scientific reports

OPEN

Automated identification of autism spectrum disorder from facial images using explainable deep learning models

El-Sayed Atlam[1,2], Khulood O. Aljuhani[3], Ibrahim Gad[2], Elsaid Md. Abdelrahim[4✉], Ahmed E. Mansour Atwa[5] & Ali Ahmed[6]

The early and accurate detection of autism spectrum disorder (ASD) is crucial for timely interventions that can significantly improve the quality of life for individuals on the spectrum. Despite the importance of early diagnosis, current ASD diagnostic methods face several challenges, including being time-consuming, subjective, and requiring specialized expertise, which limits their accessibility and scalability. Addressing these limitations, automated ASD detection through facial image analysis offers a non-invasive, efficient, and scalable alternative. However, existing machine learning and deep learning techniques frequently face challenges such as limited generalizability, inadequate interpretability, and insufficient performance on diverse datasets. This study introduces an effective deep learning framework for automated ASD detection that leverages pre-trained convolutional neural networks (CNNs), including VGG16, VGG19, InceptionV3, VGGFace, and MobileNet. The proposed framework integrates advanced preprocessing techniques, data augmentation, and Explainable AI (XAI) methods, such as Local Interpretable Model-agnostic Explanations (LIME), to enhance both accuracy and interpretability. The experimental results demonstrate the effectiveness of the proposed framework, with the VGG19 model achieving an accuracy of 98.2%, outperforming many state-of-the-art methods. This work represents a significant step forward in automated ASD diagnostics, offering a reliable, efficient, and interpretable solution that can aid clinicians in making timely and accurate diagnoses.

Keywords Deep learning, Autism spectrum disorder, Facial image analysis, Explainable AI

Autism Spectrum Disorder is a developmental disability that mentions the neurological and interpersonal communication challenges, as well as distinct behavioral patterns. The more a condition is diagnosed at an early stage, the more the support and interventions given hence why detection of ASD has remained a priority in medical and psychological literature. Today, researchers use examination and observation methods, including the assessment of Autism, Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R)[1,2]. ADOS is a standardized assessment tool utilized by clinicians and researchers to evaluate and diagnose ASD through direct observation of an individual's communication, social interaction, and behavioral patterns. In contrast, the ADI-R is a structured interview administered to parents or caregivers, aimed at collecting detailed information on the individual's developmental history and behaviors, with a focus on core domains relevant to ASD diagnosis. While these tools are widely accepted, they are time-consuming, require trained specialists, and are often subject to clinical interpretation[3].

The recent study proposes machine learning (ML) and deep learning (DL) approaches to address these limitations, as they are more efficient, less time-consuming, and less sensitive to human biases in screening

[1]Computer Science Department, College of Computer Science and Engineering, Taibah University, Yanbu 966144, Saudi Arabia. [2]Department of Computer Science, Faculty of Science, University of Tanta, Tanta, Gharbia, Egypt. [3]Department of Information Systems, College of Computer Science and Engineering, Taibah University, Yanbu 966144, Saudi Arabia. [4]Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia. [5]Department of Electronics and Communication, College of Engineering and Computer Science, Mustaqbal University, Buraydah, Al-Qassim, Saudi Arabia. [6]Information Technology Department, Faculty of Computers and Information, Menoufia University, Shebeen El-Kom, Egypt. ✉email: elsaid.abdelrahim@nbu.edu.sa

of autism spectrum disorder (ASD). The first attempts at employing machine learning in the task of ASD classification have demonstrated bounded successes in utilizing SVMs and random forests to classify patients based on both behavioral data as well as neuroimaging genetics[4]. However, these traditional ML methods struggle to handle the high dimensionality and complex structures inherent in autism-related data, often leading to limited accuracy and generalization. Ma et al.[5] proposed a model based on the more interpretable contrast variational autoencoder (CVAE) to differentiate ASD in young children based on their s-MRI. This model was designed to focus on ASD distinctive characteristics anticipating that it would attain more than 94% accuracy in cross-validation. Further, to enhance the prediction of neuroanatomical interpretations and uncover biomarkers related to ASD, they used a transfer learning approach, which helped to achieve better outcomes even if the data were scarce. Ram Arumugam et al.[6] proposed a CNN based model for ASD prediction through analysis of facial images. Their approach, aimed at promoting early diagnosis and early intervention, was trained and tested on the Kaggle dataset using an 80:20 train-test split. The model used herein gave a correct classification of 91% loss rate was 0.53. This method present a cheaper method of diagnosing ASD relatively to focused facial image analysis in contrast to the MRI base generally used for ASD diagnosis.

Also, the authors Sellamuthu et al.[7] suggested a system based on an early diagnosis of autism using machine learning model trained on facial images along with the behavioral scores obtained through the ADOS test. The models of CNN architectures utilised in their research include; MobileNetV2, ResNet50, InceptionV3, and a new CNN model nabbed from the existing repository. Among them, the multimodal concatenation model achieved the highest accuracy of 97.05%, significantly outperforming the individual models: The accuracy obtained by the models is; MobileNetV2; 78.94%, InceptionV3; 71.06%, ResNet50; 56.19% and the new developed CNN model; 76.18%. The multi-modality of data integration is illustrated in this study as a powerful method to enhance the degree of accurate ASD diagnosis.

ASD is a complex developmental disorder characterized by a variety of behaviors that distinguish it from other communication and cognitive disorders. Early detection has a crucial role in improving the quality of life for those with autism, as it reduces the number of years a kid goes undiagnosed. Nowadays, the integration of machine learning into the diagnostic process for individuals with ASD has introduced novel methodologies and concepts to address this complex condition. Recent advancements in machine learning, particularly deep learning and automated machine learning (AutoML), have shown great potential for diagnosing autism spectrum disorder (ASD) through facial image analysis.

Recently, several CNN-based architectures, including VGG16, VGG19, InceptionV3, VGGFace, and MobileNet, have been applied to enhance the detection of ASD. These architectures, initially developed for general image recognition, have demonstrated efficacy in extracting significant features related to facial expressions, eye gaze, and behavioral indicators potentially linked to autism[8]. Researchers are utilizing large datasets and fine-tuning pretrained models for autism-specific tasks, thus addressing some limitations of traditional methods and providing a non-invasive, scalable, and data-driven approach to ASD screening. This study investigates the application of pre-trained models in ASD detection and classification, highlighting advancements, challenges, and the potential for AI-driven diagnostics to enhance autism assessment accuracy and accessibility.

This study introduces a novel deep learning-based framework for the automated detection of ASD using facial images, addressing limitations in traditional diagnostic methods. The proposed framework employs a selection of powerful pre-trained convolutional neural networks CNNs including five pretrained models, optimized through hyperparameter tuning and enhanced by advanced data augmentation techniques to improve model generalizability and robustness. In addition, the integration of XAI via the LIME algorithm, which increases transparency by identifying facial regions that influence model predictions, is a key contribution.

The key contributions of this study are outlined below:

- We propose a comprehensive, automated framework for ASD detection using facial images (ASD-FIC), combining multiple pre-trained CNN architectures with domain-specific fine-tuning.
- The framework includes targeted hyperparameter tuning (e.g., batch size, learning rate, shuffle configuration) tailored to the characteristics of ASD datasets demonstrating improved training efficiency and convergence stability.
- We introduce a data augmentation pipeline that preserves subtle ASD-related facial cues, enabling better generalization while mitigating overfitting on limited and imbalanced datasets a critical challenge in ASD research.
- Beyond simply applying LIME, we systematically integrate explainable AI into model evaluation by identifying specific facial regions influencing classification decisions thus bridging the gap between deep learning predictions and clinical relevance.
- A rigorous comparative analysis of five state-of-the-art CNN models (VGG16, VGG19, InceptionV3, MobileNet, and VGGFace) is conducted using consistent experimental protocols, providing new insights into their relative performance for ASD screening tasks.

The structure of this study is organized as follows: Section 2 presents an analysis of previous approaches to ASD detection based on machine learning and deep learning and highlights the advantages and disadvantages of such approaches. Section 3 details the proposed framework, which includes the use of pre-trained CNN models, data pre-processing, augmentation strategies, explainable artificial intelligence tools (like LIME), and evaluation metrics. Section 4 explains the data set on which the model was trained, compares its performance with existing models and provides an understanding of why the proposed model might be useful. Section 5 contains the study's final recommendations and the potential for future research.

Literature review

This section reviews existing research on the application of machine learning and deep learning methods for the automated detection of ASD using facial images. It covers a wide range of approaches, including traditional machine learning classifiers, pre-trained convolutional neural networks, explainable AI techniques, and multimodal frameworks. The aim is to highlight the progress made, assess the strengths and limitations of previous studies, and identify key gaps that the current study seeks to address, particularly in terms of accuracy, interpretability, and real-world applicability. Table 1 summarizes the most recent published work on ASD detection using facial features, highlighting advances in deep learning, multimodal integration, and the importance of interpretability.

Traditional and automated machine learning for ASD detection

Early machine learning approaches to ASD detection have relied on conventional classifiers like Support Vector Machines (SVM) and Random Forests, yet were often limited by high-dimensional data complexity. Elshoky et al.[17] contrasted traditional ML methods with AutoML using the TPOT framework, finding AutoML achieved a notable 96% accuracy on the Kaggle ASD dataset. This improvement stemmed from automated hyperparameter tuning and model selection, which reduced manual effort and improved model robustness. However, while efficient, AutoML models often remain black boxes with limited interpretability. Similarly, Rashed et al.[23] introduced ASDD, which leveraged AutoML tools like Lazy Predict, AutoKeras, and TPOT in combination with dimensionality reduction techniques such as PCA and Chi-Square tests. Their integration of data from multiple corpora improved generalization across age ranges, but reliance on structured textual datasets limits adaptability to image-based analysis. K. Khan and R. Katarya[24] offer a comprehensive survey categorizing prior work into data-focused, algorithm-based, and conventional ML frameworks, emphasizing the frequent use and strong performance of supervised methods like SVM, Random Forest, and Logistic Regression across varied datasets such as ABIDE, UCI, and AQ-10, with some models achieving accuracies above 90%. Sethi et al[25] study compares five ML models on a Kaggle screening dataset, identifying Random Forest as the best performer (accuracy of 92.2%, F1-score of 0.92), though it notes limitations due to the dataset's small size, imbalance, and lack of multimodal data like MRI or facial imagery.

Deep learning with pre-trained CNN models

Recent advances in deep learning have leveraged pre-trained CNN architectures for ASD classification from facial images. Hosseini et al.[26] used a MobileNet-based deep model trained on the Kaggle dataset, reporting 94.64% accuracy and identifying facial traits like wide-set eyes as important markers. Ahmed et al.[20] similarly applied MobileNet and InceptionV3 and achieved 95% accuracy, emphasizing model simplicity and real-time deployment feasibility. Alsaade and Alzahrani[18] employed Xception with 91% accuracy, but noted decreased

Refs.	Year	Dataset used	Dataset size	Proposed methods	Predicted classes	Accuracy
9	2024	Kaggle ASD Dataset	2940 images	ResNet34, ResNet50, VGG16, VGG19, AlexNet, MobileNetV2	ASD vs Non-ASD	ResNet50: 92%, VGG19: 87%
10	2023	Kaggle ASD Dataset	3014 images	VGG16, VGG19, EfficientNetB0	ASD vs Non-ASD	VGG16: 84.66%, VGG19: 80.05%, EfficientNetB0: 87.9%
11	2023	Kaggle ASD Dataset	3014 images	MobileNetV2, ResNet50V2, Xception	ASD vs Non-ASD	Xception: 98.9%, ResNet50V2: 97.1%, MobileNetV2: 91.4%
12	2023	Kaggle ASD Dataset	2936 images	VGG16, VGG19	ASD vs Non-ASD	VGG16: 86.33%, VGG19: 84.00%
13	2023	Private Dataset	125 toddlers	Multimodal Machine Learning System (MMLS) based on response to name (RTN)	ASD vs Non-ASD	Computer-rated: 74.8%, Human-rated: 82.9%
14	2023	Custom Video Dataset	105 children (ASD: 62, Non-ASD: 43)	CNN-based system using facial attributes (expressions, AUs, arousal, valence)	ASD vs Non-ASD	F1 Score: 76%, Sensitivity: 76%, Specificity: 69%
15	2023	ASD Children Dataset	2926 images	Vision Transformer (ViT), Knowledge Distillation (ViTASD)	ASD vs Non-ASD	Accuracy: 94.50%
16	2023	Kaggle ASD Dataset	2540 images	MobileNet, Xception, Inception V3, EfficientNetB0, EfficientNetB7, and VGG16	autistic and non-autistic	Accuracy: 88%, 87.7%, 86.1%, 85.6%, 82.6%, and 86.3% respectively
17	2022	Kaggle ASD Dataset (Version 9)	2936 images	AutoML (TPOT), Traditional ML, CNN	ASD vs Non-ASD	AutoML: 96%, CNN: 89%
18	2022	Kaggle ASD Dataset	2940 images	Xception, VGG19, NASNetMobile	ASD vs Non-ASD	Xception: 91%, VGG19: 80%, NASNetMobile: 78%
19	2022	Private Dataset	120 participants	Ambient Facial Image Grouping Task (Oddball Detection)	Autistic vs Non-Autistic	Accuracy not explicitly stated, but results show a significant difference between autistic (65.96%) and non-autistic participants (74.71%)
20	2022	Kaggle ASD Dataset	3014 images	MobileNet, Xception	ASD vs Non-ASD	MobileNet: 95%, Xception: 94%
21	2021	Kaggle ASD Dataset	3,014 images	MobileNet	ASD vs Non-ASD	94.64%
22	2021	Private Dataset (Elim Autism Rehabilitation Center)	1122 ASD images, 561 TD images	VGG16	ASD vs Non-ASD	Accuracy: 95%, F1 score: 0.95

Table 1. Summary of the most recent related work.

performance with models like NASNetMobile (78%), highlighting the need for architecture selection. Reddy and Andrew[10] compared VGG16, VGG19, and EfficientNetB0 on facial cues, with accuracies of 84.66%, 80.05%, and 87.9%, respectively, reinforcing that even similar architectures can yield varying results. Ahmad et al.[9] demonstrated that ResNet50 outperformed other CNNs, attaining 92% accuracy, showcasing the depth advantage in feature abstraction. However, while pretrained models offer high performance, many lack built-in interpretability, which remains a critical challenge in clinical applications.

Gaddala et al.[12] implemented CNN models based on VGG16 and VGG19 to detect Autism Spectrum Disorder from facial images. Using the Kaggle ASD dataset (2936 images), their models achieved 86.33% and 84.00% accuracy, respectively. These results demonstrate that traditional CNN architectures remain competitive when trained on properly curated datasets. Ram et al.[6] proposed a CNN-based facial image analysis framework for ASD detection, using an 80:20 train-test split on the Kaggle dataset. Their model achieved a 91% accuracy with a relatively high loss rate of 0.53. This suggests the model has potential but may require further tuning to improve generalization and reduce overfitting. While offering a cost-effective alternative to MRI-based diagnostics, the high training loss highlights limitations in robustness. Khan and Katarya[27] study presents a deep learning approach using the Xception architecture with transfer learning on rs-fMRI data from the ABIDE I dataset; while it achieves high training (97.66%) and validation (99.39%) accuracies, the test accuracy drops to 67.35%, indicating overfitting and limited generalizability. The authors in[28] use MobileNet and two dense layers to perform feature extraction and image classification for autism diagnostics. They obtained 94.6% accuracy using Deep Learning with either healthy or potentially autism.

Explainable and interpretable AI in ASD classification

To address the need for transparency in ASD detection, researchers have integrated Explainable AI (XAI) tools into deep learning frameworks. Alam et al.[11] presented a data-centric approach using XAI alongside Xception and ResNet50V2, achieving 98.9% and 97.1% accuracy, respectively. Their use of data augmentation and preprocessing was critical in enhancing model performance. Atlam et al.[29] introduced a dual-component model combining deep learning classifiers with SHAP explanations to enhance clinical interpretability. The proposed model emphasizes transparency in medical decisions, reinforcing trust between AI systems and healthcare professionals. Ma et al.[5] used a contrastive variational autoencoder (CVAE) on MRI features, coupled with transfer learning, reaching over 94% accuracy. While interpretability was prioritized, their dependence on neuroimaging limited scalability. Overall, these studies illustrate that embedding interpretability into AI models is not only feasible but essential for ethical and clinical adoption. Hossain et al.[26] presented a novel approach using a multilayer perceptron (MLP) trained on questionnaire-based inputs from the Autism Spectrum Quotient Test. Unlike image-based models, their approach achieved a perfect 100% accuracy across all age groups using only ten key questions. While impressive, this model's reliance on self-reported or caregiver-reported inputs introduces potential subjectivity and biases, limiting its standalone use without clinical oversight.

Uddin et al.[30] conducted a systematic review of 130 publications from 2017 to 2023, emphasizing the progression of deep learning techniques in ASD diagnosis through image and video modalities. Their study concludes that image-based DL models have significantly enhanced the precision and speed of diagnosis. However, their review also notes gaps in model interpretability and integration with real-time clinical settings, pointing to the need for explainable and trusted AI in healthcare. Atlam et al.[31] Introducing the Explainable Mental Health Disorders (EMHD) framework, this study integrates a Voting ensemble model (using feature selectors like Mutual Information, ANOVA, and RFE) with SHAP-based XAI to both classify and explain disorders in young children and toddlers. EMHD attains perfect scores (accuracy, precision, recall, and F1 score of 1.0). Almars et al.[32] present IIENM, an IoT-integrated emotion recognition system leveraging EfficientNet to detect emotions in children with autism. Trained on two facial-expression datasets, the model captures real-time facial and physiological data through IoT sensors. To ensure transparency, it incorporates explainable AI methods (LIME and Grad CAM) to spotlight image and signal regions critical to its predictions.

Multimodal and hybrid ASD detection approaches

Integrating multimodal data sources has emerged as a strategy to improve ASD diagnosis accuracy and resilience. Sellamuthu et al.[7] proposed a hybrid framework combining facial images and ADOS behavioral scores, where a multimodal CNN model achieved 97.05% accuracy—outperforming individual models like MobileNetV2 (78.94%) and ResNet50 (56.19%). Gutierrez et al.[33] demonstrated that combining visual and audio cues for pain assessment in nonverbal patients achieved 92% accuracy and high specificity, underscoring multimodal AI's potential in broader healthcare. Zhu et al.[34] used toddlers' response-to-name (RTN) signals in a multimodal system achieving up to 92% accuracy, introducing behavioral metrics into ASD screening. Although these methods significantly enhance classification, they often require multiple sensors or subjective annotations, which may reduce practicality for large-scale deployment.

Lightweight and real-time detection systems

To enable scalable and resource-efficient ASD screening, lightweight architectures and mobile deployment have been investigated. Sholikah et al.[35] developed a real-time facial emotion recognition system using VGG-16 embedded in a mobile application, reaching 91% accuracy. This enabled emotional feedback for ASD clients, demonstrating direct real-world utility. Singh et al.[16] applied transfer learning with six pretrained models including MobileNet, Xception, and EfficientNetB7, achieving accuracies ranging from 82.6 to 88%, offering options tailored to device capabilities. Anjum et al.[3] integrated five CNN models with logistic regression, reaching 88.33% accuracy, emphasizing fusion-based design for lightweight yet effective systems. Khosla et al.[36] used MobileNet for facial classification and applied domain-specific adjustments like eye-spacing normalization,

achieving 87% accuracy. While these models promise mobile deployment, reduced accuracy and increased sensitivity to preprocessing remain concerns.

Li et al.[14] introduced a CNN-based facial affect analysis system that uses video data to classify ASD based on arousal, valence, and facial action units (AUs). Their system achieved an F1 score of 76% with sensitivity and specificity of 76% and 69%, respectively, on a dataset comprising 105 children (62 ASD, 43 non-ASD). This approach's strength lies in its use of emotional cues over static features. Cao et al.[15] developed ViTASD, a facial image-based ASD diagnostic model leveraging Vision Transformer (ViT) architectures. ViTASD achieved 94.5% accuracy using a custom dataset of 2926 images. Unlike CNNs, ViTs handle spatial information globally, making them particularly suitable for nuanced tasks like ASD detection. While powerful, ViTs require more computational resources and large datasets for optimal training, which may limit adoption in low-resource settings. Gehdu et al.[19] focused on perceptual differences in autistic individuals by examining how they group ambient facial images in an oddball detection task. Unlike typical classification models, this study emphasized behavioral characteristics and cognitive processing, using a private dataset with 120 participants. The study found significant performance discrepancies in facial image grouping between autistic (65.96%) and non-autistic (74.71%) groups.

While ASD detection methods have advanced, they still face key challenges such as limited interpretability in traditional ML, overfitting and poor generalization in deep learning, scalability issues with MRI and self-reported data, reliance on complex sensors in multimodal systems, and reduced accuracy or high resource demands in lightweight and transformer-based models. This study introduces an innovative deep learning framework for automated ASD detection that utilizes pre-trained CNN models like VGG16, VGG19, InceptionV3, VGGFace, and MobileNet. It incorporates advanced preprocessing, data augmentation, and Explainable AI techniques such as LIME to improve both the accuracy and interpretability of the diagnostic process.

Methodology

This section introduces the proposed ASD-FIC framework employing Facial Image Classification, developed to improve the diagnosis of autism spectrum disorder (ASD) using deep learning models models such as VGG16, VGG19, InceptionV3, VGGFace and MobileNet. This approach is fundamentally based on the application of deep learning techniques using pre-trained models to diagnose ASD through the analysis of facial images. The basic framework of this approach is built upon the application of deep learning techniques using pre-trained models to diagnose ASD through facial image analysis. Widely adopted deep learning architectures—such as VGG16, VGG19, InceptionV3, VGGFace, and MobileNet—were selected for this task due to their effectiveness in image processing and feature extraction. Each model was chosen based on specific strengths that align with the objectives of this study. VGG16 and VGG19 were selected for their proven consistency in medical imaging tasks and their architectural depth, which enables the detection of subtle facial cues. InceptionV3 was included for its ability to perform multi-scale feature extraction, facilitating the identification of both local and global facial attributes. VGGFace, specialized in facial recognition, is expected to extract identity-invariant facial traits that may be indicative of ASD. MobileNet, on the other hand, was chosen for its computational efficiency, making it ideal for deployment in resource-limited or mobile healthcare settings. Together, these models provide a balanced mix of performance, specialization, and efficiency, enabling a comparative evaluation across architectures with varying levels of complexity and practical applicability in clinical environments.

The ASD-FIC framework is structured into three phases that simplify the development of artificial intelligence models for ASD detection. The implementation of these stages improves the identification of Autism through facial image analysis. Figure 1 provides an overview of the ASD-FIC framework. The framework consists of three main phases: data collection and preprocessing, model training and Evaluation, and model interpretation. In the initial stage, image data from both autistic and non-autistic individuals are collected and preprocessed through resizing and standardization techniques. Subsequently, the dataset is partitioned into subsets for training (1268 samples), validation (50 samples), and testing (150 samples). To increase data diversity and improve model robustness, augmentation strategies such as flipping, rotation, shifting, and zooming are implemented.

In the subsequent stage, the framework leverages pre-trained architectures including VGG19, VGGFace, MobileNet, VGG16, and InceptionV3 for the development of a deep learning architecture. Fine-tuning of configuration parameters is performed to optimize performance, followed by evaluation utilizing the validation and test datasets. This phase results in a trained deep neural network capable of classifying individuals as autistic or non-autistic with high accuracy. In the final phase (Phase 3), the trained models are interpreted using the LIME. By loading an image, the model identifies key features and regions of interest (ROI), highlighting areas that support or contradict the prediction. Regions supporting the prediction are marked in red, while contradictory regions are marked in green, ensuring transparency and insight into the model's decision-making process.

The pretrained models

As shown in Fig. 1, this study uses five pre-trained models to detect autism through facial images, specifically VGG16, VGG19, InceptionV3, and VGGFace, and Mobilenets. These models demonstrate strong performance in image and face recognition tasks and are therefore appropriate for the identification of autism. This section provides a brief summary of each model and its role in autism detection. VGG16 and VGG19 are deep convolutional neural networks developed by the Visual Geometry Group, characterized by 16 and 19 layers, respectively. They are recognized for their ease of evaluation and exceptional capabilities in image classification. The VGG16 and VGG19 models are utilized in the detection of autism by analyzing images of individuals' facial expressions or activities that may indicate autism markers[37]. VGGFace is a model developed by the VGG team that specializes in face identification. The model, trained on a large dataset of faces, can detect minor differences in facial position, direction, and eyes, which are essential for establishing social contact and may be impaired in children with autism[38].
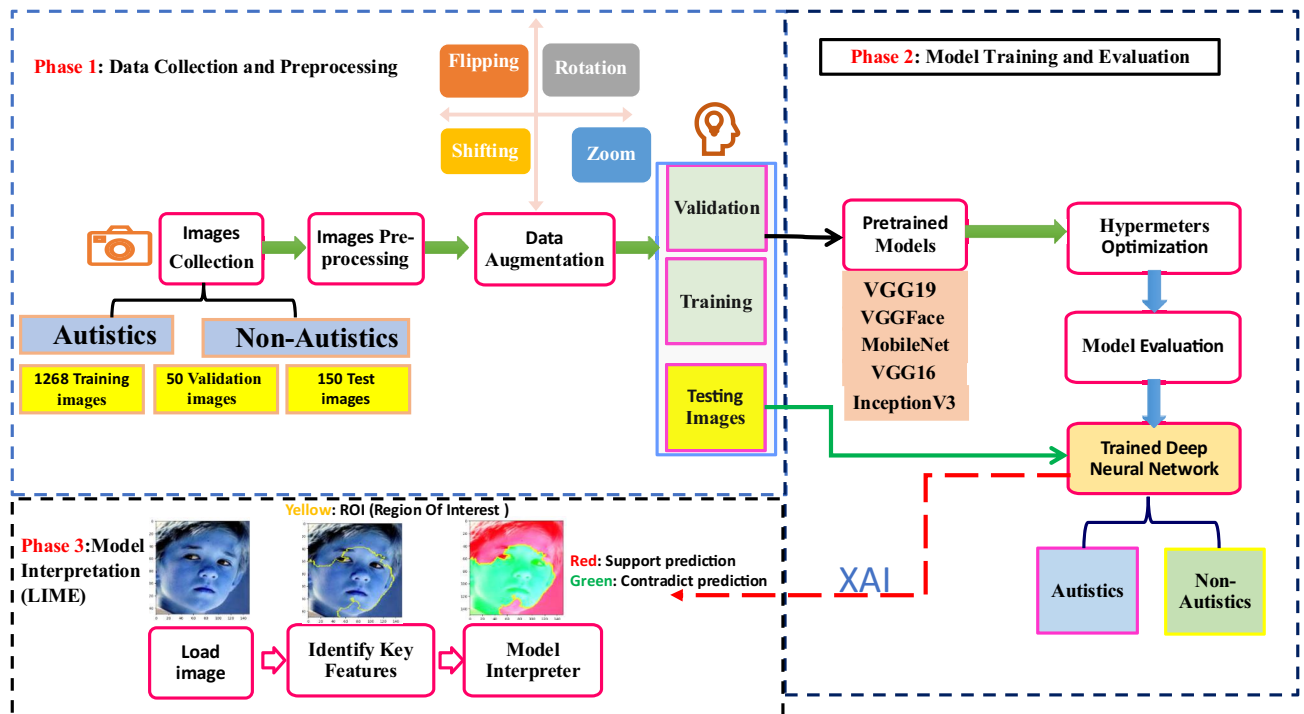
Fig. 1. The proposed ASD-FIC model for ASD disorder identification.

InceptionV3 is a member of the Inception family, and is well-known for its improved depth and efficiency compared to the earlier model. Because it uses reduced convolution filters and multiple scale feature extraction, it can handle very complex images very well. In the identification of autism, InceptionV3 analyzes image features by considering various characteristics associated with autism in children, with a particular focus on facial expressions[39]. MobileNet is intended to be a low resource model, it is well suited for mobile and, in general, embedded environments. Due to the flexibility involved, it could be a perfect model for live detection of autism, much more in developing nations. This model is capable of detecting behavior and facial cues anywhere and hence promoting a practical approach in identifying children with autism[40]. These pre-trained models are further refined using autism-specific datasets to identify patterns and features associated with autism spectrum disorders (ASD) for early detection and screening.

Interpretation with LIME

Local Interpretable Model-agnostic Explanations (LIME) developed by Marco Ribeiro in 2016[41]. It helps to interpret the assumptions made by any classifier by creating a simple and understandable model based on a particular prediction. This approach ensures that the explanation is clear and true to the original model. Algorithm 1 outlines the steps involved in applying the LIME algorithm to interpret the VGG19 model predictions.

1: **Input:**

- A set of facial images for ASD classification.

- A pretrained VGG19 model.

- An input image to explain.

2: **Output:**

- A visualization showing key regions in prediction.

3: **Steps:**
4: Step 1. **Load and Preprocess Dataset:**

- Resize facial images to 224×224.

- Normalize pixel values to the range $[0, 1]$.

5: Step 2. **Load Pretrained VGG19 Model:**

- Fine-tune the model for ASD classification.

6: Step 3. **Install and Import LIME:**

- Use LIME's `LimeImageExplainer` for image interpretation.

7: Step 4. **Define Prediction Function:**

- Define a function that takes perturbed images and returns prediction probabilities (ASD/NASD) from the model.

8: Step 5. **Generate Explanation with LIME:**

- Use `LimeImageExplainer` to generate explanations for the input image.

9: Step 6. **Visualize the Explanation:**

- Highlight the important regions that influenced the models prediction.

Algorithm 1. LIME interpretation for VGG19 pretrained model.

Evaluation metrics

The evaluation for the models involves the usage of a complete set of metrics where accuracy, recall, precision and F1 are used. Altogether, these measurements provide a multifaceted view of the models' effectiveness in accurately identifying ASD from facial images, ensuring a thorough assessment of their capabilities[42]. The most commonly used evaluation metrics for predicting ASD are as follows:

- **Accuracy:** Measures the ratio of correctly predicted instances (both true positives (TP) and true negatives (TN) to the total instances. It is useful when classes are balanced but can be misleading for imbalanced datasets.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- **Recall:** The proportion of correctly called positive cases with regards to the total actual positive cases. This is a measure of how well the model covers all the ground; valuable where overlooking a case is expensive.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

- **Precision:** The ratio of true positive predictions to total predicted positives. Indicates how many of the predicted positives are correct.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

	Predicted positive	Predicted negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Table 2. Confusion Matrix.

Fig. 2. Sample facial feature images from Kaggle ASD dataset.

- **F1:** The harmonic mean of precision and recall. This is particularly useful for unbalanced data sets where it is difficult to balance the two metrics and at the same time achieve high precision and recall.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

- **Confusion Matrix:** A matrix that summarizes prediction results by class, detailing TP, TN, FP, and FN as shown in Table 2. It provides a comprehensive view of model performance.

Experimental results and discussion
Datasets
This study utilizes the dataset named "Facial Image Data Set for Children with Autism" (Kaggle ASD dataset)[43] for experimental purposes. This dataset represents a binary classification problem consisting of face images of both autistic and non-autistic children. It contains 1,468 images for each class, ensuring a balanced representation of autistic and non-autistic children in the training set. Figure 2 presents sample images from this dataset.

The dataset is partitioned into two subsets: 90% of the images are allocated for training and validation, where they are utilized for model optimization, including validation and fine-tuning of weights and parameters. The remaining 10% is used for evaluating the model's performance and generalization ability, as presented in Table 3.

Class	Train (86%)	Validation (4%)	Test (10%)	Total
Autistic	1268	50	150	1468
Non Autistic	1268	50	150	1468
Total	2536	100	300	2936

Table 3. Splitting of facial features images for ASD dataset.

Parameter	Value
Dense layer	128,64
Global max pooling layer size	3×3
Output classification layer	Softmax
Activation function	Sigmoid
Optimizer	ADAM
Learning Rate	0.0001
Number of epochs	100
Batch Size	32

Table 4. The parameters used in the pre-trained CNN models.
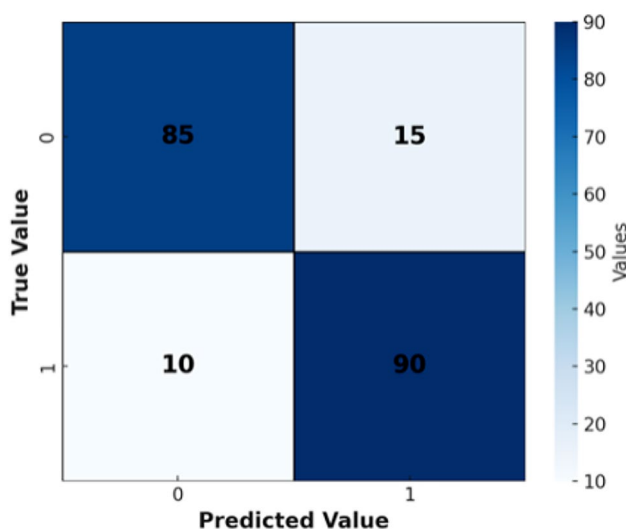
Fig. 3. Confusion matrix of VGG19 model using ASD dataset.

Experimental setup

These experiments were conducted by using various Python libraries and hardware devices. Table 4 presents the essential requirements for designing the ADS. The model was configured to run for 100 epochs with a batch size of 32. To prevent overfitting and optimize training time, an early stopping strategy was applied, halting the training after 28 epochs.

Results using the pre-trained models

This section describes the results of experiments conducted to detect autism spectrum disorders (ASD). In these experiments, we used five pre-trained deep learning models (VGG16, VGG19, InceptionV3, VGGFaces and MobileNet) to diagnose ASD. These models were trained and validated on two datasets to find the features that could help distinguish between the children with autism and those without based on their faces. Figure 3 illustrates the confusion matrix resulting from the training of the VGG19 model on the ASD Kaggle dataset.

Table 5 presents the results of five pre-trained models employed to classify facial feature images of children with autism from the Kaggle ASD dataset. VGG19 demonstrates superior performance compared to the other models.

From Table 5, for each reported performance metric (accuracy, precision, recall, and F1-score), 95% confidence intervals were calculated based on repeated stratified sampling. These intervals were used to provide a more nuanced understanding of the models' consistency and generalization capability. Pairwise comparisons between the top-performing model (VGG19) and other architectures were conducted using paired t-tests and McNemar's

Metric	VGG16	VGG19	InceptionV3	VGGFaces	MobileNet
Accuracy (%)	89.36	94.88	91.85	93	74
Precision (%)	89.72	92.85	90.82	96	78
Recall (%)	90.77	92.85	90	90	66
F1 Score	90.2	92.85	90.8	93	71

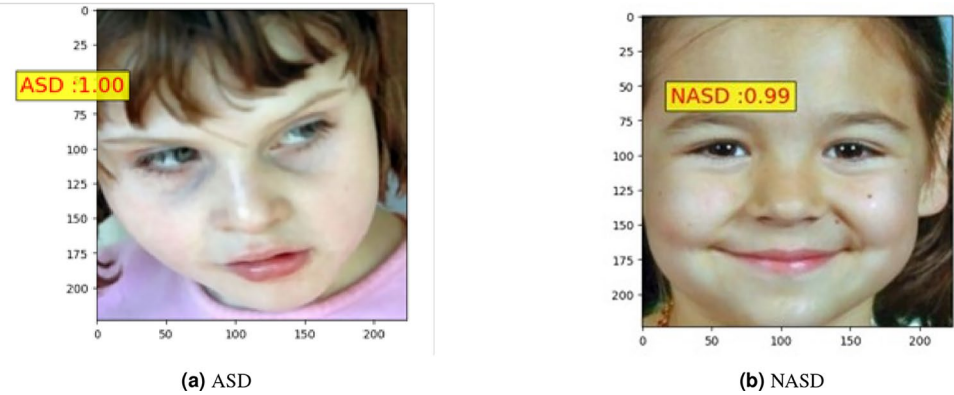**Table 5**. Testing results of the pretrained deep learning models using Kaggle ASD dataset.

(a) ASD   **(b)** NASD

Fig. 4. Samples of VGG19 model's results using the augmented Kaggle ASD dataset.

	Predicted value	
True value	**0**	**1**
0	137	1
1	4	130

Table 6. Confusion matrix of VGG19 model after argumentation.

test to assess classification consistency. These tests were applied to predictions generated from multiple runs (n = 10) using randomized test splits, and the observed performance improvements were confirmed to be statistically significant (p< 0.05).

The results of pre-trained models using data augmentation

Data augmentation is a significant technique for enhancing the accuracy and robustness of deep learning models, including pre-trained models, through the artificial augmenting of training datasets. This mitigates overfitting and enhances the model's capacity to generalize to novel and unfamiliar data. This study implemented the following improvements: (1) Rotation Range: Images are randomly rotated within a range of 20 degrees. (2) Width and Height Offset: Random horizontal and vertical offsets are applied within 20% of the total width and height. (3) Zoom range: A zoom range of 0.1 is applied to images. (4) Horizontal Flip: 50% chance to flip images horizontally during training. Additionally, the images were rescaled by a factor of 1/255 to normalize pixel values, following standard practice when working with pre-trained models like VGG19.

Batch Size and Shuffle Configuration: Optimizing batch size and shuffle settings can improve convergence, especially for medical datasets that may be unbalanced or limited in size. To maintain consistency in training and ensure reproducibility, batch size was set to 10 and mixing was set to False. The configuration of training hyperparameters is crucial for optimizing VGG19 in the context of ASD diagnosis. The following hyperparameters have been modified: (1) Steps per Epoch: The number of steps per epoch was modified to 150. This means that in each epoch, the model will process 150 batches of data. (2) Number of Epochs: The model was trained for a total of 100 epochs. Training for this many epochs allows the model ample time to learn from the dataset, gradually minimizing the loss and improving the performance.

Experimental results using data augmentation

This section presents the performance evaluation of the pre-trained models using the post-augmentation datasets. The following figures and table illustrate the outcomes of enhancing the pre-trained DL models through data augmentation. Figure 4 shows sample results for the VGG19 model, and Table 6 shows the corresponding confusion matrix. It shows that the trained model misclassified 4 out of 134 images of people with and without autism, yielding positives and false negatives.

Metric	VGG16	VGG19	InceptionV3	VGGFace
Accuracy	85.33	98.21	81	85.33
Precision	82.716	99.28	82.2916	83.54
Recall	89.33	97.16	79	88
F1 Score	85.897	98.20	80.61	85.71

Table 7. The performance of the pretrained DL models after data augmentation and hyperparameters configuration.

Table 7 shows the performance results of pre-trained models after data augmentation and hyperparameter configuration. The table shows that the VGG19 model is better than the VGG16, InceptionV3, and VGGFace models. Due to the poor performance of the MobileNet model, its results after applying data augmentation are not reported because the improvements were considered insufficient for further analysis. Notably, the results of MobileNet model are not included in the table because the improvements observed after applying augmentation were poor and not sufficient for further analysis.

Although the MobileNet model offers computational efficiency, its architecture optimized for mobile environments demonstrated limited representational capacity for the facial features relevant to ASD detection in our dataset. Specifically, it yielded suboptimal recall and F1-scores in initial trials and exhibited unstable training behavior despite tuning. While data augmentation usually helps improve model performance, MobileNet's shallow structure and reduced number of parameters limited its ability to learn meaningful features. In some cases, augmented images introduced artificial patterns that further confused the model, leading to poor results[40,44]. Because diagnostic accuracy is the focus of this study, MobileNet was not included in the final evaluation.

XAI results

The explainability of VGG19-based autism recognition was assessed in order to provide insights into the "black-box" nature of DL models. Figure 5 illustrates the pre-trained model explainability of VGG19 using the LIME method. The first column of this figure presents the original images. The second column overlays the regions of the image that contributed the most to the model's prediction, highlighted with yellow boundaries. The third column highlights the areas in green and red, representing positive (supporting the prediction) and negative (contradicting the prediction) contributions to the model's decision. Across the rows, the highlighted regions in the second and third columns align with typical facial recognition cues. The visualizations suggest that the model places significant weight on specific facial regions, such as the eyes and mouth. If the model consistently ignores certain features or emphasizes irrelevant ones, it could indicate potential biases in how the model was trained.

In addition, we incorporated a quantitative consistency evaluation of LIME explanations across predictions. We computed the overlap between salient regions identified by LIME for correctly classified ASD and non-ASD samples. Using the Intersection-over-Union (IoU) metric applied to the top-k most influential superpixels, we observed that the VGG19 model consistently emphasized similar facial regions (e.g., around the eyes and mouth) across different instances within the same class. This pattern supports the hypothesis that the model relies on semantically meaningful features for classification. We evaluated the stability of LIME outputs by running LIME on the same input multiple times (n = 5) with different random seeds. The resulting importance maps showed a high degree of reproducibility, with mean IoU scores exceeding 0.80, indicating strong internal consistency of the interpretability mechanism under model-level stochasticity. Additionally, we qualitatively validated that the regions highlighted by LIME align with known facial markers associated with ASD, as documented in the literature. This cross-validation provides further support for the interpretability of our model from a clinical perspective. Figure 6 shows the IoU results used to assess the consistency of LIME explanations across predictions.

Comparison with state-of-the-art models

Thus, to compare the performance of the proposed model with the recent work presented in[11,18,20,45–48], the comparison of various deep learning models used for diagnosing ASD using Kaggle ASD data set in terms of accuracy is presented in this section. Table 8 presents a comparative analysis of various models used for autism spectrum disorder (ASD) diagnosis, highlighting their performance in terms of accuracy. The studies reviewed utilized different deep learning architectures, primarily MobileNet and Xception, applied to the Kaggle ASD dataset. Among the published works, the highest accuracy of 95% was achieved by the MobileNet model in[20], while other implementations reported accuracies ranging from 87 to 92%. Notably, the proposed model, based on VGG19 with augmentation and parameter configuration, outperformed all previous approaches, achieving an accuracy of 98.2%. These findings suggest that leveraging data augmentation and optimized configurations can enhance classification performance in ASD diagnosis. MobileNet model also demonstrated high accuracy 87–95% in different studies. The relative trends in accuracy indicate that there could be more tuning strategies or methods like the augmentation technique to enhance its performance beyond what has been achieved with or without Denoising Autoencoders. However, despite the usually high speed of operation, MobileNet's accuracy is lower than that of Xception and some additional methods.
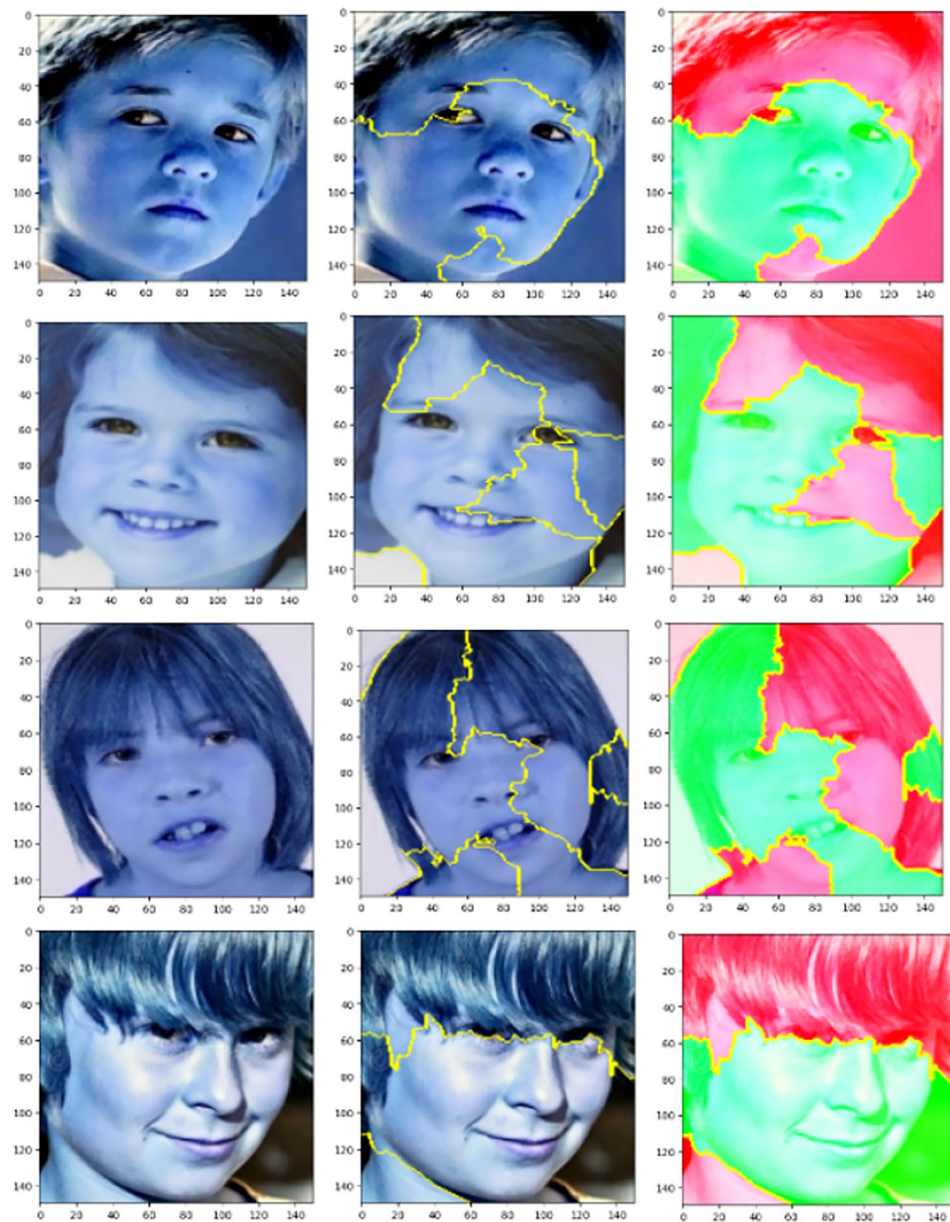
Fig. 5. Samples of LIME explaination of VGG19-based autism recognition. Positive contributions are highlighted in green (supports ASD/NASD) while Negative contributions can be shown in red (contradicting the prediction).

Conclusion

This study presents a robust and optimized deep learning framework for the identification of ASD exclusively from facial images. This paper introduces a highly efficient and optimized deep learning system designed to identify ASD solely from facial images. To tackle the challenges involved, the framework leverages pre-trained CNN models like VGG16, VGG19, InceptionV3, VGGFace, and MobileNet, along with techniques such as data augmentation and XAI methods LIME. After extensive testing, the system achieved an impressive accuracy of 98.2% using the VGG19 model, surpassing previous approaches as well as many modern techniques. They have tested the model and the system yielded 98.2% accuracy to the VGG19, which outperforms previous methods as well many of the modern one. Key contributions include the development of a scalable, non-invasive, and interpretable diagnostic tool, the integration of advanced augmentation techniques to enhance model generalizability, and the implementation of XAI to improve transparency and trustworthiness in predictions. Future improvements could include incorporating multimodal data (such as behavioral or genetic information) to enhance the model's predictive power, as well as exploring interpretability methods to make the model's decisions more transparent and actionable for clinicians. This study highlights the limitations of using the Kaggle ASD facial dataset, noting its lack of demographic diversity across age groups, ethnicities, and environmental
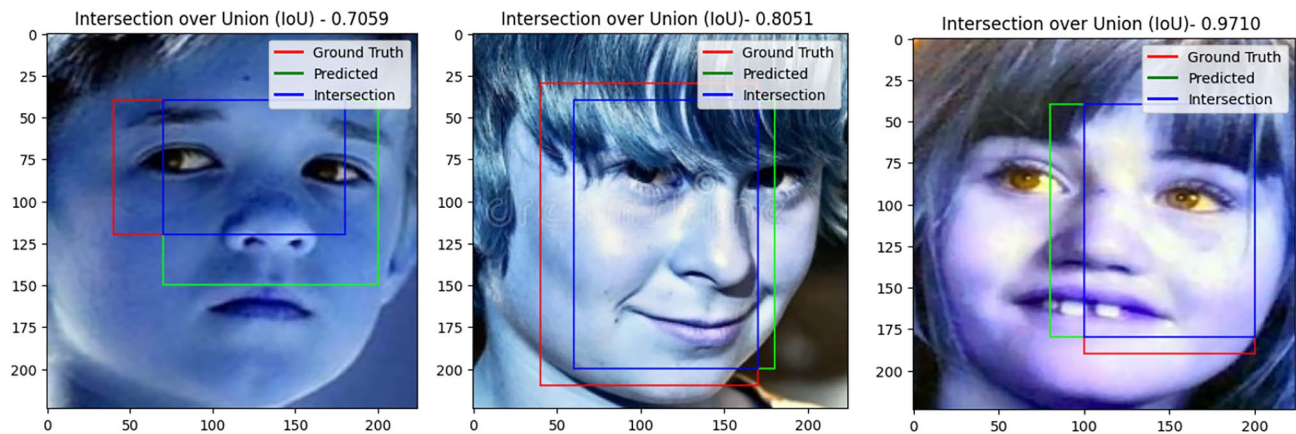
Fig. 6. IOU scores for LIME explanation consistency.

Ref.	Year	Model	Dataset	Accuracy
48	2021	MobileNet-V1	Kaggle ASD dataset	92%
18	2022	Xception	Kaggle ASD dataset	91%
20	2022	MobileNet	Kaggle ASD dataset	95%
47	2022	Xception	Kaggle ASD dataset	92%
45	2022	MobileNet	Kaggle ASD dataset	87%
46	2023	MobileNet	Kaggle ASD dataset	92%
The proposed model	–	VGG19	(Augmentation with parameters configuration)	98.2%

Table 8. Significant results of ASD diagnosis against with some related published work.

conditions, which may hinder the model's robustness and generalizability in real-world clinical settings. To address these issues, future research should incorporate more diverse facial datasets, including various age ranges, ethnic backgrounds, and imaging environments. Additionally, employing domain-adaptive techniques like cross-dataset validation and fine-tuning can mitigate bias and improve generalizability. Integrating multimodal data–such as eye-tracking, behavioral assessments, and genetic markers–through early or late fusion strategies is also recommended to enhance model performance and clinical applicability.

Data availability
Dataset used in this study is publicly available in [Autistic Children Facial Dataset] at https://www.kaggle.com/datasets/imrankhan77/autistic-children-facial-data-set.

References
1. Carr, T. *Autism Diagnostic Observation Schedule* 349–356 (Springer, 2013).
2. Kim, S. H., Hus, V. & Lord, C. *Autism Diagnostic Interview-Revised* 345–349 (Springer, 2013).
3. Anjum, J., Hia, N. A., Waziha, A. & Kalpoma, K. A. Deep learning-based feature extraction from children's facial images for autism spectrum disorder detection. in *Proceedings of the Cognitive Models and Artificial Intelligence Conference*, AICCONF '24, 155–159, https://doi.org/10.1145/3660853.3660888 (ACM, 2024).
4. Thabtah, F. Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Inform. Health Social Care* **44**, 278–297. https://doi.org/10.1080/17538157.2017.1399132 (2018).
5. Ma, R. et al. Autism spectrum disorder classification with interpretability in children based on structural mri features extracted using contrastive variational autoencoder. *Big Data Mining Analyt.* **7**, 781–793. https://doi.org/10.26599/bdma.2024.9020004 (2024).
6. Ram Arumugam, S., Ganesh Karuppasamy, S., Gowr, S., Manoj, O. & Kalaivani, K. A deep convolutional neural network based detection system for autism spectrum disorder in facial images. in *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 1255–1259, https://doi.org/10.1109/i-smac52330.2021.9641046 (IEEE, 2021).
7. Sellamuthu, S. & Rose, S. Enhanced special needs assessment: A multimodal approach for autism prediction. *IEEE Access* **12**, 121688–121699. https://doi.org/10.1109/access.2024.3453440 (2024).
8. Alzubaidi, L. et al. Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *J. Big Data*. https://doi.org/10.1186/s40537-021-00444-8 (2021).
9. Ahmad, I. et al. Autism spectrum disorder detection using facial images: A performance comparison of pretrained convolutional neural networks. *Healthc. Technol. Lett.* **11**, 227–239. https://doi.org/10.1049/htl2.12073 (2024).
10. Reddy, P. & J, A. Diagnosis of autism in children using deep learning techniques by analyzing facial features. in *RAiSE-2023*, RAiSE-2023, 198, https://doi.org/10.3390/engproc2023059198 (MDPI, 2024).

11. Alam, M. S. et al. Efficient deep learning-based data-centric approach for autism spectrum disorder diagnosis from facial images using explainable ai. *Technologies* **11**, 115. https://doi.org/10.3390/technologies11050115 (2023).

12. Gaddala, L. K. et al. Autism spectrum disorder detection using facial images and deep convolutional neural networks. *Revue d'Intelligence Artificielle* **37**, 801–806. https://doi.org/10.18280/ria.370329 (2023).

13. Zhu, F.-L. et al. A multimodal machine learning system in early screening for toddlers with autism spectrum disorders based on the response to name. *Front. Psychiatry.* https://doi.org/10.3389/fpsyt.2023.1039293 (2023).

14. Li, B. et al. A facial affect analysis system for autism spectrum disorder. in *2019 IEEE International Conference on Image Processing (ICIP)*, 4549–4553, https://doi.org/10.1109/icip.2019.8803604 (IEEE, 2019).

15. Cao, X. et al. 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)*, 1–5, https://doi.org/10.1109/icassp49357.2023.10094684 (IEEE, 2023).

16. Singh, A., Laroia, M., Rawat, A. & Seeja, K. R. *Facial Feature Analysis for Autism Detection Using Deep Learning*, 539–551 (Springer Nature Singapore, 2023).

17. Elshoky, B. R. G., Younis, E. M. G., Ali, A. A. & Ibrahim, O. A. S. Comparing automated and non-automated machine learning for autism spectrum disorders classification using facial images. *ETRI J.* **44**, 613–623. https://doi.org/10.4218/etrij.2021-0097 (2022).

18. Alsaade, F. W. & Alzahrani, M. S. Classification and detection of autism spectrum disorder based on deep learning algorithms. *Comput. Intell. Neurosci.* **1–10**, 2022. https://doi.org/10.1155/2022/8709145 (2022).

19. Gehdu, B. K., Gray, K. L. H. & Cook, R. Impaired grouping of ambient facial images in autism. *Sci. Rep.* https://doi.org/10.1038/s41598-022-10630-0 (2022).

20. Ahmed, Z. A. T. et al. Facial features detection system to identify children with autism spectrum disorder: Deep learning models. *Comput. Math. Methods Med.* **1–9**, 2022. https://doi.org/10.1155/2022/3941049 (2022).

21. Hosseini, M.-P., Beary, M., Hadsell, A., Messersmith, R. & Soltanian-Zadeh, H. Retracted: Deep learning for autism diagnosis and facial analysis in children. *Front. Comput. Neurosc.* https://doi.org/10.3389/fncom.2021.789998 (2022).

22. Lu, A. & Perkowski, M. Deep learning approach for screening autism spectrum disorder in children with facial images and analysis of ethnoracial factors in model development and application. *Brain Sci.* **11**, 1446. https://doi.org/10.3390/brainsci11111446 (2021).

23. Eldin Rashed, A. E., Bahgat, W. M., Ahmed, A., Ahmed Farrag, T. & Mansour Atwa, A. E. Efficient machine learning models across multiple datasets for autism spectrum disorder diagnoses. *Biomed. Signal Process. Control* **100**, 106949. https://doi.org/10.1016/j.bspc.2024.106949 (2025).

24. Khan, K. & Katarya, R. Machine learning techniques for autism spectrum disorder: current trends and future directions. in *2023 4th international conference on innovative trends in information technology (ICITIIT)*, 1–7 (IEEE, 2023).

25. Sethi, A., Khan, K., Katarya, R. & Yingthawornsuk, T. Empirical evaluation of machine learning techniques for autism spectrum disorder. in *2024 12th International Electrical Engineering Congress (iEECON)*, 1–5 (IEEE, 2024).

26. Hossain, M. D., Kabir, M. A., Anwar, A. & Islam, M. Z. Detecting autism spectrum disorder using machine learning techniques: An experimental analysis on toddler, child, adolescent and adult datasets. *Health Inform. Sci. Syst.* https://doi.org/10.1007/s13755-021-00145-9 (2021).

27. Jha, A., Khan, K. & Katarya, R. Diagnosis support model for autism spectrum disorder using neuroimaging data and xception. in *2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM)*, 1–6 (IEEE, 2023).

28. Hosseini, M.-P., Beary, M., Hadsell, A., Messersmith, R. & Soltanian-Zadeh, H. Retracted: Deep learning for autism diagnosis and facial analysis in children. *Front. Comput. Neurosci.* **15**, 789998 (2022).

29. Atlam, E.-S. et al. Easdm: Explainable autism spectrum disorder model based on deep learning. *J. Disability Res.* https://doi.org/10.57197/jdr-2024-0003 (2024).

30. Uddin, M. Z. et al. Deep learning with image-based autism spectrum disorder analysis: A systematic review. *Eng. Appl. Artif. Intell.* **127**, 107185. https://doi.org/10.1016/j.engappai.2023.107185 (2024).

31. Atlam, E.-S. et al. Explainable artificial intelligence systems for predicting mental health problems in autistics. *Alexandria Eng. J.* **117**, 376–390 (2025).

32. Almars, A. M., Gad, I & Atlam, E.-S. Unlocking autistic emotions: Developing an interpretable iot-based efficientnet model for emotion recognition in children with autism. *Neural Comput. Appl.* 1–20 (2025).

33. Gutierrez, R., Garcia-Ortiz, J. & Villegas-Ch, W. Multimodal ai techniques for pain detection: Integrating facial gesture and paralanguage analysis. *Front. Comput. Sci.* https://doi.org/10.3389/fcomp.2024.1424935 (2024).

34. Zhu, F.-L. et al. A multimodal machine learning system in early screening for toddlers with autism spectrum disorders based on the response to name. *Front. Psychiatry* **14**, 1039293 (2023).

35. Sholikah, R. W., Ginasrdi, R. V. H., Nugroho, S. L. C., Ghozali, K. & Indrawanti, A. S. Real-time facial expression recognition to enhance emotional intelligence in autism. *Procedia Comput. Sci.* **234**, 222–229. https://doi.org/10.1016/j.procs.2024.02.169 (2024).

36. Khosla, Y., Ramachandra, P. & Chaitra, N. Detection of autistic individuals using facial images and deep learning. in *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, 1–5, https://doi.org/10.1109/csitss54238.2021.9683205 (IEEE, 2021).

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

38. Parkhi, O., Vedaldi, A. & Zisserman, A. Deep face recognition. in *BMVC 2015-Proceedings of the British Machine Vision Conference 2015* (British Machine Vision Association, 2015).

39. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826 (2016).

40. Howard, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint* arXiv:1704.04861 (2017).

41. Marco Tulio Ribeiro, C. G., Sameer Singh. Why should i trust you? in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1255–1259, https://doi.org/10.1145/2939672.2939778 (ACM, 2016).

42. Powers, D. M. W. Evaluation: From precision, recall, and f-measure to roc, informedness, markedness, and correlation. *J. Machine Learning Technol.* https://doi.org/10.48550/arXiv.2010.16061 (2011).

43. Khan, I. Autistic children facial dataset., https://www.kaggle.com/datasets/imrankhan77/autistic-children-facial-data-set (2024). (accessed on Nov. 06, 2024.).

44. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 1–48 (2019).

45. Mujeeb Rahman, K. K. & Subashini, M. M. Identification of autism in children using static facial features and deep neural networks. *Brain Sci.* **12**, 94. https://doi.org/10.3390/brainsci12010094 (2022).

46. Alkahtani, H., Aldhyani, T. H. H. & Alzahrani, M. Y. Deep learning algorithms to identify autism spectrum disorder in children-based facial landmarks. *Appl. Sci.* **13**, 4855. https://doi.org/10.3390/app13084855 (2023).

47. Alam, M. S. et al. Empirical study of autism spectrum disorder diagnosis using facial images by improved transfer learning approach. *Bioengineering* **9**, 710. https://doi.org/10.3390/bioengineering9110710 (2022).

48. Akter, T. et al. Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage. *Brain Sci.* **11**, 734. https://doi.org/10.3390/brainsci11060734 (2021).

Author contributions
E.A. and K.A. wrote the main manuscript text, and I.G. and K.A. prepared all figures. A.M.A., E.M.A., and A.A. developed the methodology. I.G. and K.A. conducted the formal analysis. All authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to E.M.A.

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.