scientific reports

Check for updates

OPEN

Leveraging artificial intelligence for diagnosis of children autism through facial expressions

Mahmood A. Mahmood[1,2], Leila Jamel[3]✉, Nazik Alturki[3] & Medhat A. Tawfeek[4,5]

The global population contains a substantial number of individuals who experience autism spectrum disorder, thus requiring immediate identification to enable successful intervention approaches. The authors assess the detection of autism-related learning difficulties in children by evaluating deep learning models that use transfer learning methods along with fine-tuning methods. Using autism spectrum disorder (ASD) diagnosed child RGB images data, researchers evaluated six prevalent deep learning structures: DenseNet201, ResNet152, VGG16, VGG19, MobileNetV2, and EfficientNet-B0. ResNet152 reached the highest accuracy rate of 89% when functioning independently. This paper develops a hybrid deep-learning model by integrating ResNet152 with Vision Transformers (ViT) to achieve better classification performance. The ViT-ResNet152 model's convolutional and transformer processing elements worked together to improve the accuracy of the diagnosis to 91.33% and make it better at finding different cases of autism spectrum disorder (ASD).The research outcomes demonstrate that AI tools show promise for delivering highly precise and standardized methods to detect ASD at an early stage. Future research needs to include multiple data types as well as extend dataset variability while optimizing hybrid architecture systems to elevate diagnostic forecasting. The incorporation of artificial intelligence in ASD evaluation services holds promise to transform early therapy approaches, which leads to better results for autistic children all around the globe.

Keywords Deep learning, Autism, Autism in children images, Artificial intelligent, Autism identification

ASD is a developmental disorder with the primary defining features being difficulty in social interaction, verbal and nonverbal communication, and limited interests and repetitive behaviors. This disorder occurs in about every 54 children around the globe, and thus early detection and early intervention can help minimize the effects[1,2]. Standard diagnostic methods that are partly used for diagnosing autism entail behavioral methods that take a considerable amount of time and are moralistic in that they are often open to bias. There is therefore increasing interest in using deep learning from AI to flag early learning difficulties linked to autism more objectively and at a larger scale[3]. The detection of ASD through facial expression analysis represents a promising method because it offers non-obtrusive testing along with easy implementation capabilities and potential for identifying ASD early. Biomedical assessment methods, including behavioral assessments and speech analysis, involve expensive expert evaluation together with subjective analysis, while facial analysis through deep learning delivers both objective measurements and scalable solutions. Behavioral tracking techniques show merit in ASD diagnosis but need prolonged observation times, which makes them difficult to access and time-consuming. Language barriers, together with the individual variations in speech development among children with ASD, limit the effectiveness of speech analysis. The analysis of facial elements demonstrates uniformity among all demographics and permits exact measurement through automated AI processing. Deep learning models have proven effective at detecting facial characteristics in ASD children because studies confirm these patients show unique facial patterns with abnormal eye movements and variations in facial shape along with muscle control. Through facial image analysis, this research develops a method to detect autism spectrum disorder early using reliable technology that supports current diagnostic practices without requiring human opinions.

[1]Department of Information Systems, College of Computer and Information Sciences, Jouf University, 72341 Sakaka, Aljouf, Kingdom of Saudi Arabia. [2]Department of Information Systems and Technology, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt. [3]Department of Information Systems, College of Computer and Information Science, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, 11671 Riyadh, Kingdom of Saudi Arabia. [4]Department of Computer Science, College of Computer and Information Sciences, Jouf University, 72341 Sakaka, Aljouf, Kingdom of Saudi Arabia. [5]Department of Computer Science, Faculty of Computers and Information, Menoufia University, 32511 Shebin Elkom, Egypt. ✉email: lmjamel@pnu.edu.sa

The progress made in the field of AI and its subspecialty of machine learning has been instrumental in practicing improvements in areas such as neurodevelopmental disorders that include autism spectrum disorder. It found that deep learning, as a branch of machine learning, has significant applications of analyzing intricate data structures such as images and behavioral data for diagnosis and classification of various health conditions[4]. Of particular importance, CNNs have been applied in the study of medical images with remarkable accuracy in tasks such as the detection of tumors and brain imaging analysis[5,6]. Such successes have created enthusiasm in the use of deep learning approaches in the diagnosis of ASD with the hopes of identifying pointers of learning impairments related to autism at a very tender age[7].

There are numerous discussions addressing the implementation of deep learning in the situation of ASD. For instance, Heinsfeld et al.[7] employed a deep learning model of the identification of ASD using rs-fMRI data with a success rate of 70%. In the same vein, Duda et al.[8] trained the machine learning algorithms to extract the ASD-related behaviors from electronic health records, mapping the applicability of the AI for early diagnosis. Strong evidence from such studies points towards how AI can help to process and analyze intricate data that may be difficult for experts in the ASD field to process and analyze, hence leading to more accurate diagnosis[9,10].

Deep learning research focused on ASD detection techniques demonstrates strong findings, yet practical implementation remains challenging because of identified technical boundaries. CNN-based models show strong results for medical image diagnosis, including ASD screening, yet they face difficulties in processing extended dependencies within facial information. The ability to learn global representations is a transformer-based model's strength, while it fails to capture CNN-like spatial feature extraction abilities. Most previous research studies examined single-model architecture that did not allow complete exploitation of cross-model advantages. The small number of relatively imbalanced ASD detection datasets poses problems for deep learning models to reach reliable generalization.

This paper evaluates the identification of autism-related learning difficulties in children using transfer learning and fine-tuning on six deep learning models, which include DenseNet201, ResNet152, VGG16, EfficientNet-B0, MobileNetV2, and VGG19. The evaluation utilized selected deep-learning models because they demonstrated strong results in image classification duties and showed potential when analyzing ASD-related behavioral data and image information. The dense connections in DenseNet201 advance both information exchange and gradient update, and EfficientNet-B0 delivers optimized precision and computational performance. Analyzing these models concerning ASD diagnosis can identify the most suitable artificial intelligence tools for clinical use. Previous research examined individual neural network models or conducted minimal architectural identity evaluations but failed to conduct thorough evaluations regarding multiple deep learning structures for ASD diagnosis.

The researchers chose deep learning models for this study because they showed proven success in image classification while demonstrating potential for ASD diagnostic work. CNNs such as VGG16 and VGG19 with DenseNet201, along with EfficientNet-B0 as well as MobileNetV2, find extensive medical imaging usage by handling hierarchical spatial feature extraction. DenseNet201 solves both feature gradient problems and computational efficiency challenges, but EfficientNet-B0 finds the optimum accuracy-efficiency trade-off. MobileNetV2 functions best with lightweight applications, thus making it appropriate for real-time ASD screening. Vision Transformer architecture enables self-attention operations to detect subtle features linked to ASD because it evaluates lengthy image relationships effectively. The deep residual connections in ResNet152 improve the flow of gradients to stop vanishing gradients from impacting deep networks. The proposed model unites the ViT architecture with ResNet152 to exploit their complementary strengths, which allows global feature extraction from ViT and hierarchical spatial representation from ResNet152 for better classification outcomes and diagnostic generalization. The integrated approach solves current ASD detection model challenges since these models depend solely on CNNs or transformers without utilizing their separate strengths completely.

The application of AI diagnosis in ASD stands at a critical juncture because of substantial hurdles when it comes to identifying early signs through facial expressions. Current diagnosis approaches depend on behavioral observations from humans that require extended periods and cost a lot of money while demonstrating human judgment issues. Deep learning successfully analyzes medical images, yet its use for detecting autism spectrum disorder through facial recognition faces several challenges because ASD facial indicators are difficult to detect and available data remains limited and imbalanced. The current research landscape features primarily CNN-based structures or transformer-based models yet fails to combine spatial characteristics extraction with contextual information learning capabilities. The proposed hybrid ResNet152-ViT model emerges as a solution to overcome the challenges presented by diagnosing autism spectrum disorder from facial images because it improves both classification performance and model stability while maximizing generalization potential.

This paper develops a hybrid deep learning framework that merges Vision Transformer (ViT) with the ResNet152 architecture to boost classification precision and model generalizability. The proposed combination of ViT transformer self-attention mechanisms with ResNet152 hierarchical features achieves higher performance in autism learning difficulty detection between children's groups. The combined deep learning system achieves improved diagnostic precision as well as generalizing its performance on a wide range of ASD cases. The hybrid ResNet152 with Vision Transformer (ViT) system aims to produce better ASD diagnosis results by using the advantages of each architecture from facial image analysis. ResNet152 demonstrates outstanding performance in extracting multi-level spatial features to detect subtle facial cues that indicate ASD manifestations in images. When deployed in certain situations, CNNs demonstrate restrictions in their ability to detect both long-range dependencies and global contextual information. The self-attention processes of ViT identify relationships between elements across complete images because they effectively discover structural patterns that single CNN networks cannot detect. Combining ResNet152 features with ViT features leads to an improved feature representation, which both prevents model overfitting and adds to system accuracy levels. A synergistic approach

analyzes both facial image texture details and broader contextual relationships, which results in a resilient ASD classification model.

The main contribution of this research seeks to develop healthcare AI through comprehensive deep learning model effectiveness evaluation, including the investigation of recommended hybrid architecture for identifying ASD-related learning difficulties at early stages. This research produces results to help scientists create more efficient diagnostic methods that allow timely detection of ASD while maintaining patient safety and comfort to enable better early intervention and enhance ASD patient quality of life.

The paper is structured in the following manner. A review of the literature and information pertinent to the investigation are found in Section "Related work". The materials and methods of deep neural network details are provided in Section "Materials and methods". The experimental results are covered in Section "Experimental results". A discussion of results is presented in Section "Discussion". A summary of the conclusions and suggestions for additional research are included in the conclusion of Section "Conclusion".

Related work

Diagnosis of autism spectrum disorder with the aid of deep learning has turned out to be effective in recent times. Heinsfeld et al. (2018) in[7] showed that using deep learning algorithms, a classification accuracy of 70% is possible with the help of resting state functional MRI (rs-fMRI) data for differentiating between ASD and individuals with ASD. Many approaches of using the application of AI in neurodevelopmental disorder diagnosis proved that deep learning can efficiently process large data of images data. Besides images, machine learning has also been used more and more often with behavioral data traceable with ASD. In a similar vein, Duda et al. (2016) in[8] mined EHR for behaviors that can be used in identifying patients with ASD through machine learning. This approach thus highlighted how AI could improve early diagnostic capability through the analysis of extensive behavioral data that can be cumbersome for a clinician to perform manually. Based on such preparation, Voulodimos et al. (2018) in[11] compared different sorts of CNNs for medical image analysis. Moreover, they projected DenseNet, ResNet, and VGG having high accord in tasks that include medical imaging. They were instrumental in bringing strong methodology in using such architectures for diagnosing ASD, particularly when working with image data. Newer and sustainable developments in deep learning are new models like EfficientNet-B0 that are accurate and efficient at the same time. In[12], Tan and Le introduced EfficientNet-B0, which makes better use not just of depth and width at the same time–previous architectures have scaled width and depth inefficiently. This is quite innovative and has broad implications for medical diagnostics, at least where identification of learning challenges linked to ASD is concerned.

In the same way, Howard et al. (2019) in[13] came up with MobileNetV2, a model designed to improve its efficiency in the mobile and edge stations. MobileNetV2 is a lightweight deep learning framework that can be used in real-time diagnostic applications and therefore has the potential of facilitating the early diagnosis of ASD. As with other tasks of medical image analysis, the model can be useful for ASD diagnosis too. Hence, there's no doubt that undertaking a diagnosis for ASD at the right time and with precision is vital. As mentioned by Wallace et al. (2020) in[14], early identification and early intervention positively impact the quality of the lives of children with ASD. In their study of the intervention implemented during the early developmental stage, they underscored the importance of valid diagnostic instruments that can diagnose ASD at an early age, hence appropriate treatment. Chen et al. (2017) in[15] employed deep learning systems in identifying measures of connectivity in the brain of people with ASD. They used a CNN model to analyze the fMRI-derived functional connectivity matrices with a high level of accuracy for classifying individuals with ASD from TD. This work supports the notion of appreciating the depth of learning in trying to demystify the neural circuitry of ASD. Lombardo et al. (2019) in[16] did not attempt to derive a new model for the diagnostic criterion but used genetics and images to enhance the diagnostic validity of the ASD diagnosis. They employed a multimodal deep learning framework that incorporates genetic information of the patient along with brain image information, and the authors showed that the utilization of such methodology could improve the diagnostic accuracy. This work serves as an example of how more than one kind of data can be used in the ASD diagnostic process. However, there are certain limitations associated with the incorporation of AI in the diagnosis of ASD. The application of AI in healthcare regarding the use of robots in neurodevelopmental disorders, such as ASD, was described by Abdelnour et al.[17] with a focus on the ethically informed aspect of AI use. They advocate for transparency in the training of the AI models, interpretability, and fairness to ensure that the tools that are clinical are used appropriately.

Li et al.[18] also looked at the nature of AI in managing individual therapy for persons with ASD. Their research was applied to the aspect of utilizing machine learning algorithms with the aim of estimating the part-specific reaction to several intercessions in order to design and implement specific treatment plans. Such an approach might help improve the treatment outcomes and provide a better matching of the patient needs in cases of ASD. The use of deep learning in diagnosing and differentiating other ailments also enhances its usage in ASD. Litjens et al.[19] summarized deep learning applied in radiology, where the authors showed deep learning to be highly accurate for tasks such as nodule detection and organ segmentation. Such successes imply that the same strategies can be borrowed in the handling of image data in the diagnosis of ASD. Another related work by Zhao et al.[20] used eye-tracking data and analyzes them using a deep learning approach to extract features that are linked to the disorder. By using this machine learning model to decode looking at time and direction for each stimulus, they were able to classify ASD and TD individuals and thus proved that AI has the capability to analyze non-conventional forms of data for diagnosing ASD. Moreover, in diagnosing ASD, the use of images and behavioral data reports, as well as an analysis of speech, has also been done. Deep learning was employed by Palkovics et al.[21] in identifying speech patterns in children with ASD and found a very high correct classification rate between the children with ASD and those who are typically developing. Based on this research, deep learning can be considered more generally across different areas linked to the diagnosis of ASD. This has been

supported by the work of Shen et al.[22], whereby they used a CNN model to detect Alzheimer's disease from MRI scans. Their studies evidenced the specificity of the model to identify various phases of Alzheimer's, and a similar approach could be used in identifying the signs of ASD at the early beginning. Eslami et al.[23] have acknowledged the enhancement of AI with conventional diagnosis approaches. They suggested an integration of deep learning with the work of experts to improve the accuracy of ASD diagnosis. This approach stresses the synergy between AI and professionals work to get the best results in clinical practice.

Hayder and Amir[24] investigated the potential effectiveness of combining Vision Transformers with Squeeze-and-Excitation blocks for early autism spectrum disorder diagnosis through facial image analysis. Static facial characteristics serve as biomarkers that help the proposed model detect delicate ASD-related signals more effectively. The evaluated data demonstrates deep learning has great potential for ASD screening because it correctly identifies autistic patients alongside typically developing children.

The authors of Karthik et al.[25] underline that discovering autism spectrum disorder (ASD) during early phases is necessary for effective intervention application. The research examines ASD recognition in facial pictures using three model combinations that incorporate Vision Transformer (ViT) with PCA-SVM, ViT with CatBoost, and SHAP along with VGG16 with XGBoost. The developed models enhance picture processing through error reduction and better detection of subtle facial features linked to ASD. The experimental findings show deep learning methods produce superior outcomes compared to conventional models, which strengthens their importance in early ASD diagnosis. Mujeeb and Subashini[26] state that neurological autism spectrum disorder disrupts human social interactions and cognitive processing in individuals. VGG16, VGG19, and EfficientNetB0 deep learning models identified autistic faces from non-autistic faces for research that resulted in accuracies of 84.66%, 80.05%, and 87.9% successively. The models received training from 3,014 facial images available on Kaggle, which showed both autistic and non-autistic childhood emotions. Pranavi and Andrew[27] emphasize that spotting autism spectrum disorder (ASD) early leads to enhanced quality of life for children with ASD together with their families. The research team conducted ASD biomarker detection through facial image analysis using MobileNet along with three other pre-trained CNN models (Xception, EfficientNetB0, EfficientNetB1) and a DNN classifier structure. The tests determined Xception to be the most effective model, which detected ASD with 96.63% AUC and 88.46% sensitivity.

Materials and methods

This section provides an analysis of the use of deep learning algorithms used in this work to diagnose learning difficulties associated with autism in children. We utilized six state-of-the-art convolutional neural network (CNN) architectures: DenseNet201, ResNet152, VGG16, EfficientNet-B0, MobileNetV2, VGG19, and the hybrid proposed model. All these models were trained and tested on a dataset involving images and behavior data to gauge their ability to identify and distinguish students having ASD-related learning difficulties.

DenseNet201

DenseNet201 is a complex CNN where every layer receives inputs from the previous layer as well as from all the previous layers existing in the network. This means there are many more shared parameters through the network, meaning issues such as the vanishing gradient problem are minimized. The DenseNet201 model is made up of several dense blocks, where each block is made of several convolution layers, as shown in Fig. 1. These blocks relate to transition layers that are the same as down sampling sections that create pool layers to let the network decrease the spatial size of the feature maps while preserving important data. One of the peculiarities of DenseNet201 is that the growth rate is another parameter that determines the number of new feature maps introduced by each layer. This growth rate assumes significant importance in preventing the network capacity and the computational capability from getting to extremes. When there is a complex pattern in the data and/or there is not much data that can be used for training, then DenseNet201 is very useful because deep supervision is provided and fine-grained feature connections are ensured through several layers[28–30].

ResNet152

ResNet152 is a deep convolutional neural network (CNN) that falls into the family of Residual Networks (ResNet) architectures that revolutionized deep learning by introducing residual learning. One of the major ideas of ResNet is so-called shortcut connections, or skips, that let the model omit one or several layers as shown in Fig. 2. These connections enable the network to learn identity mappings and hence provide a solution to the degradation issue, which results from adding more layers to a deep neural network's training by enhancing the training error[31]. ResNet152 has more layers compared to other counterparts in ResNet, with an outstanding 152 layers for identification of sophisticated features in large databases. The architecture utilizes convolutional layers that are immediately followed by batch normalization, ReLU activation, and the identity blocks, which incorporate the skip connections that ensure that the gradient flow is not compromised during the backpropagation. This design also makes it possible to train very deep networks and helps to improve the ability of the generalization across various problems[32,33].

VGG16

VGG16 is a deep convolutional neural network (CNN) known for its simplicity and effectiveness, particularly in image classification tasks. The architecture comprises 16 layers, including 13 convolutional layers followed by three fully connected layers, as shown in Fig. 3. Each convolutional layer employs small 3×3 receptive fields, which enables the model to capture fine details and intricate patterns in input images. This use of small filters allows VGG16 to build a deep hierarchy of features, progressively capturing more complex patterns as the data passes through the layers. Following the convolutional layers, max-pooling layers are introduced to reduce the spatial dimensions of the feature maps while retaining the most salient features. Despite its relatively

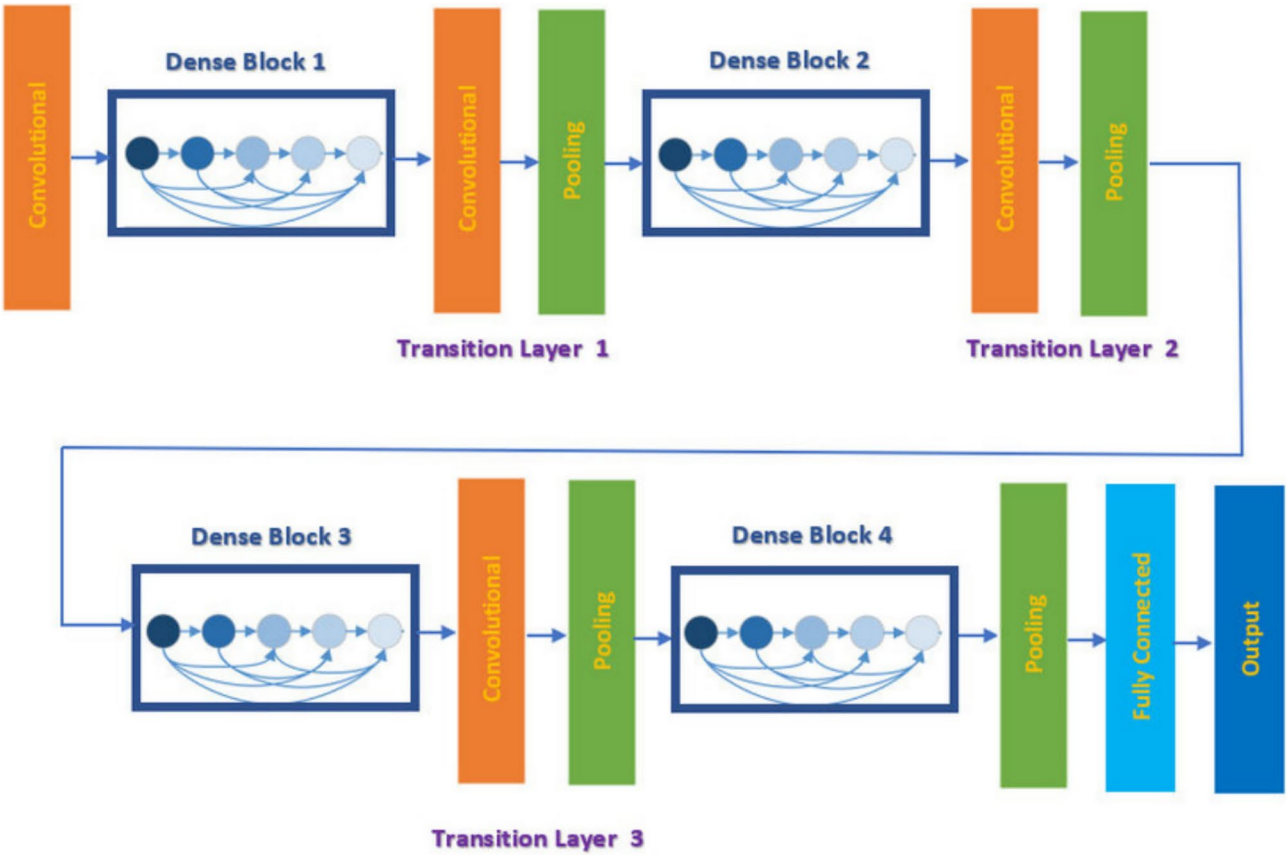

Fig. 1. DenseNet-201 architecture.

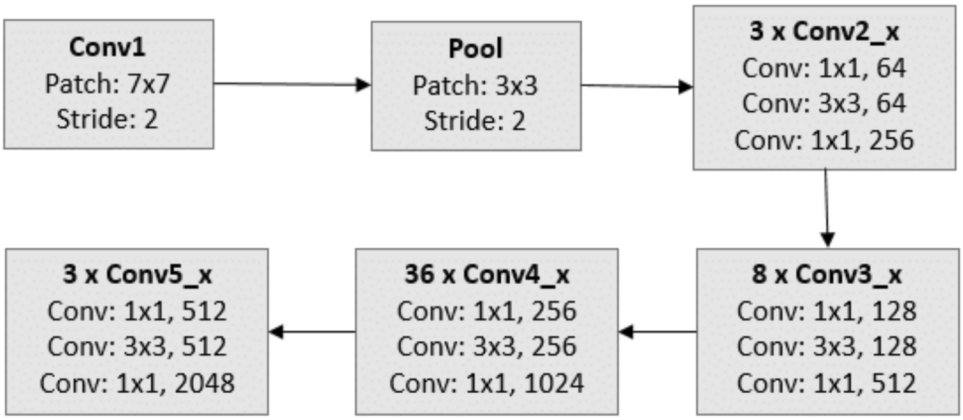

Fig. 2. Basic ResNet architecture.

straightforward architecture, VGG16 achieves high performance by stacking multiple layers and relying on small filters, which makes it effective for various computer vision tasks[34,35].

VGG19

VGG19 is yet another architecture of CNN, which is like VGG16, except it is deeper as it possesses 3 more convolution layers than VGG16 to make a total of 19 layers, as shown in Fig. 4. These additional layers provide an increase in the expressive ability and provide the model with the capacity to discern additional levels of information and characteristics in the images, which might help to increase the results of classification of images. As seen in VGG16, the overall network structure of VGG19 still remains the same along with the parameter of using small 3×3 convolutions so that the model can learn the details of the input images. Like what was observed in VGG16, the convolutional layers of the VGG19 network are succeeded by the max-pooling layers, whose

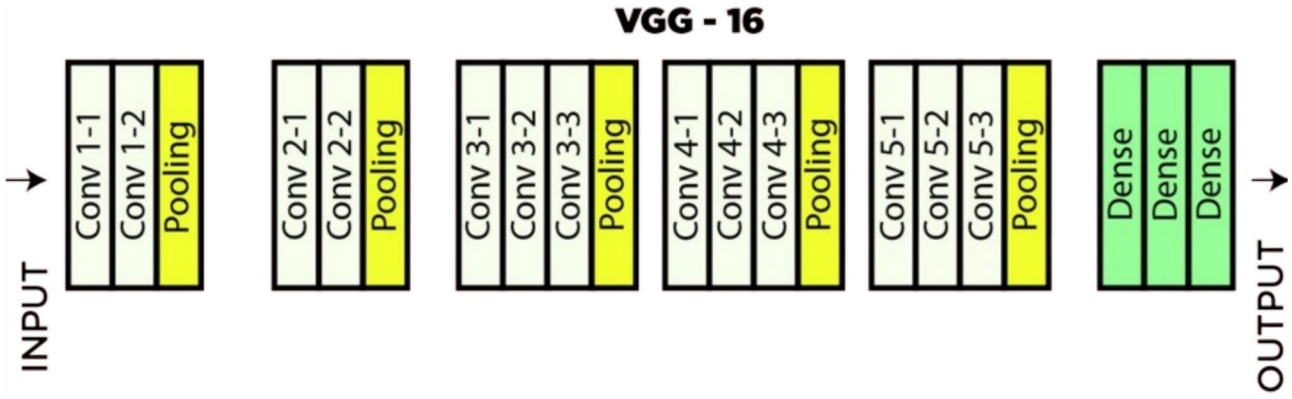

Fig. 3. VGG-16 architecture.

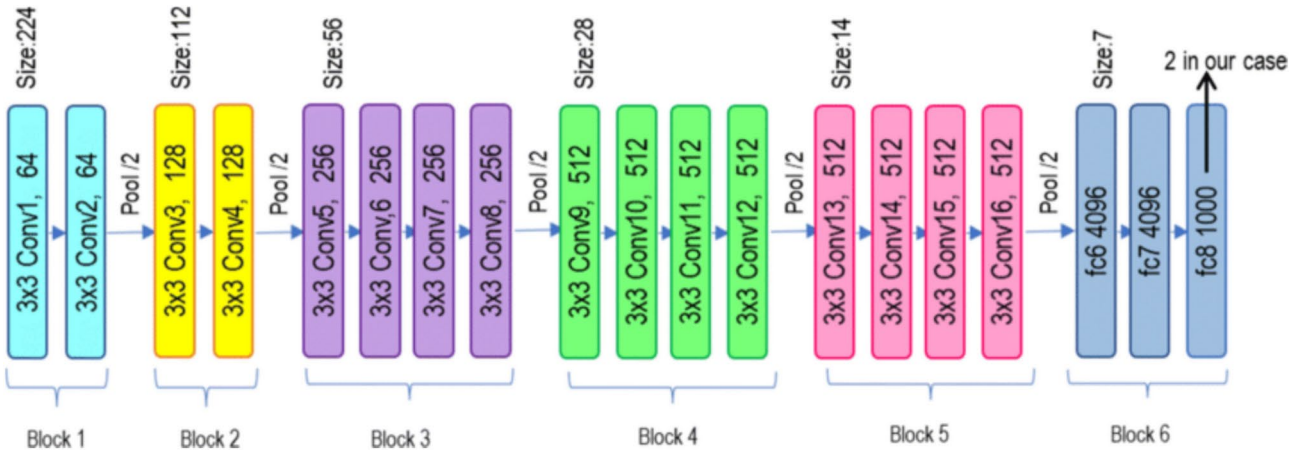

Fig. 4. VGG-19 architecture.

role is to decrease the size of the feature map while at the same time preserving the most crucial features. These reductions in the spatial dimension are also useful to manage the computational complexity of the network. At the end of the network, there are fully connected layers that are used to execute classification on the learned features[36,37].

EfficientNet-B0

EfficientNet-B0 is a widely employed deep learning model that optimizes for both accuracy and resources, allowing it to be one of the most powerful architectures for the image classification task, as shown in Fig. 5. Depth-wise, width-wise, and resolution-wise, the scaling of the network's dimensions is also formulated uniformly by a powerful and simple formula that the model offers. This makes it possible to scale up the model's capacity to other, larger versions, which, in effect, produce a hierarchy of models, each more complex and accurate than the last. EfficientNet-B0 is the fundamental model from which all others in the EfficientNet series are derived. The architecture follows mobile inverted bottleneck convolution (MBConv) layers that are light on parameters and have a great capability in feature extraction. Also, it incorporates the depth-wise separable convolutions, which help to cut the computational burden that is normally caused by the standard convolutional layers into the depth-wise and point-wise CV or convolutional layers. Such deliberations allow EfficientNet-B0 to achieve accuracy on par with other state-of-the-art designs at much lower parameter and FLOPs costs relative to the CNN counterparts[12,38].

MobileNetV2

MobileNetV2 is a convolutional neural network specially designed for optimizing the network and computational resources for mobile and IoT devices and thus is very suitable for scalable autism diagnosis systems. The architecture is based on depth-wise separable convolutions that drastically decrease computational requirements as standard convolution is divided into two layers: one for feature extraction and another for feature merging shown in Fig. 6. This significantly reduces the quantity of parameters and Floating-Point Operations (FLOPs); hence, the model is non-gargantuan and proficient. Another significant change incorporated in MobileNetV2 is that of inverted residuals with linear bottlenecks. In this structure, the shortcut connections that are linking the

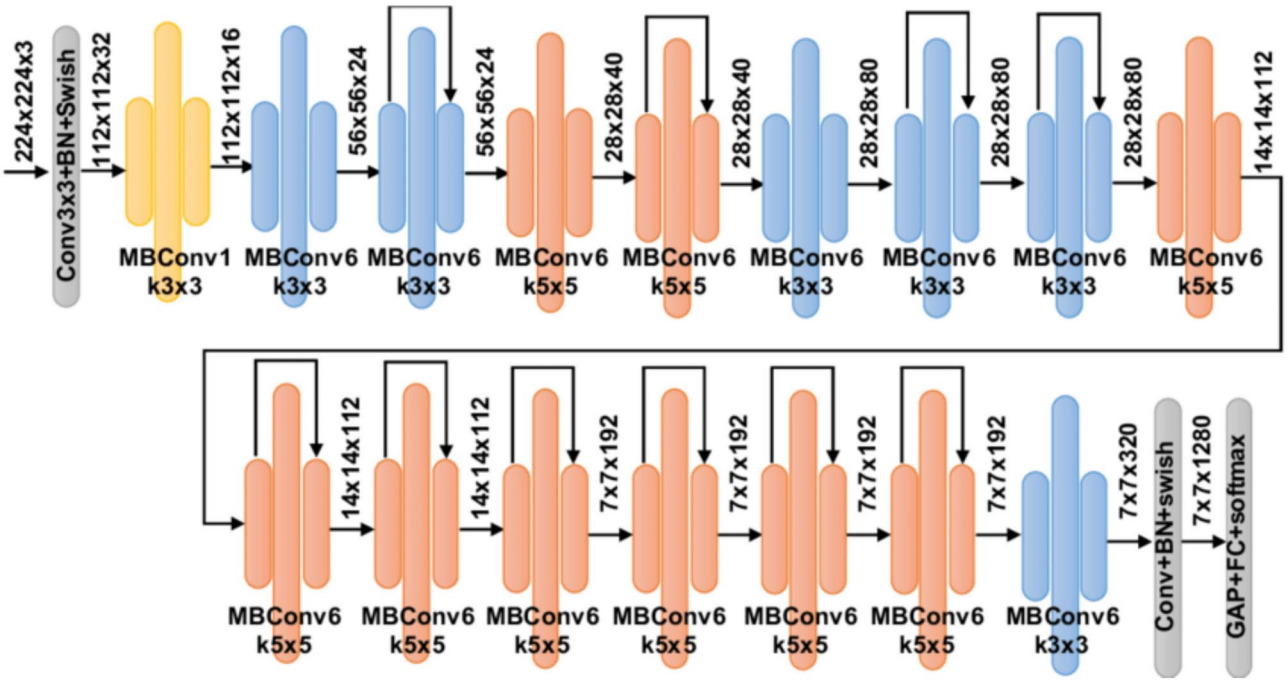

Fig. 5. EfficientNet-B0 architecture.

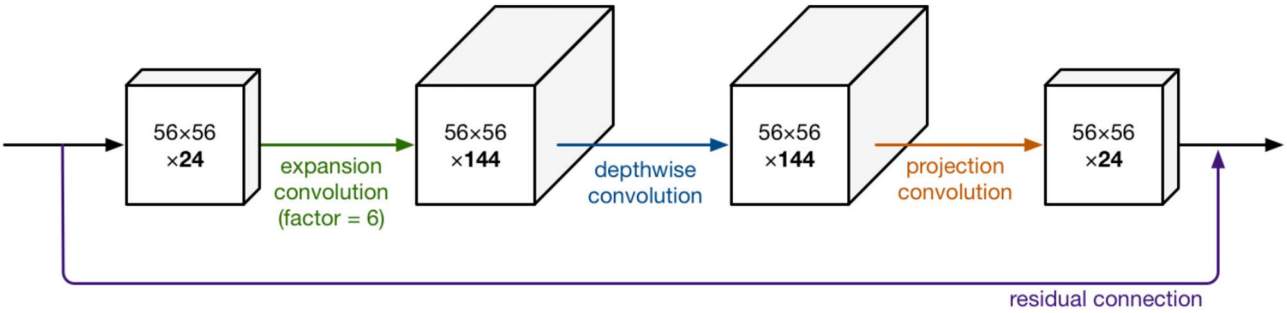

Fig. 6. MobileNet-V2 architecture.

layers can exclude full-dimensional feature maps, but the important parts of the previous layers are able to be preserved and at the same time keep the model very compact. Not only does this design optimize computation, but also the significant attributes that must be preserved across layers are maintained efficiently[36,39].

Vit transformer

The Vision Transformer (ViT) employs transformer-based models that originally operated in natural language processing to perform image classification tasks. ViT operates differently than standard convolutional neural networks (CNNs) since it processes images as consecutive patches yet transforms them through transformer encoder layers. The image segmentation process starts by dividing it into fixed-sized non-intersecting blocks, after which these blocks become linearly arranged for positional encoding maintenance. The transformer structure of ViT allows the system to recognize image dependencies throughout the entire visual space, which enhances its performance in dataset-based image classification tasks. Scientific research demonstrates that ViT delivers better results than CNN-based models for big data analysis, according to Dosovitskiy et al. and Vaswani et al.[40,41]. Research demonstrates how pre-training large datasets before fine-tuning specifically for tasks shows ViT's capacity for adaptable visual tasks. ViT has become a common choice in computer vision applications because it brings important performance and efficiency enhancements through modifications like DeiT (Data-efficient Image Transformer)[42], which improves outcomes in smaller datasets by enhancing training data use.

Proposed hybrid model

The proposed model combines Convolutional Neural Networks with Vision Transformers as shown in Fig. 7 to enhance facial image classification in autistic children. The CNN component utilizes ResNet152 as its base pre-trained on ImageNet[43] to extract hierarchical spatial features. The model uses the fine-tuning approach that

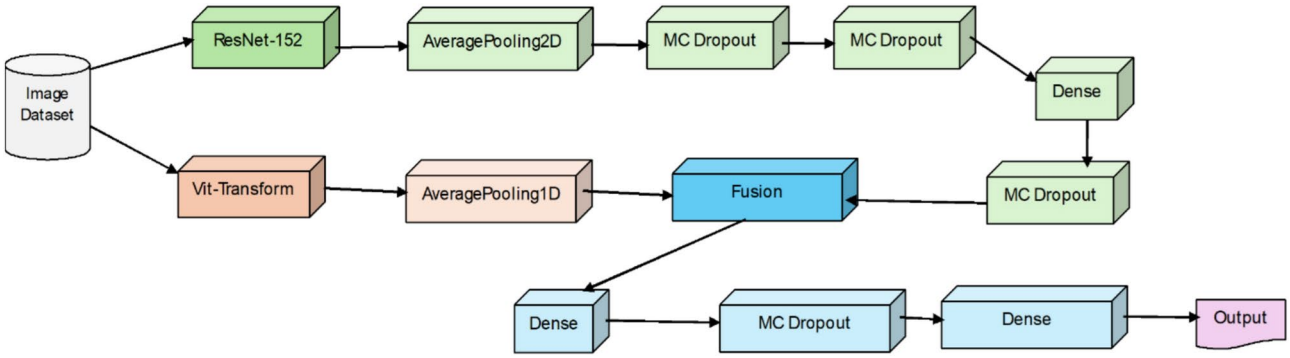

Fig. 7. Proposed hybrid model.

activates the last 20 layers of ResNet152 for parameter adjustment while keeping the previous layers static. The extracted CNN features pass through Global Average Pooling before entering two Monte Carlo (MC) Dropout layers, which bring randomization for correct feature selection. Google's ViT-Base-Patch16 functions as the ViT model that conducts transformer-based tokenization on the same image while performing normalization and channel reordering procedures. The global average pooling process collects representative feature embeddings from the ViT's final hidden state.

The fusion process begins when both network features combine in a concatenation layer and then proceed to a dense layer using ReLU activation to refine the features. A single sigmoid-activated neuron serves as the classification head after preceding it with an MC Dropout layer. The training process utilizes the Adam optimizer at a learning rate set to 0.0001 together with binary cross-entropy used as the loss function. The uncertainty estimation during inference depends on the use of MC Dropout because it generates multiple stochastic forward passes. This combined model architecture combines CNN spatial abilities with ViT long-range contextual understanding to enhance the recognition performance of autistic faces.

Figure 8 presents diverse object images selected from the extensive ImageNet database that consists of various categories, including thousands of object classes. ImageNet functions as a fundamental reference database for deep learning investigations because it contains images for both classification and recognition objectives. This database provides researchers with a versatile choice of images that include both animal subjects and household items and natural scenes together with human actions and urban city views for educational purposes in deep learning model training and evaluation.

Experimental results

In this section, we report the findings of the study with regards to the classification of autism-related learning difficulties in children using DenseNet201, ResNet152, VGG16, EfficientNet-B0, MobileNetV2, VGG19, and the proposed hybrid model. The criteria for assessment include accuracy, sensitivity, specificity, F1 score, and computational complexity. We also present the statistical values of the outcome to compare the significance of the variations between models. All the experiments were conducted with a learning rate of 0.0001, the epoch of 80 and 100, a batch size of 32, a seed of 42, and an image size = 224 × 224. The experiment computer has a 64-bit operating system, an × 64-based processor, the Intel(R) Core (TM) i7-3612QM CPU 2.10 GHz and 2.1 GHz, and 10 GB RAM. Deep learning parameters used for all models are declared in Table 1.

Performance metrics

Each model was evaluated using a range of performance metrics to assess their ability to correctly classify instances of autism-related learning challenges. The metrics used include accuracy, sensitivity, specificity, and F1-score. These metrics are defined as follows[44]:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN}) \tag{1}$$

$$\text{Precision} = \text{TP} \ / \ (\text{TP} \ + \ \text{FP}) \tag{2}$$

$$\text{Recall} = \text{TP} \ / \ (\text{TP} \ + \ \text{FN}) \tag{3}$$

$$F1\text{-measure} = 2\text{TP}/(2\text{TP} + \text{FP} + \text{FN}) \tag{4}$$

$$\text{Specificity} = \text{TN} \ / \ (\text{FP} \ + \ \text{TN}) \tag{5}$$

where, true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are critical as they represent how effective the model is. Accuracy quantifies the number of correct predictions out of all the birth outcomes. Recall, or sensitivity is the precision of correctly predicted positive cases. Recall shows the model's sensitivity for the actual positives, expressed in the percentage of correctly identified cases. F1-measure unites precision and recall, which makes it possible to evaluate the model's accuracy in terms of positive and negative instance identification. Specificity in data mining mainly addresses the aspect of accuracy in the model's identification of negative cases.

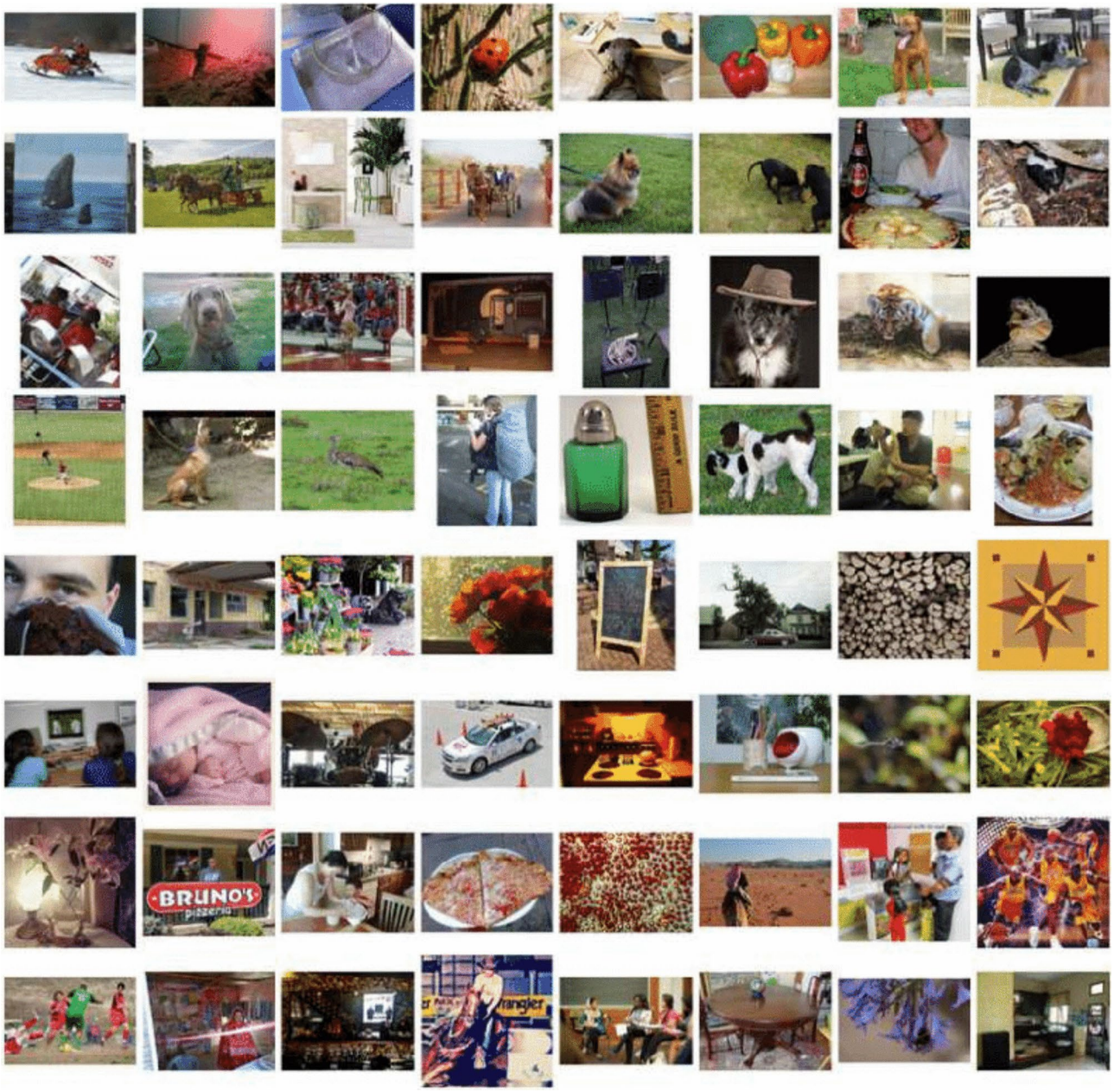

Fig. 8. Samples of ImageNet database[43].

Hyperparameter name	Value
Batch size	32
Seeds	42
min_lr	0.0001
Epochs	80 and 100

Table 1. Sample of hyperparameters values.

Dataset

The dataset used for this research comes from Kaggle under the provision of[45]. The dataset includes primary visual data through facial images of children with and without autism that most officials obtained from websites linked to autism and Facebook pages. The predominant focus of data collection included European and United States children with minor representation from other global regions. The data distribution consists of identical

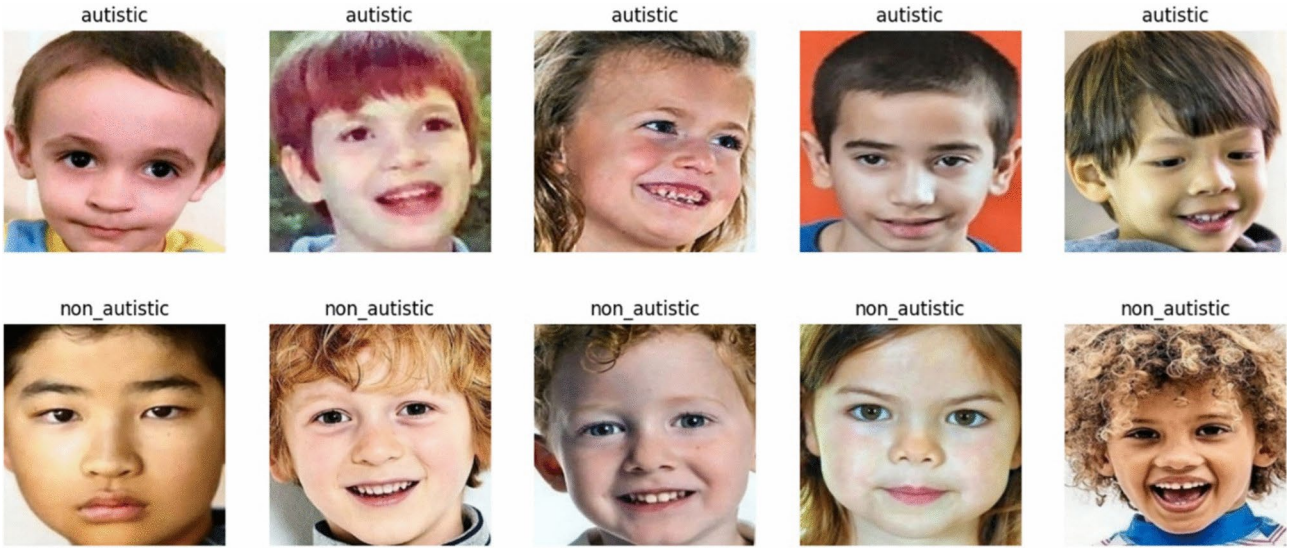

Fig. 9. Sample of dataset images[45].

Hyperparameter	Value
image size	(224, 224)
Batch size	16
Epochs	100
Learning rate	0.0001
Dropout rate	0.3
Augmentation	Rotation, flipping, contrast adjustment
Optimizer	Adam
Loss function	Binary cross-entropy

Table 2. Values of hyperparameters used in the proposed model.

quantities for male and female images from autistic children and non-autistic children. Uniform specifications must be applied to standardize the images of different sizes before the model can receive training.

Each image received three processing steps, including pixel normalization to the [0,1] scale and a 224×224 pixel resizing operation. The model received data augmentation through techniques that included image rotation and flipping as well as adjustments to image contrast, which improved model generalization while minimizing overfitting. The dataset division allocated 2540 images for training, 100 images for validation, and 300 images for testing. The feature extraction phase included using ResNet152 and Vision Transformer (ViT) models, which received additional processing methods optimized to work with each model. Standard normalization was applied for CNN-based models, yet per-image standardization served the ViT-based approach to maintain consistency. The preprocessing operations produced an organized dataset that succeeded in creating effective autism classification results.

The images in Fig. 9 showcase different facial images depicting children in two distinct categories for autism identification purposes. The classification separates autistic images into the top row and non-autistic images into the lower row. The presented samples showcase the differences across both categories regarding facial traits and expressions, thus demonstrating dataset variation.

Table 2 presents the key hyperparameters utilized in the proposed model for autism classification. These hyperparameters were carefully selected to optimize the model's performance while maintaining computational efficiency. The image size was standardized to (224, 224) pixels to ensure consistency across all inputs. A batch size of 16 was chosen to balance memory constraints and training stability. The model was trained for 100 epochs with a learning rate of 0.0001, employing the Adam optimizer to enhance convergence. Dropout was applied at a rate of 0.3 to mitigate overfitting, with Monte Carlo (MC) Dropout performing 20 passes for uncertainty estimation. Fine-tuning was restricted to the last 20 layers of ResNet152 to preserve pre-trained features while allowing domain adaptation. Data augmentation techniques, including rotation, flipping, and contrast adjustments, were applied to enhance generalization. The binary cross-entropy loss function was used to guide the model's learning process.

Results

Several confusion matrices appear in Fig. 10 that evaluate deep learning model classification performance for autistic and non-autistic identification using DenseNet-201, ResNet-152, VGG-16, EfficientNet-B0, MobileNet-V2, VGG-19, and a proposed hybrid model. The visual organization in the confusion matrix indicates true positive, positive, negative, and false negative scores of different prediction models. DenseNet-201 identifies 123 true positives, 7 true negatives, 27 false positives, and 23 false negatives; however, ResNet-152 outperforms slightly with 126 true positives, 143 true negatives, 7 false negatives, and 24 false positives. The two versions of VGG show matching trends as VGG-16 delivers 116 true positives along with 133 true negatives while misclassifying 34 instances as false positives and 17 instances as false negatives, yet VGG-19 shows 115 true positives combined with 122 true negatives alongside 35 false positives and 28 false negatives. The classification results show EfficientNet-B0 achieves better performance than MobileNet-V2 because it detects 125 true positives and 132 true negatives and produces 25 false positives and 18 false negatives, while MobileNet-V2 shows 116 true positives and 125 true negatives alongside 34 false positives and 25 false negatives. The proposed hybrid model surpasses every single model in tests by reaching maximum accuracy through 136 true positives together with 138 true negatives while keeping false positives at 14 instances and false negatives at 12 instances. The proposed hybrid model shows better classification results since it optimizes sensitivity and specificity to provide a more dependable solution for autism identification.

Table 3 demonstrates an extensive overview of deep learning model metrics and computational capacity by analyzing DenseNet-201, ResNet-152, VGG-16, EfficientNet-B0, MobileNet-V2, and VGG-19 together with the proposed hybrid system. All performance measures from the proposed hybrid model demonstrate the highest achievement, with accuracy reaching 91.33%, precision reaching 90.67%, and recall reaching 91.89%, along with an F1 measure of 91.28% and specificity of 90.79%. The model's classification performance exceeds the

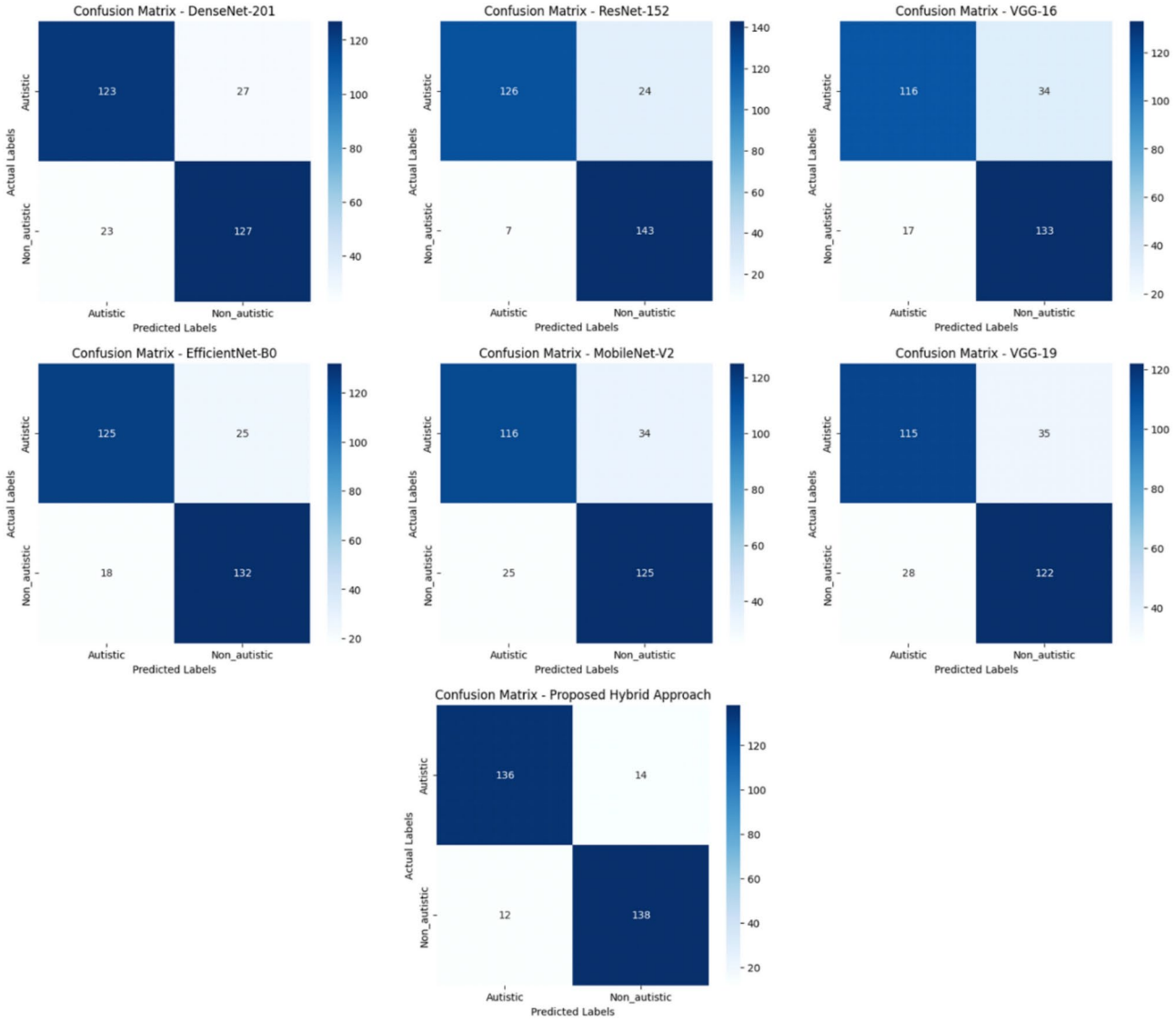

Fig. 10. Confusion matrices.

Model	Precision	Recall	F1 measure	Accuracy	Specificity	T-test *p*-value (actual vs. predicted)	ANOVA *p*-value	Real time in milliseconds
DenseNet-201	0.82	0.8425	0.8311	0.8333	0.8247	0.012	0.001	50
ResNet-152	0.84	0.9474	0.8905	0.8967	0.8563	0.045	0.001	40
VGG-16	0.7733	0.8722	0.8198	0.83	0.7964	0.003	0.001	70
EfficientNet-B0	0.8333	0.8741	0.8532	0.8567	0.8408	0.021	0.001	15
MobileNet-V2	0.7733	0.8227	0.7973	0.8033	0.7862	0.001	0.001	10
VGG-19	0.7667	0.8042	0.785	0.79	0.7771	0	0.001	80
Proposed Hybrid Model	0.9067	0.9189	0.9128	0.9133	0.9079	0	0.001	15

Table 3. Classification Report on pretrained model's vs proposed hybrid model.

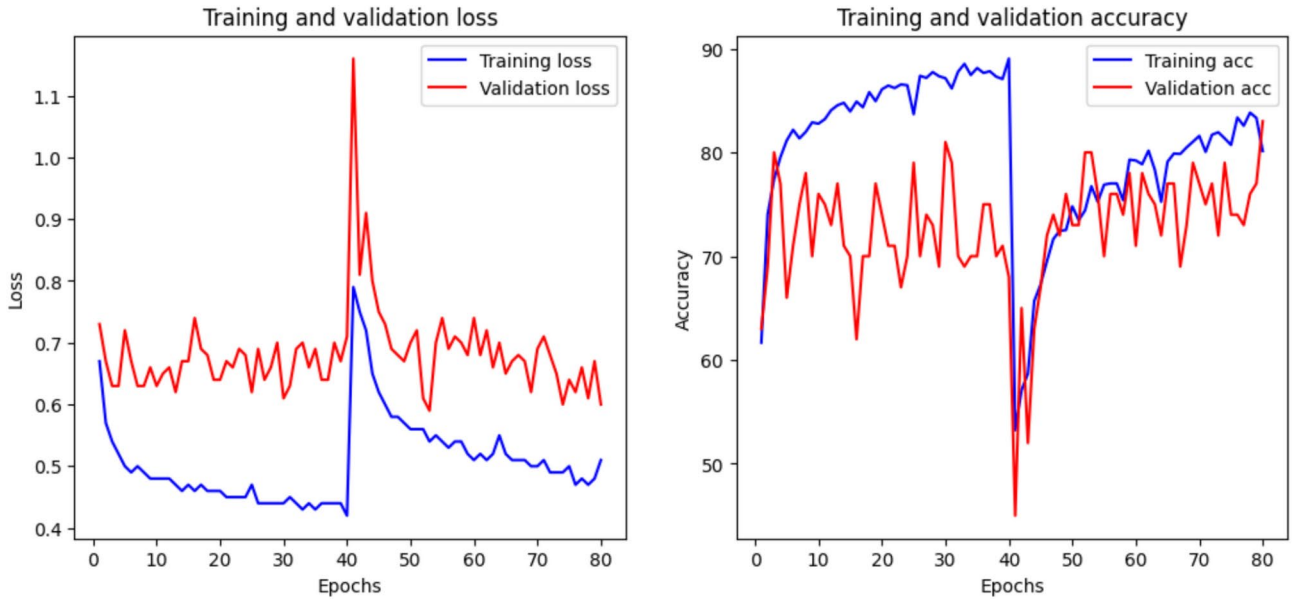

Fig. 11. Accuracy and loss for DensNet-201.

levels of alternative models, which contributes to its superior ability to conduct this task. The proposed model demonstrates superior performance reliability based on its statistically significant results through T-test *p*-values (all 0.000) and ANOVA *p*-value (0.001). The hybrid model demonstrates a real-time inference speed of 15 ms, which enables its implementation for real-time clinical applications.

The evaluation reveals different trade-offs occur between the speed of computation and performance results between the various implemented models. The inference times exceed 70 ms for VGG-16 and 80 ms for VGG-19 despite achieving 83.00% and 79.00% accuracy levels, which limits their practical utilization. MobileNet-V2 and EfficientNet-B0 trade less precision (80.33% and 85.67%, respectively) for faster response times (10 ms and 15 ms, respectively). The proposed hybrid model reaches its peak performance by surpassing other approaches both in accuracy and F1 measure while maintaining speeds near those of EfficientNet-B0. The proposed model presents itself as a suitable solution for clinical workflows because it meets both performance standards and processing speed requirements.

Figure 11 shows two graphs that depict the distinction between metrics of training and validation processes over 80 epochs for a machine learning-based solution. The left plot represents the training and validation loss, while the right plot represents the training and validation accuracy. Thus, we can observe that the training loss is continuously reducing over the epoch, signifying that the model is learning from the training data. On the other hand, the validation loss curve is also declining up to a certain epoch, after which the curve levels off, which indicates that the model is just beginning to over-train the training set.

The right plot shows the training and validation accuracy; in this graph, the training accuracy is consistently improving and reaches up to 90%, which means the model fits very well to the training set. However, subsequently the validation accuracy rises and oscillates in the range of 70–83% which shows that the models' ability to generalize on unseen data does not increase after this stage, suggesting signs of overfitting. The fact that there is a gap between the training and the validation sets means that we need to employ techniques such as regulation to enhance the generality of the models.

Figure 12 contains two plots, where one plot illustrates the training and the validation metrics of a model on 80 epochs. The left plot is for training and validation loss where training loss is in a blue curve, and it traces a decreasing trend and moves closer to the zero line, which points out that the model is learning well from the

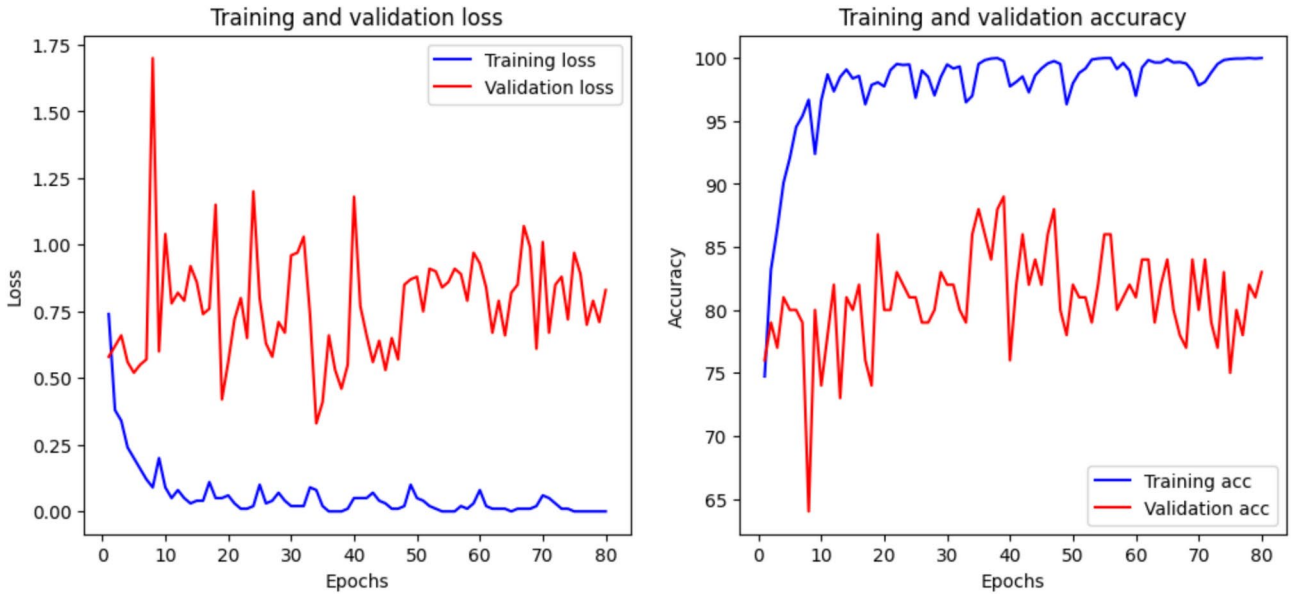

Fig. 12. Accuracy and loss for ResNet-152.

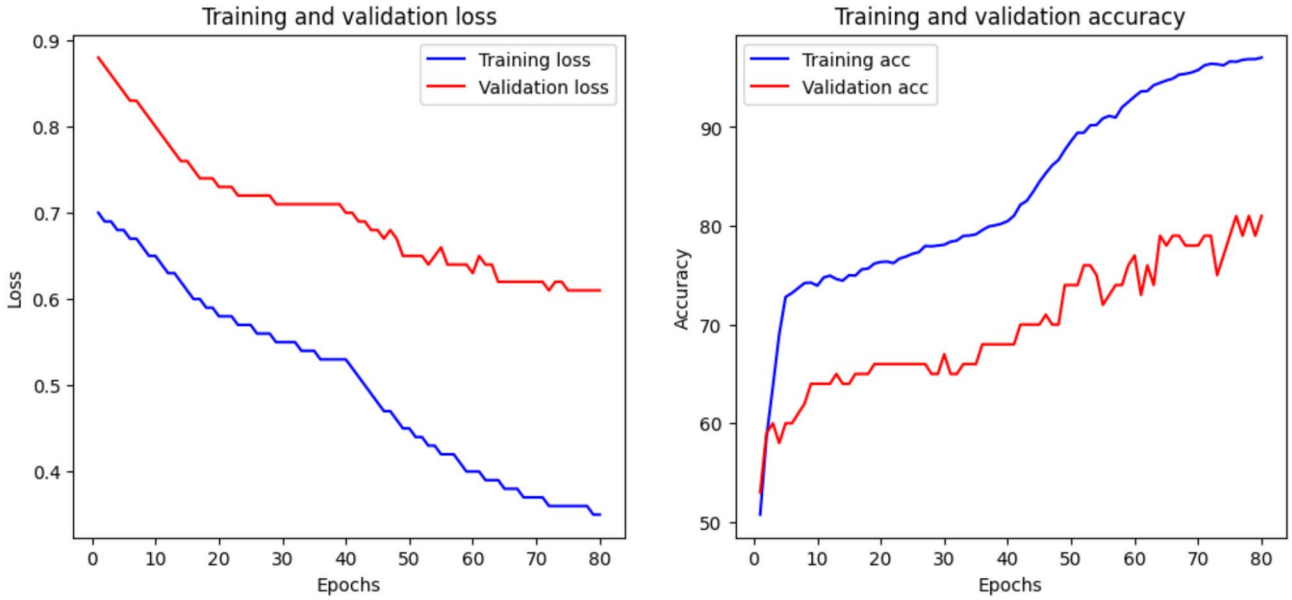

Fig. 13. Accuracy and loss for VGG-16.

training data. However, validation loss is still high and has great oscillations during the training, which indicates the model has a poor ability to predict new unseen data points and is probably overfitting. The fluctuating and jagged curve of validation loss reveals instability whereby the performance of the model tested on the validation set is unpredictable.

The right plot depicts training and validation accuracy. The blue color of the curve shows that training accuracy is rapidly achieving near to the perfection level or above 95%, which implies the model is doing very well within the training set. On the other hand, the validation accuracy has much lower values that hover between 65 and 89% and does not seem to enhance as the training progresses beyond the first epochs. The difference between the training and the validation accuracy shows that even though the model performs well at 'memorizing' the training data, it does not generalize well to unseen data, which is further evidence of overfitting.

Figure 13 shows two plots demonstrating the performance of the training and validation of a model indicating epochs to 80. The left plot captures training and validation loss; the training loss plot in blue shows that our model is learning from the training data through epochs. The same can be said for the validation loss, which also has a decreasing trend but at a slower rate that flattens out at around 0.7 after 50 epochs. This behavior implies that the

model is able to learn and generalize well in the first phase, but at some point, it stalls in terms of learning from unseen data. The problem might, however, be slight overfitting. The right plot is called training and validation accuracy. As for training accuracy, these values grow constantly and are higher than 90% at the end of training, which means that the model works very well with the training set. On the other hand, the validation accuracy increases at a slower rate and stabilizes at around 81% and points to a difference in the performance between the training and the validation set.

Figure 14 clearly shows two curves depicting the training as well as validation of a model for each of the 80 epochs. The left plot is about training and validation loss. The training loss curve demonstrates the model's ability to tackle the training dataset through the successive epochs of training. On the other hand, the validation loss stays high and varies significantly over the course of the training process. From the plot of validation loss, I observed that there is no significant downward pattern that should give indication that the model has good generalization on unseen data; hence, the model might be overfitting in the training data. It becomes even sharper, and fluctuations in the validation loss make the picture of the model's performance on the validation set even more unstable.

The right plot indicates training and validation accuracy. The blue line represents training accuracy, which rises rapidly and gets very close to 100, meaning the model has very high accuracy with the training data set. Validation accuracy is much lower and unstable, always ranging between 70 and 85 percent, and does not rise even after the first few iterations. The difference between train accuracy and validation accuracy reveals overfitting is a major problem of the model.

Figure 15 contains two graphs representing changes in the training and validation metrics of a model depending on the epoch. The left plot, Training and Validation Loss, reveals the fact that the training loss is getting smaller than epochs ranging from 0.65 to below 0.3, which shows that the model has an adequate capacity for learning from the training set. However, the validation loss remains high and fluctuates after periods indicating that the model fails to enhance the validation loss, instead revealing poor performance on unseen data. This constant difference between training and validation loss may raise an overfitting problem.

The right plot illustrates training and validation accuracy. Training accuracy increases gradually and goes above 95%, which means the model's accuracy is quite high if it's applied to the training set. Nevertheless, validation accuracy, or the percentage of correct predictions, never goes close to 79 percent and varies in a non-concentric range of 65–79 percent. The difference in the training and validation accuracy, as well as the oscillatory nature of the model's validation performance, also holds with the assessment that the model is overfitting the training examples.

Figure 16, the first subplot represents training accuracy while the second subplot represents the validation accuracy, and both are represented over the number of epochs to 80. The left graph is about training and validation loss, in which training loss gradually decreases from roughly 0.7 to below 0.4, which shows that the model can learn from the training data used in this study. The validation loss also follows in their footsteps and decreases, but this happens at a slower pace and stabilizes around 0.65 after 40 epochs. The decline of both losses shows that this model is performing an impressive learning rate in the initial epochs of training; however, the stabilization of the validation loss signifies that it requires more boosting of the model to improve the generalization of unseen data.

The right plot is for training and validation accuracy. The training accuracy line, in blue color, is constantly rising, and it is more than 95 percent at the end of the training phase, showing that the model has a good interlace with the training data. Instead, the validation accuracy starts to increase at the beginning but then

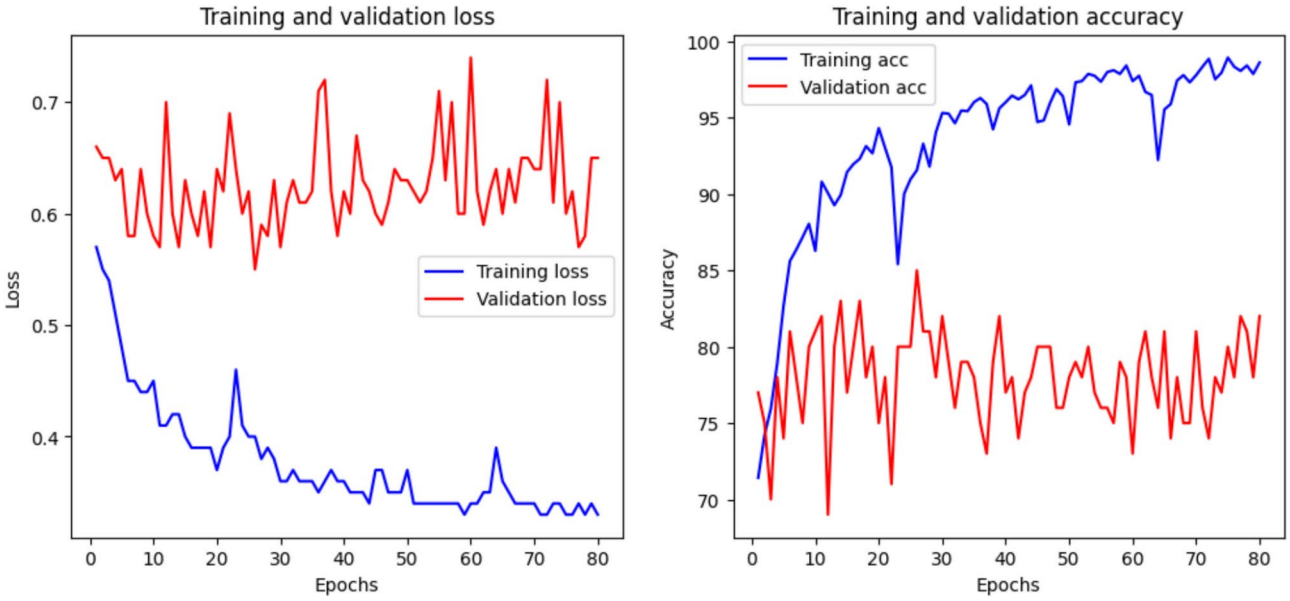

Fig. 14. Accuracy and loss for EfficientNet-B0.

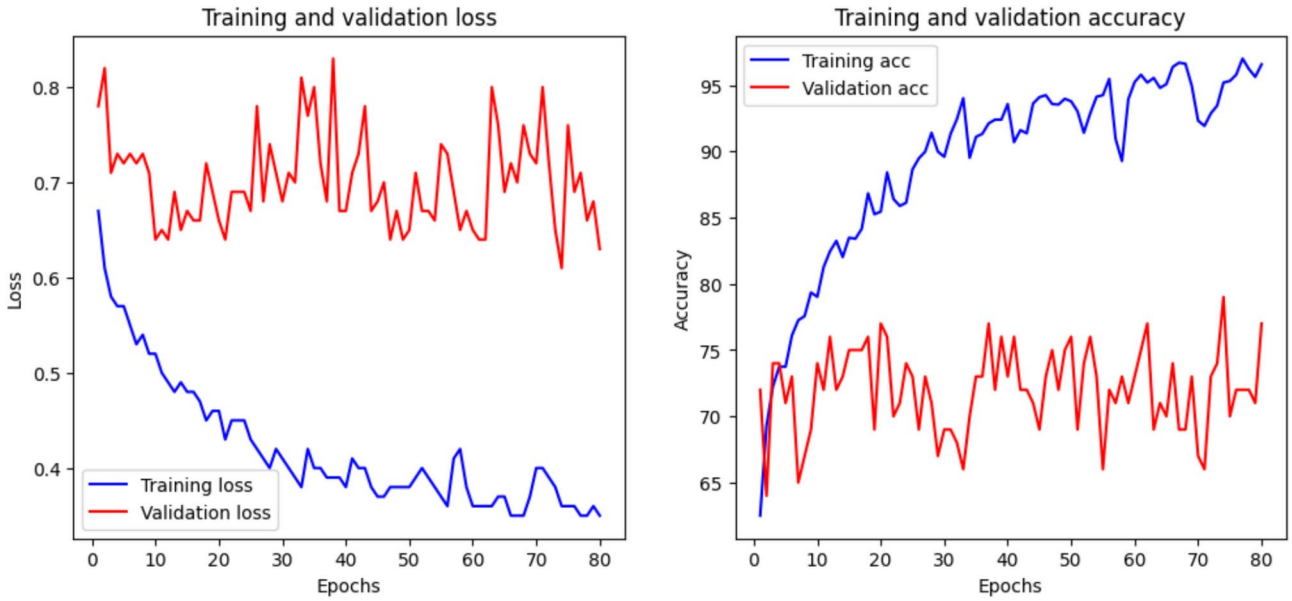

Fig. 15. Accuracy and loss for MobileNet-V2.

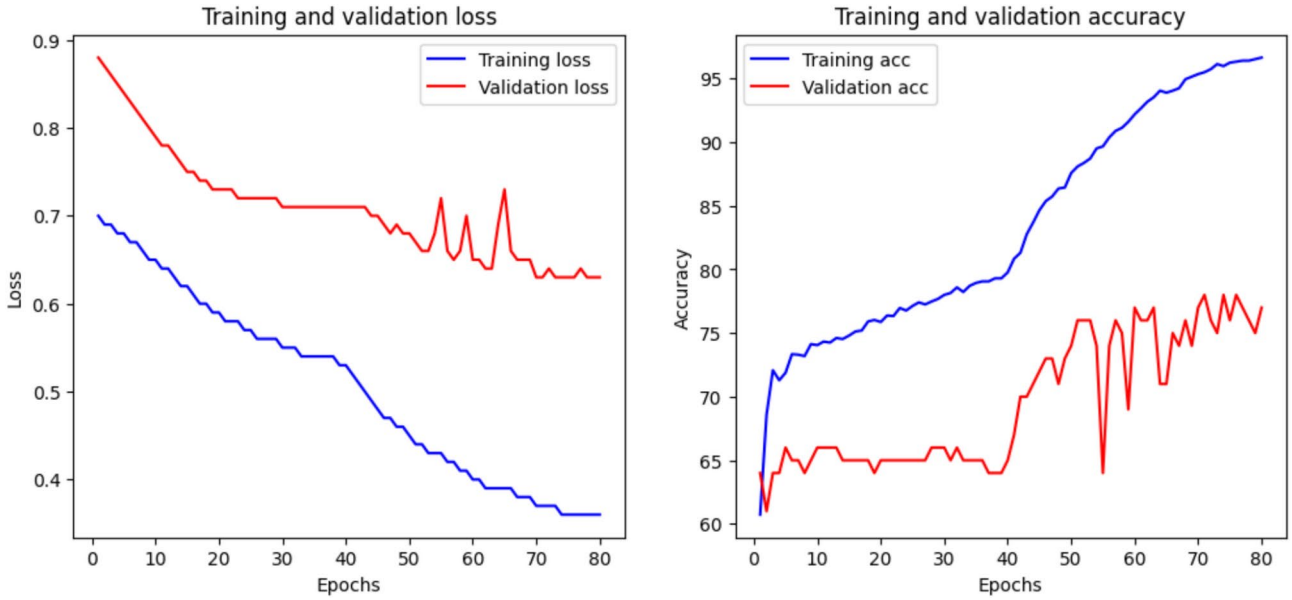

Fig. 16. Accuracy and loss for VGG-19.

oscillates around 70–78%, which means that the model is not able to generalize as well as to the new data. The difference between the training accuracy and the validation accuracy shows signs of overfitting, where the model performs very well in training data but poorly in the new data.

Figure 17 shows how a deep learning model performs on training and validation data throughout 100 epochs through accuracy trends on one graph and loss trends on another graph. The accuracy graph shows the training accuracy (blue line) rising quickly and then achieving stability near the value 1.0 in initial epochs, which demonstrates efficient model learning of training data. The validation accuracy (orange line) shows significant changes during the training period, which indicates that the model is overfitting because it fails to apply learned patterns to new unseen data points. The effort of the model toward minimizing training errors is confirmed by the decrease of the blue training loss line in the loss graph. Validation loss shows large volatility with multiple abrupt spikes across the training period, especially from the beginning until the middle stages of training. The model demonstrates overfitting behavior through its training of the data points effectively, yet its inability to predict unseen validation data. The ongoing increase in validation loss demonstrates that the model maintains excessive learning behavior during the latter training epochs. The constant descent of training loss (blue line)

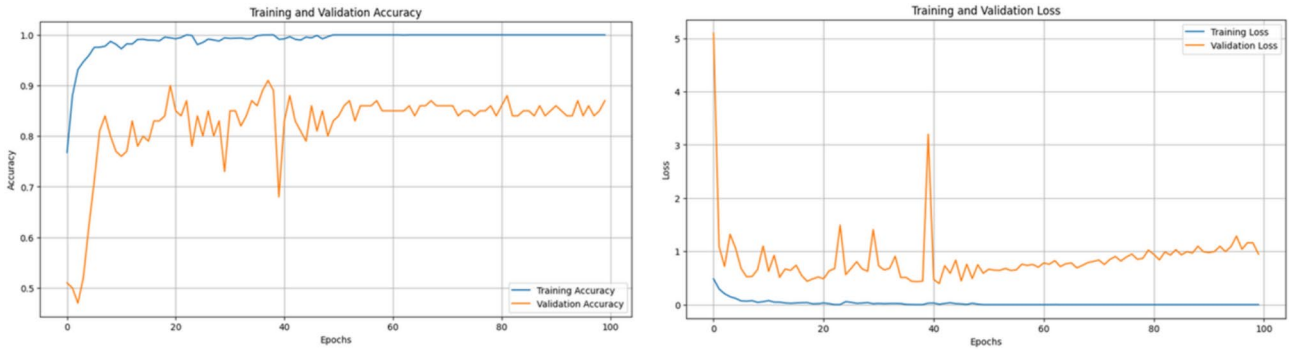

Fig. 17. Accuracy and loss for proposed hybrid model.

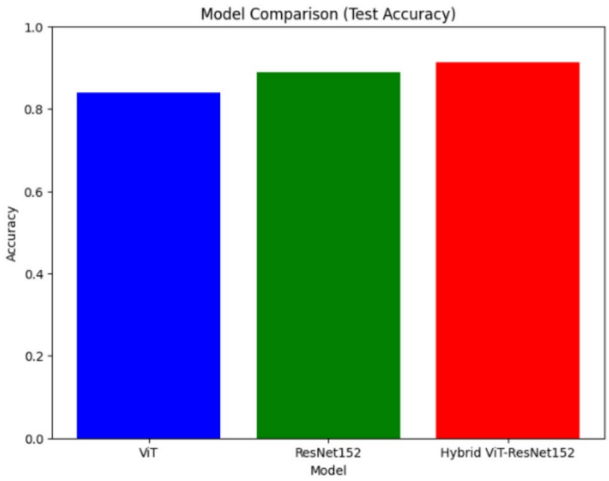

Fig. 18. Comparison vit, ResNet152 and hybrid for test accuracy.

shows continuous progress while validation loss fails to show consistent improvement, thus proving that the network requires extra generalization capabilities.

Our approach added regularizations through dropout layers, batch normalization, and early stopping to block the model from memorizing random patterns in training data. The data augmentation methods used image rotation along with flipping and normalization functions to improve both robustness and data variability. The modified model produced enhanced validation loss patterns that displayed improved stability and decreased oscillations since the implemented strategies reinforced generalization strength. The revised model demonstrated enhanced validation accuracy consistency that matched training performance better, as it indicated improved adaptability to new data.

Figure 18 display demonstrates how the test accuracy stands between three different models, which consist of Vision Transformer (ViT), ResNet152, and a combination of ViT and ResNet152. Each unit along the y-axis shows accuracy measurement between 0 and 1, yet the x-axis differentiates between the three models. The blue bar indicates the ViT model offers relatively strong accuracy yet does not reach the same level as the green bar depicting ResNet152 accuracy performance achievement. The combination of the ViT and ResNet152 models (shown as a red bar) reaches the best accuracy score among the three available models. The combination of ViT models with ResNet152 results in superior feature extraction abilities and classification skills, which generates better results when testing real data.

Figure 19 shows which parts of the input photos affected the model prognosis most significantly. The model highlights important clinical indicators like facial features and symmetrical regions in the face during autism spectrum disorder classification. The model-based diagnostic decisions rest upon anatomical patterns that become visible through visualization methods, thus enabling trustworthy clinical deployment.

The proposed hybrid deep learning system performance stands against traditional ASD diagnosis procedures and simpler machine learning models through the Table 4 evaluation. The current ASD diagnosis process depends on both the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R), which need expert professionals and amount to lengthy examination times. Support Vector Machines (SVM) deliver subpar performance measurements because they do not work well across different situations. The model is tested against established diagnostic procedures for evaluation. The hybrid modeling

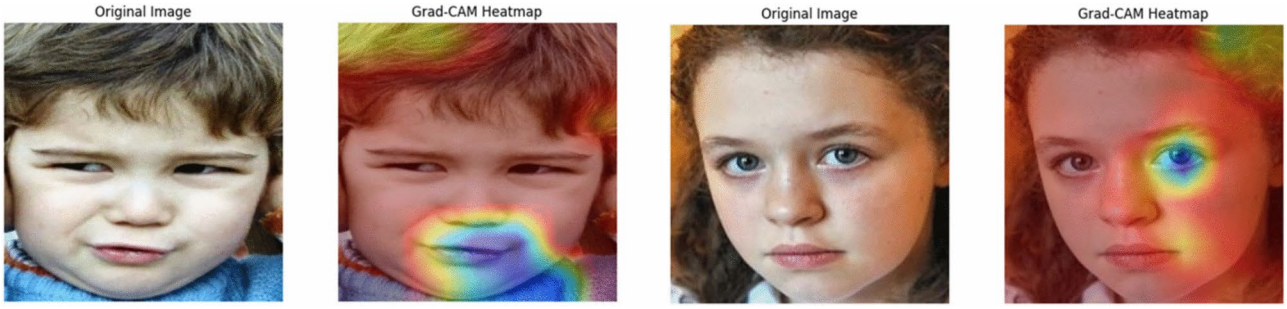

Fig. 19. Grad-CAM heatmap.

Method	Precision	Recall	F1 measure	Accuracy	Specificity
Traditional ASD diagnostic (ADOS, ADI-R)[46]	0.79	0.81	0.80	0.82	0.80
SVM prediction of ADOS[47]	0.85	0.87	0.86	0.88	0.86
Proposed hybrid model	0.9067	0.9189	0.9128	0.9133	0.9079

Table 4. Comparison between the proposed hybrid deep learning model and traditional ASD diagnostic approaches.

Refs.	Year	Methods	Accuracy (%)
Hayder and Amir[24]	2025	Enhanced vision transformers	87.67
Karthik et al.[25]	2024	ViT-XGBoost-SHAP	91.30
Pranavi and Andrew[27]	2024	VGG16	84.66
		VGG19	80.05
		EffiecientNetB0	87.90
Rahman and Subashini[26]	2022	MobileNet	84.67
		Xception	90
		EfficientNetB0	86.64
		EfficientNetB1	89.67
		EfficientNetB2	**88.67**
Proposed hybrid model	–	ResNet152 + ViT	91.33

Table 5. comparison table summarizing related studies on autism detection using deep learning models for publicly available dataset from Kaggle.

system outperforms traditional techniques in all metrics by delivering higher accuracy as well as precision alongside recall and F1 scores and specificity.

Discussion

Deep learning models are a potential tool for detecting autism-related learning disabilities in children at an early stage. By studying DenseNet201, ResNet152, VGG16, EfficientNetB0, MobileNetV2, and VGG19 with transfer learning and fine-tuning, the author realized the strengths along with the limitations of each model for clinical use and its shortcomings.

ResNet152 showed excellent performance with 89% accuracy, proving slightly better than VGG16, VGG19, MobileNetV2, EfficientNetB0, and DenseNet201 models in identifying children with autism-related learning disabilities from those without. The composite scaling method allows ResNet152 to achieve improved depth, accuracy, and width, resulting in its effective performance along with its efficacy. The harmonious combination of accuracy and efficiency makes ResNet152 an ideal solution for clinical purposes, especially within resource-strapped healthcare systems.

The analysis through Table 5 examines deep learning model performance in autism spectrum disorder (ASD) detection through facial image assessments. The proposed hybrid model of ResNet152 with Vision Transformers (ViT) reaches a 91.33% accuracy level, surpassing the ViT-XGBoost-SHAP model (91.3%) presented by Karthik et al. (2024). Rahman and Subashini (2022) observed significant success when Xception delivered 90% accuracy, which outperformed MobileNet (84.67%) as well as EfficientNet variants (86.64–89.67%) in older research works. According to the research of Pranavi and Andrew (2024), EfficientNetB0 (87.9%) displayed superior performance compared to VGG16 (84.66%) and VGG19 (80.05%) in their evaluation of CNN architectures. The

research paper by Hayder and Amir (2025) presented Enhanced Vision Transformers, which reached an 87.67% success rate in their tests. The analysis demonstrates that combined models with transformer capabilities offer superior performance for facial image diagnosis of ASD through their implementation.

The proposed research developed a hybrid deep learning model that integrates Vision Transformer (ViT) with ResNet152 to improve classification accuracy. ViT-ResNet152 works by combining two architectures that enable ViT to exploit image-based relationships while ResNet152 applies its hierarchical nature to extract features. Experimental results showed that the combined vision model outperformed the results of the standalone network, indicating that the integration of Vision Transformer with traditional convolutional networks increases the accuracy of detecting autism-related learning disabilities. The successful implementation of this hybrid architecture demonstrates that the combination of multiple architectures produces better classification accuracy results with an accuracy of 91.33% along with improved dataset generalization capabilities and robust performance. The results from the AI model need external database tests together with inspections in authentic clinical settings to validate its final performance. These initial findings create a solid starting point, while researchers need to test the model further with various clinical settings to determine its generic use. Upcoming research requires collecting larger multi-site clinical data for testing practical clinical usage alongside real-time doctor assessments. The study demonstrates the future potential of deep learning models in ASD detection yet suggests more research should be conducted in this direction. AI tools need further development and validation before becoming valuable assets for early autism diagnosis.

This study has some limitations because the dataset used does not include all potential autism patients. This scientific work emphasizes the value of artificial intelligence algorithms to reshape how clinicians identify autism-related learning disabilities at an early stage. These excellent diagnostic tools show sufficient accuracy and sensitivity, which means that this technology should be adopted by clinical programs to provide objective and timely diagnoses that reduce subjective assessments. Early detection and accurate diagnosis of learning disabilities are essential to implement the right interventions that lead to improved developmental outcomes for children with ASD. Moreover, geographic bias emerges because the primary participants in the dataset come from children in U.S. and European areas, while also reducing the model's suitability toward diverse populations. The model's widespread application becomes limited when its usefulness is reduced due to developmental differences between regions that are affected by genetic and environmental factors, along with cultural elements. The absence of representation from underrepresented global areas could produce unfavorable diagnostic performance results when the model processes children from non-Western countries.

Although they provide a powerful set of examples. ASD is characterized by diverse manifestations and multiple symptoms, creating complexity in diagnosing the condition, while model performance can vary based on different data types beyond basic imaging and different patient populations. Transfer learning requires the use of pre-trained models with biases from the ImageNet dataset that may fail to match the RGB images and behavioral dataset of this study. Performance improves when we train our models with expanded and diverse dataset information. The analysis in this work performs a general model comparison by evaluating six deep learning models but fails to identify potential improvements from newer models or hybrid architecture solutions.

Conclusion

The paper indicates that deep learning technology successfully identifies autism-related learning disabilities in young children at an early stage. ResNet-152 proved to be the most successful autonomous model according to research outcomes after implementing transfer learning methods through fine-tuning six pre-trained convolutional neural networks consisting of DenseNet-201, ResNet-152, VGG16, VGG19, MobileNet-V2, and EfficientNet-B0. The model reached 89% accuracy. Safer and more effective approaches can be built from Vision Transformer (ViT) models united with ResNet-152 structures, even though these approaches show potential yet possess two critical problems stemming from pre-trained ImageNet dataset biases alongside the restriction to identify diverse autism spectrum disorder conditions. The ViT-ResNet152 model achieves better dataset generalization with increased classification accuracy by blending the Vision Transformer global attention model with ResNet-152 hierarchical feature extraction characteristics. Experimental tests confirmed that uniting both structures led to maximum accuracy, reaching 91.33% in detection results. As demonstrated by this hybrid network design, transformers integrated into traditional convolutional models improve the accuracy and reliability of ASD detection algorithms. Medical assessments find precise, scalable support from AI ensemble learning tools that work with specific approaches to reduce dependence on clinical judgment. In future work, the development of ensemble frameworks and expanded autism spectrum disorder dataset collection should receive attention from future research initiatives to improve both accuracy and eliminate biases in generalization outcomes. The diagnostic precision will increase through the addition of speech patterns, eye-tracking, behavioral assessments, and image-based features, which offer extra autism-related indicators. The validation of AI-based autism detection tools requires medical expert partnership and clinical professional collaboration to ensure their real-world reliability and effectiveness. Real-world healthcare deployment of clinical trial-tested models in hospitals will close experimental research-practical application gaps, making these technologies more dependable and impactful in real-world healthcare environments.

Data availability

Available online: https://www.kaggle.com/datasets/cihan063/autism-image-data.

Received: 5 February 2025; Accepted: 25 March 2025
Published online: 08 April 2025

References

1. Lord, C., Elsabbagh, M., Baird, G. & Veenstra-Vanderweele, J. Autism spectrum disorder. *The Lancet* **392**(10146), 508–520 (2018).
2. Maenner, M. J. et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 Sites, United States, 2018. *MMWR Surveill. Summ.* **70**(11), 1–16 (2021).
3. Thabtah, F. Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Inform. Health Soc. Care* **44**(3), 278–297 (2019).
4. Hosseini-Asl, E., Gimel'farb, G. & El-Baz, A. Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. (2016). arXiv preprint arXiv:1607.00556.
5. Lundervold, A. S. & Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.* **29**(2), 102–127 (2019).
6. Dinsdale, N. K., Bluemke, E., Smith, S. M. & Namburete, A. I. L. Learning patterns of the developing brain from neonates to young adults using convolutional networks. *Neuroimage* **238**, 118201 (2021).
7. Heinsfeld, A. S., Franco, A. R., Cameron Craddock, R., Buchweitz, A. & Meneguzzi, F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage Clin.* **17**, 16–23. https://doi.org/10.1016/j.nicl.2017.08.017 (2018).
8. Duda, M., Ma, R., Haber, N. & Wall, D. P. Use of machine learning for behavioral distinction of autism and ADHD. *Transl. Psychiatry* **6**(2), e732 (2016).
9. Thomas, M. S. C. & Davis, R. The promise of AI in detection, diagnosis, and treatment of neurodevelopmental disorders. *Psychol. Bull.* **146**(6), 533–559 (2020).
10. Arbabshirani, M. R., Plis, S. M., Sui, J., Calhoun, V. D. & Silva, R. F. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage* **145**, 137–165 (2017).
11. Voulodimos, A., Doulamis, N., Doulamis, A. & Protopapadakis, E. Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, 1–13. https://doi.org/10.1155/2018/7068349 (2018).
12. Tan, M. & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946.
13. Howard, A. G., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V. & Le, Q. V. (2019). Searching for MobileNetV3. arXiv preprint arXiv:1905.02244.
14. Wallace, S., Fein, D., Rosenthal, M. & Barton, M. Early intervention and long-term outcomes for children with autism spectrum disorder. *J. Autism Dev. Disord.* **50**(7), 2461–2473 (2020).
15. Chen, H., Duan, G. & Zhang, L. Brain connectivity patterns in ASD: A deep learning approach. *Front. Hum. Neurosci.* **11**, 481 (2017).
16. Lombardo, M. V., Moon, H. M. & Mandelli, M. J. A multimodal deep learning framework for ASD diagnosis integrating genetic and neuroimaging data. *Mol. Autism* **10**(1), 7 (2019).
17. Abdelnour, N., Varoquaux, G. & Thirion, B. Ethical challenges in the use of AI for neurodevelopmental disorders. *Nat. Rev. Neurol.* **14**(6), 392–393 (2018).
18. Li, Y., Wang, X. & Gao, J. Personalized intervention strategies for ASD: A machine learning approach. *J. Child Psychol. Psychiatry* **60**(10), 1123–1132 (2019).
19. Litjens, G. et al. A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88. https://doi.org/10.1016/j.media.2017.07.005 (2017).
20. Zhao, Y., Wang, S. & Gao, H. Eye-tracking data analysis for ASD identification using deep learning. *IEEE Access* **6**, 43889–43898 (2018).
21. Palkovics, J., Kolozsvári, L. & Iványi, P. Speech analysis in children with ASD using deep learning. *Cogn. Comput.* **12**(3), 560–570 (2020).
22. Shen, D., Wu, G. & Suk, H. I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **19**, 221–248 (2017).
23. Eslami, T., Saa, J. F. & Tassabehji, M. A hybrid AI approach combining deep learning and expert knowledge for ASD diagnosis. *Artif. Intell. Med.* **96**, 118–130 (2019).
24. Ibadi, H. & Lakizadeh, A. ASDvit: Enhancing autism spectrum disorder classification using vision transformer models based on static features of facial images. *Intell.-Based Med.* **11**, 100226. https://doi.org/10.1016/j.ibmed.2025.100226 (2025).
25. Karthik, M. D., Jeba Priya, S. & Mathu, T. Autism detection for toddlers using facial features with deep learning. In *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India* 726–731 https://doi.org/10.1109/ICAAIC60222.2024.10575487 (2024).
26. Mujeeb Rahman, K. K. & Monica Subashini, M. Identification of autism in children using static facial features and deep neural networks. *Brain Sci.* **12**(1), 94. https://doi.org/10.3390/brainsci12010094 (2022).
27. Reddy, P. Diagnosis of autism in children using deep learning techniques by analyzing facial features. *Eng. Proc.* **59**, 198. https://doi.org/10.3390/engproc2023059198 (2023).
28. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4700–4708 https://doi.org/10.1109/CVPR.2017.243 (2017).
29. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* 11–19 https://doi.org/10.1109/CVPRW.2017.156 (2017).
30. Attallah, O. MB-AI-His: Histopathological diagnosis of pediatric medulloblastoma and its subtypes via AI. *Diagnostics* **11**, 359. https://doi.org/10.3390/diagnostics11020359 (2021).
31. He, K., Zhang, X., Ren, S. & Sun, J. (Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778 https://doi.org/10.1109/CVPR.2016.90 (2016).
32. He, K., Zhang, X., Ren, S., & Sun, J. Identity mappings in deep residual networks. In *European Conference on Computer Vision* 630–645. (Springer, 2016). https://doi.org/10.1007/978-3-319-46493-0_38
33. Nguyen, L., Lin, D., Lin, Z. & Cao, J. Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation. 1–5. https://doi.org/10.1109/ISCAS.2018.8351550 (2018).
34. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556
35. Geeks for Geeks Organization. VGG-16 | CNN model, https://www.geeksforgeeks.org/vgg-16-cnn-model/ (Last Update March 21, 2024)
36. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4510–4520 https://doi.org/10.1109/CVPR.2018.00474 (2018).
37. Khattar, A. & Quadri, S. Generalization of convolutional network to domain adaptation network for classification of disaster images on twitter. *Multimed. Tools Appl.* https://doi.org/10.1007/s11042-022-12869-1 (2022).
38. Alhichri, H., Alsuwayed, A., Bazi, Y., Ammour, N. & Alajlan, N. Classification of remote sensing images using EfficientNet-B3 CNN model with attention. *IEEE Access* https://doi.org/10.1109/ACCESS.2021.3051085 (2021).
39. Seidaliyeva, U., Akhmetov, D., Ilipbayeva, L. & Matson, E. Real-time and accurate drone detection in a video with a static background. *Sensors* **20**, 3856. https://doi.org/10.3390/s20143856 (2020).
40. Dosovitskiy, A. et al. Discriminative unsupervised feature learning with exemplar convolutional neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(9), 1734–1747 (2014).

41. Vaswani, A., Shazeer, N., Parmar, N. Uszkoreit, J. Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. Attention is All You Need. arxiv, 30, (2023).
42. Touvron, H., Cord, M., Douze, M. Massa, F., Sablayrolles, A. & Jegou, H. *Proceedings of the 38th International Conference on Machine Learning, PMLR* vol. 139, pp. 10347–10357 (2021).
43. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA* 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848.
44. Mahmood, M. A. & Alsalem, K. Olive leaf disease detection via wavelet transform and feature fusion of pre-trained deep learning models. *Comput. Mater. Continua* **78**(3), 3431–3448 (2024).
45. Khan, I. Autistic Children Facial Dataset. (2022). Available at: https://www.kaggle.com/datasets/cihan063/autism-image-data (Last Accessed 17 February 2025).
46. Kamp-Becker, I. et al. Is the combination of ADOS and ADI-R necessary to classify ASD? Rethinking the "Gold Standard" in diagnosing ASD. *Front. Psychiatry* **24**(12), 727308. https://doi.org/10.3389/fpsyt.2021.727308 (2021).
47. Zhang, X. et al. Support vector machine prediction of individual Autism Diagnostic Observation Schedule (ADOS) scores based on neural responses during live eye-to-eye contact. *Sci. Rep.* **14**, 3232. https://doi.org/10.1038/s41598-024-53942-z (2024).

Acknowledgements
This research was funded by the Deanship of Scientific Research and Libraries at Princess Nourah bint Abdulrahman University, through the Research Funding Program, Grant No. (FRP-10-1445).

Author contributions
Conceptualization, M.A.M. and M.A.; Methodology, M.A.M.; Software, L.J., M.A., and N.A.; Validation, M.A. and M.A.M.; Resources, M.A.M., M.A, L.J., and N.A.; Data curation, M.A, N.A., M.A.M, and L.J.; Formal analysis, N.A., M.A., and L.J.; Investigation, M.A.M.; Project administration, L.J.; Supervision, L.J. and N.A.; Visualization, M.A.; Writing—original draft, M.A.M, and M.A.; Writing—review & editing, L.J., N.A. and M.A.M. All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Informed consent
Not applicable.

Additional information
Correspondence and requests for materials should be addressed to L.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.