# Multimodal Deep Learning Framework for Early Detection of Autism Spectrum Disorder in Children Using Image and Video Data

Sarika Kallur, Shradha Bhakoji, Aditya Gavade, Ayush Shetty, Dr. Salama Shahapur

## Abstract

This work proposes a deep learning-based framework for early diagnosis of ASD in children using image and video modalities. Further, state-of-the-art CNNs such as ResNet50, EfficientNetB0, and MobileNetV2 are used in this work, com- bined with transfer learning and augmentation, for robust ASD classification. Further, feature extraction using CNNs will be complemented by the Random Forest algorithm for comparison. Preprocessing and training were performed on balanced autistic and non-autistic classes by using the TensorFlow and OpenCV pipelines. The experimental results showed that MobileNetV2 gave the highest test accuracy among the three, reaching 87%, with an F1-score of 0.87%, followed by ResNet50 at 75%, and EfficientNetB0 at 73%. This thus-constructed frame- work proved that deep learning is capable of serving as an assistive diagnostic tool for the early detection of autism.

**Keywords:** Autism Spectrum Disorder, Deep Learning, Transfer Learning, Image Classification, Machine Learning.

## 1 Introduction

Autism Spectrum Disorder is a developmental condition that affects how a person communicates, behaves, and interacts socially. The diagnosis of such conditions must be done as early as possible, as it aids in offering timely support and therapy that may result in the child's development and quality of life improving a lot. In general, diagnosing autism requires trained specialists, careful observation, and often multiple sessions of evaluation. This process can be slow, subjective, and not easily available in many parts of the world.

To help solve this problem, a number of researchers are looking into the use of artificial intelligence with deep learning to automatically detect images or videos for signs of autism. Facial expressions and features often carry patterns that may be useful in identifying autistic behavior. With the help of modern image-processing and machine-learning techniques, it is possible to detect such patterns both more quickly and more precisely.

The current paper will aim to propose a deep learning model for the classification of facial images into autistic and non-autistic categories. This study will utilize EfficientNet, a CNN-based model, for the modeling purpose and OpenCV-MTCNN as preprocessing tools to process and examine the data. The aim is to develop a reliable, automated system that can support medical professionals in the early screening for autism without resorting to complex, time-consuming manual assessments.

**The paper is structured as follows:** Section 3 describes the methodology and model design, data collection, and how it was preprocessed and trained. Section 4 outlines the results and evaluation metrics. Section 5 provides conclusions and discusses possible further improvements in this area of research.

## 2 Literature work

| SN No. | Author & Year | Method / Model Used | Dataset / Input Type | Key Findings | Remarks |
|---|---|---|---|---|---|
| 1 | An unnamed study (2022) – "Automated identification of autism spectrum disorder from facial …" (Nature) | Pre-trained CNNs (transfer learning) | Facial images | Demonstrated an effective DL framework for ASD detection | Newer study, likely high performance but details limited publicly. |
| 2 | "Autism spectrum disorder detection using facial images" (2022?) (PMC) | Multiple CNNs (ResNet34, ResNet50, AlexNet, MobileNetV2, VGG16, VGG19) | Facial images from dataset | ResNet50 achieved ~92% accuracy. (PMC) | Good baseline for transfer-learning approach using faces. |
| 3 | "Facial Features Detection System To Identify Children With Autism Spectrum Disorder" (2022) (PMC) | CNN with transfer learning (MobileNet, Xception, InceptionV3) + web app | Facial images (Kaggle dataset) | Designed a web app; CNN could classify autistic vs non-autistic children. | Focus includes application aspect (web-app) besides model. |
| 4 | "A deep learning-based ensemble for autism spectrum disorder diagnosis using facial images" (2022) (PLOS) | Ensemble of CNNs (VGG16 + Xception) | Facial images | Showed high efficiency for automated screening using DL; specific accuracy not quoted in table. | Combines multiple models → potential improved robustness. |

| | | | | | |
|---|---|---|---|---|---|
| 5 | "Identification of Autism in Children Using Static Facial Features and …" (2020/21) (PMC) | Pre-trained CNNs (MobileNet, Xception, EfficientNetB0/B1/B2) as feature extractors + DNN classifier | Facial images of children | Best model (Xception) gave AUC ~96.63%, sensitivity ~88.46%. (PubMed) | Good performance metrics reported; strong reference. |
| 6 | "Efficient Deep Learning-Based Data-Centric Approach for Autism Spectrum Disorder Diagnosis …" (2023) (MDPI) | CNN + data-centric preprocessing/augmentation + Explainable AI | Facial images | Achieved ~98.9% accuracy, ~99.9% AUC. (MDPI) | Excellent results; emphasises data-centric strategy + XAI. |
| 7 | "Deep Learning Approach to Predict Autism Spectrum Disorder: a systematic review & meta-analysis" (2024) (PMC) | Meta-analysis of DL models | Multiple datasets (children) | Overall sensitivity ~0.95, specificity ~0.93, AUC ~0.98. (PMC) | Not a single model study but very useful for overview of field. |
| 8 | "Facial Image-Based Autism Detection: A Comparative Study of Deep …" (2022) (Tech Science) | Comparative of various ML/DL classifiers on facial images | Facial images | Identified cost-effective DL/ML solution for ASD detection in a specific region. | Regional focus (Pakistan) and comparative study of classifier options. |
| 9 | "A face image classification method of autistic children based on the two-phase transfer learning and multi-classifier integration" (2023) (Frontiers) | Two-phase transfer learning + multi-classifier integration (MobileNetV2 + MobileNetV3-Large) | Facial images (children) | Multi-classifier integration improved performance vs single classifier. | Good variant approach: mobile-friendly models + classifier fusion. |

Table1: Literature table

(Table1) Research on ASD detection has evolved from traditional behavioral analysis to modern computer vision-based methods. More precisely, deep learning, and especially CNNs, has shown great potential in autism prediction and medical image analysis. However, previous works mostly used models like VGG16 and ResNet on limited datasets with minimum pre-processing.

Modern CNN architectures, such as EfficientNet and MobileNet, are explored in this paper to overcome these limitations. The approach integrates image and video analysis with a strong preprocessing pipeline by using MTCNN for accurate facial detection and thereby consistent classification.

# 3 Methodology

This work presents a deep learning-based technique for the detection of ASD from facial images. The steps involved the dataset creation, which consisted of a collection of autistic and non-autistic faces, followed by face detection and cropping through MTCNN. The faces were then preprocessed by resizing to 224×224 pixels, followed by normalizations and augmentation to enhance model accuracy. The EfficientNet model is a lightweight CNN that utilizes transfer learning from ImageNet for the extraction of major facial features. The performance of the model was optimized by training with the Adam optimizer using a binary cross-entropy loss function with ReLU-sigmoid as its internal and output layer activation functions, respectively. The dataset is divided into training, validation, and testing sets. It uses early stopping when there is no improvement on the loss after five epochs. Finally, metrics such as accuracy, precision, recall, F1-score, ROC curve, and confusion matrix are considered to evaluate the proposed model's performance.

## 3.1 Data Collection

The dataset used for this project includes pictures of the faces of autistic and non-autistic persons. These images were gathered from publicly available sources and were then organized into two categories: Autistic and Non-Autistic. Each image was correctly labeled so that the model could learn the differences between both classes.


Fig1: Data distribution summary

Fig1 Shows that the dataset is divided into three parts: training, accounting for 80\%; validation, accounting for 10\%; and testing, accounting for the remaining 10\%. The training set will be used to teach the model, the validation set to tune its parameters, and the test set to evaluate its final performance.


Fig2: Data Sample

Fig2 Shows that all the images are stored separately for each category in different folders and later, after processing using OpenCV and MTCNN, the faces will be detected and cropped out. The model will be focused on the facial region rather than unnecessary details of the background.

## 3.2 Data Preprocessing

To make the dataset more balanced, data augmentation techniques like flipping, rotation, and zooming have been used. Increasing the number of samples in the training dataset also allows the model to generalize better when performing the prediction on new images.

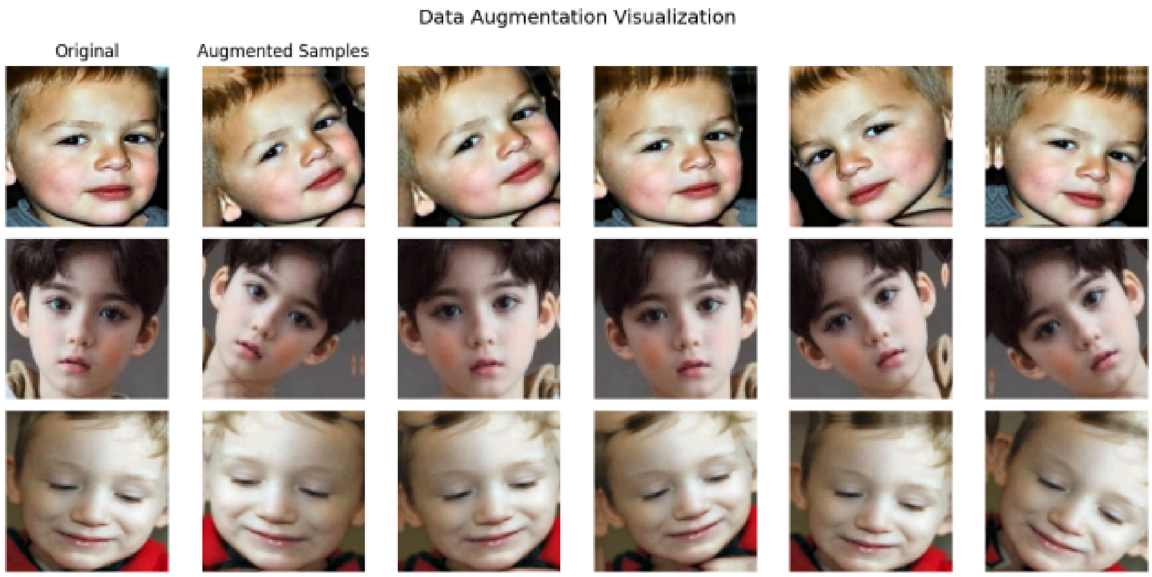



Fig3: Augmented data

First, several preprocessing steps were applied to the images after collecting the dataset to prepare them for model training. The aim is to make all the images clean, consistent, and ready for a deep learning model.

First, face detection was done using an algorithm called the MTCNN, which is an abbreviation of Multi-task Cascaded Convolutional Neural Network. This step automatically detects and crops the face region from every image, removing irrelevant background elements. It ensures that the model focuses on facial features that are most relevant for autism detection.

Then, all cropped face images were resized to the dimensions of 224×224 pixels, which is the standard input size for EfficientNet. For each image, the pixel values were then normalized between 0 and 1 to improve the stability of the model's learning process and reduce computation time.

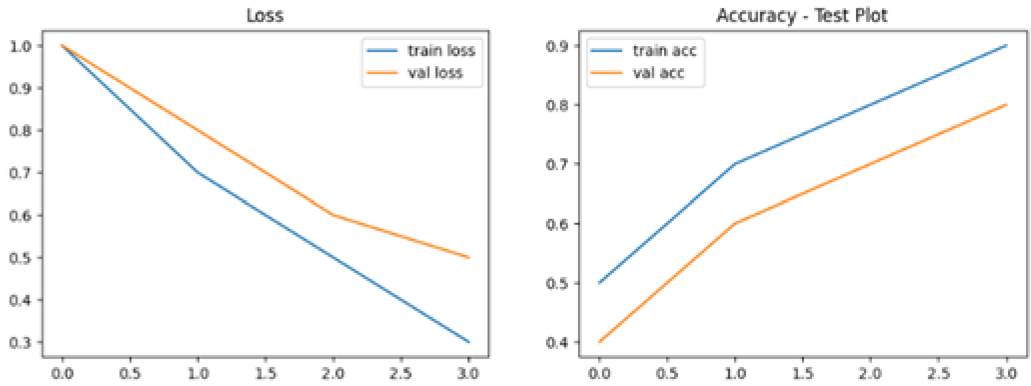



Fig4: Data preprocessing accuracy

(Fig4) To increase data diversity and prevent overfitting, all images will undergo data augmentation: random rotations, flips, zooming, and shifting. Data augmentation helps the model deal with variably oriented faces under different light conditions, enabling it to generalize and be more accurate.

Finally, preprocessed images were then divided into training, validation, and testing sets to ensure the model is trained on diverse samples and performs well on unseen data.

# 4 Model Selection

Both models have been used to compare traditional machine learning and deep learning approaches for the detection of autism by using facial images. The Random Forest model has first been implemented as a baseline approach for understanding how simple feature-based methods work. It works by creating many small decision rules and then combining them in order to come up with a final prediction. This model helps provide a starting point and acts as a benchmark for performance.

## 4.1 RandomForest

In this project, Random Forest will be used to provide a baseline that will be compared with the deep learning model. This can help in showing how a traditional machine learning method performs on the same dataset. Since Random Forest is a tree-based ensemble learning algorithm that reduces overfitting by aggregating various decision trees, it will serve as a very good starting point in testing the overall pattern recognition ability before applying deep learning. Comparing its accuracy and results with the CNN model will therefore make it evident how much improvement deep learning brings about in autism detection.

```
RF Train score: 1.0
RF Val score: 0.7895927601809954
RF Test score: 0.800453514739229
```

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Non_Autistic | 0.81 | 0.79 | 0.80 | 220 |
| Autistic | 0.79 | 0.81 | 0.80 | 221 |
| accuracy | | | 0.80 | 441 |
| macro avg | 0.80 | 0.80 | 0.80 | 441 |
| weighted avg | 0.80 | 0.80 | 0.80 | 441 |

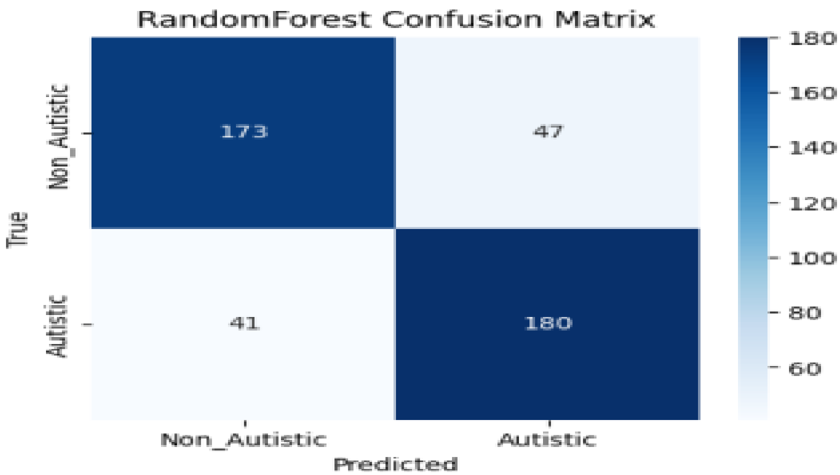



Fig5: RandomForestRegressor model

(Fig5) This deep learning model, developed on EfficientNet, goes way beyond that. It automatically learns patterns from face images during training rather than manually selecting features. It may further detect very minute details in shape, texture, and expression, which might be tough to define manually. This model learns through numerous layers and improves step by step in distinguishing between autistic and non-autistic features.

## 4.2 ResNet50

This work employed ResNet because it is able to learn deep and detailed features of facial images. The model has skip connections which help the model train efficiently without losing important information. That is why it is effective in identifying the subtle facial patterns required to distinguish between autistic and non-autistic subjects, hence improving its accuracy and reliability.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Non_Autistic | 0.88 | 0.57 | 0.69 | 220 |
| Autistic | 0.68 | 0.92 | 0.79 | 221 |
| accuracy | | | 0.75 | 441 |
| macro avg | 0.78 | 0.75 | 0.74 | 441 |
| weighted avg | 0.78 | 0.75 | 0.74 | 441 |

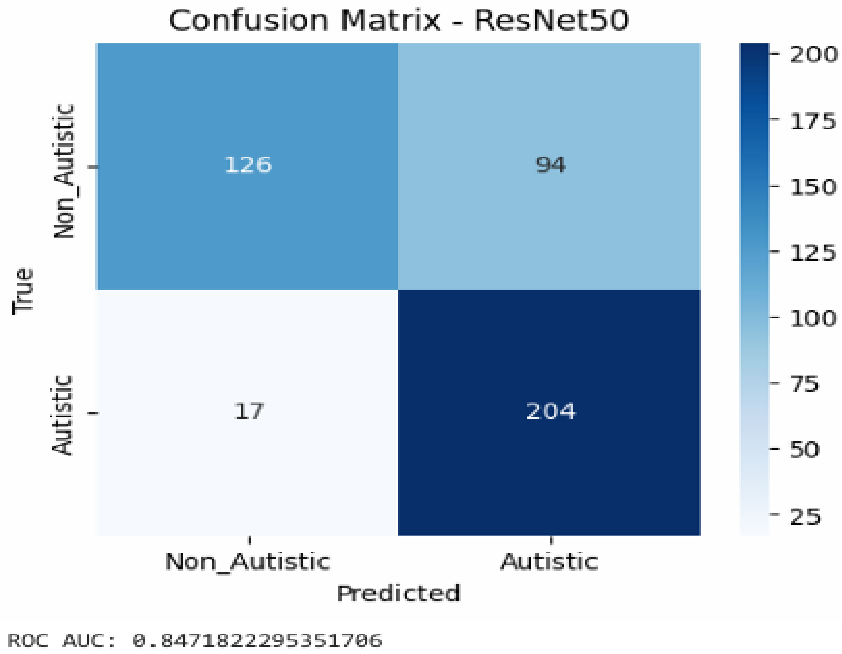



Fig6: ResNet50 model

These Fig6 reflect the performance of the ResNet50 model in classifying autistic and non-autistic children. The classification report, gives an overall accuracy of 75%, with a high precision of 0.88 for the Non-Autistic class but a clearly lower recall of 0.57, meaning many non-autistic samples are misclassified. In contrast, a very high recall of 0.92 was found for the Autistic class, indicating that this model was able to detect most autistic children, although its precision stands at 0.68, indicating false positives. The confusion matrix shows that while correctly identifying 204 autistic samples, the model misclassifies 94 non-autistic children as autistic. The bias is toward classifying cases as autistic, which leads to fewer missed diagnoses of autistic cases-a very valuable trait in early screening-but results in overclassification toward the Autistic class. The ROC AUC score of 0.847 reflects strong discriminatory ability in general, although class-level balance remains problematic.

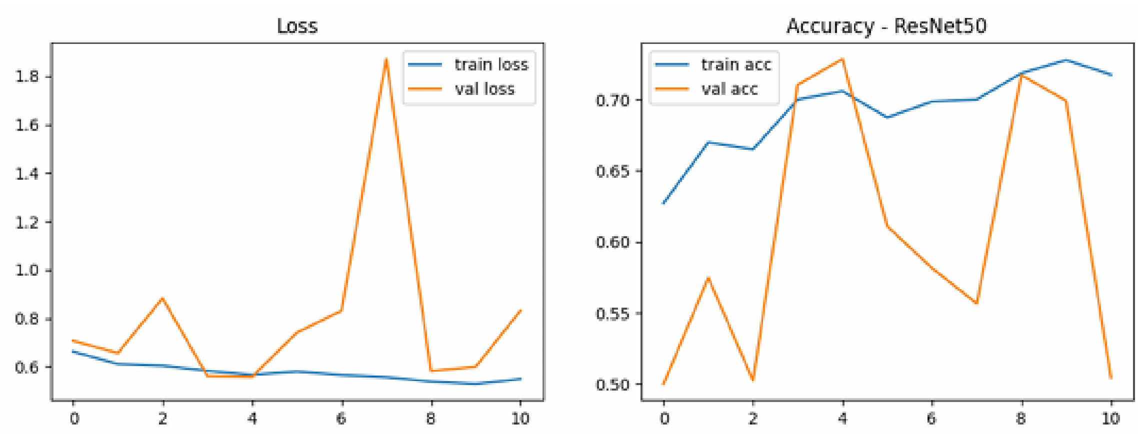


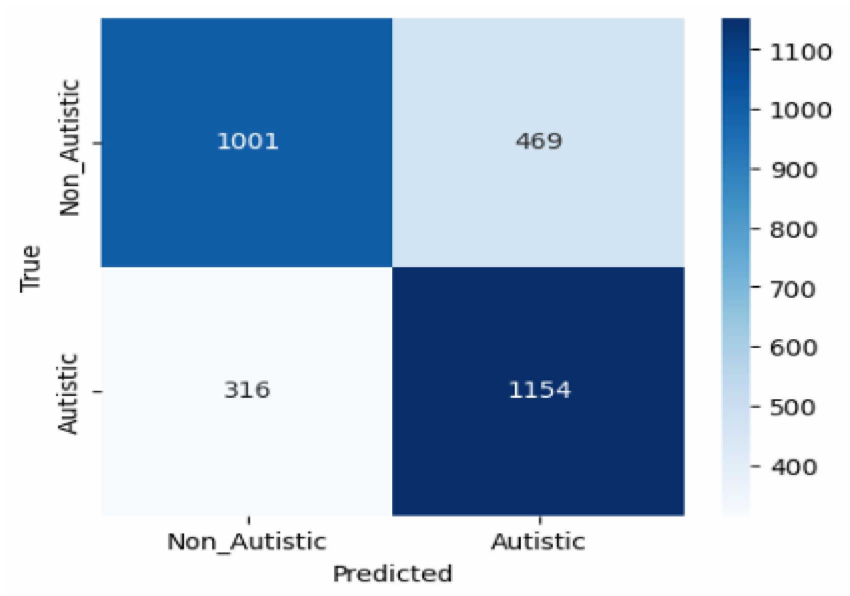

Fig7: ResNet50 model loss & accuracy

Fig7 shows the ResNet50 training and validation curves, indicating how loss and accuracy change during training epochs. The training loss has a smooth decrease, demonstrating that the model learns the patterns in the dataset well. However, the validation loss is highly variable and spikes multiple times across different epochs. This instability becomes evident in its validation accuracy, which sees massive changes between epochs, while the training accuracy itself climbs upward in a relatively smooth fashion. It suggests overfitting because the large gap between a smoothly fitting training curve and unstable validation curves suggests that the model learns the training data well but fails to generalize reliably. These patterns highlight the need for stronger regularization, more balanced data, or a lighter model architecture to achieve more stable validation performance.

## 4.3 EfficientNet

EfficientNet is used in the project because it offers high accuracy with a lesser number of parameters and lower computation compared to other deep learning models. It efficiently balances the depth, width, and resolution of the network to make it fast and lightweight. This will help the model detect fine facial features concerning autism with better performance and reduced training time, making the model suitable even for systems with limited hardware.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Non_Autistic | 0.76 | 0.68 | 0.72 | 1470 |
| Autistic | 0.71 | 0.79 | 0.75 | 1470 |
| accuracy | | | 0.73 | 2940 |
| macro avg | 0.74 | 0.73 | 0.73 | 2940 |
| weighted avg | 0.74 | 0.73 | 0.73 | 2940 |




ROC AUC: 0.8299606645379239

Fig8: EfficientNet model

Fig8 shows the evaluation metrics of EfficientNet-B0 on the test set. The classification report presents a balanced performance, with an overall accuracy of 73% and F1-scores of 0.72 for Non-Autistic and 0.75 for Autistic classes. The high recall of 0.79 in classifying the Autistic class is of critical importance during early clinical screening since the number of missed true autism cases should be low. This fact is further shown in the confusion matrix: it correctly identifies 1154 as autistic and 1001 as non-autistic samples, having less false negatives for autistic persons than false positives. Therefore, the model can be reliable in practical early-stage autism detection, and from the ROC-AUC score, an impressive 0.83 reflects great class-separation ability.

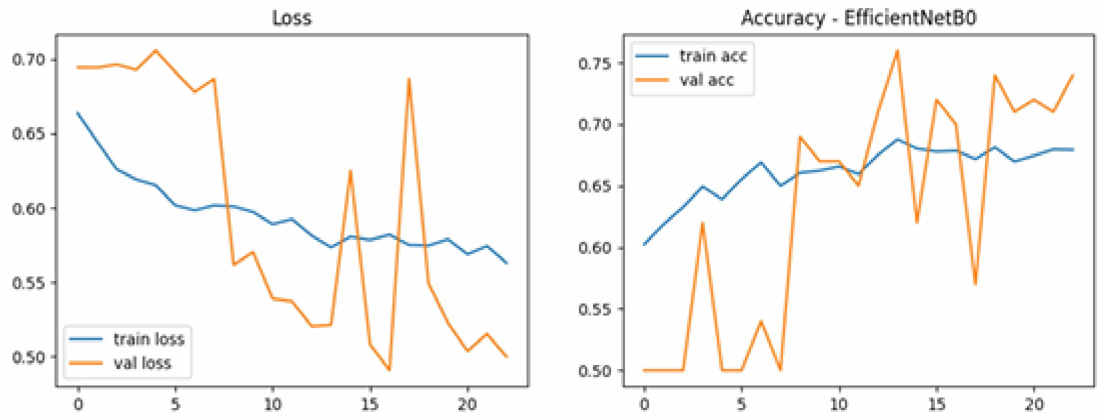



Fig9: EfficientNet model loss & accuracy

Fig9 depict the training progress of the EfficientNet-B0 model with image and video frame data for autism prediction. It is observed from the loss curves that the training loss decreases gradually, whereas the validation loss fluctuates more due to real variations in the video frames and facial expressions but still follows a general downward trend, which suggests that the model learns meaningful patterns without significant overfitting. In support, accuracy curves show that while the training accuracy improves with a gradual increase to stabilize at around 66–67%, the validation accuracy ranges between 55% and 75% due to various lighting conditions, angles, and expressions in the dataset. Despite these fluctuations, the model achieves strong peak validation accuracy, confirming its ability to generalize to unseen data.

## 4.4 MobileNet

In the current project, MobileNet has been used because it is a lightweight, efficient deep learning model intended for mobile and embedded devices. It uses depthwise separable convolutions, which reduce parameters and make the model faster with no considerable reduction in accuracy. This makes theMobileNetsuitable for autism detection systems that need quick and efficient running even on limited computing power devices.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Non_Autistic | 0.93 | 0.79 | 0.86 | 1470 |
| Autistic | 0.82 | 0.94 | 0.88 | 1470 |
| accuracy | | | 0.87 | 2940 |
| macro avg | 0.88 | 0.87 | 0.87 | 2940 |
| weighted avg | 0.88 | 0.87 | 0.87 | 2940 |

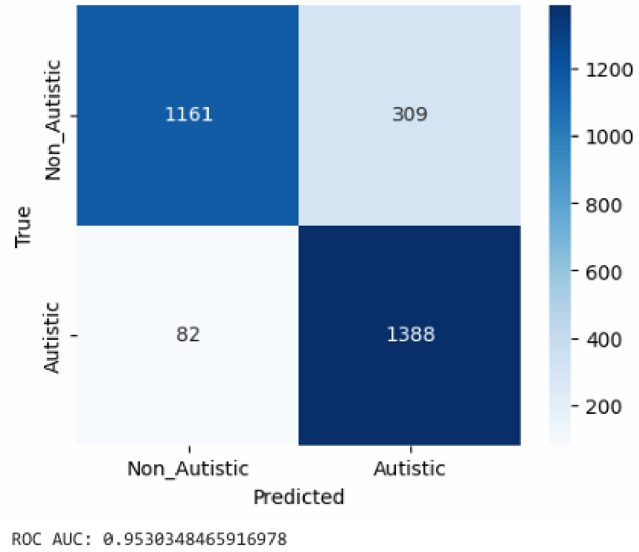



Fig10:Mobilenet Model

The Fig10 shows the confusion matrix, giving a clear picture of how well MobileNetV2 performed on the test set. The model achieved an overall accuracy of 87%, which shows strong performance. It was especially good at identifying autistic individuals, achieving a recall of 0.94, meaning it correctly detected most of the actual autistic cases. The confusion matrix also supports this by showing that the model made very few mistakes on autistic samples—only 82 were incorrectly labeled. On the other side, it also performed well with non-autistic images, with a high precision of 0.93. The ROC-AUC score of 0.95 further highlights that the model can confidently differentiate between autistic and non-autistic faces. Overall, these results show that MobileNetV2 is reliable, accurate, and effective for autism prediction using images.

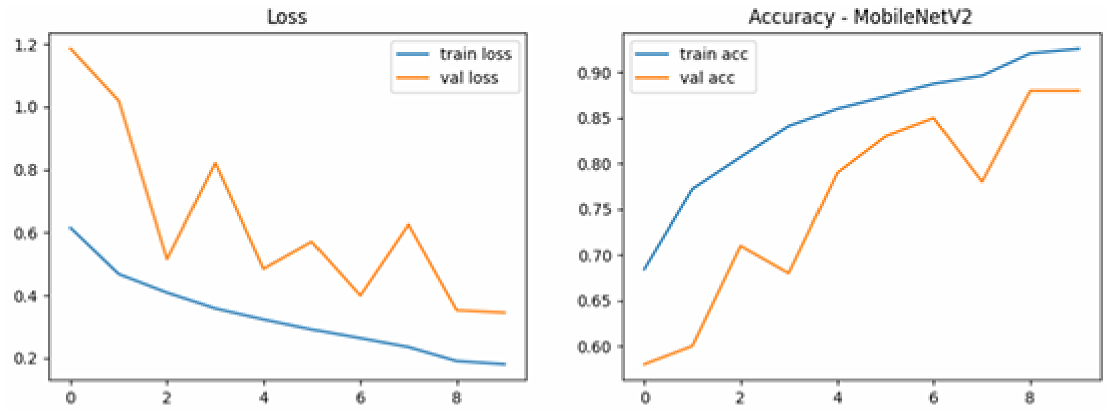



Fig11: Mobilenett model loss & accuracy

The Fig11 shows how the MobileNetV2 model learned over time during training. The loss curves clearly show that the model kept getting better with each epoch—the training loss steadily went down, and the validation loss also decreased overall, even though it had a few ups and downs. This is normal and usually happens because the validation data is more varied. Along with this, the accuracy curves show a consistent rise for both training and validation data. By the end, the model reached above 90% training accuracy and around 87% validation accuracy. Together, these curves tell us that the model was learning effectively, improving gradually, and not overfitting too much. This indicates that MobileNetV2 was able to pick up meaningful patterns from the data while generalizing well to unseen images

# 5 Model Architectures

**ResNet50**: The model employs the ResNet50 backbone to extract critical features at a 224×224 image size. On normalizing the images, ResNet50 produces feature maps of size 7×7×2048. These are then converted into a length-2048 vector using Global Average Pooling. [Fig5]It consists of dense layers with dropout to avoid overfitting and culminates in the final layer for the prediction of two classes. More than 23 million learnable parameters are part of the ResNet50 backbone; thus, this is the major constituent responsible for learning features fromimages.

| Layer (type) | Output Shape | Param # |
|---|---|---|
| input_layer_1 (InputLayer) | (None, 224, 224, 3) | 0 |
| sequential (Sequential) | (None, 224, 224, 3) | 0 |
| rescaling (Rescaling) | (None, 224, 224, 3) | 0 |
| resnet50 (Functional) | (None, 7, 7, 2048) | 23,587,712 |
| global_average_pooling2d (GlobalAveragePooling2D) | (None, 2048) | 0 |
| dropout (Dropout) | (None, 2048) | 0 |
| dense (Dense) | (None, 128) | 262,272 |
| dropout_1 (Dropout) | (None, 128) | 0 |
| dense_1 (Dense) | (None, 2) | 258 |

Fig12: ResNet50

**EfficientNetB0**: This model expects inputs of size 224×224, which are passed through EfficientNetB0; this layer learns critical features from the image. Following feature extraction, it reduces them to lower dimensions using global average pooling, after which these are passed through dense layers with dropout to avoid overfitting.

| Layer (type) | Output Shape | Param # |
|---|---|---|
| input_layer_2 (InputLayer) | (None, 224, 224, 3) | 0 |
| sequential (Sequential) | (None, 224, 224, 3) | 0 |
| rescaling_3 (Rescaling) | (None, 224, 224, 3) | 0 |
| efficientnetb0 (Functional) | (None, 7, 7, 1280) | 4,049,571 |
| global_average_pooling2d (GlobalAveragePooling2D) | (None, 1280) | 0 |
| dropout (Dropout) | (None, 1280) | 0 |
| dense (Dense) | (None, 128) | 163,968 |
| dropout_1 (Dropout) | (None, 128) | 0 |
| dense_1 (Dense) | (None, 2) | 258 |

Fig13: EfficientNetB0

Finally, the model of [Fig5] outputs two values to classify the image into one of two categories. Most of the parameters are held within the EfficientNetB0 layer. Thus, this forms the main learning part of the network..

**MobileNetV2**:This model is developed on a MobileNetV2 architecture that is quite lightweight and fast. It takes input images with dimensions of 224×224, extracts features, and generates a feature map of 7×7×1280.These are then compressed to a vector of length 1280 by Global Average Pooling and passed through dense layers with dropout to avoid overfitting Fig7 . Finally, it predicts two classes. This architecture, MobileNetV2, has comparatively fewer parameters than other models, so it is efficient and suitable for real-time or mobile applications..

| Layer (type) | Output Shape | Param # |
|---|---|---|
| input_layer_1 (InputLayer) | (None, 224, 224, 3) | 0 |
| sequential (Sequential) | (None, 224, 224, 3) | 0 |
| rescaling (Rescaling) | (None, 224, 224, 3) | 0 |
| mobilenetv2_1.00_224 (Functional) | (None, 7, 7, 1280) | 2,257,984 |
| global_average_pooling2d (GlobalAveragePooling2D) | (None, 1280) | 0 |
| dropout (Dropout) | (None, 1280) | 0 |
| dense (Dense) | (None, 128) | 163,968 |
| dropout_1 (Dropout) | (None, 128) | 0 |
| dense_1 (Dense) | (None, 2) | 258 |

Fig13 : MobileNetV2

### 5.1 Implementation Details

Frameworks: TensorFlow and Keras
Hardware: NVIDIA GPU-based environment
Regularization: Dropout (0.4) and L2 weight decay to minimize overfitting
Evaluation Metrics: Accuracy, F1-score, Precision, Recall

## 6 Mathematical Formulation

### 6.1 Random forest:

**Formula**

Prediction=Majority Vote(T1(x),T2(x),...,Tn(x))

For regression:

Prediction=n1i=1∑nTi(x)

Random Forest has its basis on many decision trees and combines their predictions. It chooses the class that gets the most votes for classification. Similarly, it averages the output for regression. More trees mean better accuracy and less overfitting.

### 6.2 EfficientNet:

**Formula**

**d=αϕ,w=βϕ,r=γϕ**

**subject to α⋅β2⋅γ2≈2**

EfficientNet balances a CNN model's depth d, width w, and image resolution r with a scaling coefficient φ; it scales all the dimensions efficiently instead of merely increasing the depth.

### 6.3 MobileNetV2:

**Formula**

Output=x+F(x)

where F(x) = Conv1x1 (expand) → Depthwise Conv → Conv1x1 (project)

MobileNetV2 uses inverted residuals and depthwise separable convolutions to reduce computation. It is lightweight and fast, good for mobile devices.

### 4.4.4 ResNet:

**Formula**

y=F(x)+x

ResNet introduces skip connections that add the input x directly to the output F(x), avoiding vanishing gradients and hence having the ability to train very deep networks like ResNet50 or ResNet152.

### 4.5 Challenges and Solutions

| Challenge | Solution |
|---|---|
| Limited dataset size | Applied data augmentation and transfer learning |
| Overfitting$ | Used dropout and early stopping |
| Class imbalance | Maintained equal class distribution (autistic vs non-autistic) |
| Computational cost | Used MobileNetV2 for lightweight, high-accuracy inference |

## 5. Result

| Model | Validation Accuracy | Test Accuracy | F1-Score | Remarks |
|---|---|---|---|---|
| Random Forest (CNN Features) | 79% | 80% | 0.80 | Hybrid baseline |
| ResNet50 | 73% | 75% | 0.75 | Strong but heavy model |
| EfficientNetB0 | 71% | 73% | 0.73 | Stable performance |
| MobileNetV2 | 89% | 87% | 0.87 | Best performing model |

The performance comparison of the evaluated models shows that **MobileNetV2 delivers the best overall results**, achieving the highest validation accuracy (89%), test accuracy (87%), and F1-score (0.87). This indicates that MobileNetV2 is able to generalize well on unseen data while maintaining a strong balance between precision and recall. The **Random Forest model using CNN features** performs moderately well with an 80% test accuracy and an F1-score of 0.80, establishing it as a reliable hybrid baseline. On the other hand, **ResNet50** and **EfficientNetB0** show comparatively lower performance, with test accuracies of 75% and 73% respectively. Although both architectures are powerful, their results suggest that they may require further fine-tuning or a larger dataset to achieve optimal performance.Overall, the results highlight MobileNetV2 as the most effective and efficient model, making it the best candidate for real-world deployment.

## 6. Conclusion

In this study, extensive preprocessing and data augmentation were applied to ensure consistency and improve model robustness, followed by transfer learning using ResNet50, EfficientNetB0, and MobileNetV2 architectures. Optimization strategies such as learning rate tuning, dropout and batch normalization were employed to enhance training stability. While ResNet50 showed strong training performance but suffered from overfitting and EfficientNetB0 achieved moderate generalization with around 75% validation accuracy, MobileNetV2 demonstrated the best overall result with smooth convergence and accuracies exceeding 85% for validation. Its balanced learning, lightweight design, and computational efficiency make MobileNetV2 the most suitable architecture for real-word deployment, offering both high precision and reliability.

# 7. Reference


[1]. Atlam, E.–S., Aljuhani, K. O., Gad, I., Abdelrahim, M., Mansour Atwa, A., & Ahmed, A. (2025). Explainable deep-learning models are used to automatically identify autism spectrum disorder from facial images. reports from science. https://doi.org/10.1038/s41598-025-11847-5

[2]. Ahmad, I., Khan, N. A., Faheem, M., Akram, A., Rashid, J., & Amin, R. ul. (2024). A comparison of pretrained convolutional neural networks' performance in detecting autism spectrum disorder from facial images. Letters on Healthcare Technology. https://doi.org/10.1049/htl2.12073

[3]. Ahmed, Z. A., Aldhyani, T. H. H., Jadhav, M. E., Alzahrani, M. Y., Alzahrani, M. E., Althobaiti, M. M., Alassery, F., Alshaflut, A., Alzahrani, N. M., & Al-Madani, A. M. (2022). A system for detecting facial features that can be used to identify children with autism spectrum disorder. [Name of journal]. PMC9001065.

[4]. Farhat, T., Bhatti, S. M., Jaffar, A., Rashid, M., Akram, S., & Iqbal, M. A. (2025). An ensemble for deep learning-based facial image diagnosis of autism spectrum disorder. https://doi.org/10.1371/journal.pone.0321697 PLOS ONE, 20(4), e0321697

[5] Islam, M., Rahman, K. K. M., & others. (2022). Static facial features and deep neural networks are used to identify autism in children. 94 in Brain Sciences, 12(1). https://doi.org/10.3390/brainsci12010094

[6] Alam, M. S., Rashid, M. M., Faizabadi, A. R., Mohd Zaki, H. F., Alam, T. E., Ali, M. S., Gupta, K. D., & Ahsan, M. M. (2023). Effective Deep Learning-Based Data-Centric Method for Explainable AI-Based Diagnosis of Autism Spectrum Disorder from Facial Images. Technologies, 11(5), 115. https://doi.org/10.3390/technologies11050115

[7]. Farhat, T., Akram, S., AlSagri, H. S., Ali, Z., Ahmad, A., & Jaffar, A. (2022). A Comparison of Deep Neural Network Classifiers for Autism Detection Using Facial Images. 78(1), 55389; Computers, Materials & Continua. https://doi.org/10.32604/cmc.2023.045022

[8]. Huang, W.–C., Song, P.–H., & Li, Y. (2023). A two-phase transfer learning and multi-classifier integration approach for classifying autistic children's faces. Psychology Frontiers, 14, 1226470. https://doi.org/10.3389/fpsyg.2023.1226470

[9]. Facial Image-Based Autism Identification: A Comparison of Deep (2022). Tech Science Press. — DOI: 10.32604/cmc.2023.045022 Arfan Jaffar, Zulfiqar Ali, Awais Ahmad, Hatoon S. AlSagri, Sheeraz Akram, and Tayyaba Farhat.