Fine-Tuned Hybrid EfficientNetB2–ViT-B16 Architecture Model on fMRI Brain Data for ASD Classification

Fedia Bahloul[1] , Raouia Mokni[2] and Monji Kherallah[3]

Abstract— **Autism, or autism spectrum disorder (ASD), is a neurodevelopmental condition characterized by atypical development of neurons in the brain. Diagnosing ASD remains a major challenge due to the complexity and heterogeneity of the symptoms associated with this disorder. Early detection is essential, as rapid treatment significantly improves patients' quality of life. However, traditional diagnostic methods are often subjective, time-consuming, and resource-intensive. To defeat these limitations, this study proposes an improved deep learning-based framework for the early detection of ASD from functional magnetic resonance imaging (fMRI) data. For this study, a pretrained Convolutional Neural Network (CNN) model EfficientNetB2 and a pretrained Vision Transformer (ViT) model ViT-B16 have been fine-tuned and then get combined to an hybrid model. All these models will be evaluated its performance on the pre-processed ABIDE dataset. Data augmentation techniques were also implemented to overcome the limitations of the dataset and improve the model's generalization ability. The experimental results show that the fine-tuned EfficientNetB2 model achieves 96% accuracy, while the fine-tuned ViT-B16 model realises 95% and the hybrid model reaches 97%. These results were promising, highlight the potential of transfer learning models optimized for clinical applications, demonstrating their effectiveness to aid in the early and accurate diagnosis of ASD. This study thus contributes to the development of diagnostic systems compatible with the state of the art, paving the way for better management of neurodevelopmental disorders.**

I. INTRODUCTION

ASD is a neurodevelopmental disorder that usually appears in early childhood and persists throughout life. It is mainly characterized by difficulties in communication and social interaction, as well as repetitive behaviors and restricted interests, such as rigid routines, specific and intense interests, or stereotypical movements. These difficulties often encompass deficits in understanding others' intentions and emotions, exhibiting appropriate eye-contact behaviors and facial expressions, and adapting to expected social behaviors [1] [2].

In 2020, the prevalence of ASD per 1,000 children aged 8 years ranged from 23.1 in Maryland to 44.9 in California. The total prevalence of ASD was 27.6 per 1,000 (one in 36) children aged 8 years and was 3.8 times as prevalent among boys as among girls (43.0 versus 11.4) [3]. As technologies become increasingly integrated in daily activities, they offer great opportunities to improve quality of life [4]. However, recent advances in computing and deep learning, particularly transformer-based architectures such as Vision Transformers (ViT) and Swin Transformers [5], have reinforced the importance of brain imaging in autism research. Thanks to machine learning and, more recently, deep learning, researchers are now able to analyze complex fMRI (functional magnetic resonance imaging) data more effectively in order to identify characteristics specific to people with autism and make earlier and more accurate diagnoses. Convolutional Neural Networks (CNN) and ViT are particularly notable for their ability to automatically learn different levels of representation from brain images, surpassing traditional methods.

Despite these advances, the implementation of reliable diagnostic systems remains a challenge, particularly due to the limited size of fMRI databases, their heterogeneity, and significant inter-individual variations. Transfer learning, which involves reusing models pre-trained on large image sets, is a promising approach for improving performance in medical imaging. In this study, we examine the effectiveness of a fine-tuned hybrid model for ASD detection based on cerebral fMRI data. Specifically, we evaluate the pre-trained CNN model EfficientNetB2 merged with fine-tuned transformer ViT-B16 architecture on the pre-processed Autism Brain Imaging Data Exchange (ABIDE) dataset. To optimize performance, we apply data augmentation techniques and implement class balancing strategies, with the goal of improving both classification accuracy and model robustness. The main objective is to demonstrate the contribution of optimized transfer learning approaches to enable reliable and early ASD detection through non-invasive methods, thereby helping to support clinical decision-making and improve patient care.

The main contributions of this work can be summarized as follows:

- Application of data augmentation techniques to the pre-processed and publicly available ABIDE dataset, containing brain fMRI slices divided into two categories: ASD and NON-ASD. Data augmentation was used to expand the dataset, followed by class balancing to ensure an even distribution between the two groups.
- Implementation of the hybrid model that integrates the pretrained EfficientNetB2 CNN model and the pre-trained ViT-B16 model in order to benefit from both

*This work was not supported by any organization

[1] Fedia Bahloul , Advanced Technologies for Environment and Smart Cities Unit (ATES Unit), University of Sfax, Tunisia, Faculty of National School of Engineers of Sfax (ENIS), University of Sfax, Tunisia `bahloul.fedia@gmail.com`

[2] Raouia Mokni, Advanced Technologies for Environment and Smart Cities Unit (ATES Unit), University of Sfax, Tunisia ; Computer Science Department, Higher Institute of Management of Gabes, University of Gabes, Tunisia `raouia.mokni@isggb.u-gabes.tn`

[3] Monji Kherallah, Advanced Technologies for Environment and Smart Cities Unit (ATES Unit), University of Sfax, Tunisia, Faculty of Sciences of Sfax, University of Sfax, Tunisia `monji.kherallah@fss.usf.tn`

local and global representations for improved ASD classification.
- Performance evaluation through extensive experiments conducted to analyze the fine-tuned hybrid model, highlighting the effectiveness of transfer learning in improving detection accuracy and robustness. The results underscore the potential of deep learning approaches as reliable tools to support early ASD diagnosis and advance neuroimaging-based clinical applications.

The rest of the article is organized as follows: Section 2 presents the state of the art, while Section 3 describes the proposed methodology, including the adjusted pre-trained ViT-B16 model. Section 4 presents the experimental results while, section 5 describes thediscussion and comparison with previous work. This will be followed by a conclusion and perspectives for future research.

II. Related Work

In this section, a summary and analysis of research on the detection of ASD using deep learning models were presented.

Nawghare and Prasad [6] introduced an hybrid framework for early ASD diagnosis in children by integrating two data modalities: facial images (ACD Image Dataset) and clinical questionnaire data (ACD Questionnaire Dataset). Visual features are extracted using a pre-trained VGG16 network, while the questionnaire data are analyzed using multiple machine learning algorithms, with Random Forest yielding the best performance. A late fusion strategy is then applied to combine predictions from both modalities through a final decision layer, enhancing system robustness. Experimental results indicate that VGG16 achieves 86.9% accuracy on image data, whereas Random Forest reaches 95.74% on questionnaire data. The proposed hybrid model attains an overall accuracy of 88.34%, surpassing the performance of individual CNN or ML models. Evaluation is conducted using standard metrics, including accuracy, precision, recall, and specificity. These findings highlight the effectiveness of multimodal fusion in improving the reliability of automated ASD diagnosis

In other study, Qiu and Zhai [7] propose an hybrid approach for enhancing ASD diagnosis through the integration of neuroimaging and behavioral data. Utilizing the ABIDE dataset, which includes resting-state fMRI and Social Responsiveness Scale (SRS) scores from 821 participants (379 with ASD and 442 controls), the authors extract both static and dynamic functional connectivity matrices from the fMRI data. These matrices are processed using a CNN enhanced with an attention mechanism to capture deep features, which are subsequently combined with SRS behavioral scores. A support vector machine classifier is then employed for binary classification (ASD vs control). Evaluated via ten-fold cross-validation, the model achieves notable performance metrics: 94.30% accuracy, 92.87% sensitivity, 95.47% specificity, and an F1-score of 94.73%, outperforming several standalone classical and deep learning models. The authors conclude that fusing brain and behavioral data leads to more accurate and robust automated ASD diagnosis.

Alharthi, et al, [8] propose an hybrid framework for ASD diagnosis using neuroimaging data from the ABIDE dataset. Their methodology incorporates both fMRI and structural MRI (sMRI). For sMRI, the authors generate two-dimensional slices from three or four dimensional volumes, followed by augmentation and normalization. They evaluate several ViTs, specifically ConvNeXt, MobileNet, Swin Transformer, and ViT, on these sMRI slices. For fMRI, a 3D CNN is applied directly to the volumetric data. Experimental results indicate that transformer models on sMRI yield a maximum accuracy of approximately 77% with an F1-score near 76%. In contrast, the 3D-CNN applied to the 50 central fMRI slices achieves superior performance, reaching 87% accuracy and an 82.6% F1-score. The study concludes that selecting informative fMRI slices and leveraging functional information are critical for accurate automated ASD diagnosis. Furthermore, the findings suggest that hybrid architectures combining CNNs and transformers can enhance classification robustness.

Parvathy et al,[9] present an automated deep learning framework for ASD diagnosis using structural MRI data. The proposed pipeline begins with preprocessing steps to enhance image quality and remove irrelevant information. For feature extraction, the authors combine a ViT with a Residual DenseNet architecture. Classification between autistic individuals and healthy controls is performed using a Long Short-Term Memory (LSTM) network. To further improve model performance, the parameters are tuned via the Modified Zebra Optimization Algorithm (MZOA). The method is evaluated on the ABIDE dataset, comprising 1,112 MRI scans from 539 ASD patients and 573 control subjects. Experimental results demonstrate an accuracy of approximately 94.64%, indicating strong discriminative capability and highlighting the model's potential as a supportive tool for clinical ASD diagnosis.

While the methods proposed by the authors attain high accuracy and improve the efficiency of the models, the underlying architectures remain comparatively complex. In this context, we propose a simpler alternative based on the use of a fine-tuned hybrid model (EfficientNetB2+ViT-B16), capable of delivering high performance for the classification task.

III. PROPOSED METHODOLOGY

As mentioned in Section 2, numerous approaches have been devoted to diagnosing ASD using deep learning methods. Although some have achieved remarkable performance, they often rely on complex architectures. In contrast, our work is based on a fine-tuned hybrid model (EfficientNetB2+ViT-B16) and aims to improve classification accuracy through targeted adjustments to its architecture. As illustrated in Fig 1, our approach to ASD classification comprises six main steps: (1) data collection, (2) data augmentation, (3) data preprocessing, (4) implementation of the proposed deep learning-based model, and finally

(5) decision-making to classify subjects into two categories (ASD or NON-ASD).

A. Data Description

We used the publicly accessible ABIDE dataset available on the Kaggle platform, dedicated for ASD research. This dataset is part of the ABIDE initiative, which aims to collect fMRI data from individuals diagnosed with ASD as well as healthy control subjects [10]. The database provides preprocessed fMRI data that facilitating the identification and analysis of ASD. It is well organized and includes both data from the ABIDE phase along with a CSV file containing the participants' clinical information.

The dataset is organized as follows:

- Data sources: a main folder containing all the data.
- Preprocessed ABIDE dataset: a dataset designed to simplify analysis.
- ABIDE-Preprocess: folder containing preprocessed fMRI data in NIfTI (.nii) format.
- Abide-data.csv: CSV file describing the diagnosis associated with each participant.

The fMRI volumes in NIfTI (.nii) format were converted into PNG images for use in the classification pipeline.

B. Data Augmentation

Data augmentation is an essential step in order to improve the performance of deep learning models in the classification of ASD. Indeed, ASD-specific datasets are generally limited in size due to the small number of patients available, which can lead to overfitting and poor generalization ability. Data augmentation helps overcome this constraint by artificially increasing the size and diversity of the training set, thereby promoting the learning of more robust features. The initial dataset, sourced from the open platform Kaggle, comprises 539 images with ASD and 572 images of NON-ASD subjects. In order to enrich this data, various image transformations were applied while preserving the essential characteristics related to the diagnosis of autism. The following augmentation techniques were used:

- Rotation: Random rotations between [-20°, 20°] were implemented to simulate different head orientations.
- Translation: Horizontal and vertical displacements of up to 20% were introduced to reproduce slight variations in position.
- Shearing: A shearing transformation was performed to introduce realistic geometric deformations.
- Zoom: A random zoom of ±20% was applied to simulate scale variations.
- Flipping: Horizontal inversions were performed to increase data variability while preserving essential diagnostic features.
- Filling: Empty areas generated by these transformations were filled by interpolating neighboring pixels.

  Thanks to the application of these augmentation techniques, the image database was significantly enriched, reaching a total of 3,773 images for the ASD class and 4,002 images for the NON-ASD class.

C. Data Preprocessing

This section presents the main preprocessing steps applied to prepare the data for model training. First, the images were grouped into batches of 32 samples. then, they were resized to a resolution of (224 × 224) pixels and their pixel values were normalized. To improve generalization ability and the model's performance, the dataset was divided into three distinct subsets: 75% of the data was respectively reserved to training, while 12.5% was dedicated for testing and validation. Finally, the SMOTE method was implemented to balance the dataset so that the ASD and NON-ASD classes each contained 4,002 images.

D. Application of Deep Learning-Based Proposed Model

This study investigates the use of deep learning models for the ASD classification, focusing on a fine-tuned EfficientNetB2 model, the fine-tuned ViT-B16 and the fine-tuned hybrid (EfficientNetB2+ViT-B16) architecture, adjusted and optimized for this task. We adopt a transfer learning and fine-tuning strategy to adapt those models, now referred to as Fine-Tuned-EfficientNetB2 (FT-EfficientNetB2 model), the Fine-Tuned-ViT-B16 (FT-ViT-B16 model) and the Fine-Tuned hybrid model (FT-Hybrid model). The models are retrained on the preprocessed ABIDE dataset to enhance their capacity to learn discriminative features relevant to ASD detection.

1) EfficientNetB2 model: EfficientNetB2 belongs to the EfficientNet family, where network depth, width, and input resolution are jointly scaled to optimize feature extraction. This compound scaling strategy balances accuracy and computational efficiency, an advantage for medical imaging tasks like ASD detection. We retain the original convolutional backbone but replace the classification head. Specifically, global average pooling is substituted with global max pooling, as max pooling emphasizes the most discriminative features, which is beneficial for detecting subtle abnormalities in brain MRI scans. A custom head is then attached, consisting of: (i) a dense layer with 256 neurons, ReLU activation, and L2 regularization; (ii) a dropout layer ($p = 0.4$) to mitigate overfitting; and (iii) a single-neuron output layer with sigmoid activation for binary ASD vs. non-ASD classification. The architecture of the proposed fine-tuned EfficientNetB2 model is illustrated in fig2

2) ViT-B16 model: The fig2 shows the architecture of the FT-ViT-B16 proposed in this study for ASD and NON-ASD classification. The model processes 224 × 224 × 3RGB images through a pre-trained ViT-B16 backbone, which includes patch embedding, multi-head self-attention mechanisms, MLP blocks, and layer normalization within its transformer encoders. To adapt the model while reducing overfitting, we adopt a partial fine-tuning strategy: roughly 50% of the lower transformer layers are frozen, while the upper layers remain trainable. This strikes a balance between preserving general knowledge from pre-training and specializing to the medical domain. At the backbone output, global average pooling aggregates token-wise representations. The resulting vector is passed to a custom classification head

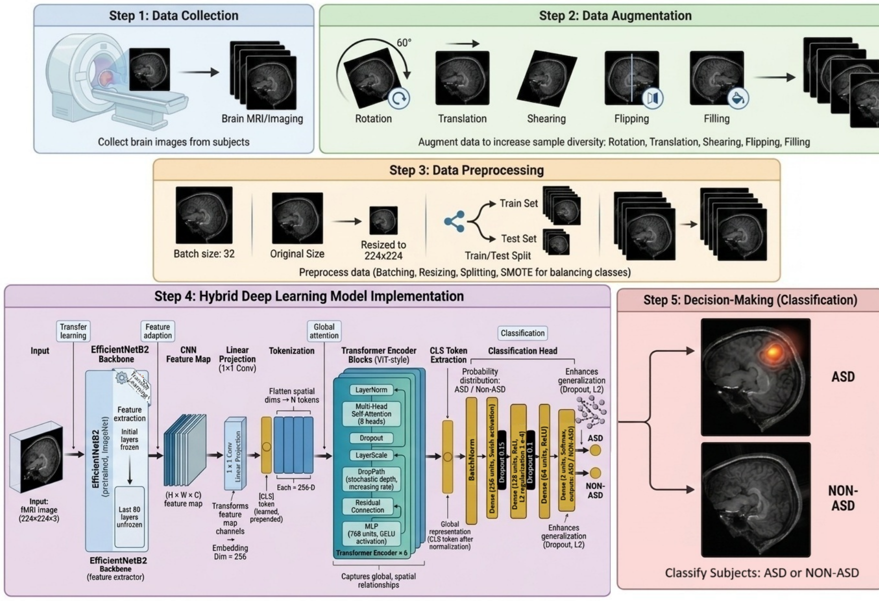


Fig. 1. The Overview of the Proposed Model

consisting of two fully connected layers (256 and 128 neurons) with GELU activation, augmented by L2 regularization, batch normalization, and dropout to improve generalization. A final dense layer with two neurons and Softmax outputs probabilities for the ASD and NON-ASD classes. This fine-tuned design leverages the global representations learned by ViT while enhancing robustness and suitability for clinical classification tasks.

3) Hybrid model: To address small-sample ASD classification from fMRI data which resized to 224×224×3, we introduce a fine-tuned hybrid of EfficientNetB2 and ViT-B16. This design jointly leverages convolutional networks for local feature extraction and transformers for modeling global dependencies. The model uses an ImageNet-pretrained EfficientNet without its top classification layer. Following a transfer learning strategy, we initially freeze all backbone layers, then partially fine-tune only the last layers. This preserves lower layers that extract generic patterns (e.g., contours, textures) while adapting higher representations to the ABIDE dataset. The backbone outputs a feature map of size (H, W, C). A 1 × 1 convolution linearly projects the channels to an embedding dimension of 256 (EMBED-DIM), acting as an affine transformation to match the attention mechanism's required format. The spatial dimensions are then flattened into N tokens of size 256, where each token corresponds to a spatial position, analogous to patches in a standard ViT. A learnable CLS token is prepended to the sequence to aggregate global information across tokens. The transformer consists of 6 blocks (NUM-TRANSFORMER-BLOCKS), each containing: Multi-head self-attention (8 heads), where each head captures distinct subspaces of global dependencies, particularly relevant for complex spatial correlations in medical imaging, an MLP (dimension 768, GELU activation), a LayerNorm before each submodule and residual connections around each to stabilize learning and prevent gradient vanishing, and Regularization includes Dropout, DropPath (stochastic depth with blockwise increasing rate), and LayerScale. After the final transformer block, the CLS token is normalized and extracted as the global image representation. This vector passes through a classification head with dense layers incorporating Batch Normalization, Swish/ReLU activations, Dropout, and L2 regularization. The output layer uses two neurons with Softmax to produce ASD vs NON-ASD probabilities. EfficientNet captures local patterns and fine textures, while the transformer models long-range dependencies. This synergy yields a more discriminative representation than either architecture alone, particularly beneficial for complex medical images like those from the ABIDE dataset.

Fig 2 illustrates the architecture of the hybrid model.

IV. EXPERIMENTAL RESULTS

As part of this study, all data preprocessing and classification steps were performed in Python. Model training was carried out on the Kaggle platform, using GPU-accelerated computing resources to optimize efficiency and processing time. In order to evaluate the contribution of deep learning models to the classification of ASD, a series of experiments was conducted using the proposed models FT-EfficientNetB2 FT-ViT-B16 and the FT-Hybrid. The evaluation of these models was then further refined through the analysis of confusion matrices, allowing predictions to be compared with actual labels, as well as through the study of learning curves, illustrating the evolution of accuracy on the training and validation sets. Table I shows the performance achieved by the proposed models. The FT-EfficientNetB2 model achieved 96% accuracy, with comparable performance in terms of recall, F1 score and precision. The FT-ViT-B16 model attained 95% accuracy, accompanied by strong and stable

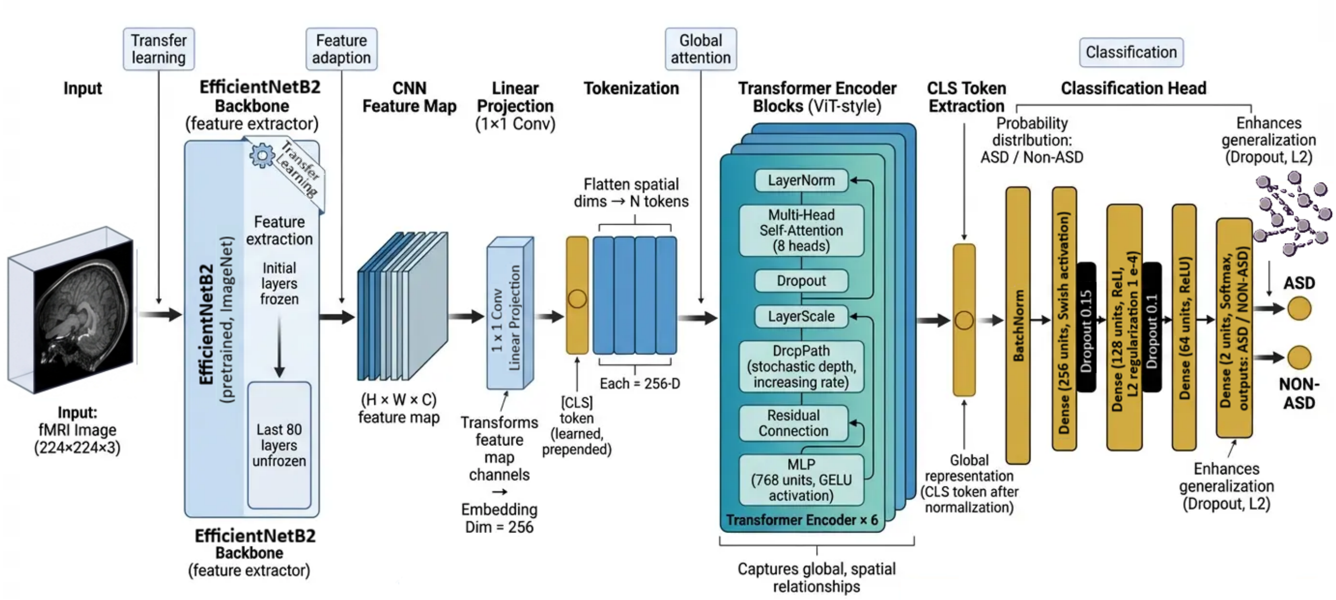


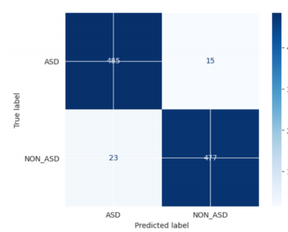
Fig. 2. Hybrid Model Architecture

performance in terms of precision, recall, and F1-score. And the FT-Hybrid model yielded an an accuracy of 97%, while maintaining high and balanced values for recall and F1-score, with precision reaching 98%. These results demonstrate the high reliability of the model in detecting and differentiating between the two classes studied. In these experiments, fMRI images were divided into two classes, ASD and NON-ASD, based on the pre-processed and publicly available ABIDE dataset. The performance of the refined models were then analyzed in detail using confusion matrices, which allowed for comparison of predicted labels with actual labels, as well as learning curves illustrating the evolution of accuracy on the training and validation sets.

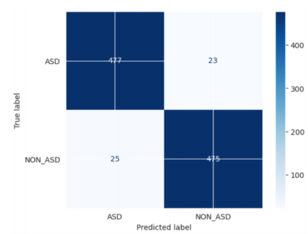
TABLE I
PERFORMANCE ACHIEVED BY THE PROPOSED MODELS

	Accuracy	Precision	Recall	F1 score
FT-EfficientNetB2 Model	96%	96%	96%	96%
FT-ViT-B16 Model	95%	95%	95%	95%
FT-Hybrid Model	97%	98%	97%	97%

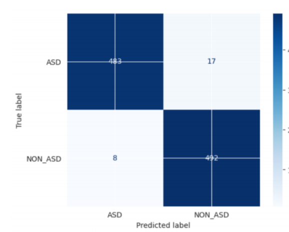
Fig 3 presents the confusion matrix obtained for the 3 trained models. The results indicate that the hybrid model correctly identified 483 ASD images and 497 NON-ASD images, while the FT-EfficientNetB2 model correctly classified 485 ASD images and 477 NON-ASD images and the FT-ViT-B16 model correctly classified 477 ASD images and 475 NON-ASD images. these results demonstrating a strong ability to discriminate between the two categories. Although overall performance was high, a few minor classification errors between classes were nevertheless observed by the 3 proposed models.



Confusion Matrix Obtained by FT-EfficientNetB2 Model



Confusion Matrix Obtained by FT-ViT-B16 Model



Confusion Matrix Obtained by FT-hybrid Model

Fig. 3. Confusion Matrices corresponding to FT-EfficientNetB2, FT-ViT-B16 and FT-Hybrid models

Fig 4 illustrates the evolution of accuracy on the training and validation sets for the proposed 3 models over 100 epochs. A gradual increase followed by stabilization of the 3 curves indicates effective learning and good generalization ability of the model. Conversely, a divergence between the curves characterized by training accuracy that continues to increase while validation accuracy stabilizes may reveal overfitting, indicating that the models are learning features that are too specific to the training data and generalizing less well to new data.

Based on theses experimental results, we can see that the hybrid model, created by combining the EfficientNetB2 and ViT-B16 architectures, outperforms each of the models

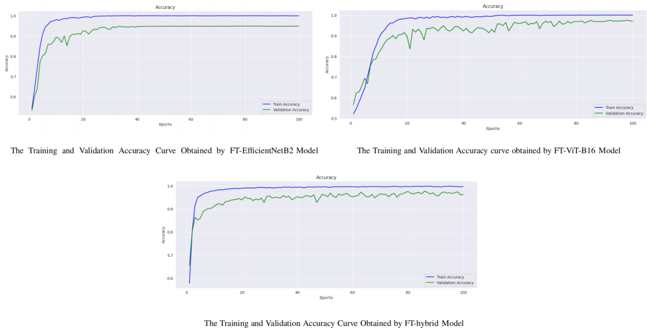


Fig. 4. The Training and Validation Accuracy Curves corresponding to FT-EfficientNetB2, FT-ViT-B16 and FT-Hybrid models

TABLE II
COMPARATIVE ANALYSIS OF PROPOSED WORK WITH THE PREVIOUS STUDIES

Authors	Datasets	Models	Accuracy (%)
Proposed	ABIDE	FT-Hybrid (EfficientNetB2-ViT)	97%
Nawghare, P., & Prasad, J., [6]	ACD Data & ACD Questionnaire Data	Hybrid (VGG16-Random Forest)	88.34%
Qiu, L., & Zhai, J., [7]	ABIDE	Hybrid CNN-SVM	94,3%
Alharthi et al., [8]	ABIDE I	3D-CNN-ViT	77%

individually in terms of accuracy. This improvement confirms that merging the representations extracted by the two models allows for more efficient use of the information present in the images, leading to more accurate and reliable classification of ASD.

V. DISCUSSION AND COMPARATIVE EVALUATION

This section presents a comparison of the performance of the proposed fine-tuned model with that of previous work on the classification of ASD based on the ABIDE dataset. The FT-Hybrid model was evaluated in a binary classification setting (ASD vs NON-ASD), and the corresponding results are summarized in Table II.

The proposed FT-Hybrid model achieved an accuracy of 97%, surpassing various existing approaches, including that of Nawghare, P., & Prasad, J. (2025) [6] with a rate of 88.34%, that of Qiu, L., & Zhai, J., (2024) [7] with 94.3%, and that of Alharthi, et al., (2023) [8] with 77%. These results demonstrate a real significant improvement of up to 20% over previous methods. The model's strong performance stems from a balanced architectural strategy that jointly optimizes network depth, width, and input resolution. This design enables more efficient extraction of discriminative features from fMRI data, enhancing class separation. As a result, the model achieves higher classification accuracy and greater robustness to data variability. These findings underscore the importance of carefully optimized deep architectures for ASD detection.

VI. CONCLUSION AND FUTURE WORK

ASD is a complex neurodevelopmental disorder where early diagnosis enables prompt and appropriate intervention. Recent deep learning advances in medical imaging offer new opportunities to improve diagnostic accuracy using fMRI data. In this study, we proposed three fine-tuned models, FT-EfficientNetB2, FT-ViT-B16, and a FT-Hybrid, combining both architectures, for binary classification of ASD and NON-ASD. The hybrid model achieved the best performance, with 97% accuracy, surpassing the individual models. Complementary metrics (precision, recall, F1-score) confirmed its reliability in distinguishing ASD from control subjects. These results highlight the potential of fine-tuned deep architectures for fMRI-based ASD diagnosis. Nevertheless, certain limitations exist. The dataset size is relatively small. Additionally, simple Federated Averaging may not be optimal for non-IID data variability across clinical sites, potentially affecting system robustness. Despite these constraints, the results support potential clinical application as a diagnostic aid for ASD, offering valuable support for healthcare professionals in interpreting neuroimaging data. To enhance clinical interpretability, future work will incorporate explainable AI methods such as Grad-CAM and attention maps to identify brain regions most influential in classification decisions. We also plan to enrich the image dataset and integrate multimodal neuroimaging data to improve model robustness and generalization.

REFERENCES

[1] Haoues, M., & Mokni, R. (2023). Toward an autism-friendly environment based on mobile apps user feedback analysis using deep learning and machine learning models. PeerJ Computer Science, 9, e1442.
[2] Mokni, R., & Haoues, M. (2020). Deep classifier model for autism spectrum disorder prediction. In Knowledge Innovation Through Intelligent Software Methodologies, Tools and Techniques (pp. 138-149). IOS Press.
[3] Singer, A. (2021). Autism Science Foundation. In Encyclopedia of Autism Spectrum Disorders (pp. 513-515). Cham: Springer International Publishing.
[4] Haoues, M., & Mokni, R. (2024, November). Hybrid Deep Learning Model for Predicting Mental Health States from Digital Media Content. In International Conference on Management of Digital (pp. 159-173). Cham: Springer Nature Switzerland.
[5] Tlili, W., & Mokni, R. (2025, December). ILD-SwinViT: Hybrid Transformer Networks for Interstitial Lung Disease Diagnosis from HRCT Images. In 2025 26th International Arab Conference on Information Technology (ACIT) (pp. 76-84). IEEE.
[6] Nawghare, P., & Prasad, J. (2025). Hybrid CNN and random forest model with late fusion for detection of autism spectrum disorder in Toddlers. MethodsX, 14, 103278.
[7] Qiu, L., & Zhai, J. (2024). A hybrid CNN-SVM model for enhanced autism diagnosis. Plos one, 19(5), e0302236.
[8] Alharthi, A. G., & Alzahrani, S. M. (2023). Multi-slice generation sMRI and fMRI for autism spectrum disorder diagnosis using 3D-CNN and vision transformers. Brain Sciences, 13(11), 1578.
[9] Parvathy, R., Arunachalam, R., Damodaran, S., Al-Razgan, M., & Ali, Y. A. (2025). An intellectual autism spectrum disorder classification framework in healthcare industry using ViT-based adaptive deep learning model. Biomedical Signal Processing and Control, 106, 107737.
[10] Kumarashat, S. (n.d.). ABIDE Preprocessed Dataset [Data set]. Kaggle. Retrieved 2025, from https://www.kaggle.com/datasets/sahilkumarashat/abide preprocessed dataset