Journal of Disability Research

2024 | Volume 3 | Pages: 1–8 | e-location ID: e20240025
DOI: 10.57197/JDR-2024-0025

Facial Classification for Autism Spectrum Disorder

Maram Fahaad Almufareh[1], Samabia Tehsin[2], Mamoona Humayun[1,*] and Sumaira Kausar[2]

[1]Department of Information Systems, College of Computer and Information Sciences, Jouf University, Sakakah 72388, Saudi Arabia
[2]Department of Computer Science, Bahria University, Islamabad, Pakistan

Correspondence to:
Mamoona Humayun*, e-mail: mahumayun@ju.edu.sa, Mobile: 00966536839205

Received: February 26 2024; Revised: March 9 2024; Accepted: March 10 2024; Published Online: March 29 2024

ABSTRACT

Autism spectrum disorder (ASD) is a mental condition that affects people's learning, communication, and expression in their daily lives. ASD usually makes it difficult to socialize and communicate with others, and also sometimes shows repetition of certain behaviors. ASD can be a cause of intellectual disability. ASD is a big challenge in neural development, specially in children. It is very important that it is identified at an early stage for timely guidance and intervention. This research identifies the application of deep learning and vision transformer (ViT) models for classification of facial images of autistic and non-autistic children. ViT models are powerful deep learning models used for image classification tasks. This model applies transformer architectures to analyze input image patches and connect the information to achieve global-level information. By employing these techniques, this study aims to contribute toward early ASD detection. ViT models are showing good results in identifying facial features associated with ASD, leading toward early diagnostics. Results show the ViT model's capability in distinguishing the faces of autistic and non-autistic children.

KEYWORDS

intellectual disability, artificial intelligence, disability diagnosis, autism, facial classification

INTRODUCTION

Autism spectrum disorder (ASD) is a neural-developmental condition that results in challenges with social communication, expression, and interaction. These can include difficulty in understanding and responding to social interactions, talking effectively and spontaneously, and learning nonverbal communication (Hannon et al., 2023). Recognizing these symptoms early in a child's behavior and development is very important for providing the necessary support and guidance. This study investigates the potential of modern technologies in artificial intelligence and computer vision to analyze the facial images of children. These models are targeted to identify complex patterns in facial expressions and features that are associated with ASD. By analyzing these patterns, the study targets to contribute to the development of an early, easy, and accurate tool for ASD identification, paving the path for specialized support for each child.

Around 30% of individuals with ASD also meet the criteria for an intellectual disability (ID). The presence of an ID cannot be predicted solely based on an ASD diagnosis (De Rham and Marco, 2016). Other factors, like genetic and environmental influences, also play a role. Facial expressions and features are used as fundamental clues from old ages, to identify mental disorders such as autism. Vision transformer (ViT) models, initially introduced for processing natural languages, have shown remarkable performance in analyzing complex structures, relationships, and arrangements. This study aims to use ViT models for analyzing the facial expressions and features unique to children with ASD.

Early detection of autism in kids is very important for availing their maximum potential. Watching for different clues for delays in responses and unusual behaviors can make the early detection of ASD possible. Early diagnosis can lead to timely therapies, and support programs can significantly improve a child's development and quality of life. So this study can really provide doctors and parents with a clue toward early diagnosis.

Human faces tell countless stories. Can computers "read" these stories to help us understand autism? This study explores a new approach using powerful "computer vision" tools called ViT models. These models analyze facial features of children with and without autism, searching for patterns that might offer clues about the condition. While acknowledging the complex nature of autism and its variations, this study aims to develop a tool for early identification. This paper describes the existing research on facial cues in autism, explains how we collected and prepared data, and details the design and training of the ViT model.

The primary research objectives of this paper are as follows:

- To investigate and identify specific facial expressions and features that serve as potential indicators of ASD in children.
- To assess the effectiveness and adaptability of ViT models in the classification of facial images, specifically in the context of distinguishing between autistic and non-autistic children.
- To contribute to the field of early autism detection by developing a robust and accurate classification model.

This paper is organized into five sections. The second section provides the literature review on this topic; the third section explains the applied research methodology; the fourth section provides the analysis of results; and the fifth section concludes the paper.

LITERATURE REVIEW

Extensive research has investigated the potential of facial expressions and features as biomarkers for ASD in children (Golarai et al., 2006). Studies have identified atypicalities in various aspects. Individuals with ASD often demonstrate decreased eye gaze duration and frequency compared to neurotypical children (Griffin et al., 2023). Research suggests altered smiles in ASD, characterized by reduced intensity, symmetry, and genuine appearance (Clements et al., 2023). Some studies report subtle differences in facial features like intercanthal distance or philtral length in children with ASD, though findings remain inconsistent (Hartston et al., 2023). It is crucial to acknowledge the limitations of relying solely on facial cues for ASD diagnosis. The heterogeneity of the disorder, cultural variations in expression, and potential overlap with other conditions necessitate a multifaceted approach.

Artificial intelligence has been extensively used in many domains (Aleem et al., 2022; Mehran et al., 2023; Sahu et al., 2024). But in the healthcare sector, it is very crucial and helpful (Almufareh et al., 2023a, b, c, d, 2024a, b).

Deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful tool for analyzing medical images. Previous studies have employed CNNs to classify facial images for ASD detection, achieving promising results (Celard et al., 2023; Dhar et al., 2023; Li et al., 2023). Autism detection is also done using deep learning methods (Jeyarani and Senthilkumar, 2023; Talaat, 2023; Zhang et al., 2023). However, CNNs primarily focus on local features and struggle to capture long-range dependencies and contextual information within images. This limitation can hinder their ability to accurately identify complex and difficult markers of ASD in facial expressions.

Autism detection is gaining the focus of research community due to its significant impacts (Alharthi and Alzahrani, 2023; Cao et al., 2023; Deng et al., 2024).

ViT models represent a recent advancement in image classification, representing superior performance compared to traditional CNNs (Arkin et al., 2023). ViT models are performing very well in the healthcare industry, including diagnosis, prognosis, and treatments (Oukdach et al., 2024; Pacal, 2024). ViT models operate directly on image patches, which can help them to effectively capture global features, relationships, and context throughout the image. This capability makes them particularly well-suited for tasks requiring analysis of complex spatial relationships, potentially offering advantages in discriminating between slight facial cues associated with ASD.

Because ViT models can work with different types of data and tasks, they might be able to identify autism even when it looks different in different individuals. This is important because autism can show up in many different ways.

The selection of ViT models over alternative deep learning architectures for ASD classification is very important. ViT models work on image patches for classification tasks. By this, they can capture both local and global features. In most of the other conventional deep learning methods, the focus is mostly on the local features only. Therefore, ViT models provide enhanced accuracy and reliability in ASD diagnosis.

It is to be noted that using facial recognition for medical diagnosis can be tricky. We need to be careful about biasness in the technology, protecting people's privacy, and avoiding discrimination based on how they look. In healthcare, especially, it is important to be open about how these tools work, make sure they are fair, and develop them responsibly.

We have found some promising ways to spot autism by looking at faces, but there is still more to learn. Subtle clues and changes in faces have not been studied much, and the methods we use now might not work for everyone. This study addresses these gaps by (i) utilizing the powerful capabilities of ViT models to capture both local and global features in facial images; (ii) employing a diverse and properly annotated dataset to limit potential biases and improve generalization; and (iii) rigorously validating the model's performance to ensure its accuracy and reliability for real-world applications.

By effectively addressing these research gaps, this study has the potential to contribute significantly to the development of a robust and accurate tool for early ASD detection using facial image analysis.

METHODOLOGY

Data preprocessing

The dataset consists of facial images of children diagnosed with ASD and non-autistic children. Each image, denoted as X_i, is associated with a binary label y_i indicating the class and the class can be autistic or non-autistic. The data are resized to the same size for uniform processing. The facial images undergo a few preprocessing steps, so that the applied model can perform effectively. Let N represent the total number of images, X_i denote the ith image, and y_i denote its corresponding label.

Normalization, resizing, and augmentation are sequentially applied to each image as follows:

$$X_i = \mathcal{N}(\mathcal{R}(\mathcal{A}(X_i))),$$

where X_i represents the ith image, A denotes augmentation, R represents resizing, and N denotes normalization. This research employs rotation and flipping techniques to augment the size of the training data. The mathematical representation of rotation can be expressed as follows:

$$Img_{\{rot\}} = Img(scos(\gamma) - tsin(\gamma), ssin(\gamma) + tcos(\gamma)), \qquad (1)$$

where $Img(s,t)$ represents the original image, $Img_{\{rot\}}$ represents the rotated image, and γdenotes the angle of rotation.

The mathematical representation of vertical flipping can be represented as follows:

$$Img_{Vflipped}(s,t) = Img(s, H - t) \qquad (2)$$

and the mathematical representation of horizontal flipping can be given as follows:

$$Img_{Hflipped}(s,t) = Img(W - s, t). \qquad (3)$$

Feature extraction

The ViT model architecture is used for feature extraction for children's facial images. The feature extraction module is represented in Figure 1. Facial images are taken as the input, divided into fixed-size patches, linearly embedded, and then processed through a series of transformer blocks. Let X_{patch} represent the linear embedding of patches and $X_{\text{transformer}}$ denote the final output after transformer blocks:

$$X_{\text{patch}} = \text{LinearEmbedding}(X_i)$$

$$X_{\text{transformer}} = \text{TransformerBlocks}(X_{\text{patch}})$$

Classification

After processing information in the transformer blocks, the model uses a technique called "global average pooling" to summarize the most important features into a single, consistent representation. The pooled representation X_{pooled} is then passed through a fully connected linear layer. This layer has softmax as the activation function for binary classification.

$$X_{\text{pooled}} = \text{GlobalAvgPooling}(X_{\text{transformer}})$$

$$y_{\text{pred}} = \text{Softmax}(\text{LinearClassifier}(X_{\text{pooled}}))$$

Here, y_{pred} is the predicted probability distribution over the two classes (autistic and non-autistic) obtained through the softmax function.

Loss function

The model is trained using cross-entropy loss, denoted as L_{CE}, which measures the dissimilarity between the predicted and true class distributions:

$$L_{\text{CE}} = -\sum_{i=1}^{N} y_i \cdot \log(y_{\text{pred,i}})$$

The entire model, including the ViT architecture and the classification head, is trained end-to-end using backpropagation

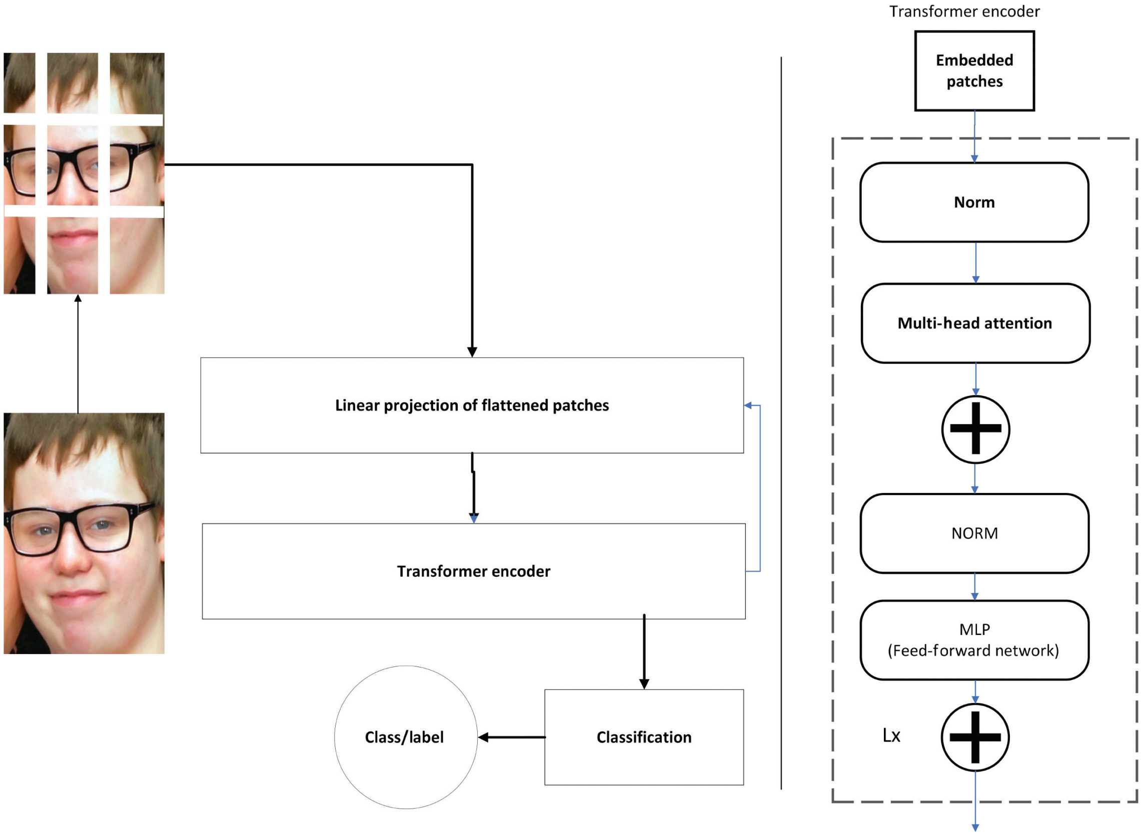


Figure 1: Feature extraction module.

and optimization techniques such as stochastic gradient descent or Adam. The model is fine-tuned on the training dataset to minimize the cross-entropy loss.

RESULTS AND ANALYSIS

Evaluation

The trained ViT model is evaluated on a test dataset. It is a binary classification problem; therefore, metrics such as accuracy, precision, recall, and F1-score are used to assess the performance of the proposed methodology.

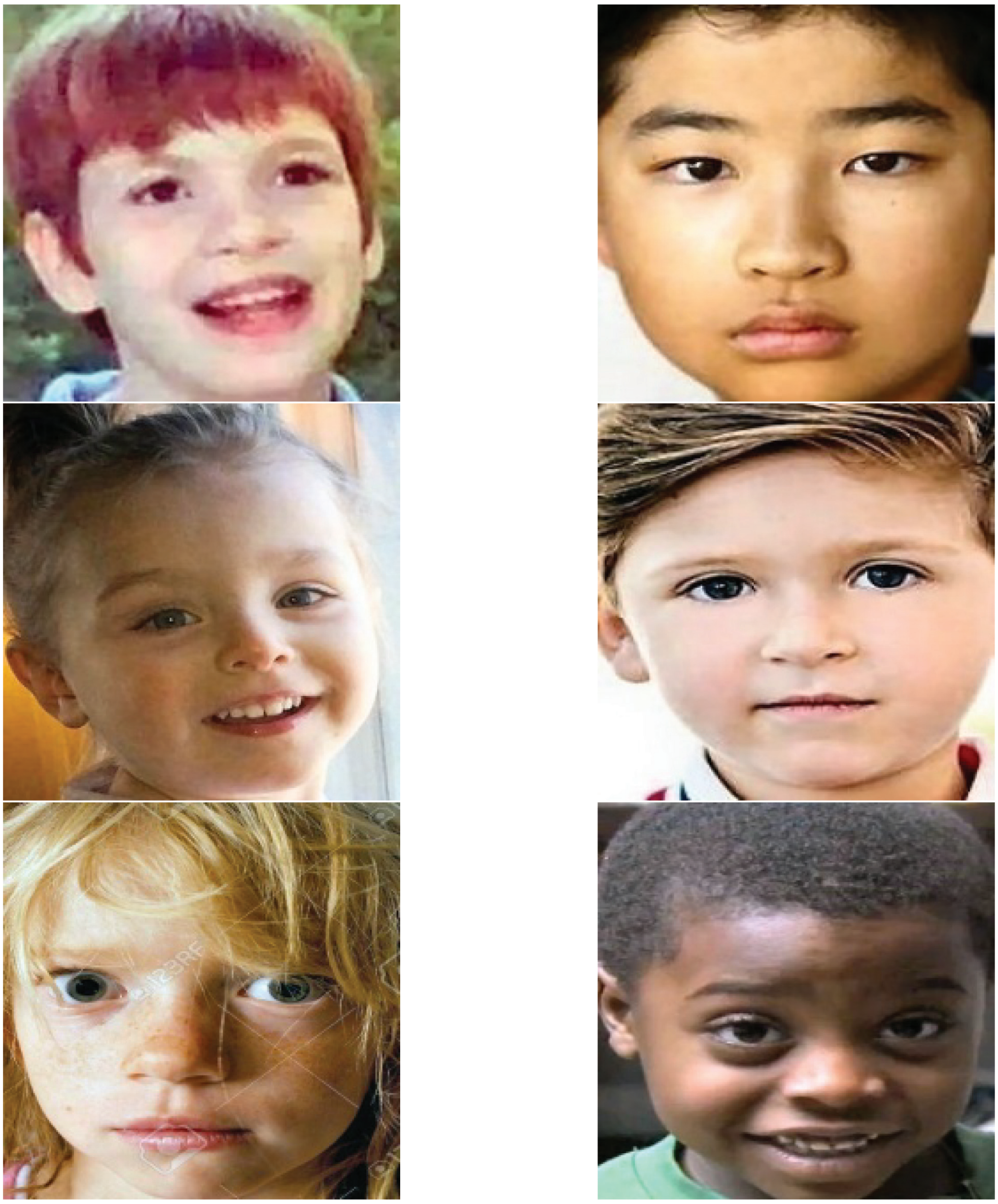
The hyperparameters, including learning rate, batch size, and the number of transformer blocks, are optimized to achieve the optimal solution for the proposed method.

Dataset

In this study, the dataset titled "Facial Image Data Set for Children with Autism" has been employed for experimentation purposes (Kaggle, n.d.). This dataset poses a binary classification problem. The dataset comprises facial images of both autistic and non-autistic children, consisting of 1327 images for each class, i.e. faces of autistic children and non-autistic children. This ensures a balanced representation in the training set. A few images of this dataset are shown in Figure 2. T-SNE visualization of the dataset is shown in Figure 3.

Experimental setup

We ran the experiments on a powerful computer system with GPU acceleration to speed up training the models. We used



(a) Autistic Faces

(b) Non-Autistic Faces

Figure 2: Kaggle dataset image samples: (a) autistic faces and (b) non-autistic faces.

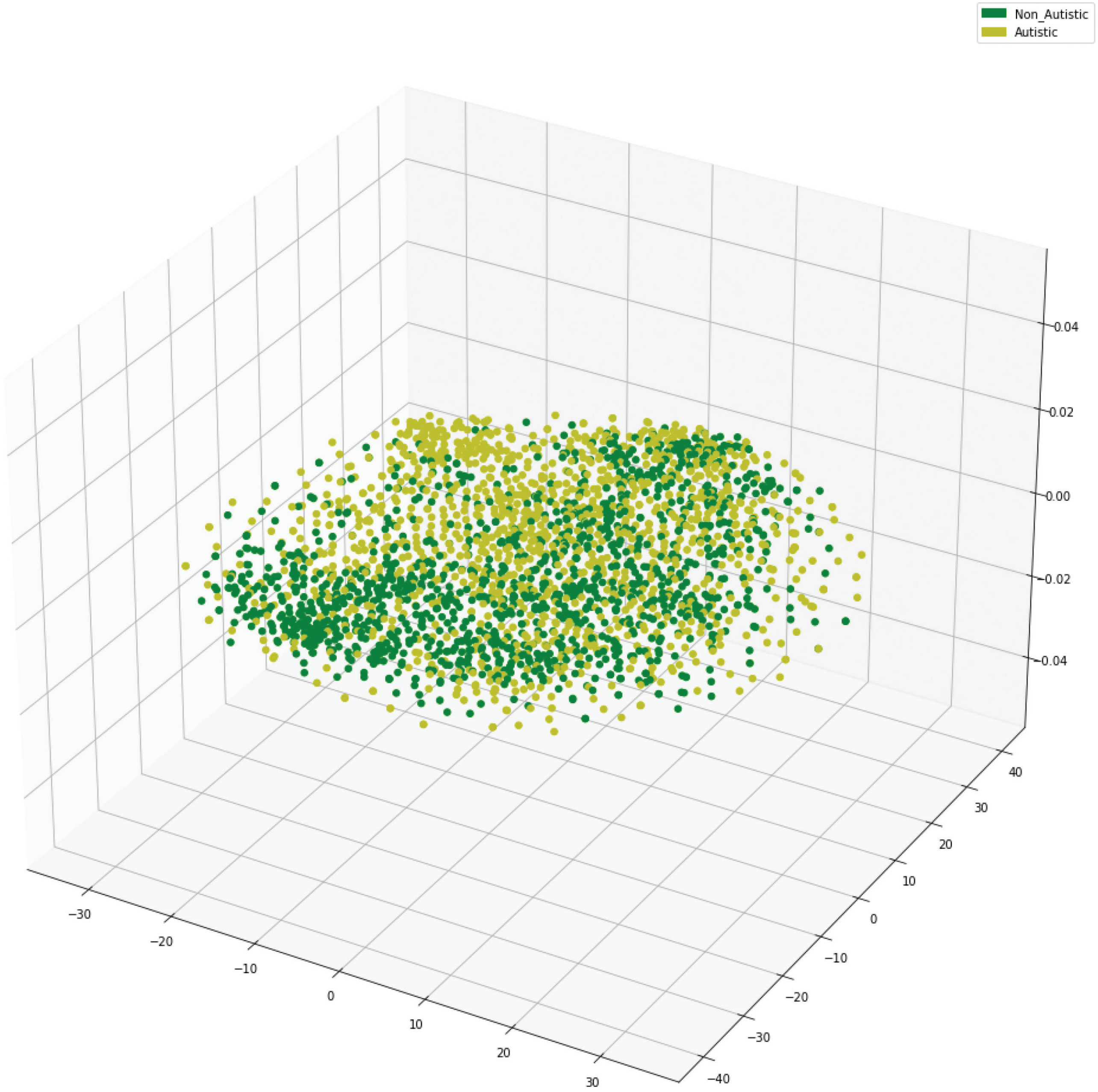


Figure 3: T-SNE visualization of the dataset.

software tools like TensorFlow and Keras, along with libraries for working with images and evaluating the results.

To guarantee the reliability and robustness of the trained model, we implemented a rigorous validation approach. This strategy partitions the dataset into distinct training, validation, and testing subsets in the ratio of 0.8:0.1:0.1. Hyperparameter tuning and model selection are guided by the performance on the validation set, while the real-world performance is ultimately assessed on the unseen testing set.

Experimental results

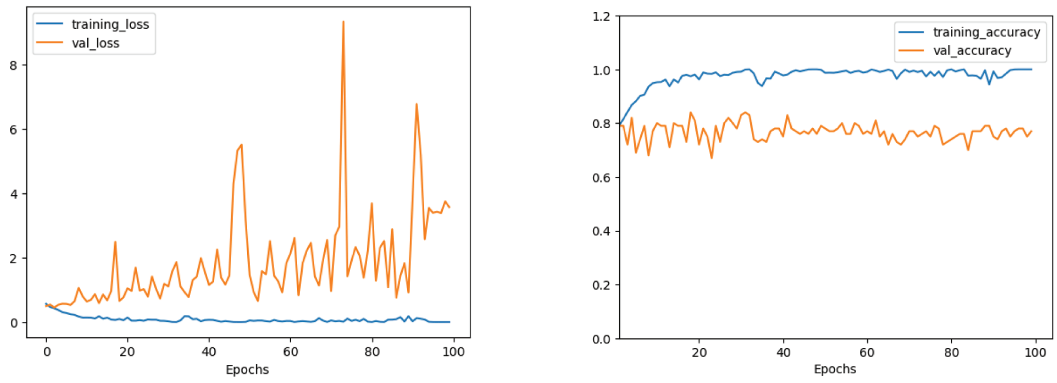
This section explains different experimental results on the specified dataset. Table 1 presents a comparative analysis of the experimental results obtained from two different models applied to the task of autistic face classification. The models evaluated in this comparison are ViT and VGG-16.

Table 1: Comparison of experimental results.

Model	Validation accuracy	Validation loss
Proposed	0.7700	0.5646
VGG-16 (Simonyan and Zisserman, 2014)	0.7292	0.5309

In terms of validation accuracy, the ViT model achieved an accuracy of 0.7700, while the VGG-16 model attained a slightly lower accuracy of 0.7292. Regarding validation loss, the ViT model showed a loss value of 0.5646, whereas the VGG-16 model showed a slightly lower loss value of 0.5309.

More evaluation metrics are also tested for the analysis of the proposed method. The proposed model demonstrated



(a) Loss curve

(b) Accuracy curve

Figure 4: Loss and accuracy curves: (a) loss curve and (b) accuracy curve.

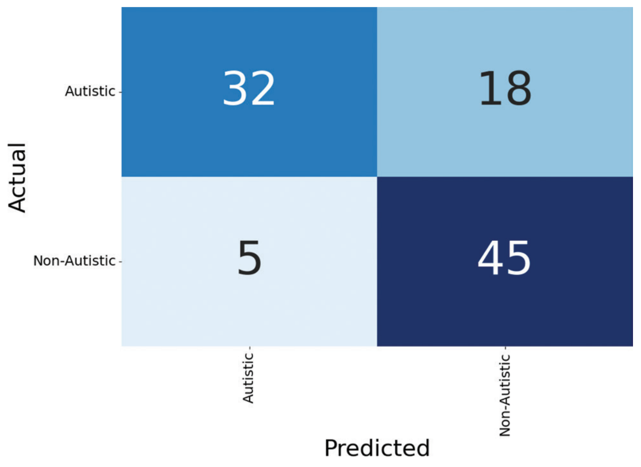


Figure 5: Confusion matrix.

remarkable performance; it achieved a recall of 0.7700 and a precision of 0.7700, presenting its ability to effectively identify true-positive cases while minimizing false positives. These metrics are crucial in the context of autism diagnosis, where accurate classification of individuals is critical.

While the ViT model achieved respectable recall and precision scores, its overall accuracy remained relatively moderate. The reason is the complexity of the presented problem. It is very difficult to classify the disease just based on facial cues. But this can be used as an early indicator of ASD diagnosis. Moreover, the differences and diversity among people with ASD make it even harder to achieve accurate classification results.

The loss curve in Figure 4 illustrates the trends in loss and accuracy during the training and validation processes. These curves are used to analyze and observe the performance of applied machine learning models.

Figure 5 shows the confusion matrix for the proposed method. It presents true positives, false positives, true negatives, and false positives in a tabular form. This can be helpful in the calculation of recall, precision, and accuracy.

The box plot in Figure 6 shows the visual representation of the distribution of performance metrics such as accuracy,

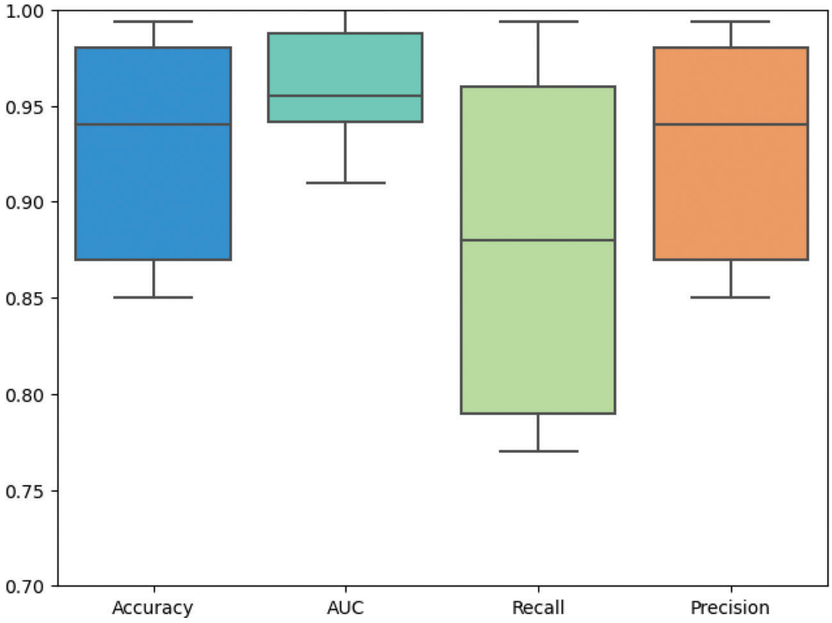


Figure 6: Box plot of data distribution. Abbreviation: AUC, area under the curve.

area under the curve, recall, and precision. These show the performance during the training process.

The use of facial recognition technology in medical diagnosis, particularly for autism detection using facial images, raises important ethical considerations that need careful examination. Although this technology aids in the early identification of ASD, it also presents potential risks and challenges. Few concerns regarding privacy, data security, and consent arise, as the collection and analysis of facial data may interfere with privacy rights. Moreover, there is also a need to address issues of biases in training or algorithm. As such, it is crucial for researchers and practitioners in the field to engage in communication and converse about the ethical implications and concerns of using facial recognition technology in ASD diagnosis and to try hard to limit potential risks while keeping maximum benefits for individuals and the society.

CONCLUSION AND FUTURE PROSPECTS

This research work explains the innovative application of ViT models for the classification of facial images, classifying autistic and non-autistic children. By employing computer vision and deep learning techniques, this study reveals that ViT models can pick up on subtle facial expressions and features that might be linked to ASD, potentially opening doors for earlier diagnosis. The findings show encouraging progress in recognizing unique characteristics associated with ASD, leading to improved identification and moving toward early diagnosis.

This research can provide several practical applications and potential clinical significance. Primarily, the utilization of ViT models enables the early detection of ASD in children through the analysis of facial images. By employing deep learning techniques, this approach supports healthcare professionals in identifying potential indicators of ASD, which could otherwise go missing. The practical application extends to aiding in the timely intervention and support for children with ASD, as early diagnosis facilitates access to person-specific treatments and therapies. From a clinical perspective, this research offers a noninvasive and effective tool for ASD assessment. Furthermore, the analysis of facial features using ViT models contributes to a deeper understanding of ASD.

This research can be enhanced in many directions. Incorporating multimodal data sources, such as textual or behavioral information, alongside facial images, could enhance the robustness and accuracy of ASD classification models. To make these ViT models even more helpful, we need to see if they work just as well for people from different backgrounds and cultures. This would allow them to be used more widely and help more people. Overall, continued interdisciplinary collaboration and integration of cutting-edge methodologies are essential for advancing the field of early ASD detection and intervention.

FUNDING

The authors extend their appreciation to the King Salman Center for Disability Research for funding this work through Research Group no. KSRG-2023-085.

ACKNOWLEDGEMENTS

The authors extend their appreciation to the King Salman center for Disability Research for funding this work through Research Group no. KSRG-2023-085.

REFERENCES

Aleem A., Tehsin S., Kausar S. and Jameel A. (2022). Target classification of marine debris using deep learning. *Intell. Autom. Soft Comput.*, 32(1), 73-85.

Alharthi A.G. and Alzahrani S.M. (2023). Do it the transformer way: a comprehensive review of brain and vision transformers for autism spectrum disorder diagnosis and classification. *Comput. Biol. Med.*, 167, 107667.

Almufareh M.F., Tariq N., Humayun M. and Almas B. (2023a). A federated learning approach to breast cancer prediction in a collaborative learning framework. *Healthcare*, 11(24), 3185.

Almufareh M.F., Tehsin S., Humayun M. and Kausar S. (2023b). Intellectual disability and technology: an artificial intelligence perspective and framework. *J. Disabil. Res.*, 2(4), 58-70.

Almufareh M.F., Tehsin S., Humayun M. and Kausar S. (2023c). Artificial cognition for detection of mental disability: a vision transformer approach for Alzheimer's disease. *Healthcare*, 11(20), 2763.

Almufareh M.F., Tehsin S., Humayun M. and Kausar S. (2023d). A transfer learning approach for clinical detection support of monkeypox skin lesions. *Diagnostics*, 13(8), 1503.

Almufareh M.F., Imran M., Khan A., Humayun M. and Asim M. (2024a). Automated brain tumor segmentation and classification in MRI using YOLO-based Deep Learning. *IEEE Access*, 12, 16189-16207.

Almufareh M.F., Kausar S., Humayun M. and Tehsin S. (2024b). A conceptual model for inclusive technology: advancing disability inclusion through artificial intelligence. *J. Disabil. Res.*, 3(1), 20230060.

Arkin E., Yadikar N., Xu X., Aysa A. and Ubul K. (2023). A survey: object detection methods from CNN to transformer. *Multimed. Tools Appl.*, 82(14), 21353-21383.

Cao X., Ye W., Sizikova E., Bai X., Coffee M., Zeng H., et al. (2023). Vitasd: robust vision transformer baselines for autism spectrum disorder facial diagnosis. In: *ICASSP 2023-2023 IEEE International Conference on Acoustics*, *Speech and Signal Processing (ICASSP)*, IEEE, pp. 1-5.

Celard P., Iglesias E.L., Sorribes-Fdez J.M., Romero R., Vieira A.S. and Borrajo L. (2023). A survey on deep learning applied to medical images: from simple artificial neural networks to generative models. *Neural Comput. Appl.*, 35(3), 2291-2323.

Clements C.C., Ascunce K. and Nelson C.A. (2023). In context: a developmental model of reward processing, with implications for autism and sensitive periods. *J. Am. Acad. Child Adolesc. Psychiatry*, 62(11), 1200-1216.

De Rham A. and Marco E.J. (2016). Genetics and autism spectrum disorder. *Dev. Neurobiol.*, 76(5), 633-652.

Deng A., Yang T., Chen C., Chen Q., Neely L. and Oyama S. (2024). Language-assisted deep learning for autistic behaviors recognition. *Smart Health*, 32, 100444.

Dhar T., Dey N., Borra S. and Sherratt R.S. (2023). Challenges of deep learning in medical image analysis—improving explainability and trust. *IEEE Trans. Technol. Soc.*, 4(1), 68-75.

Golarai G., Grill-Spector K. and Reiss A.L. (2006). Autism and the development of face processing. *Clin. Neurosci. Res.*, 6(3-4), 145-160.

Griffin J.W., Azu M.A., Cramer-Benjamin S., Franke C.J., Herman N., Iqbal R., et al. (2023). Investigating the face inversion effect in autism across behavioral and neural measures of face processing: a systematic review and Bayesian meta-analysis. *JAMA Psychiatry*, 80, 1026-1036.

Hannon B., Mandy W. and Hull L. (2023). A comparison of methods for measuring camouflaging in autism. *Autism Res.*, 16(1), 12-29.

Hartston M., Avidan G., Pertzov Y. and Hadad B.S. (2023). Weaker face recognition in adults with autism arises from perceptually based alterations. *Autism Res.*, 16(4), 723-733.

Jeyarani R.A. and Senthilkumar R. (2023). Eye tracking biomarkers for autism spectrum disorder detection using machine learning and deep learning techniques. *Res. Autism Spectr. Disord.*, 108, 102228.

Kaggle. (n.d.). Autism_Image_Data. https://www.kaggle.com/datasets/cihan063/autism-image-data.

Li X., Li M., Yan P., Li G., Jiang Y., Luo H., et al. (2023). Deep learning attention mechanism in medical image analysis: basics and beyonds. *Int. J. Netw. Dyn. Intell.*, 2, 93-116.

Mehran A., Tehsin S. and Hamza M. (2023). An effective deep learning model for ship detection from satellite images. *Spat. Inf. Res.*, 31(1), 61-72.

Oukdach Y., Kerkaou Z., El Ansari M., Koutti L., Fouad El Ouafdi A. and De Lange T. (2024). ViTCA-Net: a framework for disease detection in video capsule endoscopy images using a vision transformer and convolutional neural network with a specific attention mechanism. *Multimed. Tools Appl.*, 67, 1-20.

Pacal I. (2024). MaxCerVixT: a novel lightweight vision transformer-based approach for precise cervical cancer detection. *Knowl.-Based Syst.*, 289, 111482.

Sahu M., Dash R., Mishra S.K., Humayun M., Alfayad M. and Assiri M. (2024). A deep transfer learning model for green environment security analysis in smart city. *J. King Saud Univ.-Comput. Inf. Sci.*, 36(1), 101921.

Simonyan K. and Zisserman A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Talaat F.M. (2023). Real-time facial emotion recognition system among children with autism based on deep learning and IoT. *Neural Comput. Appl.*, 35(17), 12717-12728.

Zhang J., Feng F., Han T., Gong X. and Duan F. (2023). Detection of autism spectrum disorder using fMRI functional connectivity with feature selection and deep learning. *Cogn. Comput.*, 15(4), 1106-1117.