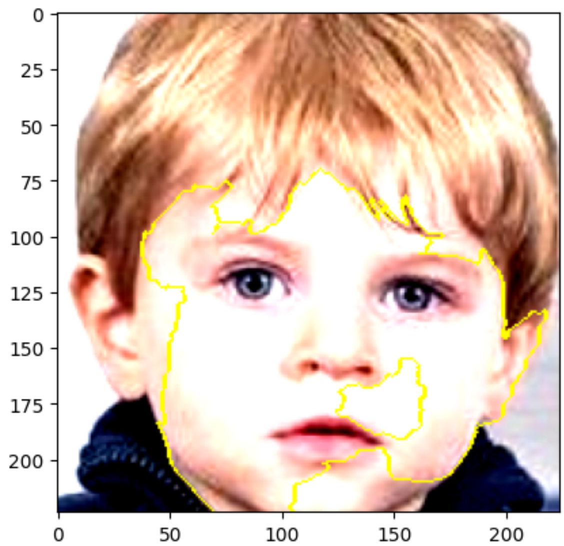
(h) SqueezeNet



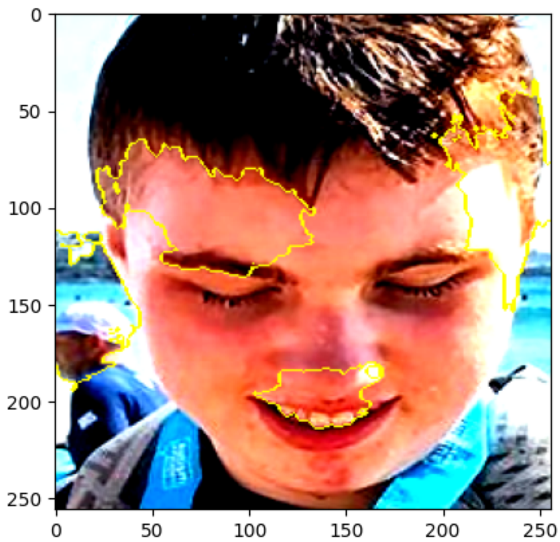
(i) EffecientNetB0



(j) DenseNet121



(k) DenseNet201



(l) VGG16

Fig. 7. LIME visualization of the deep learning models

The comparative analysis in Table 2 highlights the effectiveness of the proposed DenseNet121 model over previously reported approaches for Autism Spectrum Disorder (ASD) classification using facial images. The DenseNet121 model achieved the highest accuracy of 90.33%, surpassing both Tamilarasi *et al.,* [26] (89.00%) and Jahanara *et al.,* [27] (84.67%).

Tamilarasi *et al.,* [27] employed ResNet50, achieving an accuracy of 89.00%. While ResNet50 is a powerful feature extractor, the DenseNet121 architecture leverages dense connectivity to enhance feature reuse and mitigate gradient vanishing, which likely contributed to its superior performance. Additionally, no explanation method was provided in Tamilarasi *et al.,*'s [26] approach, limiting its interpretability and trustworthiness for deployment in real-world scenarios. Jahanara *et al.,* [27], using VGG19, reported an accuracy of 84.67%, which is significantly lower than the proposed model. This performance gap highlights the limitations of the VGG architecture in handling complex feature representations in facial image data for ASD classification. Furthermore, similar to Tamilarasi *et al.,* [26], no explainability mechanism was implemented, making it challenging to understand the model's decision-making process.

In contrast, the proposed DenseNet121 model not only outperformed the comparative models in accuracy but also integrated Explainable AI (XAI) techniques, namely Grad-Cam and LIME. These methods enable the identification of critical facial regions that influence predictions, thereby enhancing transparency and user trust. This added interpretability is especially valuable in the healthcare domain, where understanding model decisions is essential for ethical and clinical acceptance. Overall, the proposed DenseNet121-based approach demonstrates its superiority in both performance and explainability, making it a robust and reliable framework for ASD identification. By combining high accuracy with interpretability, this work sets a new benchmark for ASD diagnosis using facial images, particularly in contexts requiring accessible and interpretable solutions.

Table 2
Comparison between the proposed work and previous works

Model Used	Accuracy (%)	Model Explanation
Tamilarasi *et al.,* [26] (ResNet50)	89.00	N/A
Jahanara *et al.,* [27] (VGG19)	84.67	N/A
DenseNet121 (Proposed Model)	90.33	Grad-Cam and LIME

* N/A: Not Appropriately defined

5. Conclusions

This research introduces a novel approach for identifying Autism Spectrum Disorder (ASD) using facial images, addressing the limitations of traditional diagnostic methods such as MRI and CT scans, which are often expensive, time-consuming, and inaccessible in underprivileged areas. By leveraging pre-trained deep learning models and Explainable AI (XAI) techniques, the study achieves both high classification accuracy and interpretability, making it a significant step forward in AI-driven healthcare solutions. Among the twelve state-of-the-art models evaluated, DenseNet121 emerged as the most effective, achieving the highest accuracy of 90.33%, along with precision, recall, and F1-score values of 92.00%, 92.00%, and 90.00%, respectively. The model's superior performance is attributed to its dense connectivity architecture, which enhances feature reuse and improves gradient flow. Explainability was achieved through Local Interpretable Model-Agnostic Explanations (LIME) and Gradient-weighted Class Activation Mapping (Grad-Cam), which highlighted critical facial regions influencing the model's predictions. This interpretability ensures transparency and fosters trust in the diagnostic process, which is crucial in medical applications. Compared to existing works, the proposed DenseNet121-based approach outperformed models such as ResNet50 and VGG19 in accuracy while providing additional interpretability. This combination of high performance and transparency makes the approach a promising tool for early and accessible ASD diagnosis, particularly in resource-constrained settings. This study not only underscores the potential of facial image-based ASD identification but also highlights the importance of integrating Explainable AI to bridge the gap between technological innovation and clinical practice. Future research could focus on expanding the dataset, exploring ensemble methods, and integrating additional multimodal data to further enhance performance and applicability. By addressing critical challenges in ASD diagnosis, this work contributes to advancing AI's role in equitable and effective healthcare.

References

[1] Thabtah, Fadi, Neda Abdelhamid, and David Peebles. "A machine learning autism classification based on logistic regression analysis." *Health information science and systems* 7, no. 1 (2019): 12. https://doi.org/10.1007/s13755-019-0073-5

[2] Wiggins, Lisa D., Ann Reynolds, Catherine E. Rice, Eric J. Moody, Pilar Bernal, Lisa Blaskey, Steven A. Rosenberg, Li-Ching Lee, and Susan E. Levy. "Using standardized diagnostic instruments to classify children with autism in the study to explore early development." *Journal of autism and developmental disorders* 45 (2015): 1271-1280.

[3] Hasan, M. S. "Autism spectrum disorder and its care in Bangladesh." *The Daily Star [Internet]* 26 (2020).

[4] Kosmicki, J. A., V. Sochat, M. Duda, and D. P. Wall. "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning." *Translational psychiatry* 5, no. 2 (2015): e514-e514. https://doi.org/10.1038/tp.2015.7

[5] Thabtah, Fadi. "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward." *Informatics for Health and Social Care* 44, no. 3 (2019): 278-297. https://doi.org/10.1080/17538157.2017.1399132

[6] Mostafa, Sakib, Lingkai Tang, and Fang-Xiang Wu. "Diagnosis of autism spectrum disorder based on eigenvalues of brain networks." *Ieee Access* 7 (2019): 128474-128486. https://doi.org/10.1109/ACCESS.2019.2940198

[7] Wang, Canhua, Zhiyong Xiao, Baoyu Wang, and Jianhua Wu. "Identification of autism based on SVM-RFE and stacked sparse auto-encoder." *Ieee Access* 7 (2019): 118030-118036. https://doi.org/10.1109/ACCESS.2019.2936639

[8] Akter, Tania, Md Shahriare Satu, Md Imran Khan, Mohammad Hanif Ali, Shahadat Uddin, Pietro Lio, Julian MW Quinn, and Mohammad Ali Moni. "Machine learning-based models for early stage detection of autism spectrum disorders." *IEEE Access* 7 (2019): 166509-166527. https://doi.org/10.1109/ACCESS.2019.2952609

[9] Dominic, Nicholas, Tjeng Wawan Cenggoro, and Bens Pardamean. "Machine Learning Approaches in Detecting Autism Spectrum Disorder." *Procedia Computer Science* 227 (2023): 1070-1076. https://doi.org/10.1016/j.procs.2023.10.617

[10] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016. https://doi.org/10.1109/CVPR.2016.90

[11] Pustokhin, Denis A., Irina V. Pustokhina, Phuoc Nguyen Dinh, Son Van Phan, Gia Nhu Nguyen, Gyanendra Prasad Joshi, and Shankar K. "An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19." *Journal of Applied Statistics* 50, no. 3 (2023): 477-494. https://doi.org/10.1080/02664763.2020.1849057

[12] Ichim, Loretta, and Dan Popescu. "Melanoma detection using an objective system based on multiple connected neural networks." *IEEE Access* 8 (2020): 179189-179202. https://doi.org/10.1109/ACCESS.2020.3028248

[13] Sandler, Mark, Andrew Howard, M. Zhu, A. Zhmoginov, and L. C. Chen. "Mobilenetv2: The next generation of on-device computer vision networks." *Google AI Blog* 3 (2018).

[14] Seidaliyeva, Ulzhalgas, Daryn Akhmetov, Lyazzat Ilipbayeva, and Eric T. Matson. "Real-time and accurate drone detection in a video with a static background." *Sensors* 20, no. 14 (2020): 3856. https://doi.org/10.3390/s20143856

[15] Qian, Siying, Chenran Ning, and Yuepeng Hu. "MobileNetV3 for image classification." In *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, pp. 490-497. IEEE, 2021. https://doi.org/10.1109/ICBAIE52039.2021.9389905

[16] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).

[17] Siddhesh Bangar, AlexNet Architecture Explained, https://medium.com/@siddheshb008/alexnet-architecture-explained-b6240c528bd5

[18] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9. 2015. https://doi.org/10.1109/CVPR.2015.7298594

[19] Hazarika, Ruhul Amin, Debdatta Kandar, and Arnab Kumar Maji. "An experimental analysis of different deep learning based models for Alzheimer's disease classification using brain magnetic resonance images." *Journal of King Saud University-Computer and Information Sciences* 34, no. 10 (2022): 8576-8598. https://doi.org/10.1016/j.jksuci.2021.09.003

[20] Iandola, Forrest N. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." *arXiv preprint arXiv:1602.07360* (2016).

[21] Alhichri, Haikel, Asma S. Alswayed, Yakoub Bazi, Nassim Ammour, and Naif A. Alajlan. "Classification of remote sensing images using EfficientNet-B3 CNN model with attention." *IEEE access* 9 (2021): 14078-14094. https://doi.org/10.1109/ACCESS.2021.3051085

[22] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. "Densely connected convolutional networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700-4708. 2017. https://doi.org/10.1109/CVPR.2017.243

[23] Ji, Qingge, Jie Huang, Wenjie He, and Yankui Sun. "Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images." *Algorithms* 12, no. 3 (2019): 51. https://doi.org/10.3390/a12030051

[24] Jaiswal, Aayush, Neha Gianchandani, Dilbag Singh, Vijay Kumar, and Manjit Kaur. "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning." *Journal of Biomolecular Structure and Dynamics* 39, no. 15 (2021): 5682-5689. https://doi.org/10.1080/07391102.2020.1788642

[25] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[26] Tamilarasi, F. Catherine, and J. Shanmugam. "Convolutional neural network based autism classification." In *2020 5th international conference on communication and electronics systems (ICCES)*, pp. 1208-1212. IEEE, 2020. https://doi.org/10.1109/ICCES48766.2020.9137905

[27] Jahanara, Shaik, and Shobana Padmanabhan. "Detecting autism from facial image." *International Journal of Advance Research, Ideas and Innovations in Technology* 7, no. 2 (2021): 219-225.

[28] Howard, Andrew G. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." *arXiv preprint arXiv:1704.04861* (2017).

[29] Sarabadani, Sarah, Larissa C. Schudlo, Ali Akbar Samadani, and Azadeh Kushski. "Physiological detection of affective states in children with autism spectrum disorder." *IEEE Transactions on Affective Computing* 11, no. 4 (2018): 588-600. https://doi.org/10.1109/TAFFC.2018.2820049

[30] Deshpande, Gopikrishna, Lauren E. Libero, Karthik R. Sreenivasan, Hrishikesh D. Deshpande, and Rajesh K. Kana. "Identification of neural connectivity signatures of autism using machine learning." *Frontiers in human neuroscience* 7 (2013): 670. https://doi.org/10.3389/fnhum.2013.00670

[31] Du, Yuhui, Hui Hao, Ying Xing, Ju Niu, and Vince D. Calhoun. "A transdiagnostic biotype detection method for schizophrenia and autism spectrum disorder based on graph kernel." In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3241-3244. IEEE, 2021. https://doi.org/10.1109/EMBC46164.2021.9629618

[32] Arumugam, Sajeev Ram, Sankar Ganesh Karuppasamy, Sheela Gowr, Oswalt Manoj, and K. Kalaivani. "A deep convolutional neural network based detection system for autism spectrum disorder in facial images." In *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pp. 1255-1259. IEEE, 2021. https://doi.org/10.1109/I-SMAC52330.2021.9641046

[33] Shill, Pintu Chandra. "An Ensemble Learning Model to Detect COVID-19 Pneumonia from Chest CT Scan." In *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-6. IEEE, 2022. https://doi.org/10.1109/ICCCNT54827.2022.9984312

[34] Shill, Pintu Chandra. "Identification of COVID-19 from Chest CT Scan Using CNN as Feature Extractor and Voting Classifier." In *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, pp. 1-6. IEEE, 2022. https://doi.org/10.1109/IICAIET55139.2022.9936837

[35] Sanim, Mostofa Shariar, and Khan Mehedi Hasan. "Identification of COVID-19 from Other Upper Respiratory Tract Infections Using Random Undersampling and LIME-based XAI Model." In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-6. IEEE, 2023. https://doi.org/10.1109/ICCCNT56998.2023.10307102

[36] Saha, Antor, Esrat Jahan Bristy, Md Rahatul Islam, Rafi Afzal, and Sadia Afrin Ridita. "LIME-based Explainable AI Models for Predicting Disease from Patient's Symptoms." In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-6. IEEE, 2023. https://doi.org/10.1109/ICCCNT56998.2023.10307223