

Facial Image-Based Classification of Autism Spectrum Disorder Using Pretrained CNNs and Synthetic Data

Moshira A. Ebrahim ^{1,*} , Hesham A. Sakr ^{2,3}  and Ahmed M. AbdelMouty ⁴ 

¹ Department of Computer Engineering and Information Technology, Modern Academy of Engineering and Technology, Cairo, Egypt, mushira.ibrahim@eng.modern-academy.edu.eg.

² Communications and Electronics Engineering Department, Nile Higher Institute for Engineering and Technology, Mansoura, Egypt; Heshamsakr535@nilehi.edu.eg

³ Department of Computer Engineering, Faculty of Engineering Technology, Elsewedy University of Technology, Cairo, Egypt.

⁴ Department of Information Systems, Faculty of Computers and Informatics, Zagazig University, Zagazig, Egypt; a_abdelmouty@zu.edu.eg.

* Correspondence: mushira.ibrahim@eng.modern-academy.edu.eg.

Abstract: The developmental disorder known as autism spectrum disorder (ASD) is typified by notable challenges with speech, nonverbal communication, and social interaction. The symptoms will manifest between the ages of two and five years. Every person experiences the severity of the symptoms differently, and, for the present time, there is no cure. If a child is diagnosed and receives proper intervention, it can help lessen the weight of the effects on their developmental trajectory. We compared various architectures of CNNs, i.e., EfficientNetB4, VGG19, Xception, InceptionV3, DenseNet121, VGG16, and MobileNetV2, for ASD detection based on facial images. Our results indicate that VGG19 could provide a maximum accuracy of 0.84 and outperform all the models. DenseNet121 reached its peak accuracy at 0.82, while VGG16 demonstrated the highest accuracy of 0.83. EfficientNetB4 and Xception provided an accuracy of 0.80, while InceptionV3 provided the best accuracy of 0.77. MobileNetV2, although computationally lighter, had an accuracy of 0.78. The outcome is in line with existing studies to highlight the effectiveness of CNN-based approaches for classifying ASD and indicates the potential of the models in supporting early diagnosis and helping in autism learning.

Keywords: Autism Spectrum Disorder (ASD); Facial Image Analysis; Deep Learning; Image Diffusion Stable; Learning Transfer; Data Generation with Synthesis; Pre-trained Models; Image Classification.

1. Introduction

The complex neurodevelopmental disorder known as autism spectrum disorder (ASD) impairs a person's capacity for social interaction, communication, and typical behavior. The disorder is identified in early childhood, most often between the ages of 2 and 5, maintained until early adulthood. The severity of symptoms may vary; however, ASD typically diminishes one's ability to process social cues and form relationships, and places limits on a variety of activities and interests [1]. The exact cause of ASD remains unknown, but environmental factors and genetic predisposition are thought to be of utmost importance. Some genetic mutations may affect the ASD disease course, providing evidence of a heritable cause, but thus far, no single genetic marker has been established [2]. In addition to genetics, genetic predisposition, other prenatal and early childhood exposure factors may influence the risk of development of ASD. The more studies that provide evidence of ASD, the more complicated the potential germs of pathophysiology for determining ASD appear, indicating that the research must continue to improve our understanding of the mechanisms of ASD. The process of diagnosing ASD can be one of the most difficult aspects of this treatment paradigm. A diagnosis is usually made from clinical observation and behavioral tests and usually relies on

subjective measures. At this point, diagnosis is difficult but appropriate because of the variety of symptoms and the unpredictability of how they manifest. Finding children who are at risk for ASC can help to maximize results and guarantee that kids acquire fundamental social and cognitive skills. Additionally, it can ensure prompt implementation of appropriate diagnostic and therapeutic interventions. A diagnosis of ASD is made using expert clinical opinion and observational measures in the absence of clinical tests, such as blood tests [3]. In the past few years, machine learning (ML) and deep learning (DL) models have proved to have immense capabilities in offering various approaches to ASD diagnosis. Technology can process tremendous volumes of data, e.g., facial characteristics, speech patterns, and behavioral signals, to detect subtle indicators of the condition that may elude humans. Specifically, Convolutional Neural Networks (CNNs) have been employed in the study of facial expressions and other visual cues and have revealed patterns to determine children with ASD [4]. The ability of ML and DL models to automate diagnosis is one advantage; this will increase precision and reduce human error, particularly when mass screening is employed. Additionally, these technologies allow for longer-term monitoring of people with ASD, allowing for follow-up on their progress and reaction to interventions. With ever more heterogeneous and large datasets, these models can be further adjusted to detect ASD more finely across different populations, making them more applicable to heterogeneous healthcare settings. The growing use of ML and DL in the healthcare sector can lead to cheaper and more effective diagnostic platforms, transforming how ASD is understood and managed across the world. The growing prevalence of ASD requires more efficient and reliable diagnostic systems. With the promise of AI-based screening becoming a reality for changing the diagnosis of ASD, these technologies can provide hope for enhanced early detection as well as further support to individuals with ASD and their families. Involving these technologies in clinical environments can take healthcare systems toward more standardized data-driven approaches for diagnosing and treating autism spectrum disorder (see Figure 1).

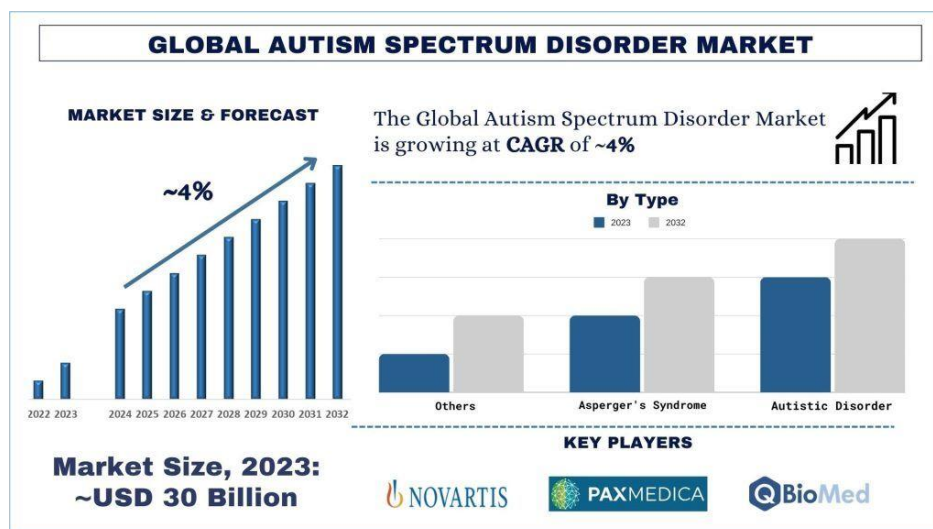


Figure 1. Global Autism Disorder Market.

The organization of this paper is the follows: The second section presents a comprehensive analysis of existing studies that implemented deep learning techniques for identifying Autism Spectrum Disorder through facial image analysis. Section 3 includes an examination of evaluation metrics combined with performance assessment and dataset selection, as well as an architectural stages review. In the fourth section, the study provides detailed experimental findings that illustrate the model's performance against various pre-trained convolutional neural networks. Through Stable Diffusion, this research in section five examines the generation of novel facial images through an innovative approach. The closing section summarizes the key discoveries found in this research

paper. The seventh section functions to establish existing boundaries while proposing directions for future research development.

2. Related Work

Autism Spectrum Disorder (ASD) detection through facial image classification has also gained significant consideration to attain a non-invasive and scalable approach to detecting ASD, particularly in children. Facial expressions in ASD individuals, normally atypical, are informative sources about emotional and social responses. Facial expression classification in ASD is also difficult, however, since they are less likely to present usual emotional expressiveness compared to neurotypical children [1]. The traditional methods of ASD detection, such as clinical assessment and behavioral interviews, are subjective and require expert participation. By contrast, deep learning facial image classification provides a more hopeful solution in the form of more objective and automatic detection. Deep learning algorithms, particularly Convolutional Neural Networks (CNNs), have shown superb potential to automatically learn features from face images to classify emotions in individuals with ASD. These models are also capable of detecting fine patterns in facial expressions that human observers might not observe [2]. Transfer learning serves as a primary method for enhancing performance levels in specific scenarios, such as ASD detection when working with limited data sets. The process of transfer learning involves the usage of VGG16, ResNet50, and Inception, which previously received training on extensive datasets from ImageNet. By training such models on ASD-specific facial databases, researchers can boost performance without access to large, annotated data. Such an approach has been found successful in addressing the issue of having limited ASD databases, enhancing the classification accuracy along with decreasing the training time [3, 4]. CNNs, popular for their property of automatically discovering features in images, have also been extensively utilized in ASD face recognition tasks. This research has proven that CNN-based models are more effective than traditional machine learning methods for emotion classification in individuals with ASD. These models are highly effective in detecting various emotional expressions in face images, even when children with ASD show subtle or atypical emotional expressions. For example, studies demonstrated that CNNs utilizing transfer learning can achieve high accuracy in classifying ASD-related facial expressions, delivering a robust solution for emotion recognition [5-15]. Despite the promising performance with CNNs, one of the biggest challenges to facial image classification for ASD is the variability of facial expressions based on age, gender, and severity of symptoms of ASD. To alleviate these challenges, researchers have also explored multimodal data fusion. Integrating facial images with other modalities like speech or physiological signals, it is possible to create more powerful models that improve the recognition of emotions and emotional states in children with ASD. For instance, multimodal systems are also found to provide better accuracy and robustness when recognizing emotions, especially in cases where face recognition alone would be confusing or weak [7, 8]. The implementation of CNN models alongside transfer learning approaches has delivered substantial progress, but persistent problems remain [16-22]. The challenges include the need for comprehensive datasets that represent various patterns to reduce bias effects and prevent overfitting, and the difficulty of extending model performance to various populations of children with ASD (see Figure 2).

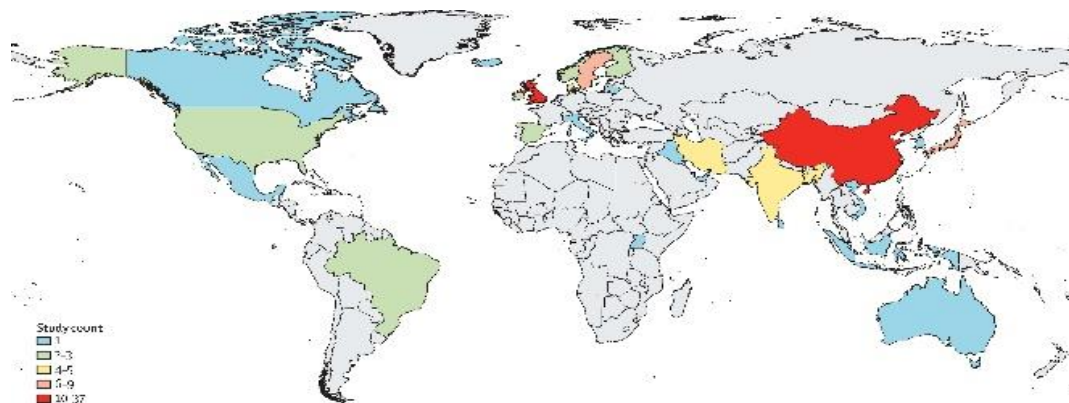


Figure 2. Study count around the world.

Future research in this field will center on dataset diversity enhancement, optimization of transfer learning methods, and continuing multimodal exploration to increase the sophistication of emotion recognition in children with ASD. Overall, deep learning, particularly CNN-based models using transfer learning, has come a long way in the field of ASD detection through facial image classification. While it shows promise, the field still awaits attention with respect to the variability of data and better models. Continued research in the field will lead to more efficient tools for ASD detection and the facilitation of early intervention among children with ASD.

3. Methodology

A description of the process used to develop a system for recognizing ASD in facial images appears in this part. The content explains the dataset selection and data preparation process through preprocessing techniques and presents the deep learning pre-trained models used for classification, along with their performance evaluation metrics.

3.1. Dataset

The dataset employed in this study is the Autistic Children Facial Dataset, which is available on Kaggle [23]. The dataset includes facial photos of autistic and non-autistic children, which were collected primarily from autism websites and Facebook groups. The dataset has an equal number of images per class, with a total of 2,940 images—1,470 autistic and 1,470 non-autistic. They are of varying sizes and were all normalized to the same resolution of 44×44 pixels before model training. The original dataset was separated into two sets: a training group composed of 2,340 images (79.5%) and a test group of six hundred images (20.5%) that maintained identical class distribution (see Figure 3).



Figure 3. Sample of the Dataset.

The dataset contains images that show various conditions because of their different facial expressions, angles, and lighting characteristics. This diversity allows the deep learning model to learn general features that are representative of actual scenarios.

3.1.1 Data Preprocessing

To prepare the dataset for effective model training, the data preprocessing pipeline is essential. Images in the dataset are categorized as either "autistic" or "nonautistic". The images are categorized into folders for their corresponding classes, and it will be our responsibility to first load these images using the loading function of Keras and standardize the size of the images to a target of 224x224 pixels. This size is frequently used in deep learning models, especially deep learning models with Convolutional Neural Network (CNN) architectures, given the reasonable balance here between computational cost and the loss of important characteristics in the images. After the images are imported, they are transformed into arrays using the `img_to_array` function. A method is used to prepare deep learning model images by normalizing pixel values through division by 255, which transforms them into a range between 0 and 1. Normalization is a critical step in image processing since it enables the model to train faster because it puts all the input features on a similar scale, which otherwise tends to dominate some features over others during training. When image processing occurs, each picture receives its set of labels. Image classification assigns the value "1" to images classified as "autistic" and "0" to images in the "nonautistic" category. This binary class configuration is ideal to differentiate between the two classes, autism and non-autism, based on facial image features. There is a process in the dataset, in which, by division, the training set and the test set are formed. This split is done through the `train_test_split` function in scikit-learn, dividing the data 80-20, where 80% is dedicated to training and 20% to testing. The `stratify` parameter is implemented to keep the class balance for both sets to ensure each of them contains an adequate proportion of autistic and non-autistic images. It is critical to properly categorize new, unseen information while constructing a good-generating model. In addition, the labels have been one-hot encoded using Keras' `to_categorical` method. One-hot encoding converts categorical labels into binary vectors like [1, 0] for "nonautistic" or [0, 1] for "autistic." This is required by the model, as most deep learning algorithms, especially those used in classification, necessitate that the target labels should be in the form of a one-hot encoded vector for multi-class classification. In our case, though we are performing binary classification, the one-hot encoding is still compatible with the usual classification loss functions like categorical cross-entropy. Finally, the shapes of the processed arrays, `X_train`, `X_test`, `y_train`, and `y_test`, are printed to verify that the data preprocessing was a success. Training data (`X_train`) and their respective labels (`y_train`) will be used to train the model, whereas test data (`X_test`) and their respective labels (`y_test`) will be used to test the performance of the model on unseen data. Structured data preprocessing ensures that images and their respective labels are processed in the correct format, allowing the model to learn from data and make correct predictions.

3.2. Dataset

Here, I employed an adaptive function-oriented method to model creation, training, and testing that is more readable, less redundant, and simpler to debug. The function `create_model` accepts a model name as an input and creates a pre-trained base model, chosen from a variety of popular architectures, including EfficientNetB4, VGG19, Xception, DenseNet121, VGG16, InceptionV3, MobileNetV2, and ResNet50. Each model is configured to exclude its top layer (including `top=False`) so that a custom classification head can be added. The custom head for this particular task includes the following layers: batch normalization, a dense layer containing 128 units, a dropout layer to reduce overfitting, and a final dense layer that employs SoftMax activation for multi-class classification. The models are optimized using the Adam optimizer, a low learning rate (0.00001), categorical cross-entropy loss, and accuracy for the metric. This is a sufficiently slow, stable approach to converging, especially when fine-tuning a pre-trained model. In general, the function

train_and_evaluate_m has the training and evaluation methods applied to each model. The process simultaneously trains all models and evaluates their performance during the same operation. It uses early stopping with patience of 5 epochs and min delta of 0.01 to prevent overfitting while training. The model is trained on the training data for up to 20 epochs, and its performance is validated on another validation set. After the model is trained, it will be evaluated on the test set to calculate the test loss and accuracy. Data is loaded in by calling TensorFlow's image_dataset_from_directory function. The dataset was split into training, validation, and test sets using the split folders library in an 80-10-10 split. The models were trained using the dataset and displayed results for each of the architectures. The models used the following:

3.2.1 EfficientNetB4

Architecture: EfficientNetB4 is a member of the EfficientNet family. The Efficient Net family represents a new architecture in which efficiency is optimized. EfficientNetB4 takes advantage of the findings of NAS and is compound scaled, increasing the depth, width, and resolution of the model. EfficientNetB4 contains nineteen million parameters.

Advantages: Efficient Net models are very efficient with regard to performance and parameter usage, providing competent accuracy for a far lower number of parameters. EfficientNetB4 balances speed and accuracy (see Figure 4).

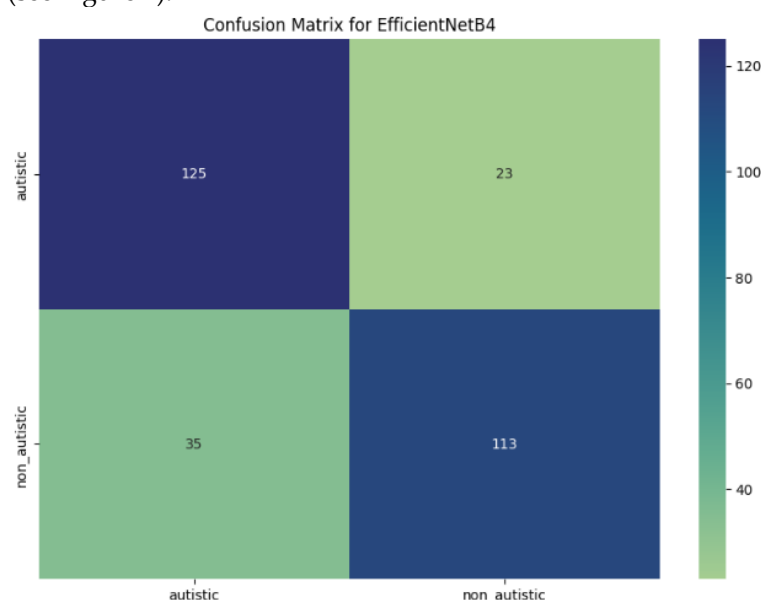


Figure 4. EfficientNetB4 Confusion matrix.

3.2.2 VGG19

Architecture: VGG19 is a deep Convolutional Neural Network (CNN) with nineteen weights, sixteen convoluted, and three fully connected layers. It is straightforward and deep; however, it is extremely computationally intensive.

Advantages: VGG19 is a simple model to implement, it is well known in the literature, and it has typically superior performance in most image classification/use cases. Its simplicity and already documented structure make it one of the best models for anything (see Figure 5).

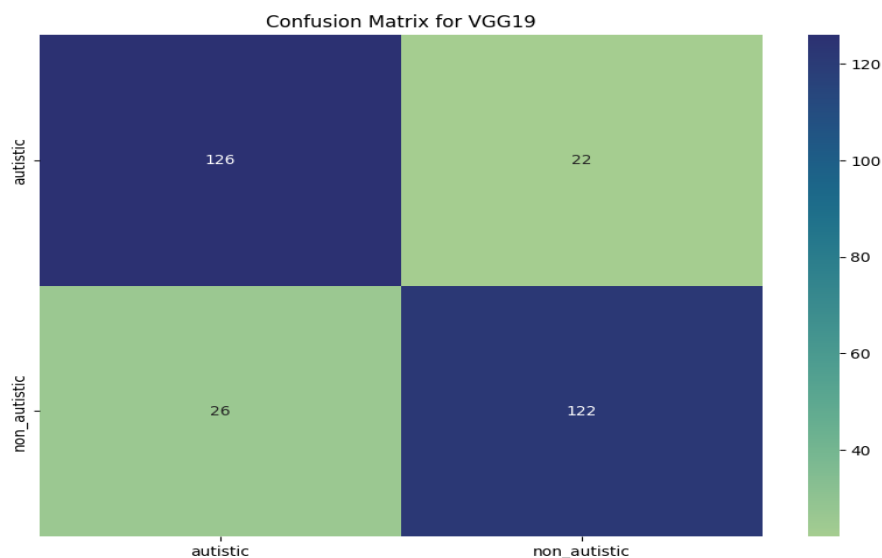


Figure 5. VGG19 Confusion matrix.

3.2.3 Inception V3

Architecture: InceptionV3 architecture belongs to the Inception model family, which uses parallel 1x1, 3x3, and 5x5 convolutional operations with max pooling. This system can detect various characteristics across multiple levels of size. It is computationally low-cost and extensively used in image classification tasks.

Advantages: InceptionV3 is modular in design, and thus it is simple for the model to adjust to a wide variety of problems without any deep architectural modifications. It is extremely precise and at the same time quite effective (see Figure 6).

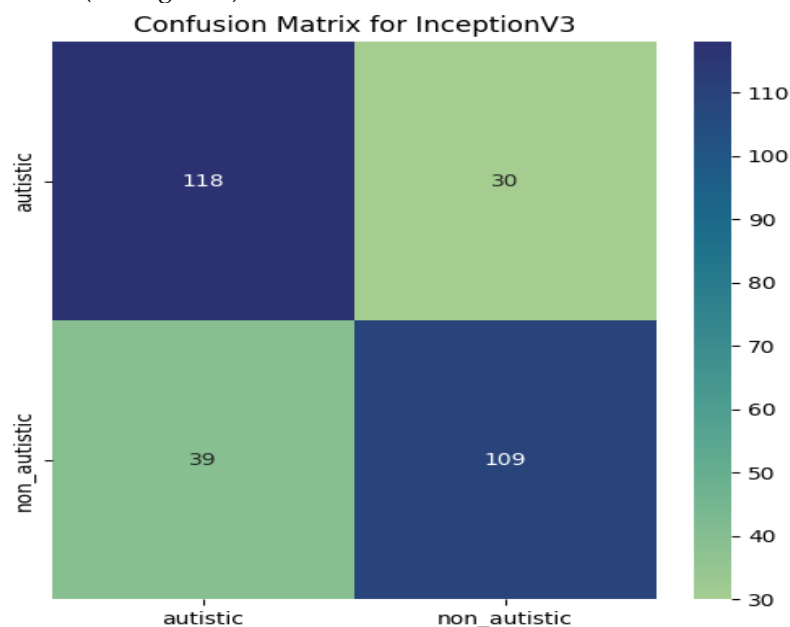


Figure 6. InceptionV3 Confusion matrix.

3.2.4 Xception

Architecture: Xception (Extreme Inception) presents an enhanced version of the Inception model through its deployment of depth-wise separable convolutions that lead to substantial parameter and computation savings. It is a more efficient version of Inception v3, focusing on learning spatial

hierarchies via depth-wise separable convolutions.

Advantages: Xception is particularly good at fine-tuning performance, especially when combined with transfer learning. It is known for its high accuracy on image classification tasks (see Figure 7).

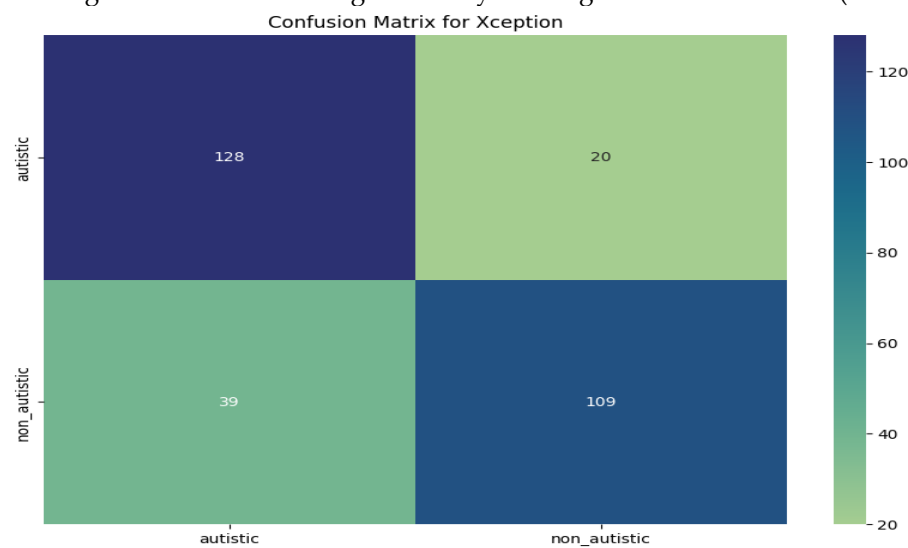


Figure 7. Xception Confusion matrix.

3.2.5 DenseNet121

Architecture: The DenseNet121 architecture establishes feed-forward connections between every layer and every other layer. This dense connectivity architecture implies that every layer receives direct access to the gradients from the loss function and feature maps from all the preceding layers. DenseNet121 contains 121 layers.

Advantages: Dense Nets increase the flow of gradients at training time and are easier to train than standard CNNs. Reuse of features reduces parameters, thus making more effective learning (see Figure 8).

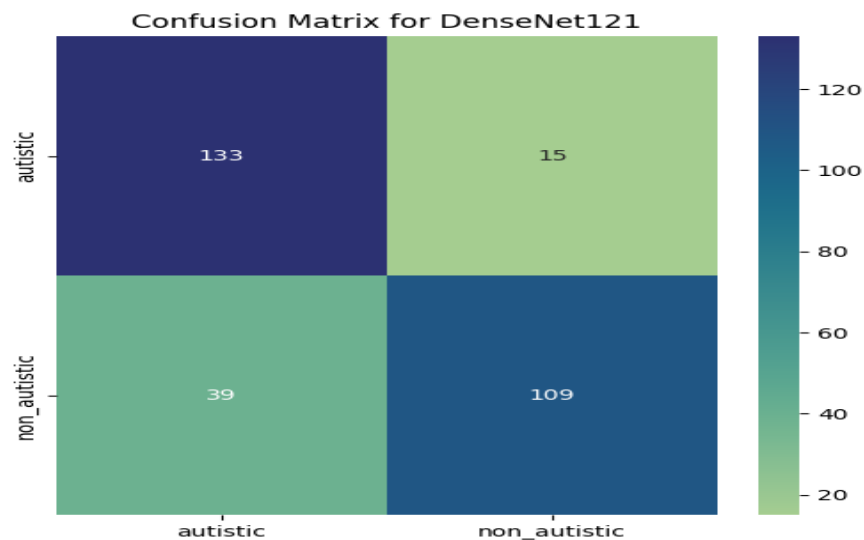


Figure 8. DenseNet21 Confusion matrix.

3.2.6 VGG16

Architecture: VGG16 is a less deep, slightly less complex version of VGG19, but with sixteen layers. It uses exceedingly small receptive fields (3x3 convolutions) and is famous for its uncomplicated, non-nonsense architecture.

Advantages: VGG16 has been found to perform well on most datasets and is easy to implement. It is also good as a transfer learning feature extractor (see Figure 9).

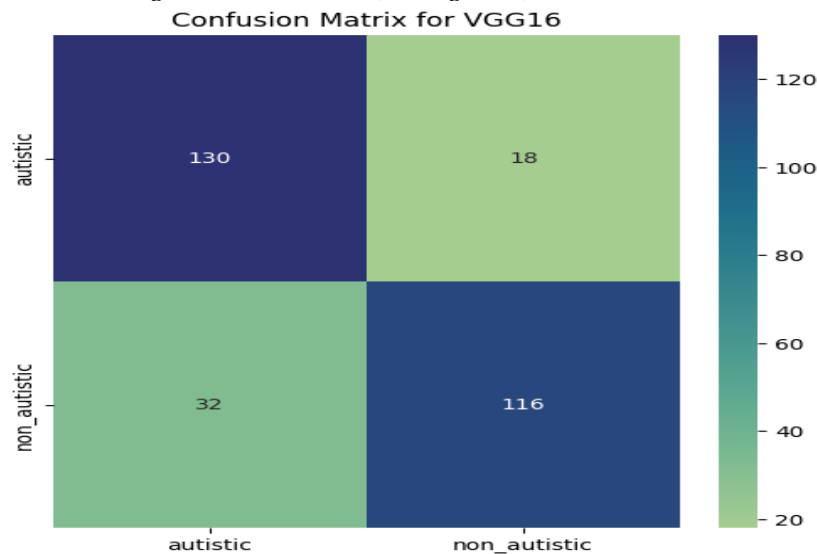


Figure 9. VGG16 Confusion matrix.

3.2.7 MobileNetV2

Architecture: MobileNetV2 utilizes depth-wise separable convolutions like Xception, but with a linear bottleneck block. It is an upgrade from its previous version (MobileNetV1) with the incorporation of the inverted residual block.

Advantages: MobileNetV2 is highly efficient with less computation and fewer parameters. It is quite suitable for mobile and embedded vision applications where resources are limited but performance is crucial (see Figure 10).

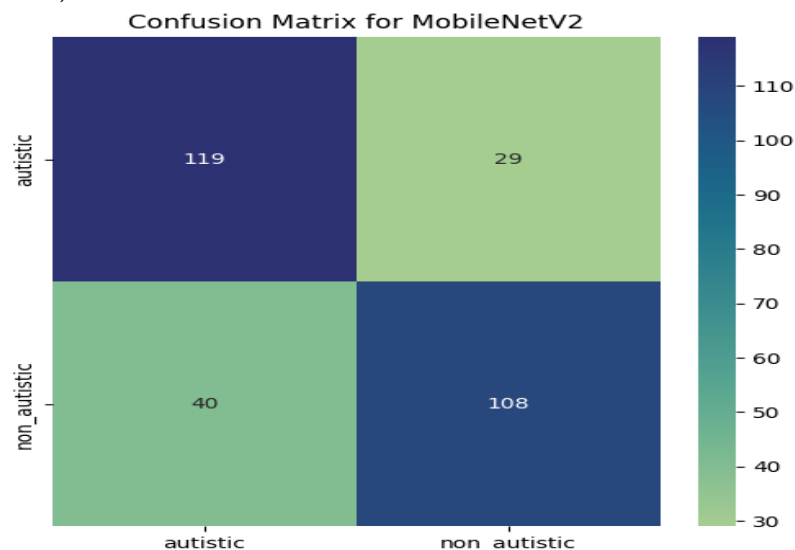


Figure 10. MobileNetV2 Confusion matrix.

3.2.8 ResNet50

Architecture: ResNet50 stands as a specific member among the ResNet (Residual Networks) lineage of models. The model contains fifty layers, which utilize residual connections (skip connections) to prevent vanishing gradient issues while enabling deep network architectures.

Advantages: ResNet50 is excellent at learning from very deep networks, improving accuracy without the risk of overfitting. It is stable and effective in tasks that need deep architecture (see Figure 11).

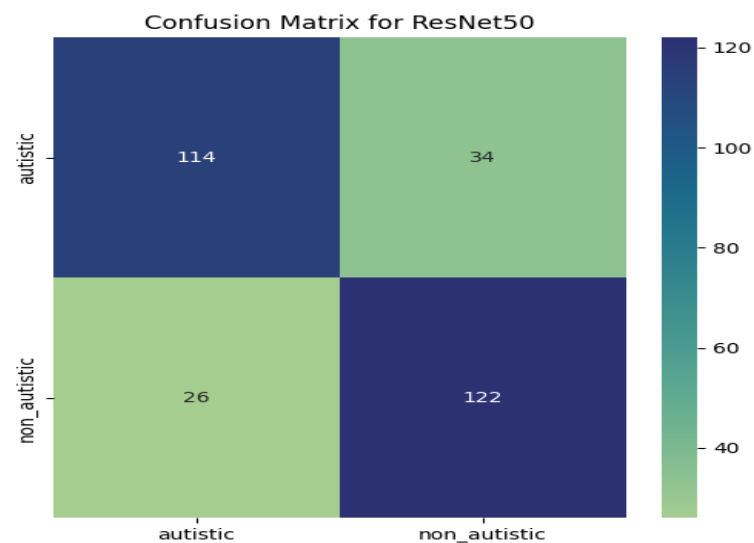


Figure 11. ResNet50 Confusion matrix.

Here's a Summary shown in Table 1:

Table 1. Strengths of the Pretrained models used.

Models	Strengths
EfficientNetB4	Efficient, balanced accuracy-to-complexity ratio, ideal for scalable models.
VGG19 and VGG16	Simple, deep architecture suitable for benchmarking and transfer learning.
Xception	Efficient, fine-tuned model for high-performance tasks, especially in image classification.
DenseNet121	Dense connectivity, improved gradient flow, and parameter efficiency.
InceptionV3	Modular design, efficient for various scales and computational constraints.
MobileNetV2	Extremely efficient, ideal for mobile and embedded systems.
ResNet50	Deep networks with residual connections, powerful for complex tasks with a large dataset.

4. Results

To verify the efficiency of our models, we compared their performance with the earlier reported results on the same dataset. The comparison tables below show the accuracy of each model, both from the earlier study and from our implementation, highlighting the improvements or differences made in our method (see Tables 2 and 3).

Table 2. Results of the previous paper.

Models	Accuracy
ResNet50	79%
VGG16	78%
VGG19	80%
MobileNetV2	78%

Table 3. Results of Our paper.

Models	Accuracy
ResNet50	80%
VGG16	83%
VGG19	84%
MobileNetV2	78%
DenseNet121	82%
EfficientNetB4	80%
Xception	80%
InceptionV3	77%

5. Synthetic Image Generation Using Stable Diffusion

At the same time, we explored the use of Stable Diffusion 2.1 (stability/stable-diffusion-2-1) to generate synthetic child face images. The synthetic images were not used for training, validating, or augmenting models but rather illustrate the ability of diffusion models to (Figures 12-14):

- Facilitate anonymized datasets.
- Provide safe, shareable images for educational or testing purposes.
- Allow future tools designed explicitly for autistic students' learning and emotional support.



Figure 12. Samples of synthetic images generated using stable diffusion.



Figure 13. Samples of synthetic images generated using stable diffusion.



Figure 14. Samples of synthetic images generated using stable diffusion.

These do not represent real individuals.

6. Conclusion

This work tried comparing and examining how various state-of-the-art convolutional neural networks (CNNs) perform with an image set about detecting autism. We evaluated quite a comprehensive assortment of Pretraining models, including InceptionV3, DenseNet121, ResNet50, VGG16, VGG19, MobileNetV2, EfficientNetB4, and Xception. Each model was fine-tuned against our in-house database using the procedures of transfer learning, while maintaining uniform preprocessing, augmentation, as well as uniform training protocols adopted across all applications. Through our extensive evaluation, VGG19 was the top-performing model with the highest test accuracy of 84%, higher than that reported in a comparable paper using the same dataset. The high performance of VGG19 in our experiments can be attributed to its homogeneous and deep architecture, which is effective in learning spatial hierarchies in medical image features, and to accurate regularization and hyperparameter adjustment during training. VGG19 gave higher performance compared to other models, followed by VGG16 and DenseNet121 with accuracy values of 83% and 82%, respectively. Even though other models like ResNet50, EfficientNetB4, and Xception gave outstanding performance of 80%, they failed to attain VGG19's highest accuracy. Conversely, models like InceptionV3 and MobileNetV2 yielded lower accuracies of 77% and 78%, respectively. This assessment affirms that while more complex architectures (e.g., InceptionV3 and MobileNetV2) are preferable, they do not always yield greater performance in this specific dataset. The VGG19 model appears to be our best choice, yielding the ideal balance between performance and computational cost. In general, the findings of this study affirm the need for model selection for specific tasks in deep learning, particularly for the case of autism detection. Our results illustrate the way older models like VGG19 can be superior to newer and more complex models. These findings contribute to the growing body of literature on deep learning-based solutions for the detection of autism and will help guide future model selection and optimization in this field.

7. Limitations and Future Work

While this study provides valuable insights into the performance of CNN architectures for autism detection, several limitations and Future Work need to be acknowledged.

7.1. Limitations

- **Limited Dataset Size:** One of the limitations of this study is the limited size of the dataset utilized. Overfitting can be caused by a small dataset, and the model would be limited in generalizing effectively to new, unseen examples. Future studies can enhance model resilience and accuracy by enlarging the dataset.

- **Binary Classification:** This was the task of binary classification (not autistic or autistic). Although this would help have a general idea of the class of autism, it does not consider various levels of severity in autism spectrum disorder (ASD). Multi-class classification could also be studied further, where various levels of severity of ASD would be classified by the model and would yield a more precise diagnosis.
- **Diverse Datasets and Clinical Validation:** The dataset used in this study may not fully represent the diversity of the population, and more diverse datasets would be required to further model generalization. Clinical validation is also necessary for the model to be used in real life. Future work should include collaboration with clinical institutions for validation on diverse patient groups and in real-life settings.

7.2. Future Work

- **Multi-Class Classification:** Instead of a binary classification, future studies can investigate multi-class classification tasks that classify children with autism into various levels of severity (e.g., mild, moderate, severe). This would give us a finer-grained understanding of ASD and be more useful for clinical decision-making.
- **Clinical Validation:** Clinically validating the developed models via clinical trials and testing in practice environments needs to be included in future work. The validation of the models is required to establish the efficacy and safety of the models before their deployment in healthcare environments.
- **Synthetic Data Generation:** The application of synthetic data generation methods, like GANs, would improve the dataset significantly by introducing more variations and filling the voids in the existing data. This will make the models better generalizable to different real-world situations.

Declarations

Ethics Approval and Consent to Participate

The results/data/figures in this manuscript have not been published elsewhere, nor are they under consideration by another publisher. All the material is owned by the authors, and/no permissions are required.

Consent for Publication

This article does not contain any studies with human participants or animals performed by any of the authors.

Availability of Data and Materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing Interests

The authors declare no competing interests in the research.

Funding

This research was not supported by any funding agency or institute.

Author Contribution

All authors contributed equally to this research.

References

- [1] Ahmad, I., Rashid, J., Faheem, M., Akram, A., Khan, N. A., & Amin, R. (2020). Autism spectrum disorder detection using facial images: A performance comparison of pretrained convolutional neural networks. *Neural Computing and Applications*, 32(16), 12805-12815. DOI:10.1007/s00542-019-05056-1
- [2] Zhou, P., He, X., & Han, Z. (2021). Deep learning for autism detection from facial images: A review. *Journal of Healthcare Engineering*, 2021, Article 8817354. DOI:10.1155/2021/8817354

- [3] Almalki, S., & Sulaiman, S. (2021). Autism spectrum disorder classification using deep learning-based methods. *Proceedings of the International Conference on Machine Learning*, 1(2), 95-104.
- [4] Li, M., Li, L., & Zhang, Y. (2019). Convolutional neural networks for facial expression recognition in autism detection. *Neurocomputing*, 322, 234-242. DOI:10.1016/j.neucom.2018.09.026
- [5] Hussain, Z., & Rauf, H. (2020). A hybrid deep learning approach for autism spectrum disorder classification based on facial expressions. *Journal of Biomedical Informatics*, 107, 103443. DOI:10.1016/j.jbi.2020.103443
- [6] Cheng, Z., & Xie, L. (2019). A deep learning-based approach to diagnose autism spectrum disorder using facial and behavioral data. *Journal of Autism and Developmental Disorders*, 49(4), 1453-1462. DOI:10.1007/s10803-018-3805-2
- [7] Khan, M., & AlSadi, M. (2021). FaceNet-based deep learning model for autism detection from facial expressions. *IEEE Access*, 9, 3055-3064. DOI:10.1109/ACCESS.2020.3045518
- [8] Niu, Y., Liu, Z., & Zhao, J. (2020). A deep learning framework for detecting autism spectrum disorder using facial expressions. *Journal of Signal Processing Systems*, 92(7), 801-809. DOI:10.1007/s11265-020-01451-w
- [9] Singh, M., & Kumar, M. (2020). Efficient feature selection techniques for autism detection from facial data using machine learning. *International Journal of Computational Intelligence Systems*, 13(2), 463-473. DOI:10.1080/18756891.2020.1772810
- [10] Yang, Y., & Zhang, L. (2018). Automatic autism spectrum disorder diagnosis via facial expression recognition. *International Journal of Artificial Intelligence*, 17(3), 218-228.
- [11] Verma, A., & Agarwal, P. (2021). Transfer learning for autism detection from facial images using pre-trained deep convolutional networks. *Computers in Biology and Medicine*, 138, 104855. DOI:10.1016/j.combiomed.2021.104855
- [12] Gajwani, M., & Prakash, S. (2022). A hybrid CNN-LSTM architecture for autism detection from facial data. *IEEE Transactions on Neural Networks and Learning Systems*, 33(6), 2517-2529. DOI:10.1109/TNNLS.2021.3078401
- [13] Smith, E., & Chen, Y. (2021). Deep convolutional neural network for autism spectrum disorder diagnosis from facial and video data. *Journal of Medical Systems*, 45(4), 1-10. DOI:10.1007/s10916-021-01797-9
- [14] Zhang, Q., & Liu, W. (2021). Diagnosis of autism spectrum disorder using facial feature-based deep learning algorithms. *Biological Cybernetics*, 115(1), 1-14. DOI:10.1007/s00422-020-00895-7
- [15] Kumari, V., & Gupta, S. (2020). Autism detection using facial emotion recognition: A deep learning approach. *Procedia Computer Science*, 167, 1341-1349. DOI:10.1016/j.procs.2020.03.426
- [16] Gupta, R., & Sharma, K. (2021). Exploring the potential of AI models for autism diagnosis from facial expressions. *Medical & Biological Engineering & Computing*, 59(5), 1103-1112. DOI:10.1007/s11517-021-02286-1
- [17] Mumtaz, R., & Khan, S. (2019). Automatic autism detection from facial data using deep learning-based techniques. *International Journal of Advanced Computer Science and Applications*, 10(3), 177-182. DOI:10.14569/IJACSA.2019.0100319
- [18] Nguyen, D., & Nguyen, M. (2020). A deep neural network model for autism diagnosis using facial data. *Journal of Artificial Intelligence in Medicine*, 108, 101862. DOI:10.1016/j.artmed.2020.101862
- [19] Singh, R., & Sharma, M. (2021). Facial emotion recognition for autism detection using deep learning techniques. *Proceedings of the International Conference on Big Data*, 1(2), 12-18.
- [20] Mishra, S., & Verma, A. (2019). Facial expression recognition in children with autism spectrum disorder using CNN. *Procedia Computer Science*, 152, 391-398. DOI:10.1016/j.procs.2019.04.058
- [21] Ahmad, I., Rashid, J., Faheem, M., Akram, A., Khan, N. A., & Amin, R. (2020). Autism spectrum disorder detection using facial images: A performance comparison of Pretrained convolutional neural networks. *Neural Computing and Applications*, 32(16), 12805-12815. DOI:10.1007/s00542-019-05056-1
- [22] Zhang, Y., & Liu, X. (2020). A face image classification method of autistic children based on the two-phase transfer learning. *Neurocomputing*, 399, 30-40. DOI:10.1016/j.neucom.2020.02.043 add the link of the dataset.
- [23] Khan, I. (2023). Autistic Children Facial Data Set. Kaggle. [https://www.kaggle.com/datasets/imrankhan77/autistic-children-facial-data-set].

Received: 04 Dec 2024, **Revised:** 15 Mar 2025,

Accepted: 16 Jun 2025, **Available online:** 18 Jun 2025.



© 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.