MobileNet-Powered Autism Spectrum Disorder Detection from Facial Images

Renuka Sahebrao Sarse[1], Kirti Wanjale[2], Shilpa Katikar[3],
[1,2,3] *Dipartment of Computer Engineering Vishwakarma Institute of Information Technology, Pune, Maharahtra*

***Abstract*—A developmental disability linked to brain development, autism spectrum disorder (ASD) results in difficulties with behavior, socialization, and communication. Children with autism exhibit significant morphological alterations to their faces that distinguish them from normally developing (TD) children. Our aim is to assist psychiatrists by identifying children with ASD at an early age, which will lessen their symptoms and enhance their cognitive functioning. In this research study, we deploy a neural network-based binary classifier on a pre-trained model of MobileNet to a publically available data set named Autism_Image_Data that we obtained from Kaggle to present a functional model using a facial imaging for autism detection. The dataset contains 2540 training images with 300 test images. This model achieves an accuracy of 89.58% on the dataset that was trained using pre-trained MobileNet architecture model through transfer learning.**

***Index Terms*—Autism; facial images; convolution neural network; binary classifier; Activation function, transfer Learning; MobileNet; ASD.**

1. INTRODUCTION

Autism Spectrum Disorder known as ASD, is one of the types of neurodevelopmental disorder. It is typified by a variety of symptoms, including repetitive patterns of behavior. It affects social interaction, communication, and conduct. Symptoms often appear in early childhood and vary in severity. Early indications of autism spectrum disorder (ASD) in some youngsters include decreased eye contact, insensitivity to their name, or seeming indifference in caretakers. On the other hand, some children may exhibit traits like aggression, withdrawal, or a loss of previously learned language abilities after initially following a regular developmental trajectory throughout some months or years of life. The average age at which ASD symptoms become apparent is two years old. It's critical to understand that every child with ASD exhibits a unique behavioral pattern, and that there are differences in severity between low and high functioning. As there is no way to prevent ASD, but early detection and intervention will provide the tools and strategies needed to effectively address these challenges. This can result in significant changes to children’s behavior, skills and language development. Machine learning has made remarkable progress in recent years and is gaining importance in various applied sciences, especially in fields of medicine and biomedicine. These techniques have been proposed to aid data interpretation in clinical decision-making and diagnosis. Therefore, disease screening methods using machine learning are widely researched. The paper in [2] demonstrates the use of various techniques such as ANN, SVM, and RF. This experiment was performed on the three ASD datasets in the UCI database [3]. The dataset used for that study is text-based dataset. But there is another image-based dataset at [22], by using which we can detect whether a child has autism or not. So this study mainly focuses of detecting the autism in children by giving the images to the CNN model. The study at [16] shows the process of diagnosing autism in children by using facial image analysis.One kind of deep learning algorithm is the convolutional neural network (CNNs), are useful for extracting key features from medical images and grouping them together. This helps researchers and medical professionals diagnose and treat diseases. Obtaining and uploading images to the model, identifying key features, building models for image datasets, and ultimately producing classified images and reports based on the analysis are the steps involved in medical image classification. This summarizes how any CNN model operates. This paper also gives the performance and accuracy for common pre-trained model such as VGG16, VGG19, MobileNet, InceptionV3, Densenet121, Xception, ResNet-50 etc.

The choice of using the pre-trained models depends on the task we are performing that is the purpose for which we are making use of it. The best way to choose a pre-trained model for a specific task is to experiment with different models and see which one performs best on the given data which can be text, images, voice, audio, video etc. So in this study we tried different CNN pre-trained models on the image dataset. The previous studies about this topic shows that out of all the models the MobileNet models gives the more accuracy and performs better than other models [11]. The other paper at [16] have shown the comparative accuracy and time for execution of different pre-trained models such as MobileNet, Densenet121, InceptionV3, Xception, ResNet-50. Out of all these mentioned models the MobileNet has highest accuracy and takes less time. The findings of this study [11] demonstrated that the MobileNet model was able to reach a maximum accuracy of 95.75%. So in most studies, the MobileNet model has achieved the highest accuracy than the other pre-trained models, and the same has happened in our study as well.

This study mainly focuses on using the MobileNet model for ASD detection. MobileNet is a specified convolutional neural network architecture designed for resource-limited devices and mobile platforms. It uses depth-separable convolutions to significantly reduce the computational cost and parameters. Key features include width and resolution multipliers, allowing control over model size and efficiency. This model is often pre-trained on large datasets like ImageNet to capture common features. In ASD detection code, MobileNet serves as a pre-trained base model and a specific layer is selected for feature extraction. To customize the model for the particular task, fine-tuning was carried out on the ASD dataset, showcasing MobileNet's adaptability in a range of computer vision applications. The following is a summary of study's contributions:

- Proposing a deep learning framework that uses facial image recognition to identify autism.
- Improving model performance by applying transfer learning strategies.
- Conducting a thorough comparative analysis and comparing models with and without transfer learning using metrics as accuracy, precision, recall, and F1 score.
- Last and main contribution is that the early ASD detection from the facial images of the patient so that early treatment can be given to the patient. And to do this we have to adapt this CNN model to real time which can be done with real-time video capture.

The fundamental framework for our study is depicted in Figure 1[15]. Starting with the collection of data and continuing through data loading and preprocessing, model preparation and training, and model performance testing. Thus, this framework demonstrates the operation of the CNN.

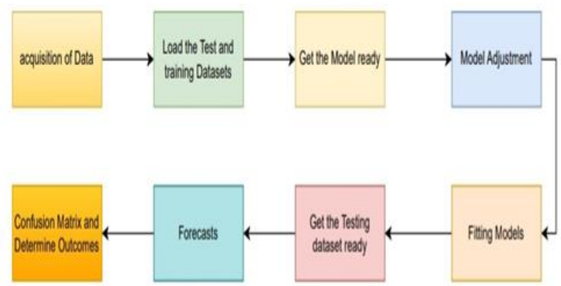

Figure 1. Fundamental structure of our study

2. LITERATURE SURVEY

A mental condition known as autism spectrum disorder (ASD) can be identified by facial expression recognition analysis. A comprehensive literature survey for ASD detection is presented in this section. M.S.Rashid et al. [1] shows that how CNN models include VGG-19, ResNet-50V2, MobileNetV2, EfficientNetB0 and Xception model perform ASD using facial image analysis. They also highlight machine learning improve accuracy. Zeyad A. T. Ahmedet al. [2] uses MobilNet, Xception, InceptionV3 models with CNN architecture to display distance between facial landmarks that distinguish autistic and non-autistic faces. They conclude that MobileNet model having 95% accuracy which is high. Angelina LU et al.[4] propose an ASD screening with facial image by using VGG-16 model. They use kaggle dataset and model gives the 95% accuracy. Deep learning based on facial images ought to produce models with lower error rates.

Suman Raj et al. [5] sought to detect ASD using variety of deep learning techniques. They use three non-clinical dataset children, adolescent and adults. When comparing outcomes, CNN classifier produced better result of 97% accuracy. K.K. Mujeeb Rahman et al. [6] use dataset of TD and ASD kids. For access test it uses five CNN based model like MobileNet, Xception, EfficientNetB0, EfficientNetB1, EfficientNetB2. They find that Xception model performs better than other models with 95% accuracy. The study at [24] compares the best pre-trained CNN models on an Indian dataset and investigates the use of machine learning for banknote classification. A new MNET architecture achieves 99.68% accuracy for banknote classification and perfect accuracy for quality classification, while VGG16 stands out with 87.50% accuracy. By focusing on both classification and quality parameters, this work fills in research gaps and offers a useful dataset for upcoming investigations.

Inzamam UL Haque et al. [7] uses a DCNN model with KDEF dataset for facial recognition utilizing the DCNN model by using angled sideview pictures and anticipate changes that will be required for face application of expression. Vinicius Silva et al. [9] use a robotic tool with Zenoh R50 Robokind platform. ZECA is used as mediator. By these 5 facial expressions recognize like (anger, happiness, sadness, surprised) and it is very comfortable to analyze.In a comparison with the most advanced 3D facial expression recognition accuracy, the suggested subsystem performed 88% better than 84% respectively. Herag Arabian, Verena et al. [10] take five neutral network architecture of VGG-16, ResNet-50, GoogleNet, ShuffleNet, EfficientB0. They take three datasets OLU-CASIA, JAFFE, FACES.In this paper, performance measured by Grad-CAM prediction. On the based-on performance ResNet-50 has better performance. Ugur Erkan et al. [12] proposed best way to classify ASD data and give support to daignosis. They use three datasets (Children, adolescences and adults). KNN method, Support Vector Machine (SVM) and Random Forest method (RF) were used for testing. By using confusion matrix test the performance from that they conclude that RF technique classified data with 100% accuracy.

This is evident from above literature survey that facial image recognition is the best way to analyze the autisum spectrum disorder (ASD). For that there are many pretrained models like MobileNet, ResNet, VGG-16, VGG-19, ImageNet, Xception, InceptionV3, EfficientNetB0, EfficientNetB1, EfficientNetB2. Datasets are easily available on kaggle platform or some common dataset as ck+, KDEF JAFFE etc.There are many classifiers which improved the performance and mostly performance measured by confusion matrix.

Review Matrix

Various studies have explored the use of machine learning and deep learning models to detect autism spectrum disorder (ASD) using diverse datasets. Many leveraged datasets from Kaggle and other sources such as ABIDE, UCI, and EEG recordings, with image counts ranging from around 3,000 to larger, more complex collections. Preprocessing techniques varied from simple image resizing and normalization to more advanced methods like batch normalization, segmentation algorithms, and noise removal. Widely used pretrained models included VGG16/VGG19, ResNet variants, MobileNet, EfficientNet (B0–B2), and Xception, often combined with classifiers such as decision trees, SVMs, CNNs, and DNNs. Performance metrics typically included accuracy, precision, recall, specificity, sensitivity, AUC, and confusion matrices, with many models achieving high classification rates—some reporting over 95% precision and AUC values near 98%. These studies highlight the growing effectiveness of deep learning in supporting early ASD detection.

3. PROPOSED METHODOLOGY

A. Dataset:

The data set for this research is Autism_Image_Data which we gathered from kaggle which is freely available. The dataset is pre-dominantly divided into four sub-types, the training set is namely being train.It comprises 2540 training images for the model which are in a valid .jpg format. The consolidate file directly is further sub-divided into two sub-directories namely Autistic (1470 files) and non-autistic (1470 files). The testing set contains 300 files for testing of the executed model. The last one being the valid directory containing two further sub-directly namely autistic containing 50 files and non-autistic containing 50 files.On a summary of the overall data set it contains a massive of 5880 .jpg images.

Dataset

Images	Training	Consolidated	Test	Valid
Autistic	1270	1470	150	50
Non-Autistic	1270	1470	150	50

B. System Architecture

An overview of the technologies and methodologies used in this study is provided in this section. This section provides an overview of general procedure, as shown in Figure 2. The whole process mainly includes image acquisition, training and evaluation. The first step is to prepare the face image database, which includes the dataset for each of the three stages namely training, validation, and testing. First, pictures from the training and validation datasets will be used to train the model. Preprocessing techniques such as augmentation, rescaling, and feature

extraction are used during training. After that, the idea of transfer learning will be put into practice., which is nothing but the use of MobileNet pre-trained model. Lastly, to confirm the effectiveness of our methods, the model will be assessed using the testing dataset.

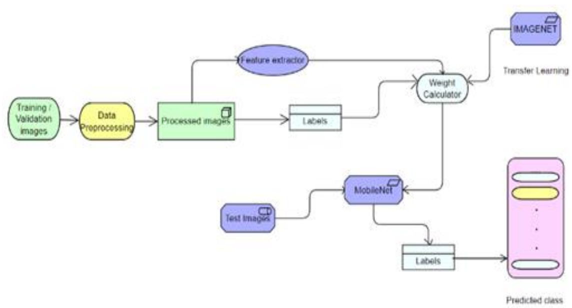

Figure 2. Architecture of the System

A. Convolutional Neural Networks (CNN)

A Convolutional Neural Networks (CNN) is a deep learning algorithm that are used to recognize images, process natural language, identify objects, detect objects in videos, and recognize voices [23]. CNN architecture consist of Convolutional, pooling and fully connected layers. In feature extraction and image processing, these layers have different functions. Of them, the convolutional layers are the most important because they are essential parts of the CNN architecture.

Convolution Layer: It is the fundamental component of CNN, applying filters to the input image to extract various features, including edges, textures, and shapes. This layer uses convolutional operations with learnable filters to identify patterns and features in the input images. The following line of code represents the convolutional layer from MobileNet model.

```
last_layer = pre_trained_model.get_layer('conv_pw_13_relu')
```

Fully Connected layer: This layer is used to make a prediction or classify the image. That is whether the given image is autistic or non-autistic will be finally decided by this layer. The following lines represents the fully connected layers.

```
x = Dense(512, activation='relu')(x)
```

It is the first fully connected layer.

```
x = Dense(1, activation='sigmoid')(x)
```

It is last fully connected layer.

Activation Functions:

It acts as a decision-maker for each neuron, determining whether it should activate based on a weighted sum of inputs plus bias. By introducing non-linearity, it helps the network to understand intricate patterns and reach complex conclusions. Popular activation functions are Tanh, Sigmoid, and ReLU. This study uses the ReLU and Sigmoid activation functions. The output layer uses Sigmoid to squash the final output between 0 and 1 and the hidden layer uses ReLU as an activation function to introduce non-linearity.

```
x = Dense(512, activation='relu')(x)
x = Dense(1, activation='sigmoid')(x)
```

C. MobileNet

It is a deep learning pre-trained model commonly used for real-time image processing on mobile devices.

Working of MobileNet in code:

- Weights from the 'imagenet' dataset are fed into the MobileNet model that has already been trained.
- Next, the code indicates the activation that comes after the final depthwise separable convolution layer by selecting the layer 'conv_pw_13_relu' from the MobileNet architecture.
- Global Average Pooling (GAAP) layer is added to the model to the model to adjust it for binary classification, the Fully Connected layers (ReLU activation) are added.
- After that, this updated model is assembled and put to use for assessment and training. Leveraging MobileNet's learnt representations for the particular job of ASD detection is made possible by using it as a feature extractor.

D. Transfer Learning:

It is a technique which involves adapt a model that trained on one task to a second task that is similar. It makes use of the information gleaned from the source task to enhance the target task's learning, In this work, the MobileNet model is used to facilitate transfer learning. First, a sizable and varied dataset is utilized to trained the model for general image identification. The final layers of MobileNet are then modified to detect ASD by stacking more layers on top of them and modifying the previously learnt features for the new purpose. Through this procedure, the model is able to take advantage of the prior information that is stored in the weights of MobileNet, leading to improved performance on the ASD dataset with small sample sizes.

4. RESULTS AND ANALYSIS:

The following segment reports the result of approach that are being demonstrated and analysed. After 32 epochs, our model produces the test accuracy of 89.58 % and a loss of 38.62 % over the new unseen data. The following images are the reprensentation of model accuracy and loss analysed for 32 epochs over the training and validation dataset.

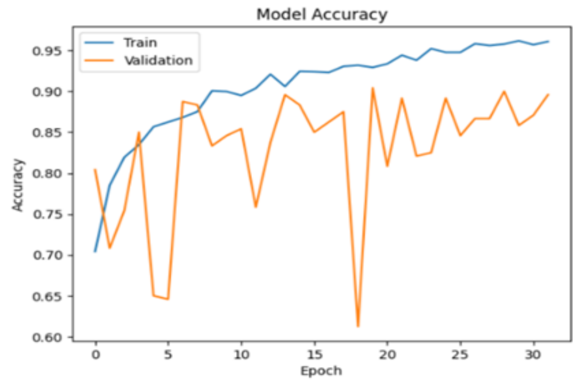

Figure 4. Model accuracy on training and

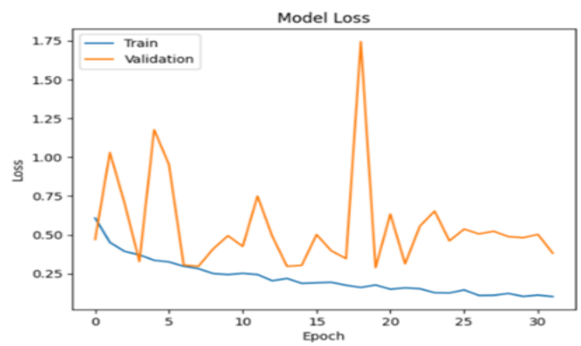

Figure 5. Model loss on training and validation dataset

The following Figure provides a more thorough representation of our suggested approaches in terms of the total number of correct and incorrect predictions through a confusion matrix. A more detailed view of our proposed methods is given in figure 6, which uses a confusion matrix to show the total number of right and wrong predictions. Figure 7 displays the Receiver Operating Characteristics (ROC) curve. The Y-axis represents the True Positive Rate (TPR), while the X-axis represents the False Positive Rate (FPR). The area under the curve (auc) for our model is 0.55.

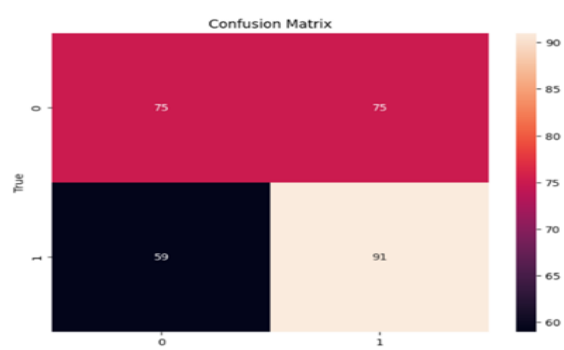

Figure 6. Confusion Matrix

Figure 7. Receiver Operating Characteristic Curve

The CNN model's visual predictions for the diagnosis of autism spectrum disorder (ASD) are displayed in Figure 8. A sample of nine test images, labeled with their corresponding predictions (0 for non-autistic and 1 for autistic), produced by the trained model, are displayed in the image grid. This figure shows how well the model performed based on facial image classification to classify people into ASD and non-ASD groups.

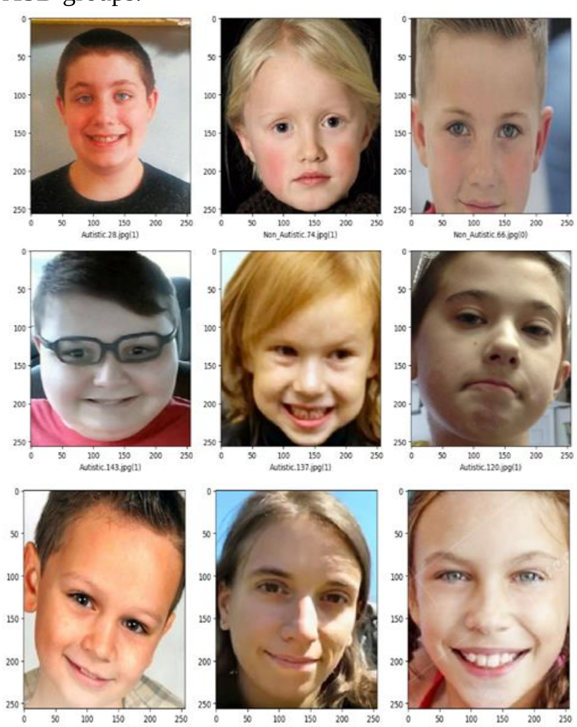

Figure 8. Model Predictions: MobileNet Results

5. CONCLUSION

This study demonstrates the identification of autism through facial expressions, employing a Convolutional Neural Network with a pre-trained MobileNet model and utilizing transfer learning techniques. The effectiveness of this experiment is measured by performance metrices, such as accuracy, precision, recall and F-measure. The application of the MobileNet model with Transfer Learning significantly improves the learning process, yielding promising outcomes with the highest recognition accuracy recorded at 88.50%.

With an impressive classification accuracy of 88.50%, utilizing the pre-trained MobileNet model through transfer learning signifies the potential of utilizing children's facial images as an economical and effective means for ASD screening. This approach presents a convincing approach for detection of autism in youngsters.

6. FUTURE SCOPE

Facial image detection has great potential for the future of autism detection. Researchers are increasingly using artificial intelligence and machine learning algorithms to analyze facial expressions and other parameters to detect autism. While current methods include structured interviews and behavioral assessments, facial image detection has the potential to provide more objective and accurate diagnoses. However, more research is needed using the techniques of convolution neural networks along with different pre-trained models to fully understand the potential of facial image detection in autism detection and to establish and further improve its reliability and validity.

REFRENCES:

[1] Alam, M. S., Rashid, M. M., Roy, R., Faizabadi, A. R., Gupta, K. D., & Ahsan, M. M. (2022). Empirical study of autism spectrum disorder diagnosis using facial images by improved transfer learning approach. Bioengineering, 9(11), 710.

[2] Ahmed, Z. A., Aldhyani, T. H., Jadhav, M. E., Alzahrani, M. Y., Alzahrani, M. E., Althobaiti, M. M., & Al-Madani, A. M. (2022). Facial features detection system to identify children with autism spectrum disorder: deep learning models. Computational and Mathematical Methods in Medicine, 2022.

[3] Jahanara, S., & Padmanabhan, S. (2021). Detecting autism from facial image. International Journal of Advance Research, Ideas and Innovations in Technology, 7(2), 219-225.

[4] Lu, A., & Perkowski, M. (2021). Deep learning approach for screening autism spectrum disorder in children with facial images and analysis of ethnoracial factors in model development and application. Brain Sciences, 11(11), 1446.

[5] Raj, S., & Masood, S. (2020). Analysis and detection of autism spectrum disorder using machine learning techniques. Procedia Computer Science, 167, 994-1004.

[6] Mujeeb Rahman, K.K.; Subashini, M.M. Identification of Autism in Children Using Static Facial Features and Deep Neural Networks. Brain Sci. 2022, 12, 94.

[7] Haque, M. I. U., & Valles, D. (2018, December). Facial expression recognition from different angles using dcnn for children with asd to identify emotions. In 2018 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 446-449). IEEE.

[8] Sivasangari, A., Ajitha, P., Rajkumar, I., & Poonguzhali, S. (2019). Emotion recognition system for autism disordered people. Journal of Ambient Intelligence and Humanized Computing, 1-7.

[9] Silva, V., Soares, F., Esteves, J. S., Santos, C. P., & Pereira, A. P. (2021). Fostering emotion recognition in children with autism spectrum disorder. Multimodal Technologies and Interaction, 5(10), 57.

[10] Arabian, H., Wagner-Hartl, V., & Moeller, K. (2022, September). Network Architecture Influence on Facial Emotion Recognition. In Current Directions in Biomedical Engineering (Vol. 8, No. 2, pp. 524-527). De Gruyter.

[11] Sadik, R., Anwar, S., & Reza, M. L. (2021). Autismnet: Recognition of autism spectrum disorder from facial expressions using mobilenet architecture. International Journal, 10(1), 327-334.

[12] Erkan, U., & Thanh, D. N. (2019). Autism spectrum disorder detection with machine learning methods. Current Psychiatry Research and Reviews Formerly: Current Psychiatry Reviews, 15(4), 297-308.

[13] Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. IEEE transactions on affective computing, 13(3), 1195-1215.

[14] Husna, R. N. S., Syafeeza, A. R., Hamid, N. A., Wong, Y. C., & Raihan, R. A. (2021). Functional magnetic resonance imaging for autism spectrum disorder detection using deep learning. Jurnal Teknologi, 83(3), 45-52.

[15] Alsaade, F. W., & Alzahrani, M. S. (2022). Classification and detection of autism spectrum disorder based on deep learning algorithms. Computational Intelligence and Neuroscience, 2022.

[16] Abdulazeez Mousa, Fatih Özyurt & Shivan H. M. Mohammed (2021). Autism detection from facial images using deep learning methods. pathology diagnosis, International Conference on Advanced Engineering, Technology and Applications (ICAETA-2021), 2021.

[17] Liao, M., Duan, H., & Wang, G. (2022). Application of machine learning techniques to detect the children with autism spectrum disorder. Journal of Healthcare Engineering, 2022.

[18] Farooq, M. S., Tehseen, R., Sabir, M., & Atal, Z. (2023). Detection of autism spectrum disorder (ASD) in children and adults using machine learning. Scientific Reports, 13(1), 9605.

[19] Erkan, U., & Thanh, D. N. (2019). Autism spectrum disorder detection with machine learning methods. Current Psychiatry Research and Reviews Formerly: Current Psychiatry Reviews, 15(4), 297-308. https://www.mayoclinic.org/diseases-conditions/autism- spectrum-disorder/symptoms-causes/syc-20352928