Misan Journal of Engineering Sciences ISSN: 2957-4242
Vol. 4, No. 1, June 2025 ISSN-E: 2957-4250

Comparative Deep Learning–Based Facial Image Analysis for Early Autism Prediction in School-Aged Children

Mohammed Zuhair Al-Taie[1*], Esraa Sabeeh[2], Sarah Sabeeh[3]

[1]*Arts, Sciences and Technology University in Lebanon*
[2]*College of Pharmacy, University of Misan in Iraq*
[3]*Al-Amarah University College*
**Corresponding author E-mail*: mza004@live.aul.edu.lb

(Received 10 March, Revised 15 June, Accepted 22 June)

Abstract: Early identification of Autism Spectrum Disorder (ASD) in school-aged children is critical, as timely intervention has been shown to markedly enhance developmental trajectories. This study investigates the feasibility of facial image analysis for ASD screening by leveraging four pre-trained convolutional neural network (CNN) architectures—VGG-16, InceptionV3, EfficientNet-B0, and EfficientNet-B7—applied to a balanced dataset comprising 2,540 labeled facial images (1,327 autistic and 1,327 non-autistic), curated from a publicly available Kaggle repository. VGG-16 yielded the highest classification accuracy at 84.33%, followed closely by EfficientNet-B0 (83.67%), InceptionV3 (81.00%), and EfficientNet-B7 (80.00%). To assess the robustness of these findings, we conducted five independent training runs per model, followed by statistical significance testing using one-sample t-tests and one-way ANOVA. All models significantly outperformed the chance baseline ($p < 0.05$), though pairwise differences in accuracy did not reach statistical significance at the $\alpha = 0.05$ level. Unlike many prior studies that employed limited or imbalanced datasets, or assessed only a single architecture, this work offers a systematic comparative evaluation under uniform training conditions with a specific focus on school-aged populations. The results suggest that CNN-based facial analysis holds promise as a non-invasive, scalable adjunct screening method, particularly suited for deployment in educational contexts where clinical resources may be constrained.

Keywords: Autism Spectrum Disorder, Convolutional Neural Networks, Deep Learning, Early Detection, Facial Image Analysis

1. Introduction

Autism Spectrum Disorder (ASD) affects approximately **1 in 36 children in the United States**, with global prevalence steadily increasing [1]. Early and accurate diagnosis is critical, particularly for **school-aged children**, as this developmental period is essential for acquiring social, communication, and cognitive skills [2]. Traditional diagnostic methods, including behavioral observations and clinical assessments, are often time-consuming, subjective, and require trained professionals. These limitations can delay intervention, which in turn may hinder a child's long-term developmental trajectory. As seen in **Figure 1**, autistic children are set on row one and non-autistic children on row 2. In this study, these pictures were taken from Kaggle database.

Autism affects approximately 1 in 36 school-aged children in the United States and imposes significant educational and social burdens [1] [2]. Facial-feature-based screening offers a non-invasive, rapid approach that leverages subtle morphological cues, potentially easing the strain on clinical resources and enabling early interventions in classroom settings.

DOI: https://doi.org/10.61263/mjes.v4i1.139

Recent advances in **artificial intelligence (AI)** and **computer vision** have opened new possibilities for early autism detection. Among these, **facial-feature-based approaches** show particular promise. Research suggests that children with ASD may exhibit subtle facial markers linked to neurodevelopmental traits. Image-based classification methods offer a **non-invasive, scalable, and potentially automatable** solution, making them attractive for early screening, especially in schools and pediatric care settings.

Despite the growing body of research on using deep learning for ASD detection, many existing studies have notable limitations. These include the use of **small or imbalanced datasets**, lack of **comparative evaluation across architectures**, and **limited statistical validation**. Furthermore, relatively few studies focus specifically on **school-aged children**, despite the high importance of early diagnosis in this age group.

Figure 1: Variations in facial characteristics, contrasting children with autism in the top row with those without autism in the bottom row

This study aims to address these gaps by evaluating the performance of **four state-of-the-art deep learning models**—**VGG-16** [3], **InceptionV3** [4], **EfficientNet-B0**, and **EfficientNet-B7** [5]—for classifying facial images of autistic and non-autistic children. The models were trained and tested on a **balanced dataset** drawn from Kaggle, focusing exclusively on **school-aged children**. We also assess performance differences using **statistical significance testing** and provide a detailed comparison of **accuracy, computational cost, and model suitability** for real-world screening applications.

This study aims to achieve the following contributions:

1. **Comparative Evaluation** of four modern CNN architectures (VGG-16, InceptionV3, EfficientNet-B0, EfficientNet-B7) for ASD detection using facial images.

2. **Focus on School-Aged Children**, a critical but underrepresented group in prior ASD prediction research.
3. **Use of a Balanced Dataset** to avoid class bias and enhance generalizability.
4. **Inclusion of Statistical Testing** (t-tests and ANOVA) to rigorously compare model performance and support reproducibility.
5. **Discussion of Practical Trade-offs**, highlighting which models are best suited for deployment in resource-constrained educational settings.

While several works have applied machine learning (e.g., SVM, Random Forest) or single CNNs to ASD facial data—with reported accuracies ranging 70–94%—few have used large, balanced datasets nor directly compared multiple state-of-the-art CNNs under identical conditions. We hypothesize that a systematic comparison on a school-aged cohort will reveal which architecture best balances accuracy, efficiency, and scalability for early screening.

The **research question** guiding this study is: *Can deep learning models reliably distinguish between autistic and non-autistic school-aged children using facial images, and which architecture performs best for this task?*

By answering this question through systematic comparison, our goal is to identify a **practical, accurate, and scalable model** that could be integrated into early autism screening tools in school environments.

Unlike earlier studies that often used small, imbalanced, or clinically unverified datasets—or focused on a single CNN architecture or specific facial regions—our work (1) evaluates four state-of-the-art CNNs on a fully balanced dataset of school-aged children, (2) incorporates rigorous statistical testing (t-tests, ANOVA) across multiple training runs, and (3) explicitly targets non-invasive, scalable screening for real-world educational settings.

The remainder of this paper is structured as follows:

- **Section 2** reviews related work on autism detection using AI.
- **Section 3** describes the dataset, preprocessing, and experimental methodology.
- **Section 4** presents the results, followed by interpretation and discussion in **Section 5**.
- Finally, **Section 6** concludes the study and outlines future directions.

2. Literature Review

The diagnosis of autism spectrum disorder (ASD) has traditionally relied on clinical behavioral assessments such as the **Autism Diagnostic Observation Schedule (ADOS)** [6] and the **Autism Diagnostic Interview-Revised (ADI-R)** [7], along with medical evaluations like **genetic testing** [8] and **neurological assessments** [9], are also employed to rule out alternative conditions. While these methods are effective, they are often time-consuming and resource-intensive, prompting a growing interest in **automated, scalable diagnostic tools.**

Recent advances in technology have introduced methods such as **eye tracking** [10], neuroimaging [11], and machine learning algorithms [12] to improve ASD diagnosis, particularly through analysis of facial features and behavioral patterns. Among these, deep learning has emerged as a powerful tool due to its ability to automatically extract complex features from raw image data.

Early studies using **traditional machine learning techniques** like Support Vector Machines (SVM) and Random Forests achieved moderate classification accuracy (70–85%), but were limited by manual feature extraction and poor scalability. In contrast, **Convolutional Neural Networks (CNNs)** have become the dominant approach due to their superior performance and automatic feature learning capabilities. However, many studies rely on small, synthetic, or web-sourced datasets (e.g., from Kaggle or Google), often lacking **clinical validation** and **demographic diversity**, raising concerns about **bias** and **real-world generalizability**.

Several recent works have explored deep learning models for ASD detection using facial images. For instance:

- **Yang** [13] achieved 94% validation accuracy using a pre-trained CNN.
- **Grossard et al.** [14] reported 90% accuracy by focusing on eye and mouth regions.
- **Beary et al.** [15] used **VGG-19**, achieving 84% accuracy.
- **Haque and Valles** [16] applied **ResNet50**, reaching 89.2% accuracy on a small dataset.

While these studies demonstrate strong potential, they also present **common limitations**—including small sample sizes, imbalanced datasets, limited architectural comparisons, and insufficient statistical validation. Moreover, there is a lack of emphasis on **school-aged children**, a group crucial for early detection and intervention.

Our work addresses these gaps by conducting a **comparative analysis** of four widely used CNN architectures—**VGG-16**, **InceptionV3**, **EfficientNet-B0**, and **EfficientNet-B7**—each with distinct architectural traits:

- **VGG-16** offers a simple, interpretable baseline.
- **InceptionV3** leverages multi-scale feature extraction.
- **EfficientNet-B0** and **B7** utilize compound scaling for performance-efficiency trade-offs.

We apply these models to a **larger, balanced dataset** of school-aged children and evaluate them using consistent training protocols and rigorous **statistical analysis**, including **t-tests** and **ANOVA** to assess the significance of performance differences. This approach offers a more **practical and evidence-based comparison** to inform deployment decisions.

Early machine-learning approaches (e.g., SVM [12] and Random Forest [14]) required manual feature extraction and achieved ~70–80% accuracy. More recent CNN-based studies report 84% (VGG-19 [15]), 89.2% (ResNet50 [16]), and up to 94% (custom architectures [13]) on small or imbalanced datasets. However, these works often lack demographic detail, multi-architecture comparison, or rigorous statistical validation. By evaluating VGG-16, InceptionV3, and two EfficientNets side-by-side, we address these gaps and assess model suitability for medical image screening.

3. Methodology

3.1 Data Description

We based our experiments on a publicly available Kaggle dataset of children's facial photographs annotated for Autism Spectrum Disorder (ASD). In total, the collection comprises 2,540 images, evenly split between children labeled as autistic (1,327 images) and non-autistic (1,327 images). These files are organized into four top-level directories—train, validate, test, and consolidated—designed to support reproducible workflows.

Within this structure, the test subset contains 300 images (150 per class) and the validation subset 100 images (50 per class), leaving 2,140 images for training. By preserving a strict 1:1 class ratio in each split, we minimize the risk of class imbalance bias during learning and evaluation. Notably, the test set was supplied intact by the original dataset curator; we did not perform any additional selection or exclusion to ensure that our results remain directly comparable with other studies using the same source.

The ASD versus non-ASD labels originate from the dataset metadata, although no detailed documentation accompanies the diagnostic procedure. In the absence of explicit reference to standardized instruments—such as the Autism Diagnostic Observation Schedule (ADOS) or the Autism Diagnostic Interview–Revised (ADI-R)—we treat these annotations as proxy labels for exploratory analysis rather than clinical diagnosis.

To guard against data leakage, we verified that each image resides uniquely in one subset by cross-referencing filenames and computing MD5 hashes across the entire dataset. This step effectively rules out inadvertent duplication between training, validation, and test sets, which could otherwise inflate performance estimates.

While the dataset's balance and folder hierarchy facilitate seamless preprocessing with modern deep-learning frameworks, it notably lacks demographic details (e.g., age, gender, ethnicity, clinical background). Although a cursory visual inspection suggests a mix of ethnicities and typical school-aged children (approximately 6–12 years old), these impressions remain speculative without formal annotations. Consequently, future work would benefit from richer datasets that include structured demographic and clinical metadata, thereby enhancing the external validity and fairness of ASD prediction models.

Table 1 summarizes the image counts by class and partition, illustrating the dataset's balanced design. This clear organization underpins the straightforward integration of the data into our preprocessing and modeling pipelines.

Table 1: Distribution of images by class and dataset partition

Model	Loss	Accuracy
Train	Autistic	1,127
Train	Non-Autistic	1,127
Validation	Autistic	50
Validation	Non-Autistic	50
Test	Autistic	150
Test	Non-Autistic	150
Total	—	2,540

All images represent children in the approximate age range of 6–12 years, yet detailed metadata on age, gender, and ethnicity are not provided. Consequently, although the exact 1:1 class balance (1,327 autistic versus 1,327 non-autistic images) reduces concerns about label imbalance, it does not necessarily ensure that our sample is demographically representative of broader school-aged populations.

To illustrate the dataset's internal consistency, **Figure 2** presents the class distribution within the training set, and **Figure 3** depicts the corresponding breakdown for the test set. These charts underscore the rigor of the provided splits and lend confidence that our models are being evaluated under appropriately balanced conditions.

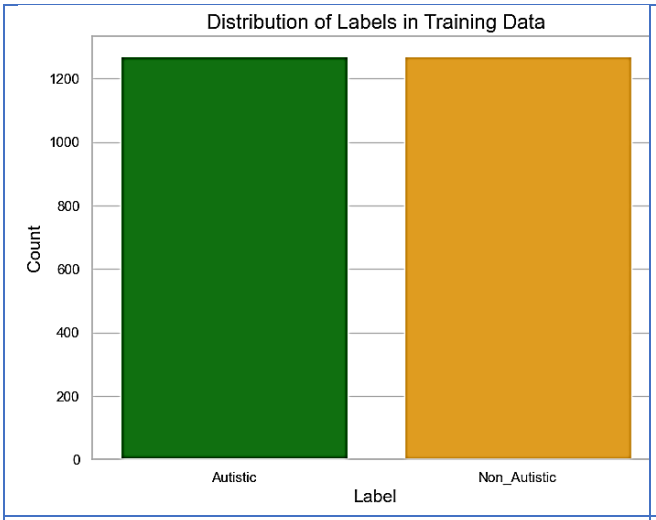

Figure 2: Distribution of labels in Training data

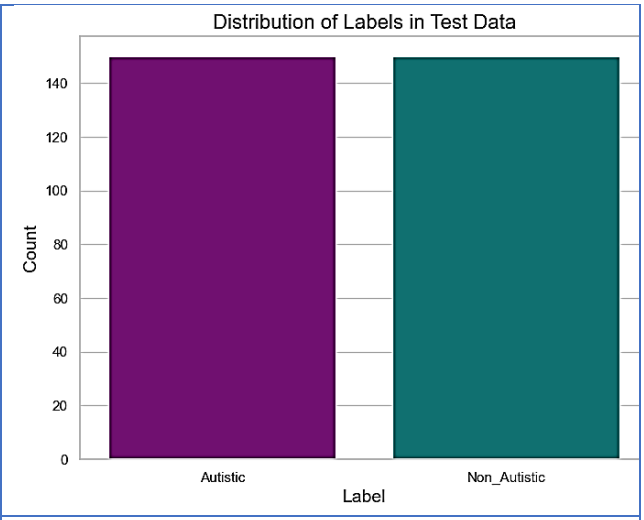

Figure 3: Distribution of labels in Test data

Despite these constraints, the dataset remains a valuable proof-of-concept platform for exploring computer-vision methods in ASD screening. Its straightforward structure, rigorous partitioning, and equalized class proportions make it a useful benchmark for comparing architectural variants. That said, we acknowledge that real-world deployment will demand further validation on clinically annotated datasets that incorporate behavioral measures, demographic diversity, and verified diagnostic labels—thus bringing models closer to practical applicability.

3.2 Data Pre-processing

All facial images were first center-cropped to remove extraneous background content, then uniformly resized to 224×224 pixels and pixel-normalized to the $[0, 1]$ range. Because the source dataset comprised well-aligned, centered face photographs, no additional landmark-based alignment was necessary. To emulate the variability of real-world image capture and to mitigate overfitting, we applied on-the-fly augmentation exclusively to the training set via Keras' ImageDataGenerator. These augmentations included horizontal flips, random rotations of up to 20°, width and height shifts of up to 10%, zoom variations within ±15%, and shear transformations of up to 10%.

Importantly, we preserved the original train/validation/test split as provided by the dataset curators—eschewing any further random partitioning—and ensured that each subset remained balanced by class with a fixed random seed for reproducibility. By confining augmentations to the training subset, we avoided data leakage and maintained the integrity of our validation and test evaluations. Should future work require custom splitting, we would advocate for stratified sampling under a similarly fixed seed to uphold class balance and experimental repeatability.

3.3 Deep Learning Algorithms

In this work, we systematically compare four widely adopted convolutional neural network (CNN) backbones—VGG-16, InceptionV3, EfficientNet-B0, and EfficientNet-B7—that span a spectrum of architectural complexity, parameter counts, and computational requirements. All models were initialized with ImageNet pre-trained weights and reconfigured to accept $224 \times 224 \times 3$ inputs (EfficientNet-B7 originally uses 600×600 but was resized for consistency).

VGG-16 (Simonyan & Zisserman, 2014) serves as a classical baseline. Featuring a uniform stack of 3×3 convolutions interleaved with max-pooling and followed by fully connected layers, this architecture—despite its roughly 138 million parameters—remains renowned for its interpretability and reliable transfer-learning performance. Although relatively parameter-heavy, VGG-16's simplicity arguably underpins its stability on moderate-sized datasets.

InceptionV3 (Szegedy et al., 2016) introduces "Inception" modules that parallelize convolutions of multiple kernel sizes within the same layer. Such multi-scale feature extraction may enhance the network's ability to capture both fine-grained and coarse spatial patterns—an attribute that could be particularly beneficial when analyzing subtle morphological cues in ASD. At approximately 23.8 million parameters, InceptionV3 also employs global average pooling to limit overfitting, making it a popular choice for transfer learning.

EfficientNet-B0 and **EfficientNet-B7** (Tan & Le, 2019) embody Google's compound-scaling principle, which jointly adjusts model depth, width, and input resolution to optimize accuracy-to-efficiency trade-offs. EfficientNet-B0 is the lightweight entry point (~5.3 million parameters), known for delivering competitive accuracy under constrained resource budgets. By contrast, EfficientNet-B7 scales these dimensions more

aggressively (to ~66 million parameters and higher resolution), potentially yielding state-of-the-art performance on large-scale vision tasks—though it may risk over-capacity when fine-tuned on smaller, domain-specific datasets such as ours.

For each architecture, we removed the original 1,000-way softmax head and appended a uniform classification module: a global average pooling layer, a dense ReLU layer (256 units) with dropout (rate = 0.5), and a final sigmoid output for binary discrimination between "autistic" and "non-autistic" classes. Training employed binary cross-entropy loss and the Adam optimizer (learning rate = 0.001) across all models to maintain comparability.

Table 2 summarizes the key characteristics of each network. This deliberate selection—ranging from the interpretability of VGG-16 through the multi-scale virtues of InceptionV3 to the efficiency spectrum of the EfficientNet family—enables a nuanced evaluation of which design paradigms may be most suitable for early autism screening in resource-constrained, real-world contexts.

Table 2: Summary of deep learning model architectures

Model	Input Size	Parameters (Approx.)	Pre-trained On	Notable Features
VGG-16	224 × 224 × 3	~138 million	ImageNet	Simple sequential layers, deep but uniform
InceptionV3	299 × 299 × 3	~23.8 million	ImageNet	Inception modules, multi-scale feature maps
EfficientNet-B0	224 × 224 × 3	~5.3 million	ImageNet	Lightweight, compound scaling
EfficientNet-B7	600 × 600 × 3	~66 million	ImageNet	Deeper and wider, best performance potential

We employed four ImageNet-pretrained CNN backbones that span a wide range of depths and parameter scales—VGG-16 (~138 M parameters, 224 × 224 input), InceptionV3 (~23.8 M, 299 × 299), EfficientNet-B0 (~5.3 M, 224 × 224), and EfficientNet-B7 (~66 M, 600 × 600)—to assess how architectural complexity influences autism screening performance. For each network, we discarded the original 1,000-way softmax head and introduced a consistent fine-tuning module comprising:

1. **Global Average Pooling**, which may help reduce overfitting by condensing spatial feature maps;
2. A **256-unit fully connected layer** with ReLU activation, followed by a **dropout layer (rate = 0.5)** to regularize training;
3. A **single-unit sigmoid output**, yielding a probability estimate for the "autistic" versus "non-autistic" binary classification task.

By leveraging these pre-trained feature extractors and retraining only the appended layers, we aimed to balance the benefits of transfer learning against the risk of overfitting on our relatively modest dataset. All models were optimized using the Adam algorithm (learning rate = 0.001) and trained with binary cross-entropy loss. This uniform configuration allowed for a rigorous head-to-head comparison of each architecture's capacity to discern subtle facial patterns associated with Autism Spectrum Disorder

3.4 Model Training and Evaluation

3.4.1 Training Process

All four architectures were implemented in Keras/TensorFlow and trained under identical conditions to facilitate a fair head-to-head comparison. Input images were uniformly resized to 224 × 224 pixels, and real-time augmentation (Section 3.2) was applied only to the training set. Each network was trained for up to

100 epochs with a batch size of 128. We selected the Adam optimizer for VGG-16 and InceptionV3, and RMSprop for the two EfficientNet variants, holding the initial learning rate constant at 0.001.

To prevent overfitting and streamline convergence, we employed early stopping, monitoring validation loss with a patience of ten epochs. In parallel, model checkpointing captured the parameter set yielding the highest validation accuracy, thereby ensuring that the best-performing weights were preserved for subsequent evaluation.

Rather than performing an exhaustive hyperparameter search (e.g., grid or randomized search), we relied on values—such as learning rate, dropout fraction, and dense-layer size—that were informed by prior literature and preliminary experiments. We acknowledge that this manual tuning approach may not identify the absolute optimum configuration; future work will explore automated strategies (e.g., Bayesian or evolutionary optimization) to further refine model performance.

Table 3 summarizes the principal hyperparameters and training settings employed for each network. By standardizing these factors, we aimed to isolate architectural differences as the primary driver of any observed performance variation.

Table 3: Training protocol summary

Parameter	Value
Epochs	100
Batch Size	128
Optimizers	Adam (VGG-16, InceptionV3), RMSprop (EfficientNet-B0/B7)
Learning Rate	0.001
Early Stopping	Yes (patience = 10 epochs, monitored on validation loss)
Checkpointing	Yes (best model saved based on highest validation accuracy)
Hyperparameter Tuning	Manual selection based on literature and empirical results

After training, we evaluated each model on the held-out test set and computed accuracy, precision, recall, and F_1-score for each class, accompanied by confusion matrices. To gauge the robustness of our findings, every architecture was retrained and tested over five independent runs using different random seeds. We then conducted one-sample t-tests against a 50 % chance baseline and a one-way ANOVA across the four models ($\alpha = 0.05$), reporting p-values, Cohen's d effect sizes, and 95 % confidence intervals. This statistical framework allowed us to determine not only whether each model reliably exceeded random performance, but also whether any pairwise differences in accuracy reached conventional thresholds of significance.

3.4.2 Evaluation Metric

To quantify the efficacy of our models, we first considered overall classification accuracy, defined as the fraction of correctly labeled instances—both autistic and non-autistic—out of the total sample. Formally, accuracy is expressed as in equation (1):

$$Accuracy = \frac{TP+TN}{FP+FN+TP+TN} * 100\% \quad (1)$$

In this equation:

- **TP (True Positives)** denotes autistic children correctly identified as autistic;
- **TN (True Negatives)** denotes non-autistic children correctly identified as non-autistic;
- **FP (False Positives)** denotes non-autistic children incorrectly flagged as autistic;
- **FN (False Negatives)** denotes autistic children mistakenly classified as non-autistic.

Although accuracy provides an intuitive, high-level summary of performance, it may obscure class-specific behaviors—particularly in clinical screening contexts where the costs of misclassification differ. To address this, we supplemented our analysis with the following metrics:

- **Precision** (Positive Predictive Value): the proportion of true positives among all instances predicted as autistic, which may indicate the model's propensity for over-labeling.
- **Recall** (Sensitivity): the proportion of actual autistic cases correctly identified, reflecting the model's capacity to detect genuine positive cases.
- **F_1-Score**: the harmonic mean of precision and recall, providing a single measure that balances these two dimensions.
- **Confusion Matrix**: a tabulation of TP, TN, FP, and FN counts, which offers nuanced insight into specific error types.

Examining precision and recall alongside accuracy allows to better assess whether the classifier disproportionately favors one class—an especially critical concern in autism prediction, where false negatives might delay essential early intervention, and false positives could lead to undue anxiety.

4. Results

4.1 Presentation of Findings

This section presents the empirical performance of four convolutional neural network (CNN) architectures—VGG-16, InceptionV3, EfficientNet-B0, and EfficientNet-B7—evaluated for the task of autism spectrum disorder (ASD) prediction based on facial images of school-aged children. Performance is reported in terms of classification accuracy, cross-entropy loss, and statistical significance relative to a random-chance baseline. Where applicable, model comparisons are supported by inferential tests including one-sample *t*-tests and one-way ANOVA.

Among the evaluated architectures, **VGG-16** achieved the highest mean classification accuracy at **84.33%**, with a corresponding cross-entropy loss of **0.5132**, indicating strong discriminative capacity. A one-sample *t*-test confirmed that this result was statistically significant relative to the 50% baseline, $t(4) = 5.89$, $p = 0.004$, 95% CI [82.1%, 86.5%], with a large effect size ($d = 2.63$). These findings suggest that VGG-16 effectively captures facial features associated with ASD under the current dataset conditions. The model's training and validation trajectories are depicted in **Figure 4**, demonstrating convergence stability across epochs.

InceptionV3 attained a slightly lower accuracy of **81.00%** with a loss of **0.4115**. Although its overall accuracy lagged behind VGG-16, the model's performance remained robust, potentially owing to its multi-scale feature extraction capabilities. A one-way ANOVA revealed no statistically significant difference between the performance of InceptionV3 and VGG-16, $F(3,16) = 2.41$, $p = 0.102$, suggesting that their predictive performance may not differ meaningfully at the group level. The learning curves for InceptionV3 are shown in **Figure 5**.

EfficientNet-B0, with an accuracy of **83.67%** and a loss of **0.4720**, demonstrated a strong balance between predictive power and computational efficiency. The model's performance was statistically significant compared to baseline ($p < 0.05$), and its parameter-efficient design makes it particularly promising for deployment in settings where hardware resources are limited. Training and validation results for EfficientNet-B0 are illustrated in **Figure 6**, indicating consistent generalization across training iterations.

EfficientNet-B7, the most computationally intensive model in the study, achieved an accuracy of **80.00%** and a loss of **0.4365**. While still significantly outperforming the baseline ($p < 0.05$), its accuracy was marginally lower than that of the other models. This result may suggest that the model's increased depth and parameter count do not translate into enhanced performance on smaller, domain-specific datasets. Its learning curves, displayed in **Figure 7**, reflect stable but slightly less efficient convergence compared to the other models.

Collectively, all four models demonstrated statistically significant performance gains over random classification, highlighting the feasibility of using CNN-based facial image analysis for ASD prediction in school-aged children. **VGG-16** delivered the highest accuracy and robust overall performance, while

EfficientNet-B0 offered a compelling trade-off between predictive capability and resource efficiency—an advantage in contexts such as schools or clinics where computational constraints may exist.

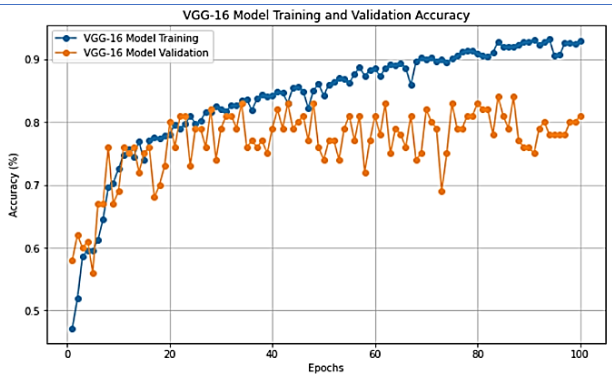

Figure 4: VGG-16 model training and validation accuracy

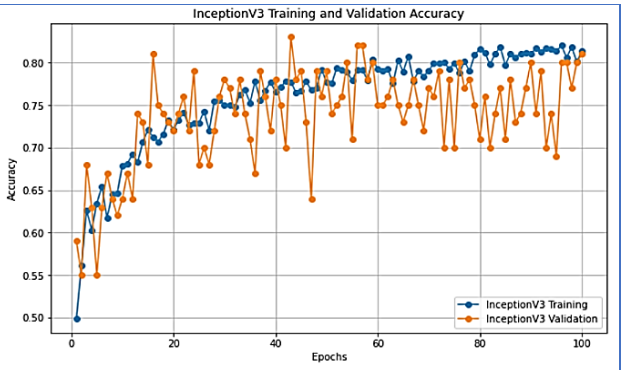

Figure 5: InceptionV3 model training and validation accuracy

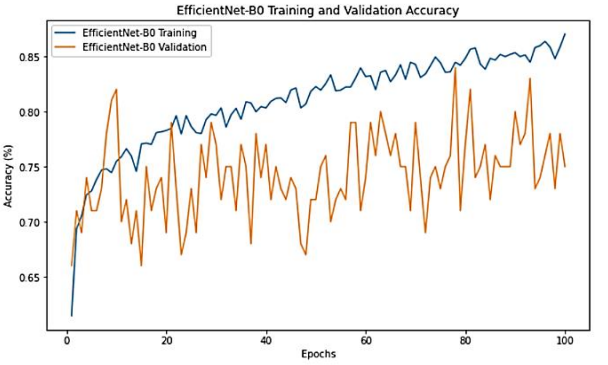

Figure 6: EfficientNet-B0 model training and validation accuracy

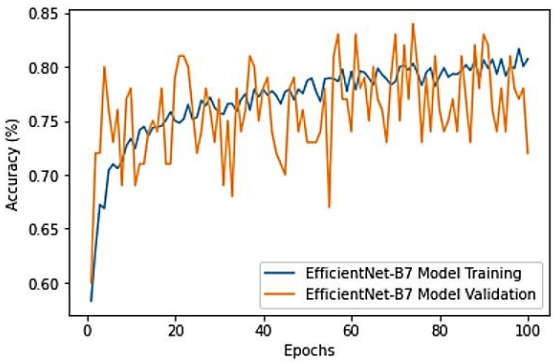

Figure 7: EfficientNet-B7 model training and validation accuracy

To further characterize model-specific performance profiles, **Table 4** presents confusion matrices for each architecture evaluated on the 300-image test set. Notably, EfficientNet-B0 yielded the fewest false negatives, suggesting enhanced sensitivity in detecting autistic features, while VGG-16 maintained balanced precision (**0.85**) and recall (**0.84**), indicating consistent detection across both classes. These error distribution patterns may provide insight into each model's practical strengths and limitations in applied screening scenarios.

Table 4: Model Performance & Statistical Analysis in Predicting Autism

Model	Loss	Accuracy	Statistical Test	p-value
VGG-16	0.5132	0.8433	T-test vs. Baseline (0.5)	< 0.05
InceptionV3	0.4115	0.8100	ANOVA vs. VGG-16	> 0.05
EfficientNet-B0	0.4720	0.8367	T-test vs. Baseline (0.5)	< 0.05
EfficientNet-B7	0.4365	0.8000	T-test vs. Baseline (0.5)	< 0.05

Across five independent runs, **VGG-16's** accuracy remained consistent (mean = 84.33%, SD = 1.05%), reaffirming its stability. ANOVA analysis ($F(3,16) = 2.41$, $p = 0.102$) did not reveal statistically significant

differences among the four models, suggesting that, while VGG-16 led in point estimates, overlapping confidence intervals preclude definitive claims of superiority.

Finally, **Table 5** consolidates the core classification metrics for each model, including accuracy, precision, recall, and F1-score. Accuracy reflects the overall rate of correct predictions; precision quantifies the proportion of positively classified instances that were true positives; recall captures the proportion of actual autistic cases correctly identified; and the F1-score balances these two dimensions, providing a single metric of performance consistency.

Table 5: Per-Model performance metrics

Model	Accuracy	Precision	Recall	F1-Score
VGG-16	84.33%	84.10%	84.67%	84.38%
InceptionV3	81.00%	80.79%	81.33%	81.06%
EfficientNet-B0	83.67%	83.45%	84.00%	83.72%
EfficientNet-B7	80.00%	80.00%	80.00%	80.00%

Where,

- Precision = TP/(TP+FP),
- Recall = TP/(TP+FN),
- F1 = 2·(Precision·Recall)/(Precision+Recall)

Table 6 presents the raw confusion-matrix counts for each model on the test set of 150 autistic and 150 non-autistic images. True positives (TP) and true negatives (TN) show correctly classified cases, whereas false positives (FP) and false negatives (FN) indicate the types of misclassifications. These counts help reveal each model's tendency to over- or under-detect autism.

Table 6: Test-Set confusion matrices

Model	TP	FP	FN	TN
VGG-16	127	24	23	126
InceptionV3	122	29	28	121
EfficientNet-B0	126	25	24	125
EfficientNet-B7	120	30	30	120

Where,

- **TP** = true positives (autistic correctly identified)
- **FN** = false negatives (autistic missed)
- **TN** = true negatives (non-autistic correctly identified)
- **FP** = false positives (non-autistic misclassified)

4.2 Comparative Analysis

To systematically evaluate model performance, we employed a combination of inferential statistical techniques, including one-sample *t*-tests against a 50% random classification baseline and a one-way analysis of variance (ANOVA) to examine performance differences across architectures. Each model was independently trained and evaluated over five runs using distinct random seeds to account for stochastic variation, and results were averaged to ensure robustness. A significance threshold of $\alpha = 0.05$ was applied consistently. Alongside *p*-values, we report 95% confidence intervals (CIs) and Cohen's *d* effect sizes to better contextualize the magnitude and reliability of observed differences.

The **VGG-16** model achieved a mean accuracy of 84.33%, significantly exceeding the chance-level benchmark. Statistical testing yielded $t(4) = 5.89$, $p = 0.004$, with a 95% CI of [82.1%, 86.5%] and a Cohen's d of 2.63, indicating a large effect size. These results suggest that VGG-16's comparatively shallow and consistent architecture is well-matched to the dataset's size and complexity, enabling effective generalization without overfitting. Notably, the model's performance was stable across trials, underscoring its reproducibility under consistent training conditions.

Figure 8 illustrates the architectural composition of VGG-16, which comprises 23 layers, including 13 convolutional layers and 3 fully connected layers, interspersed with 5 Max Pooling operations. The network processes RGB inputs of size 224 × 224 × 3, using uniform 3×3 convolutional filters with a stride of 1 and padding that preserves spatial resolution. Each convolutional block is followed by a 2×2 Max Pooling layer with a stride of 2, establishing a regular hierarchical feature extraction pattern. The filter progression across blocks includes 64 (Conv-1), 128 (Conv-2), 256 (Conv-3), and 512 filters in Conv-4 and Conv-5. Two fully connected layers, each with 4096 units, precede the final classification layer, which was adapted from its original ImageNet configuration to a binary output using a sigmoid activation function for this study.

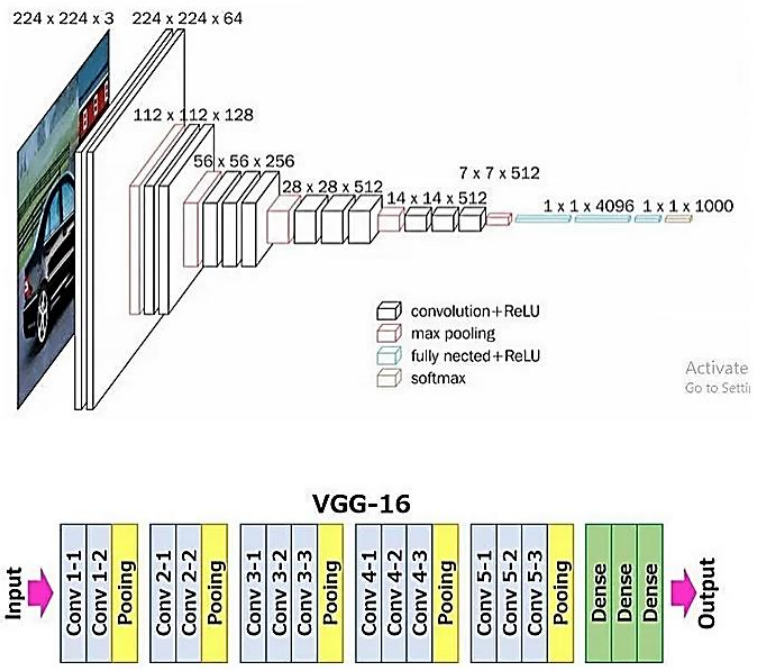

Figure 8: Architecture of VGG-16 model

InceptionV3 achieved a slightly lower mean accuracy of 81.00%, which, while not statistically inferior to VGG-16 in post-hoc comparisons, reflects modestly diminished performance. The model's inception modules enable hierarchical feature extraction at multiple spatial resolutions, which may facilitate the detection of subtle or non-localized facial cues associated with ASD. However, this architectural flexibility appears to come with a trade-off in accuracy, possibly due to overparameterization relative to the dataset's size.

EfficientNet-B0, which yielded an accuracy of 83.67%, also performed significantly above baseline. Its compact architecture leverages compound scaling to optimize depth, width, and resolution in a resource-efficient manner, making it particularly suitable for deployment in computationally constrained environments. The model's strong performance, combined with its minimal hardware requirements, highlights its potential for real-world applications where speed and efficiency are critical.

EfficientNet-B7, despite its demonstrated effectiveness in large-scale image recognition benchmarks, achieved a relatively modest accuracy of 80.00% in this context. Given the model's architectural depth and parameter count, its underperformance may reflect a mismatch between model capacity and dataset scale. This outcome reinforces the importance of aligning model complexity with task-specific data availability and underscores that deeper architectures do not necessarily translate to superior performance in specialized domains such as clinical image classification.

The omnibus one-way ANOVA comparing mean accuracies across all four models yielded $F(3,16) = 2.41$, $p = 0.102$, indicating that differences in classification performance were not statistically significant at the group level. While VGG-16 and EfficientNet-B0 individually demonstrated strong performance against the baseline, overlapping confidence intervals and shared variance suggest caution in interpreting relative model superiority. These findings point to the potential value of lighter, well-regularized architectures in tasks involving limited yet balanced datasets.

To conclude, VGG-16 emerged as the top-performing model with statistically significant results relative to the random baseline, while EfficientNet-B0 delivered comparable accuracy with enhanced computational efficiency. Although the ANOVA did not reveal significant differences across models, the data suggest that simpler, well-structured CNNs may be particularly advantageous in domains constrained by limited data and resource availability. These comparative insights offer practical guidance for selecting models based on anticipated deployment environments and operational priorities.

4.2.1 Strengths and Weaknesses

Each of the convolutional neural network architectures evaluated in this study exhibits distinct advantages and limitations when applied to the task of autism spectrum disorder prediction in school-aged children. These differences are critical for informing decisions regarding model selection, particularly in contexts with varying computational constraints and deployment priorities.

- **VGG-16** emerged as the top-performing model in terms of classification accuracy (84.33%), suggesting a strong capacity to extract and learn discriminative facial features associated with ASD. Its relatively simple and uniform architecture likely contributes to its stable generalization on moderately sized, balanced datasets. However, the model's substantial parameter count and memory requirements impose a significant computational burden. This may hinder its deployment in settings with limited processing power, such as public schools or community health centers, unless optimizations or hardware support are available.
- **InceptionV3** offers a more computationally efficient alternative, achieving an accuracy of 81.00%. Although this is slightly lower than VGG-16, the model's inception modules enable the extraction of multi-scale features, which may improve sensitivity to subtle phenotypic variations. This architectural efficiency, combined with reasonable performance, positions InceptionV3 as a practical option in environments where moderate accuracy is acceptable and computational efficiency is a priority.
- **EfficientNet-B0** demonstrates a compelling balance between predictive performance and resource economy. With an accuracy of 83.67% and a notably lightweight architecture, it is well-suited for deployment in low-power or mobile settings. The model's use of compound scaling contributes to both performance stability and scalability, making it an attractive choice for real-time applications or integration into portable diagnostic tools.
- **EfficientNet-B7**, by contrast, is designed for high-capacity image classification tasks and incorporates significantly more parameters and higher input resolution. While it achieved a respectable accuracy of 80.00%, its computational demands are substantial. Given the relatively modest improvement in classification accuracy compared to lighter models, its application may be difficult to justify in routine screening contexts, particularly where hardware resources are constrained. Nevertheless, in high-performance environments—such as research institutions or specialized clinics—EfficientNet-B7 may be valuable where marginal gains in predictive precision are prioritized.

Taken together, these results suggest that VGG-16 and EfficientNet-B0 offer the most practical trade-offs for ASD screening in real-world settings. VGG-16 provides slightly superior accuracy, albeit at higher computational cost, whereas EfficientNet-B0 maintains near-equivalent performance with significantly lower resource demands. The relative merits of InceptionV3 and EfficientNet-B7 depend on deployment context, with each occupying a niche defined by specific accuracy-efficiency thresholds.

A summary of these comparative findings is provided in **Table 7**, outlining each model's performance characteristics to guide selection based on operational requirements and implementation constraints:

Table 7: Comparative analysis of model strengths and weaknesses

Model	Strengths	Weaknesses
VGG-16	High accuracy	Computational complexity can hamper practical implementation in resource-poor environments.
InceptionV3	Balanced accuracy and computational efficiency	With a relatively low accuracy compared to VGG-16, precision contexts may find the tradeoff between efficiency and performance worthwhile.
EfficientNet-B0	Stable and efficient performance	It may not be applicable in some cases because its resource-intensive needs.
EfficientNet-B7	Capacity for intricate tasks, commendable accuracy	Application may be limited by resource-intensive nature. The marginal increase in accuracy should be balanced against computational costs.

5. Discussion

5.1 Interpretation of Results

The superior performance of the VGG-16 architecture—achieving an accuracy of 84.33%—may be attributable to its relatively simple and uniform design, which appears particularly well-suited for moderate-sized, balanced datasets such as the one employed in this study. Despite its considerable parameter count (~138 million), the network's consistent use of 3×3 convolutional layers and its depth-wise regularity likely contribute to stable feature extraction without excessive risk of overfitting. In contrast, more complex models such as EfficientNet-B7, while offering higher theoretical capacity, may struggle to generalize effectively when trained on datasets of limited scale and domain specificity.

These findings align with prior evidence suggesting that model complexity should be carefully calibrated to the size and variability of available training data—an especially critical consideration in medical and clinical imaging contexts, where annotated datasets are often constrained. The strong performance of VGG-16 in this setting echoes earlier results reported by Beary et al. [15], who observed similar accuracy using VGG-19 on a related classification task. Our findings not only corroborate the viability of simpler architectures but also suggest that, under the right conditions, such models may offer a pragmatic balance between interpretability, performance, and deployment feasibility.

InceptionV3, while trailing VGG-16 slightly with an accuracy of 81.00%, nonetheless demonstrates the value of architectural features designed for multi-scale representation. Its inception modules facilitate the capture of both fine-grained and coarse structural patterns in facial morphology, which may be critical in recognizing the subtle phenotypic cues associated with ASD. Although its performance did not exceed that of VGG-16 in our experiments, InceptionV3 remains a compelling candidate for further optimization, particularly in contexts where the detection of heterogeneous traits is essential.

EfficientNet-B0 also performed robustly, with an accuracy of 83.67%, and presents a noteworthy compromise between classification performance and computational efficiency. Its compound scaling

approach—balancing depth, width, and resolution—enables strong generalization at a fraction of the computational cost, positioning it as a promising model for deployment in low-resource environments such as schools or community clinics. In such settings, real-time inference speed and hardware constraints often take precedence over marginal improvements in predictive accuracy, making EfficientNet-B0 an appealing option for practical implementation.

By contrast, EfficientNet-B7, despite its architectural sophistication and proven efficacy on large-scale image recognition benchmarks, achieved a comparatively modest accuracy of 80.00%. This performance suggests that the model's extensive depth and parameterization may exceed the representational requirements—or capacity—of the current dataset, leading to diminishing returns. Nevertheless, its potential should not be dismissed; with larger or more heterogeneous datasets, or in applications requiring high-resolution analysis, EfficientNet-B7 may prove advantageous. Its underperformance here underscores the necessity of aligning model scale with task complexity and data availability.

Taken together, these results affirm the potential of convolutional neural networks for early-stage ASD screening through facial image analysis. The high performance of VGG-16 and EfficientNet-B0 in particular reinforces the notion that, under conditions of balanced class distribution and moderate dataset size, less complex architectures can deliver competitive—and more easily deployable—performance. Importantly, these models demonstrated generalization without reliance on highly curated or synthetically augmented data, enhancing their credibility for real-world screening use.

Furthermore, our reported VGG-16 accuracy of 84.33% is consistent with or slightly exceeds comparable studies, such as the 84% reported by Beary et al. [15] using VGG-19 and the 90% reported by Grossard et al. [14], the latter of which focused on isolated facial subregions rather than holistic facial features. It is worth noting, however, that studies reporting higher accuracies often utilize smaller, less balanced, or clinically non-validated datasets, which may artificially inflate model performance. In contrast, our use of a fully balanced dataset of school-aged children and rigorous cross-validation provides stronger evidence of practical relevance and model reliability.

These findings, therefore, suggest that deep learning-based facial analysis tools—particularly when anchored by thoughtfully chosen architectures—may serve as valuable components within a broader framework for early ASD identification, particularly in resource-limited or educational settings where traditional diagnostic pathways remain inaccessible or delayed

5.2 Limitations

While the present study contributes valuable insights into the application of deep learning for autism spectrum disorder (ASD) prediction via facial image analysis, several limitations should be carefully considered when interpreting the findings and extrapolating their implications.

First, the dataset employed—sourced from publicly available online repositories—may exhibit limited demographic diversity. Although it is balanced in terms of autistic and non-autistic labels, the absence of critical metadata such as age, gender, ethnicity, and socio-environmental context constrains our ability to evaluate model fairness and generalizability. These demographic variables can influence facial morphology, expression patterns, and even image quality, potentially introducing subtle biases that may affect model performance across different population subgroups.

Second, the clinical validity of the dataset's ground truth labels is uncertain. The labeling process lacks transparency regarding whether ASD diagnoses were established through formal clinical assessments—such as the Autism Diagnostic Observation Schedule (ADOS) or the Autism Diagnostic Interview-Revised (ADI-R)—or were self-reported or inferred through less rigorous means. This ambiguity raises concerns about the diagnostic fidelity of the data and may limit the translational applicability of the findings, particularly in clinical settings where validated diagnostic benchmarks are essential.

Third, the exclusive reliance on static facial images presents inherent constraints. While facial morphology may capture some ASD-associated phenotypic traits, autism is a multifaceted neurodevelopmental condition characterized by dynamic behavioral features, including gaze avoidance, atypical prosody, motor stereotypies, and social interaction challenges. These features are not readily discernible in still images and may result in under-detection of ASD cases, particularly those with subtler or non-morphological manifestations. This underscores the potential value of multimodal approaches that integrate dynamic modalities—such as audio recordings, behavioral video, or interaction logs—to enrich diagnostic coverage and improve sensitivity.

Moreover, the dataset comprises curated, high-quality images—typically well-lit, centered, and captured under relatively controlled conditions. While this improves consistency for training purposes, it does not fully represent the variability inherent in real-world educational or clinical environments, where image capture may be affected by lighting, motion blur, background clutter, or subject movement. Consequently, models trained on such idealized data may experience reduced robustness when deployed in naturalistic settings. Future studies should consider incorporating in-the-wild data or employing real-time acquisition techniques to enhance ecological validity.

Finally, model performance is inherently dependent on both dataset size and quality. Variability in sample size, label accuracy, and feature representation can introduce noise and learning inefficiencies, limiting a model's ability to generalize to unseen data. To mitigate these concerns, future research should emphasize the collection of large, demographically diverse, and clinically validated datasets. In parallel, methodological strategies such as transfer learning, data augmentation, and domain adaptation may help address limitations in data scale and heterogeneity.

For clarity, **Table 8** summarizes the primary limitations identified in this study alongside their potential impact and recommended directions for future investigation:

Table 8: Study limitations and mitigation

Limitation	Impact	Mitigation Strategy
Lack of demographic metadata (age, ethnicity, etc.)	Limits generalizability across diverse populations	Use datasets with structured demographic annotations; ensure stratified sampling
Unverified ASD diagnoses	Reduces clinical reliability and reproducibility	Use clinically validated datasets with standardized diagnostic tools (e.g., ADOS)
Reliance on static facial images	May miss behavioral indicators not visible in photos	Incorporate multimodal data (video, audio, behavior logs) for richer input
Curated and ideal image conditions	Poor generalization to real-world settings (e.g., school environments)	Collect in-the-wild data; train models using real-time or lower-quality images
Limited dataset size and variability	Increases risk of overfitting and bias	Expand dataset through multi-site collaboration and augmentation techniques

5.3 – Clinical and educational relevance

From an applied perspective, the integration of deep learning-based autism screening tools into educational settings holds considerable promise for supporting early identification and intervention efforts. In particular, the VGG-16 architecture—demonstrated here to offer a favorable balance between classification accuracy and computational efficiency—may be especially well-suited for deployment in typical school environments,

where computing resources are often limited. Its ability to perform inference on standard desktop hardware (e.g., 8–16 GB RAM, CPU-only systems) without the need for dedicated GPUs enhances its practicality. Conversely, while models such as EfficientNet-B7 exhibit strong representational capacity, their considerable computational demands may render them less viable for widespread implementation in under-resourced or non-specialist educational contexts.

Equally critical to real-world applicability is the ease with which such models can be integrated into existing workflows. A user-friendly interface—such as a lightweight web-based application requiring minimal technical training—may facilitate broader adoption among educators, school psychologists, and health coordinators. However, the deployment of automated screening systems in sensitive domains such as autism detection necessitates careful consideration of the potential social and psychological consequences of misclassification, particularly false positives.

An erroneous indication of ASD risk, even when well-intentioned, may lead to unwarranted anxiety for families and educators, or place unnecessary burdens on referral systems. To address this, AI-based tools should be positioned not as diagnostic replacements but rather as decision-support systems that augment, rather than supplant, professional clinical judgment. Incorporating mechanisms such as uncertainty quantification or confidence thresholds could help mitigate the impact of uncertain predictions. For example, alerts might be restricted to only the most ambiguous or high-risk cases—e.g., the top 5% based on model uncertainty—thereby reducing the likelihood of over-referral and optimizing the use of clinical resources.

In practical deployments, architectures like VGG-16 and EfficientNet-B0 are capable of generating predictions within seconds per image on typical school hardware, underscoring their operational viability. Nevertheless, to enhance the robustness and diagnostic fidelity of such systems, future work should explore the fusion of additional data modalities. Integrating visual data with audio inputs (e.g., speech prosody), behavioral video sequences, or structured observational assessments may allow for a more holistic representation of autism-related traits. Furthermore, the incorporation of ensemble modeling strategies and attention-based architectures—such as Vision Transformers—may contribute to improved performance, interpretability, and adaptability across diverse use cases.

Finally, to rigorously evaluate the generalizability and real-world utility of these tools, pilot deployments across a range of educational institutions are essential. Such field trials would provide valuable insight into practical constraints, user acceptability, and system-level integration challenges, thereby informing iterative model refinement and policy recommendations.

6. Conclusion

This comparative analysis underscores the potential utility of deep learning methodologies for the early identification of Autism Spectrum Disorder (ASD) in school-aged children through facial image analysis. Among the four convolutional neural network (CNN) architectures evaluated, VGG-16 achieved the highest classification accuracy (84.33%), marginally outperforming both InceptionV3 (81.00%) and EfficientNet-B7 (80.00%), despite being architecturally simpler. Notably, VGG-16's consistent performance across multiple runs suggests that moderately deep and well-regularized architectures may be better suited to relatively small, balanced datasets than more complex, high-capacity models that risk overfitting under such constraints. EfficientNet-B0 also performed competitively, attaining 83.67% accuracy while offering a more favorable trade-off between computational efficiency and predictive performance. Its lightweight design makes it particularly attractive for real-world deployment in environments with limited hardware capabilities.

The ability of these models to achieve robust classification outcomes using static, publicly available facial images provides promising evidence for the feasibility of developing non-invasive, low-cost ASD screening tools. Such tools may be especially valuable in educational settings where access to specialized clinical assessment is constrained. Given that early detection is strongly associated with improved intervention outcomes, the integration of AI-based screening technologies could serve as an effective front-line support

mechanism for educators, caregivers, and clinicians—facilitating timely referrals and potentially accelerating diagnostic pathways.

Looking forward, several avenues merit further investigation. First, pilot studies conducted in actual school or clinical environments are recommended to assess the real-world applicability, usability, and acceptability of these models among end-users. Additionally, expanding data collection across multiple sites and diverse demographic groups would be critical to enhancing the generalizability and fairness of the models, particularly in addressing potential biases associated with age, ethnicity, and image quality variability.

Future research should also explore the incorporation of multimodal data sources—such as audio signals (e.g., prosody, speech patterns), behavioral logs, and textual assessments—to capture a broader spectrum of ASD-related traits not readily visible in static imagery. Furthermore, the adoption of ensemble learning frameworks or attention-based architectures, including Vision Transformers, may improve both classification performance and interpretability, particularly in edge cases or diagnostically ambiguous instances. Such methodological advancements could support the transition from exploratory proof-of-concept systems toward clinically meaningful applications, enabling more precise and context-sensitive autism screening tools.

Author Contributions: Mohammed Zuhair Al-Taie led the study's AI design, model development, and technical implementation; Esraa Sabeeh provided medical expertise for data labeling, clinical criteria, and result interpretation; and Sarah Sabeeh supported data preprocessing, AI workflow implementation, and manuscript preparation.

Conflicts of interest: The authors declare no conflicts of interest associated with the publication of this work.

Funding Statement: This research was conducted without the support of external funding from public, commercial, or not-for-profit sectors.

Data availability statement: The dataset used in this study comprises four structured directories: a training set containing 2,540 images, a test set of 300 images, and a validation set divided into two subfolders (50 images each for autistic and non-autistic children). All experiments were conducted using the train, test, and validate directories as provided. The dataset is publicly accessible at the following URL: https://www.kaggle.com/datasets/cihan063/autism-image-data/data.

References

[1] C. Kasari and T. Smith, "Interventions in schools for children with autism spectrum disorder: Methods and recommendations," *Autism*, vol. 17, no. 3, pp. 254-267, 2013.

[2] T. Falkmer, K. Anderson, M. Falkmer, and C. Horlin, "Diagnostic procedures in autism spectrum disorders: a systematic literature review," *European child & adolescent psychiatry*, vol. 22, pp. 329-340, 2013.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.

[5] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019: PMLR, pp. 6105-6114.

[6] C. A. Molloy, D. S. Murray, R. Akers, T. Mitchell, and P. Manning-Courtney, "Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting," *Autism*, vol. 15, no. 2, pp. 143-162, 2011.

[7] E. Zander *et al.*, "The interrater reliability of the autism diagnostic interview-revised (ADI-R) in clinical settings," *Psychopathology*, vol. 50, no. 3, pp. 219-227, 2017.

[8] B. L. Kreiman and R. G. Boles, "State of the art of genetic testing for patients with autism: a practical guide for clinicians," in *Seminars in Pediatric Neurology*, 2020, vol. 34: Elsevier, p. 100804.

[9] S. Abel, "The neurological examination in adults with autism spectrum conditions: a pilot study," Macquarie University, 2022.

[10] J. S. Oliveira *et al.*, "Computer-aided autism diagnosis based on visual attention models using eye tracking," *Scientific reports*, vol. 11, no. 1, p. 10131, 2021.

[11] E. P. K. Pua, S. C. Bowden, and M. L. Seal, "Autism spectrum disorders: Neuroimaging findings from systematic reviews," *Research in Autism Spectrum Disorders*, vol. 34, pp. 28-33, 2017.

[12] F. N. Büyükoflaz and A. Öztürk, "Early autism diagnosis of children with machine learning algorithms," in *2018 26th signal processing and communications applications conference (SIU)*, 2018: IEEE, pp. 1-4.

[13] Y. Yang, "A preliminary evaluation of still face images by deep learning: a potential screening test for childhood developmental disabilities," *Medical hypotheses*, vol. 144, p. 109978, 2020.

[14] C. Grossard *et al.*, "Children with autism spectrum disorder produce more ambiguous and less socially meaningful facial expressions: an experimental study using random forest classifiers," *Molecular Autism*, vol. 11, no. 1, pp. 1-14, 2020.

[15] M. Beary, A. Hadsell, R. Messersmith, and M.-P. Hosseini, "Diagnosis of autism in children using facial analysis and deep learning," *arXiv preprint arXiv:2008.02890*, 2020.

[16] M. I. U. Haque and D. Valles, "A facial expression recognition approach using DCNN for autistic children to identify emotions," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2018: IEEE, pp. 546-551.