Autism Spectrum Disorder (ASD) Diagnosis Using Facial Features

Saleha Ahmed*, Aamina Binte Khurram*
*School of Electrical Engineering & Computer Science (SEECS),
National University of Sciences & Technology (NUST), 44000 Islamabad, Pakistan.
Email: *{sahmed.bese21seecs,akhurram.bese21seecs}@seecs.edu.pk

***Abstract*—Autism spectrum disorder (ASD) presents significant challenges in early diagnosis due to its complex nature and the absence of specific medical tests. This paper investigates the potential of deep learning models in aiding the diagnosis of ASD through facial image analysis. Specifically, we conduct a comparative analysis of various image classification models, including VGG16, VGG19, ResNet18, ResNet34, and ResNet50, utilizing transfer learning techniques. Our experiments involve fine-tuning these pre-trained models on a dataset comprising autistic and non-autistic facial images. We explore different hyperparameters and customization techniques to optimize the models for accuracy, precision, recall, and F1-Score. Additionally, we investigate the efficacy of the low-rank adaptation (LoRA) technique in enhancing the performance of ResNet models. Our results demonstrate promising outcomes, particularly with the ResNet50 model, which exhibits superior efficiency in ASD detection. This study contributes to the ongoing efforts in leveraging deep learning for medical diagnosis and highlights avenues for future research in refining diagnostic processes for ASD.**

***Index Terms*—Autism spectrum disorder (ASD), low-rank adaptation (LoRA)**

I. Introduction

Autism Spectrum Disorder (ASD) is a neuro-development disorder that has an impact on the social and cognitive skills of children causing repetitive behaviors, restricted interests, communication problems and difficulty in social interaction. Early diagnosis of ASD can prevent from its severity and prolonged effects [1]. The absence of a specific medical test, akin to a blood test, makes the diagnosis of autism spectrum disorder (ASD) a complex procedure. Unlike many other medical conditions that can be identified through straightforward diagnostic tests, ASD is primarily diagnosed based on behavioral observations and developmental history. However, researchers at the University of Missouri conducted a study examining the diagnosis of autism in children by analyzing their facial features. Their investigation revealed that children with autism possess a set of distinct facial characteristics that differ from those observed in non-autistic children. The features include an exceptionally expansive upper face, encompassing widely spaced eyes, and a comparatively shorter central facial region, covering the cheeks and nose [2]. Classification of autism with the help of facial images is cheaper as compared to classification with the help of a brain imaging dataset [3].

Transfer learning has become a leading method in image classification, especially in deep learning. It involves reusing a pre-trained model developed for one task in another related task. By using the knowledge and features learned from the pre-trained model, transfer learning reduces the need for large amounts of labeled data typically required to train deep neural networks from scratch. This paper conducts a comparative analysis of different image classification models, including VGG16, VGG19, ResNet18, ResNet34, and ResNet50, utilizing transfer learning. Various hyper parameters are analyzed and tuned to identify optimal values for achieving the best results in terms of accuracy, precision, recall, and F1-Score. The findings indicate that the ResNet50 algorithm demonstrates the highest efficiency among the models examined. Furthermore, the paper explores the LoRA optimization technique to further fine-tune the ResNet50 model and assess its impact on performance.

II. Literature Review

The problem of autism detection from facial images has been studied in depth by many researchers. Various machine and deep learning methods have been trained to identify autistic images from non-autistic ones. This section presents a summary of the works studied in the interest of understanding the existing knowledge in this domain.

In [2], the authors apply federated learning to train two different machine learning models, namely Support Vector Machines and Logistic Regression, to classify facial images as either autistic or non-autistic. They claim to achieve 81% accuracy in detecting ASD in adults and 98% accuracy in determining ASD in children.

The studies conducted in [3] provide a comparative analysis of MobileNet, InceptionV3 and InceptionResNetV2 pre-trained on ImageNet. According to the results training the MobileNet model on facial images gave a maximum of 87% testing accuracy out of the all three trained models.

The studies proposed in [4] use deep learning, specifically CNNs, to enhance ASD diagnosis through facial image analysis. The proposed models, based on VGG16 and VGG19 architectures, achieve an 84% accuracy rate in classifying ASD individuals, showcasing the potential of deep learning in refining diagnostic processes. Further research is recommended to optimize and validate these models for broader application. The efficacy of the VGG models in image classification has been demonstrated in the paper that introduced these models [5].

The work done in [1] applies multiple transfer-learning based deep learning image classification techniques to a dataset of facial images. The authors modify the architecture of some models studied by adding three fully connected layers and two batch normalisation layers their base architecture. They then use k-means clustering to further sub-classify autistic people into categories, using a k value of 2. The authors of this work claim to have achieved 91% test accuracy for classifying into autistic and non-autistic classes using a modified MobileNetV1, and then a 92% accuracy for predicting into the correct autism sub-types(kept binary in their work).

In [6], the authors present a comparative analysis of traditional machine learning models(including support vector machines and random forests), auto-ML, and more complex deep learning image classification models(including the VGG and ResNet families) in the classification of autism from facial images. The authors claim that the best performance came from using auto-ML models, which were able to self-identify the most appropriate machine learning model to apply to the problem.

Almost all existing work on ASD classification from facial images is based on some sort of transfer learning, an approach wherein deep learning models already trained on data concerning a specific problem are leveraged to solve a related problem by fine-tuning the model's parameters. This strategy has been predominantly utilized due to the limited availability of large-scale annotated datasets specifically tailored for autism spectrum disorder (ASD) classification from facial images.

The work demonstrated in [7] utilizes three pre-trained CNN models, VGG16, VGG19 and, EfficientnetB0, as feature extractors and binary classifiers. The suggested models were trained using a publicly available dataset from Kaggle that included 3014 images of children characterized as autistic and non-autistic. The models yielded accuracies of 84.66%, 80.05%, and 87.9%, respectively. Similarly, leveraging transfer learning, the study proposed in [8] investigates various pre-trained CNN architectures, including ResNet34, ResNet50, AlexNet, MobileNetV2, VGG16, and VGG19, to diagnose ASD. Results indicate that the proposed ResNet50 model achieves the highest accuracy of 92%, outperforming other transfer learning models and state-of-the-art approaches in terms of both accuracy and computational efficiency. Another approach using two-phase transfer learning is described in [9]. In this work, the authors divide the training of the network into distinct phases. They use the MobileNetv2 and MobileNetv3-Large models as feature extractors, classifying images into general facial features in the first phase, and then classifying into autistic and non-autistic classes in the second phase.

The authors of [10] present an in-depth analysis of the application of various transfer learning techniques to models of the MobileNet family. They claim to achieve a test accuracy of 94.6%, having added two additional dense layers to the base MobileNet models. However, the authors, who have used the same dataset used in this work, do acknowledge that this extremely high accuracy may be influenced by the data quality, as the limited dataset is probable to cause overfitting.

III. Methodology

In this work, we have carried out a comparative analysis of the ResNet and VGG image classification model families. In addition, we apply transfer learning techniques to fine-tune these models to solve the problem of the detection of autism from facial images. We have used a publically available dataset of facial images tof autistic and non-autistic images available on Kaggle. However, the size of the dataset is somewhat limited, with a total of 2936 images, of which 86.3% were used for training, 3.4% for validation, and the rest for testing. The limited quantity of images meant that all models applied were prone to overfitting, but by the application of measures such as regularisation and carefully choosing the number of epochs, we were able to identify implementations less likely to overfit. We explore the application of various customisation techniques, including adding a dropout layer after the base models, and applying a combination of batch-norm and fully-connected layers (as done in [1]). In addition, we explored the application od Low-Rank Adaptation to ResNet models, aiming to reduce the computation time required for training the model.

A. VGG Models

VGG16, short for Visual Geometry Group 16, is a convolutional neural network architecture developed by the Visual Geometry Group at the University of Oxford. It consists of 16 layers, primarily comprising a stack of convolutional layers followed by max-pooling layers, with increasing depth as the network progresses. The network concludes with three fully connected layers, including a softmax layer for classification. VGG16 is renowned for its simplicity and effectiveness, achieving state-of-the-art results in image classification tasks.

VGG19, an extension of VGG16, follows a similar architecture but with 19 layers. It includes additional convolutional layers, providing a deeper network architecture to capture more complex features in images. Similar to VGG16, VGG19 consists of convolutional layers followed by max-pooling layers, concluding with fully connected layers for classification. While VGG19 offers increased depth and potentially higher representational power, it also comes with added computational complexity and increased training time compared to VGG16.

B. ResNet Models

The basic building blocks of ResNets are residual blocks, which are contiguous layers in a neural network whose inputs are received at the outputs as well, before applying the activation on the combination of both the input and the block's output. Such connections between the input of the first layer and output of the last layer of a block are called skip connections.

Skip Connections: Adding more layers to a neural network doesn't always mean an increase in performance, and in fact too many layers in traditional networks actually cause reduced

performance. This idea does make intuitive sense since an increase in layers in the neural network is equivalent to an increase in the model's complexity as more and more parameters are added to the model to be learnt. And too complex of a model is more likely to overfit. Another way of looking at this problem is that when there are too many layers in the network, the features that later layers extract from the data may not be as useful since a considerable amount of noise will have accumulated in the later layers' weight vectors. Thus, the representation of data that later layers provide is considerably noisier than the original input and the layers may even start to become unrepresentative of the original input. One mechanism to make networks deeper without degrading their performance is to make use of skip connections.

These connections have the following benefits: 1) Less noisy gradients at earlier layers 2) Less noisy inputs to later layers in the network 3) Ability to make network much deeper

The primary difference between the ResNet18, ResNet34, and ResNet50 models is in the number of layers. ResNet18 is the least complex of the three models. ResNet34 is considerably more complex, with a total of 34 layers and 2-layer residual blocks. ResNet50 builds on this complexity further by introducing 3-layer residual blocks.

C. Figures and Tables

IV. Results

To start with the transfer learning, the convolutional layers within the pretrained networks were frozen and the models were trained for five epochs and ten epochs respectively. Following results were obtained; In the above results, the

Epochs = 5	Accuracy	Precision	Recall	F1-Score
VGG16	0.7967	.7405	.9133	.8179
VGG19	0.8200	0.8158	0.8267	0.8212
ResNet18	0.7767	0.8029	0.7333	0.7666
ResNet34	0.8067	0.7840	0.8467	0.8141
RestNet50	0.8400	0.8542	0.8200	0.8367

Fig. 1. Results obtained from frozen convolutional layers in five epochs

Epochs = 10	Accuracy	Precision	Recall	F1-Score
VGG16	0.8233	0.8299	0.8133	0.8215
VGG19	0.8200	0.8077	0.8400	0.8235
ResNet18	0.7967	0.8296	0.7467	0.7860
ResNet34	0.8200	0.8077	0.8400	0.8235
RestNet50	0.8267	0.8551	0.7867	0.8194

Fig. 2. Results obtained from frozen convolutional layers in ten epochs

models were observed to overfit as the number of epochs was increased to ten. Next, all our five models were trained by modifying the fully connected layers of the pre-trained models, adding an extra linear layer with 256 output features and ReLU activation, followed by the final prediction layer. The Adam optimizer was used to optimize all model parameters with a learning rate of 0.0001. The results obtained with and without a drop-out layer of 50 All the layers within the pretrained models were frozen and trained for five epochs to obtain the

Fig. 3. ResNet34 Training and Validation Accuracy over five epochs

Epochs = 5	Accuracy	Precision	Recall	F1-Score
VGG16	0.8067	0.7347	0.9600	0.8324
VGG19	0.8733	0.8373	0.9267	0.8797
ResNet18	0.8533	0.9732	0.7267	0.8321
ResNet34	0.8833	0.9528	0.8067	0.8736
RestNet50	0.8700	.8405	0.9133	0.8754

Fig. 4. Results obtained with a dropout layer of 50% probability

following results; From the above results, it was observed that the ResNet models, in particular ResNet50, perform relatively better as compared to the VGG16 and VGG19 models. Lastly, a few more fine-tuning techniques were also studied and applied to the ResNet models. The first one being as explained in [1], the authors modify the architecture of the models studied by adding two fully connected layers and three batch normalisation layers to the base architecture of various image classification models. The first batchnorm layer sends 1024 features into the first fully connected layer, whereas the second batchnorm layer sends 168 features into the second fully connected layer, after which another batchnorm layer is applied before giving the outputs of the model. In our work, we output the last layer into two neurons(as there are two classes). Following, training and validation accuracies and errors were recorded for ResNet18, ResNet34 and ResNet50 respectively

Epochs = 5	Accuracy	Precision	Recall	F1-Score
VGG16	0.8800	0.8562	0.9133	0.8839
VGG19	0.8767	0.8553	0.9067	0.8803
ResNet18	0.8933	0.8598	0.9400	0.8981
ResNet34	0.8400	0. 7865	0.9333	0.8537
RestNet50	0.9100	0.9184	0.9000	0.9091

Fig. 5. Results obtained without a dropout layer of 50% probability

Epochs = 5	Accuracy	Precision	Recall	F1-Score
VGG16	0.7767	0.7485	0.8333	0.7886
VGG19	0.7100	0.6507	0.9067	0.7577
ResNet18	0.8067	0.8108	0.8000	0.8054
ResNet34	0.8133	0.8052	0.8267	0.8158
RestNet50	0.8200	0.8636	0.7600	0.8085

Fig. 6. Results obtained by freezing all layers and training for five epochs

as they were trained for five epochs each. Lastly, another

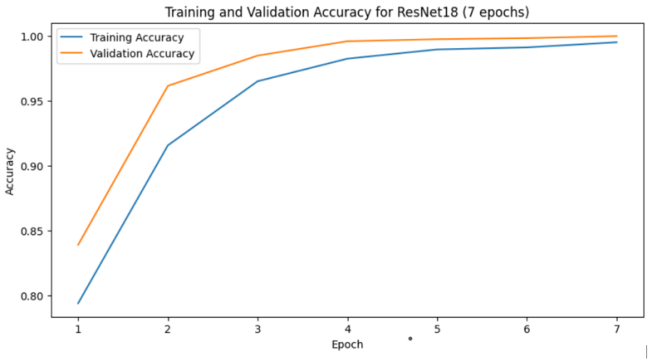

Fig. 7. ResNet18 Training and Validation Accuracy over five epochs

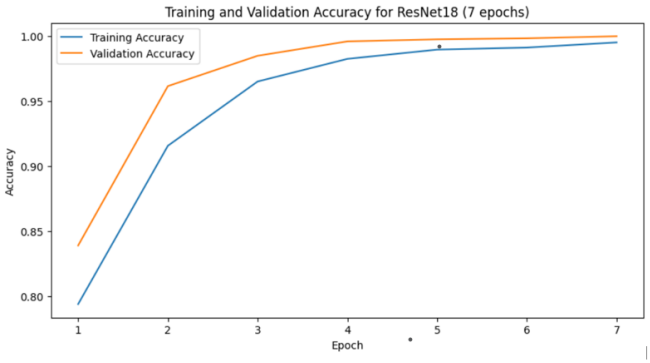

Fig. 8. ResNet18 Training and Validation Loss over five epochs

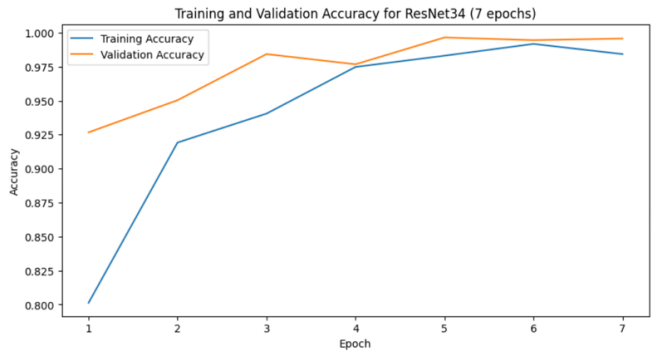

Fig. 9. ResNet34 Training and Validation Accuracy over five epochs

novel fine tuning technique Low-Rank Adaptation (LoRA), was also implemented on the ResNet models to obtain the following results:

V. Discussion and Future Directions

In this work, we have explored the application of various deep learning models to solve the problem of autism detection through facial images. We focused on the ResNet and VGG model families, and applied multiple transfer learning

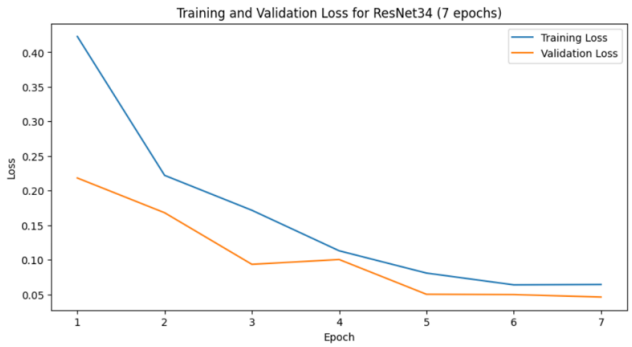

Fig. 10. ResNet34 Training and Validation Loss over five epochs

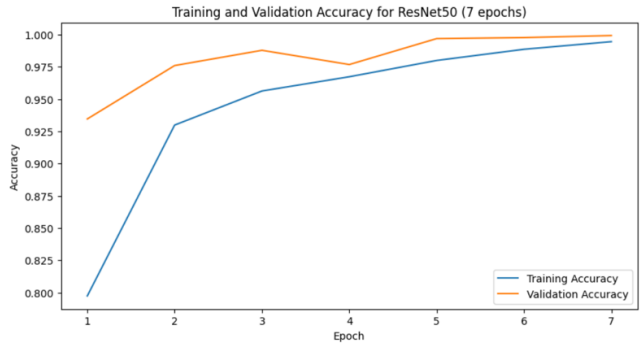

Fig. 11. ResNet50 Training and Validation Accuracy over five epochs

approaches to the models. In addition, we explored the efficacy of application of Low-Rank Adaption (a finetuning technique originally developed for large language models) on the ResNet models. The major limitation of this work is the limited size of the dataset used, which caused a tendency of models to overfit. However, given the sensitive medical nature of the problem domain, there is a dearth of richer image-based datasets of autistic faces, with most publicly available datasets gathered by people online without much regard to the quality of images

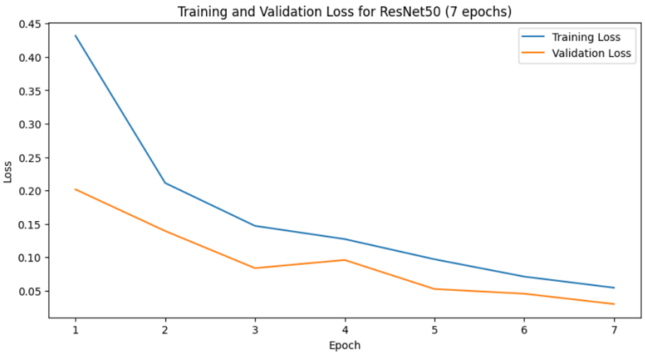

Fig. 12. ResNet50 Training and Validation Loss over five epochs

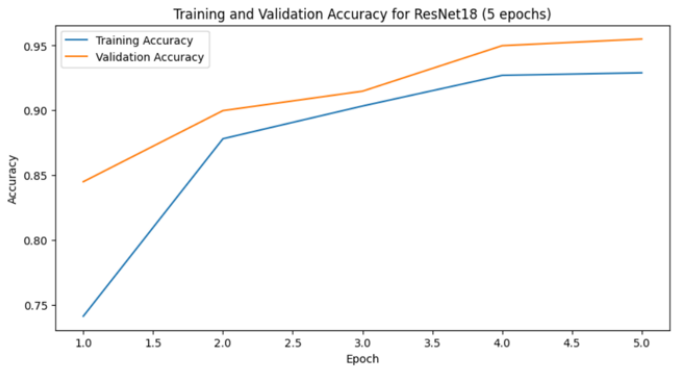

Fig. 13. ResNet18 Training and Validation Accuracy over five epochs

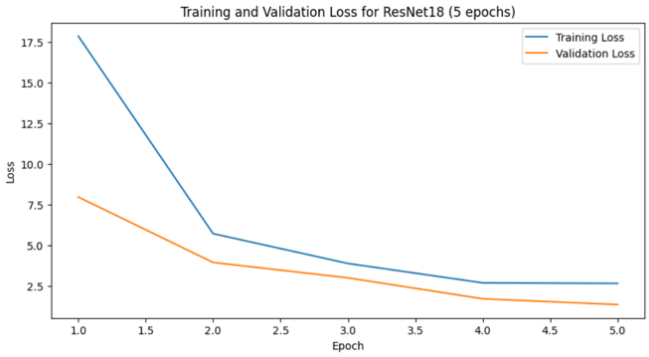

Fig. 14. ResNet18 Training and Validation Loss over five epochs

gathered. Thus better quality and greater quantity data will lead to better and more reliable results of autism classification. Another future consideration is the use of video-based data for the diagnosis of autism, which would build upon our current work involving an images-only dataset.

VI. Conclusion

In this study, we explored the efficacy of various deep learning models in detecting autism spectrum disorder (ASD) from facial images. Our experiments encompassed freezing convolutional layers, employing dropout regularization, and implementing Low-Rank Adaptation (LoRA) techniques across different ResNet architectures. Results revealed promising performance, with ResNet models consistently outperforming others, particularly ResNet50, which exhibited the highest accuracy of 90%. Furthermore, the application of dropout regularization demonstrated enhanced model performance.

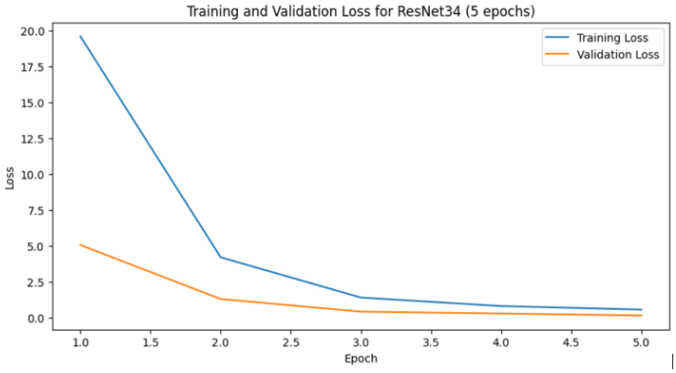

Fig. 15. ResNet34 Training and Validation Loss over five epochs

Fig. 16. ResNet50 Training and Validation Accuracy over five epochs

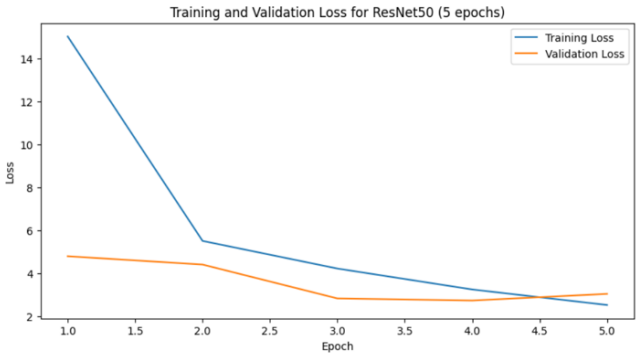

Fig. 17. ResNet50 Training and Validation Loss over five epochs

References

[1] T. Akter, M. Ali, M. I. Khan, M. Satu, M. Uddin, S. Alyami, S. Ali, A. Azad, and M. A. Moni, "Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage," *Brain Sciences*, vol. 11, p. 734, 05 2021.

[2] S. Farooq, R. Tehseen, M. Sabir, and Z. Atal, "Detection of autism spectrum disorder (asd) in children and adults using machine learning," *Scientific Reports*, vol. 13, 06 2023.

[3] A. T. A. M. Alkahtani, H., "Deep learning algorithms to identify autism spectrum disorder in children-based facial landmarks," *Appl. Sci.*, 2023.

[4] S. T. A. K. K. S. M. V. B. Gaddala, Kodepogu, "Autism spectrum disorder detection using facial images and deep convolutional neural networks," *Revue d'Intelligence Artificielle*, 2023.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[6] B. Ramdan, E. Younis, A. Ali, and O. Ibrahim, "Comparing automated and non-automated machine learning for autism spectrum disorders classification using facial images," *Etri Journal*, vol. 44, pp. 613–623, 03 2022.

[7] P. Reddy and A. J, "Diagnosis of autism in children using deep learning techniques by analyzing facial features," 01 2024, p. 198.

[8] I. Ahmad, D.-J. Rashid, M. Faheem, A. Akram, N. Khan, and R. Amin, "Autism spectrum disorder detection using facial images: A performance comparison of pretrained convolutional neural networks," *Healthcare Technology Letters*, 01 2024.

[9] Y. Li, W.-C. Huang, and P.-H. Song, "A face image classification method of autistic children based on the two-phase transfer learning," *Frontiers in Psychology*, vol. 14, 08 2023.

[10] M. Beary, A. Hadsell, R. Messersmith, and M.-P. Hosseini, "Diagnosis of autism in children using facial analysis and deep learning," 2020.