Received: 28 September 2025 | Revised: 16 March 2026 | Accepted: 27 April 2026 | Published online: 16 June 2026

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 00(00) 1–14
DOI: 10.47852/bonviewJCCE62027780

BON VIEW PUBLISHING

Robust and Interpretable Deep Learning on EEG Spectrograms for Autism Spectrum Disorder Detection

Andrew Jeyabose[1,2], Arav Chadda[1] and Venkatesh Bhandage[1,*]

[1] *Manipal Institute of Technology, Manipal Academy of Higher Education, India*

[2] *School of Medicine, University of North Carolina at Chapel Hill, USA*

Abstract: Autism spectrum disorder (ASD) manifests early alterations in neural dynamics that often precede observable behavioral symptoms, yet current diagnostic practices rely predominantly on subjective clinical assessments. In this study, we present a robust, image-based convolutional framework for automated ASD detection from resting-state electroencephalography (EEG), validated across two independent cohorts: the King Abdulaziz University ASD dataset and the Autism Centre of Excellence (ACE) dataset. Raw EEG recordings were transformed into time–frequency spectrograms via continuous wavelet transforms and then fed into three deep architectures—a custom 4-layer convolutional neural network (CNN), ResNet50, and EfficientNet—to learn discriminative features of ASD versus neurotypical patterns. To enhance clinical trust and interpretability, we integrated Smooth Grad-CAM++ to generate high-resolution activation maps that pinpoint critical spectral–temporal regions driving each classification. On the King Abdulaziz dataset, our custom 4-layer CNN achieved a test accuracy of 98.59%, with an average F1-score of 0.99, precision of 0.97, recall of 0.98, and specificity of 95.36%. On the ACE dataset, ResNet50 yielded a test accuracy of 95.23%, F1-score of 0.96, precision of 0.97, recall of 0.94, and specificity of 97.43%. Across both cohorts, all models consistently exceeded 95% accuracy and demonstrated balanced sensitivity and specificity, underscoring their generalizability. These results establish a high-performance, explainable computer-aided diagnostic system for early ASD detection using EEG spectrograms. The conjunction of deep feature learning and explainable AI visualization not only accelerates diagnostic workflows but also offers actionable insights into the neural substrates of ASD, paving the way for timely, objective interventions.

Keywords: autism spectrum disorder (ASD), EEG spectrogram analysis, convolutional neural networks (CNN), continuous wavelet transform, explainable AI

1. Introduction

Autism spectrum disorder (ASD), a neurodevelopmental disorder, is characterized by persistent deficits in social communication and interaction, alongside restricted, repetitive patterns of behavior and interests [1]. Globally, ASD affects approximately 1 in 100 children, with reported prevalence rising steadily over recent decades, an increase attributed in part to enhanced public awareness and refined diagnostic criteria [2]. Although the etiology of ASD remains multifactorial, converging evidence implicates both genetic susceptibilities and environmental exposures that disrupt early brain development. Neuropathological investigations have revealed atypical cerebellar structure and connectivity, as well as altered morphology and functional networks within the frontal, temporal, and cortical regions; limbic system abnormalities have also been documented. Moreover, individuals with ASD frequently exhibit increased cortical volume and elevated extra-axial cerebrospinal fluid [3]. Because the most rapid phase of neurodevelopment occurs in the first years of life, these early alterations may precede overt behavioral symptoms by months. Reliance on behavioral assessments alone thus risks substantial delays between the onset of underlying neural changes and clinical diagnosis. Identifying objective biomarkers such as electroencephalography (EEG)-derived spectrographic signatures could therefore enable earlier detection and intervention, improving long-term outcomes for children with ASD [4].

EEG, a method capturing brain electrical activity, noninvasively records the brain's dynamics at millisecond resolution. This makes it indispensable for characterizing neural states and detecting pathological patterns across a range of neurological and psychiatric conditions [5]. Traditionally, experts analyze vast amounts of EEG data to spot anomalies, but this conventional visual inspection of prolonged recordings is inherently time-consuming, subjective, vulnerable to intra- and inter-rater variability, and prone to error [6]. Furthermore, the complex and delicate nature of scalp EEG signals,

***Corresponding author:** Venkatesh Bhandage, Manipal Institute of Technology, Manipal Academy of Higher Education, India. Email: venkatesh.bhandage@manipal.edu

characterized by a low signal-to-noise ratio, confounding artifacts (e.g., muscle activity, eye movements), and overlapping spectral signatures, frequently complicates visual diagnosis. This often results in misinterpretations, leading to both false positives and false negatives in clinical practice. To address these challenges and streamline decision-making, there is a growing imperative to develop computer-aided diagnostic (CAD) systems. By leveraging advanced signal-processing algorithms and deep learning (DL) architectures such as convolutional neural networks (CNNs) applied to EEG spectrograms, these systems aim to better interpret EEG data, automate feature extraction, enhance classification accuracy, and accelerate clinical workflows.

Machine learning (ML) has emerged as a transformative tool in healthcare, enabling the construction of predictive models that can parse complex, high-dimensional data for tasks ranging from disease diagnosis to treatment outcome prediction. Within this paradigm, DL harnesses multilayered neural architectures to automatically learn hierarchical feature representations, yielding marked improvements in accuracy across domains such as medical imaging, genomics, and biomedical signal processing [7]. In the context of EEG analysis, ML pipelines traditionally extract handcrafted features such as spectral power, coherence, or entropy to train classifiers like support vector machines (SVMs). For example, Wadhera and Kakkar [8] demonstrated an SVM classifier using mutual information and average weighted degree features to detect neurological anomalies, while Baygin et al. combined pretrained convolutional feature extractors with an SVM to build a lightweight hybrid model [9]. More recently, researchers have shifted toward end-to-end DL frameworks, applying wavelet transforms or time–frequency representations as CNN inputs to capture subtle temporal and spectral nuances [10].

Despite these advances, existing studies suffer from two critical limitations: (1) reliance on small or single-site datasets, which hampers generalizability; and (2) the "black-box" nature of DL, which restricts clinical trust and adoption. To address these gaps, our study validates a CNN-based spectrogram classification approach across two independent ASD EEG cohorts, thereby assessing model robustness in real-world settings. Furthermore, we integrate explainable AI (XAI) techniques to illuminate the model's decision process: by employing Smooth Grad-CAM++, we generate high-resolution localization maps that identify the time–frequency regions most influential for each ASD prediction.

The primary contributions of this paper are as follows:

EEG acquisition and image generation: We collect resting-state EEG from two distinct datasets and convert raw signals into time–frequency images using continuous wavelet transforms, facilitating broad applicability to future image-based DL models.

Efficient architecture design for clinical deployment: We benchmark multiple CNN architectures, contrasting standard networks with a proposed lightweight 4-layer model. By demonstrating comparable classification accuracy at a fraction of the computational cost, this framework provides a highly efficient, deployable solution tailored for resource-constrained clinical edge devices.

Cross-dataset generalization: Addressing the single-dataset limitation prevalent in existing literature, we validate our approach across two independent cohorts. This novel cross-dataset evaluation demonstrates that the proposed classification pipeline is highly generalizable and robust against dataset-specific artifacts.

Explainable AI and channel importance: We integrate Smooth Grad-CAM++ to generate interpretable activation maps of the model's decision process. The visual results are promising, confirming that certain EEG channels hold significantly more predictive weight than others. While the framework proves it is possible to spatially localize these critical regions, an exhaustive neurophysiological analysis of the specific channels involved remains beyond the scope of this study.

The remainder of this paper is organized as follows: Section 2 reviews the related literature; Section 3 details the datasets and the proposed methodology; Section 4 presents the experimental results and discussion; and Section 5 concludes the study.

2. Literature Review

Computer-aided diagnosis of autism began with applying a wavelet chaos neural network on EEG data by Ghosh-Dastidar et al. [10]. This model is utilized for classifying EEGs into interictal, healthy, and ictal categories. Wavelet analysis decomposes the EEGs into sub-bands by frequency, and three key parameters—largest Lyapunov exponent, correlation dimension, and standard deviation—are used to represent the signal. Several classification techniques, including Levenberg–Marquardt backpropagation neural networks, linear/quadratic discriminant analysis, unsupervised k-means clustering, and radial basis function neural networks, are compared. This showed the potential of DL for autism classification. The potential for using the whole brain structure with structural MRI scans and DL was then investigated by Ecker et al. [11]. They utilized SVMs on the gray matter and white matter portions separately to conclude that spatially distributed and subtle differences in brain networks among adult ASD participants and typically developing (TD) participants can be detected.

As DL technology has evolved significantly since, better results can be expected.

Alhaddad et al. have utilized Fisher linear discriminant analysis (LDA) for the detection of ASD from EEG signals. Different preprocessing techniques for EEG signals were proposed, utilizing Fast Fourier Transform (FFT) features and raw features to get average correct rates of 90%. They concluded FFT features had higher correct rates and lower standard deviation. Following this, Ibrahim et al. [12] explored alternatives for feature extraction and found cross-correlation and discrete wavelet transform (DWT) to be most effective for epilepsy and ASD diagnosis. To restrict the frequency of the signals to a range of 0.5 Hz and 60 Hz, an elliptic band-pass filter is used. A three-model setup consisting of an Artificial Neural Network (ANN), a K-Nearest Neighbor (KNN), and a linear SVM with LDA is compared. This showed the effectiveness of using DWT for feature extraction.

Similarly, Harun et al. [13] compared the use of SVMs and ANNs for ASD detection using EEG data. They showed that ANNs were the best model with an accuracy of 90.5%, effectively showing promise for deeper networks. In more recent advances, Mohi ud Din and Jayanthy [14] have developed a methodology for the identification of ASD using second-order wavelet scattering transform coefficients as features and DL-based classification networks. VariedML classifiers, such as SVM, logistic regression, KNN, and Decision Tree, were trained and tested. They have adopted two DL architectures, one that used 1D-CNN and another that used Long Short-Term Memory (LSTM), giving accuracies of 92% and 94%, respectively. DL-based models performed better than ML-based models.

Liao et al. [15] have presented a multimodal fusion approach that combines EEG data with behavioral aspects, such as facial expression and eye fixation, for the detection of children with ASD. They have used the k-means algorithm for the extraction of eye fixation features, a CNN and a soft label for extracting facial

expression features, and 12 EEG features from different brain regions. Multimodal data fusion is performed, and an accuracy of 87.50% is obtained by using a weighted naïve Bayes model. Han et al. [16] utilize a multimodal approach consisting of eye tracking (ET) and EEG data for ASD detection. Multiscale entropy, relative power energy, and seven brain network features over five bands were extracted from the EEG signals. In total, 125 EEG and 96 ET features were extracted, and a two-layer multimodal stacked denoising autoencoder was applied for the classification. A Unimodal Feature Learning Module is applied for the higher-level features using a feature fusion module to combine both. The proposed method's accuracy of 95.56% outperformed both data models with a single modality.

Ari et al. [17] have presented an automated method utilizing the Douglas–Peucker algorithm, feature mapping technique with sparse coding, and CNN for ASD using EEG signals. EEG rhythms extrapolated by wavelet decomposition are coded in a sparse representation. They have used extreme learning machine autoencoders for augmenting data. A pretrained deep CNN model is adopted for the classification of EEG signals as ASD and healthy. The small sample of data used only consisted of 29 children, thus making for unreliable results. Tawhid et al. [18] have presented an approach for ASD detection from spectrogram images of EEG signals. The preprocessing of the EEG signals has been done using the methods of infinite impulse response (IIR) filter, common average referencing (CAR), and normalization of filtered signals. The preprocessed signals are transformed into two-dimensional spectrogram images by utilizing the short-time Fourier transform (STFT). Textural features are extracted from the images by using the Ternary CENTRIST (tCENTRIST) technique, and the feature dimension is reduced by using PCA. They have tested the extracted features for their ability to classify ASD using six different ML classifiers, and SVM has given better classification accuracy of 95.25%. They have also classified the spectrogram images by using three different models of CNN. With the DL model, they have achieved an accuracy of 99.15%. The King Abdulaziz University (KAU) dataset has been used here as well with only 16 subjects.

In a similar fashion, Tawhid et al. [19] have also presented a CAD framework for automated identification of different neurological disorders, namely, schizophrenia, Parkinson's disease, epilepsy, and autism, from an existing EEG dataset. They are utilizing a CNN model developed by them for the classification of the considered four classes of neurological disorders by applying it to the generated spectrogram images. The developed model has produced better classification results when compared to existing CNN models such as ResNet50 and AlexNet. Small data size is a drawback of this study and calls for future studies to utilize larger numbers of EEG signals.

Beyond ASD-specific research, broader advancements in medical AI highlight critical pathways for improving EEG analysis. Recent comprehensive reviews of medical image segmentation demonstrate the transformative efficacy of DL architectures such as U-Net and attention-gated networks in isolating pathological features with high spatial precision [20]. While traditionally used for structural imaging, these pixel-wise segmentation principles are highly applicable to localizing abnormal spectral events in EEG spectrograms. Furthermore, state-of-the-art frameworks in related neurological tasks showcase the power of ensemble and decentralized learning. For example, recent DL-based seizure prediction methodologies utilize ensemble classifiers and dynamic sliding windows to capture transient neural states and improve signal-to-noise ratios [21]. Similarly, fuzzy ensemble-based federated learning (FL) has been successfully deployed for EEG-based emotion recognition within the Internet of Medical Things (IoMT). This approach mitigates overfitting on high-density physiological data while enabling secure, decentralized model training across multiple institutions without compromising patient privacy [22].

The summary and comparative analysis of the recent literature review are provided in Table 1.

3. Materials and Methods

3.1. Dataset details

In this research, we experimented with the proposed approach in two publicly available datasets to validate the performance.

Dataset I: This dataset was released by KAU Hospital, located in Jeddah, Saudi Arabia [23]. Twenty children with ASD aged between 6 and 20 were included in this study, and 9 children without any neurological conditions make up the healthy group. BCI2000 software, G. Tech USB amplifiers, and G.Tech EEG cap with Ag/AgCl electrodes were utilized to collect signals from the participants in a relaxed state. Using the right ear lobe and international 10–20 systems having AFz as GND, 16 channels of data were collected: C4, Cz, C3, F4, F8, F3, Fz, F7, FP1, FP2, T3, T5, Pz, O1, Oz, and O2. All ethical approvals were guaranteed, and the dataset has been made publicly available.

Dataset II: The data collected includes EEG and phenotypic data for 142 ASD and 138 TD youth enrolled in the Multimodal Development Neurogenetics of Females with ASD (R01MH100028) project hosted by the Autism Centre of Excellence (ACE) [24]. The participants were enrolled at the Boston Children's Hospital, Yale University, the University of California, Los Angeles, and the Seattle Children's Research Institute, while the University of Southern California coordinated the data between them, ensuring ethical standards were followed across all the centers.An age range of 8–17 years was selected, with the mean being 12.8 and a standard deviation of 2.9 years.

ASD group: The participants were evaluated using the Autism Diagnostic Interview [25] and the Second Edition of the Autism Diagnostic Observation Schedule [26] and the Diagnostic

Table 1
Hyperparameter tuning details of the experimented CNN-based models

Parameter	Search space	ResNet50 Dataset I	ResNet 50 Dataset II	4-layer CNN-Dataset I	4-layer CNN-Dataset II
Batch size	[1–128]	32	32	32	32
Learning rate	[1e-6,1e-2]	5e-4	5e-4	3e-4	6e-4
Optimizer	"Adam," "SGD," "RMSProp"	Adam	Adam	Adam	Adam
Epochs	[5–100]	25	40	40	100
Weight decay	[5e-7,0.05]	0.0001	0.0001	0.0001	0.0001

and Statistical Manual of Mental Disorders, 4th edition, text revision and Differential Ability Subscales-II [27] metrics were met. For participants for whom ADI-R data were not available, thresholds present either on the Social Communication Questionnaire (Bailey, Rutley, and Lord, 2003) [28] or on the second edition of the Social Responsiveness Scale-2 [29] were fulfilled.

Neurotypical (NT) group: The NT control group comprised 138 youths (51% of whom were male) who were accepted only if they lacked any parent-reported issues, developmental or psychiatric issues, diagnoses of ASD, intellectual disabilities, schizophrenia, or learning disorders. Both groups were subject to all other basic exclusion criteria as can be found in Reference [24].

During EEG data collection, both groups watched dynamic videos resembling screensavers while in an eyes-open resting state, facilitated by EPrime. This continuous visual paradigm is utilized in pediatric cohorts to promote visual fixation and naturally minimize ocular artifacts, such as blinks and saccades. Each participant completed three recording blocks, each consisting of roughly 64 s (32 trials, each 2048 ms) of eyes-open resting, totaling around 192 s (96 trials) per experiment. High-density EEG was recorded using a standard net station acquisition template and the EGI 128-channel Net Amps 300 system with HydroCel nets. Sampling was conducted at 500 Hz, referenced to the vertex Cz, with impedances maintained under 50 kOhms. Because EEG recordings in pediatric ASD populations are inherently susceptible to motion and electromyographic artifacts, ensuring signal integrity required a sequential preprocessing approach. During acquisition, behavioral support was provided to minimize gross motor movement, while data collection staff simultaneously logged specific instances of motion or inattention. This allowed for precise epoch-level rejection and manual trimming of corrupted segments during initial post-processing. Following this structural exclusion, residual high-frequency physiological noise was further attenuated using the previously detailed common average referencing (CAR) and Chebyshev IIR filtering, ensuring a clean signal prior to spectrogram generation.

3.2. Methodology

In this section, we discuss the methodologies utilized in the proposed approach. At first, the raw EEG signals are preprocessed, followed by image generation, classification, and XAI for interpretability.

3.2.1. EEG signal preprocessing

We have used a publicly available dataset from KAU Hospital, Saudi Arabia, Jeddah, and ACE. The EEG signals were captured from a total of 16 subjects, of which 12 subjects had ASD and the remaining 4 subjects did not have any of the neurological abnormalities. The ASD group has 9 boys and 3 girls with an age group spanning across 6–20 years. The control group has only boys with the age of 9–13 years. The EEG signals were captured using 16-channel settings using the international 10–20 system. A band-pass filter with a pass frequency band (0.1–60 Hz) and a notch filter with a frequency band (60 Hz) were used during the recording time. Finally, the EEG signals were digitized at a 256 Hz sampling rate. The EEG data are preprocessed using CAR for noise removal. A low-pass Chebyshev IIR filter is applied to nullify the presence of artifacts caused by external noise and eye and muscle movements. To create the required dataset for the ASD classification, the EEG signals were divided into smaller chunks of 3.5 seconds duration. Onto the generated smaller chunks of EEG signals, STFT is applied, and EEG signals are converted to 2D matrices. The Hamming window of size 128 was used with 64 samples utilized for overlapping. This structural design permits causal inference with minimal delay; because it does not depend on future temporal context, it is particularly well-suited for the continuous analysis of live EEG data [30]. The generated dataset of spectrogram images consists of 1381 images of normal subjects and 3486 images of subjects with ASD. The resulting spectrogram images are fed to DL architectures to classify them as either ASD or control groups. Figure 1 illustrates sample spectrogram images of healthy and ASD subjects along

Figure 1
A comprehensive pipeline for the classification of autism spectrum disorder (ASD) and typically developing (TD)

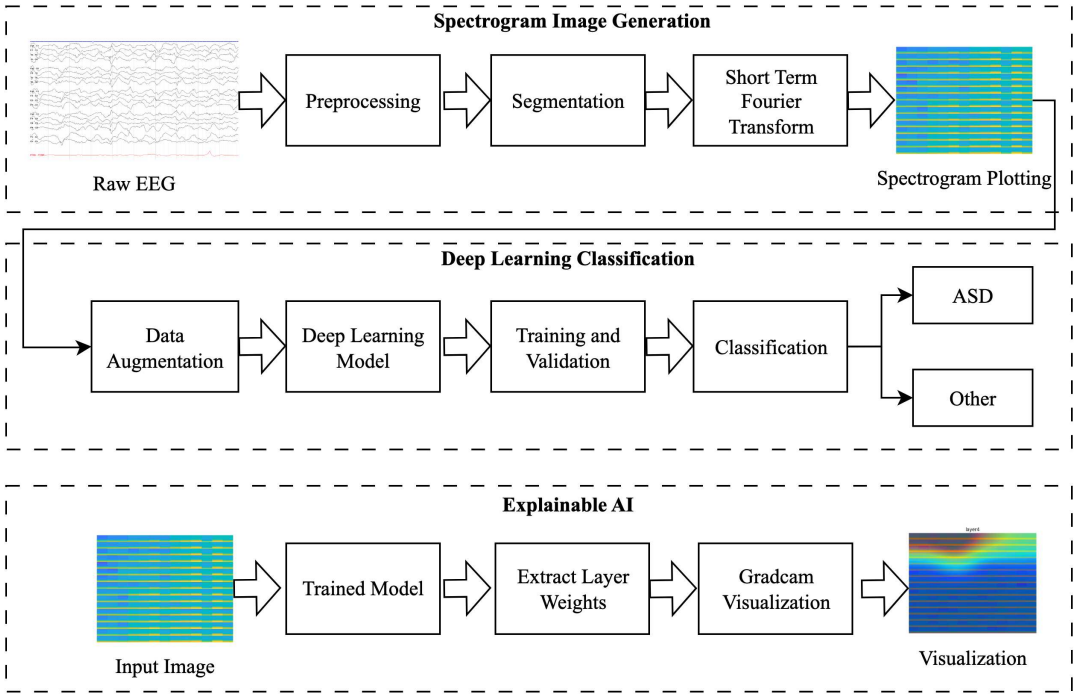
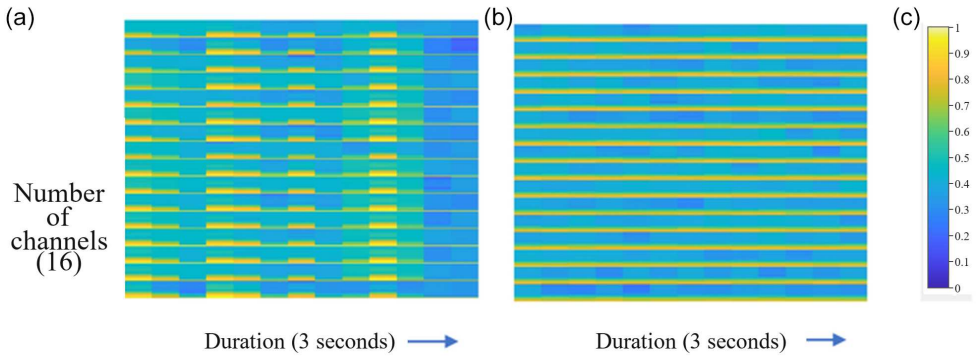


Figure 2
Sample spectrogram images generated from the preprocessed brain EEG signals: (a) subject with ASD, (b) other, and (c) the colormap



with the colormap adapted for image generation. The duration of the EEG signal chunk is indicated by the x-axis, and the number of channels used during signal acquisition is indicated by the y-axis. The number of channels is signified by the presence of horizontal yellow in Figure 2.

3.2.2. Convolutional Neural Network

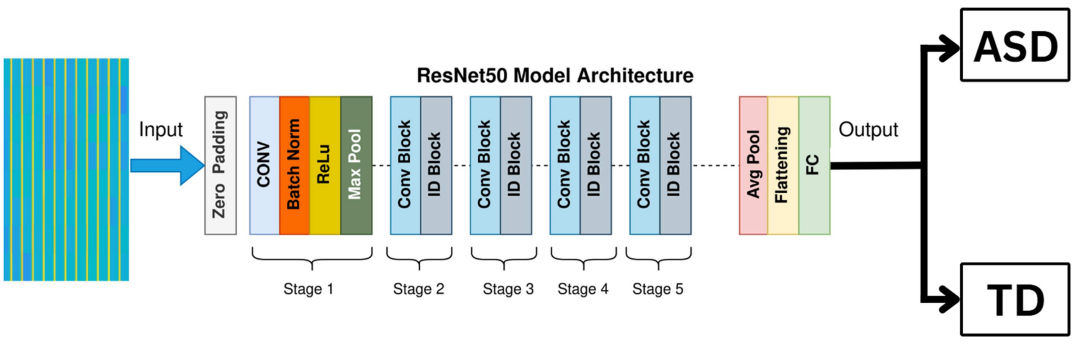
Computer vision applications using CNNs gained popularity due to their high performance in terms of accuracy and other commonly used metrics [31]. Depending on the specific application, CNNs can facilitate feature extraction or end-to-end learning for a particular task. CNN architectures typically consist of convolution, fully connected, pooling, and normalization layers. They are arranged in sequence to form a complete ConvNet. Essentially, the early convolutional layers of CNN models teach them basic features like vertices, edges, curves, and color blobs, whereas the later levels capture more abstract and unique properties. CNN training involves a lot of computation because convolutional kernels and dense layers need to have their weights adjusted. Typically, the backpropagation approach is applied to maximize the weights of a CNN. Three pretrained CNN models and the 4-layer light CNN that Ari et al. [17] recommended have been tested and compared as shown below:

EfficientNet: CNN architecture EfficientNet aims to outperform previous CNN architectures in terms of performance while requiring less computational resources and parameters [32]. Its architecture is predicated on a compound scaling technique that uses a set of predetermined scaling coefficients to scale the network breadth, depth, and resolution equally. EfficientNet is able to effectively balance model size and accuracy across a range of scales thanks to this technique. It begins with a base network and uses scaling to gradually expand the network's width and depth while controlling the number of parameters. EfficientNet is able to achieve computational efficiency and state-of-the-art performance on image classification tasks because of its scaling method.

ResNet50 is a deep CNN variation built on the ResNet (Residual Network) architecture (Figure 3), which was originally designed for image recognition applications. ResNet50's "50" stands for depth, meaning it contains 50 layers [33]. Residual connections, also known as skip connections, are the main innovation in ResNet. They allow data from earlier layers to be fed straight into deeper layers, skipping multiple levels in the process. This makes it possible to train much deeper networks and helps to solve the vanishing gradient issue. ResNet50 comprises a sequence of convolutional layers, followed by batch normalization and rectified linear unit (ReLU) activation functions. The fundamental components of the network are these tiers. ResNet50 has residual blocks, which have several convolutional layers with skip connections in addition to the standard convolutional layers. Global average pooling and a fully connected layer for classification are also included in the architecture. ResNet50's versatility has led to its extensive usage in a variety of computer vision applications. ResNet50's capacity to efficiently train extremely deep neural networks and attain high accuracy on image recognition benchmarks has made it a popular choice for a variety of computer vision workloads.

MobileNet: With depth-wise separable convolutions, MobileNet is a lightweight CNN architecture optimized for embedded and mobile devices, reducing computations and parameters [34]. MobileNetV2 introduces inverted residuals with linear bottlenecks for improved efficiency. It also features

Figure 3
Resnet50 architecture that is used for the classification of ASD and others



width and resolution multipliers to further reduce computational requirements, making it suitable for resource-constrained devices.

Custom CNN architecture: Custom CNN architecture: To provide a computationally efficient alternative suitable for real-world clinical deployment, a custom lightweight architecture was designed and evaluated. This model is composed of four convolutional layers, followed by a dense layer consisting of 256 neurons and a second dense layer with two neurons acting as the classifier. The loss function used was binary cross-entropy. Each convolutional layer is activated using the ReLU or leaky ReLU function, followed by a max-pooling layer. To prevent overfitting, dropout is utilized as a regularizing mechanism, with a 25% dropout rate inserted after convolutional layers 2 and 4, and a 50% dropout rate applied after the fully connected layer. Crucially, this custom architecture operates at merely 98 MFLOPs. Compared to the 4.09 GFLOPs required by deeper networks like ResNet50, this represents a massive reduction in computational complexity. This efficiency positions the custom 4-layer model as an ideal, deployable solution for rapid ASD screening on resource-constrained clinical edge devices.

3.2.3. Explainable AI

DL models, especially CNNs, excel at extracting complex patterns from EEG spectrograms, yet their inherent opacity can undermine clinical trust and hinder adoption. XAI techniques address this "black-box" challenge by producing visual explanations that highlight the input regions driving each prediction. Among these, Gradient-weighted Class Activation Mapping (Grad-CAM) stands out for its versatility: it computes the gradients of a target class score with respect to convolutional feature maps, aggregates these into channel-wise importance weights, and projects them back onto the input to produce a coarse localization map. Importantly, Grad-CAM requires no architectural changes unlike original CAM methods and can be applied post hoc to any CNN. To enhance resolution, Guided Grad-CAM fuses Grad-CAM's localization with guided backpropagation's fine-grained saliency via element-wise multiplication, yielding high-detail activation maps that reveal both "where" and "what" the network focuses on[35].

In medical imaging, these XAI methods have proven invaluable for validating model decisions and uncovering clinically relevant features. For example, Saleem et al. applied Grad-CAM to ResNet-based glioma classifiers, making the tumor regions transparent to radiologists and improving segmentation trustworthiness [36]. Similarly, Islametal. used Guided Grad-CAM to visualize pulmonary opacities in COVID-19 chest X-ray classification, facilitating rapid clinical interpretation[37]. In our work, we extend these techniques to 2D EEG spectrograms: by overlaying Grad-CAM and Smooth Grad-CAM++ maps onto time–frequency images, we identify the specific spectral bands and temporal segments that drive ASD versus non-ASD classification. This dual benefit of transparency and neurophysiological insight not only bolsters clinician confidence but also generates hypotheses about the spectral–temporal biomarkers underlying ASD.

The heatmap is overlaid with the image to show us the relevant parts of the image being used as can be seen in Figure 4.

The mathematical formulation for Smooth Grad-CAM++ is given below:

Let $A^k \in \mathbb{R}^{u \times v}$ denote the k-th feature map of the final convolutional layer, and let Y^c be the score for class c. For N noisy samples $\mathbf{x}^{(n)} = \mathbf{x} + \boldsymbol{\epsilon}_n$, where $\boldsymbol{\epsilon}_n \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$, the smoothed gradient moments are:

$$\mu_{ij}^{(r),kc} = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial^r Y^{c,(n)}}{\left(\partial A_{ij}^{k,(n)}\right)^r}, \quad r \in \{1, 2, 3\}$$

The pixel-wise importance coefficients are:

$$\tilde{\alpha}_{ij}^{kc} = \frac{\mu_{ij}^{(2),kc}}{2\,\mu_{ij}^{(2),kc} + \sum_{a,b} \bar{A}_{ab}^{k}\, \mu_{ij}^{(3),kc}}$$

where $\bar{A}^k = \frac{1}{N} \sum_{n=1}^{N} A^{k,(n)}$ is the mean feature map. The channel weights are:

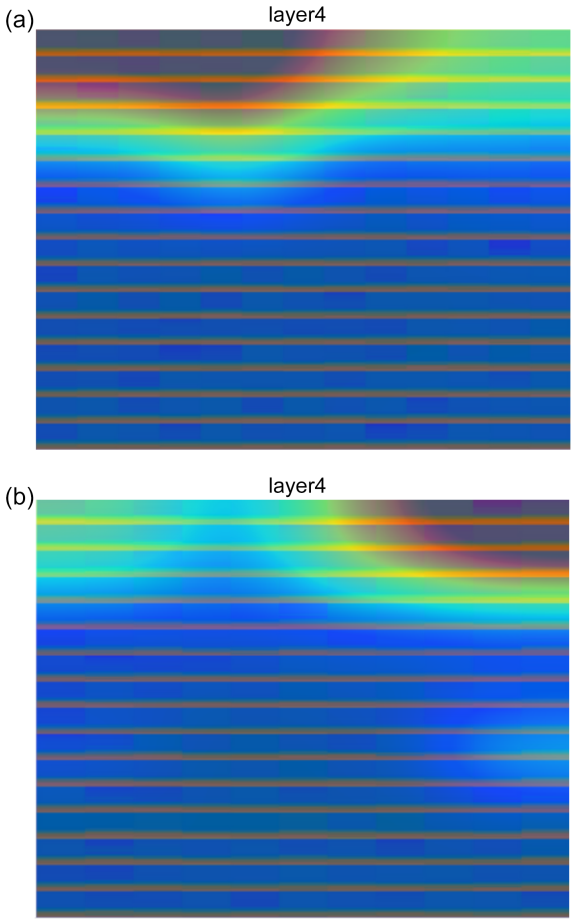
$$\tilde{w}_k^c = \sum_{i,j} \tilde{\alpha}_{ij}^{kc} \cdot \mathrm{ReLU}\left(\mu_{ij}^{(1),kc}\right)$$

The final saliency map is:

$$\mathscr{L}_{\mathrm{SmoothGradCAM++}}^{c} = \mathrm{ReLU}\left(\sum_k \tilde{w}_k^c\, \bar{A}^k\right)$$

which is subsequently upsampled to the input resolution and normalized to $[0, 1]$.

Figure 4
Smooth GradCam++ heatmap overlaid over spectrogram image as for (a) ASD patient and (b) TD participant



3.2.4. Model training and hyperparameter tuning

The dataset is split into a train and test set, with the test set comprising 20% of the total images. The train set was further split into train and validation, with 20% of the images being put in the validation set and the rest 80% being put in the train set.

Three different optimizer techniques have been tested, namely, Adam, Stochastic Gradient Descent with Momentum, and RMSProp, with Adam showing the best results, and henceforth, all the models have been trained using Adam. The details of hyperparameter tuning are provided in Table 1.

4. Results and Discussion

4.1. Experimental analysis

The models were first experimented on Dataset I, and the loss curves comparing the train loss and validation loss can be seen in Figure 5.

Upon training the MobileNet model, we could see the model had overfit as the validation loss remained high even after the train loss had reduced completely. EfficientNet also did not improve much upon training past the 5th epoch as can be seen in Figure 5(c). The ResNet model had a volatile validation loss but trained to the lowest overall validation loss as can be seen in Figure 5(d).

As we can see, the custom CNN has the most uniform training graph between the validation and training loss, although this could also be due to its relatively small number of parameters compared to the rest of the models. The Confusion matrices for the model's classification are shown in Figure 6.

When checking the results of the classification done by the four models, we can see that MobileNet and EfficientNet performed the worst on our test data as can be seen in Figure 6(b) and Figure 6(c). ResNet50 had the best performance among all the models. The custom 4-layer CNN showed close to no false positives, which could be a useful feature for confirmation testing. The 4-layer CNN is more prone to true negatives than false positives, while Resnet50 is more balanced. The more specific performance metrics have been calculated as shown in Table 2 (Dataset I) and Table 3 (Dataset II).

As can be seen in Table 3, ResNet50 shows the most promise with F1 classification scores of 0.99 and 0.98 and a significantly higher specificity than all other models. The same model configurations have been trained on Dataset II as well to see if the models can be applied to any data or if they have only captured the pattern on the KAU dataset. The models applied to Dataset II and the loss curves comparing the train loss and validation loss can be seen in Figure 7.

Upon training, it was observed that the MobileNet and EfficientNet models are prone to severe overfitting as can be seen in Figure 7(b and c). The custom 4-layer architecture once again had the most uniform training and validation loss, which was unable to fully capture the underlying curve, potentially due to limitations in model complexity, even though it was trained for 100 epochs. For ResNet50, the validation loss exhibited oscillations throughout the training process as can be seen in Figure 7(d), indicating potential instability in learning, though it still led to the lowest validation loss. The Confusion Matrices for the models performing on Dataset II are shown in Figure 8.

As we can see from Figure 8, the difference between the results of the models is much larger than that of Dataset I as

Figure 5

Loss curves comparing the training of validation losses when the models (a) MobileNet, (b) custom architecture, (c) EfficientNet, and (d) ResNet50 were trained on the KAU dataset (Dataset I)

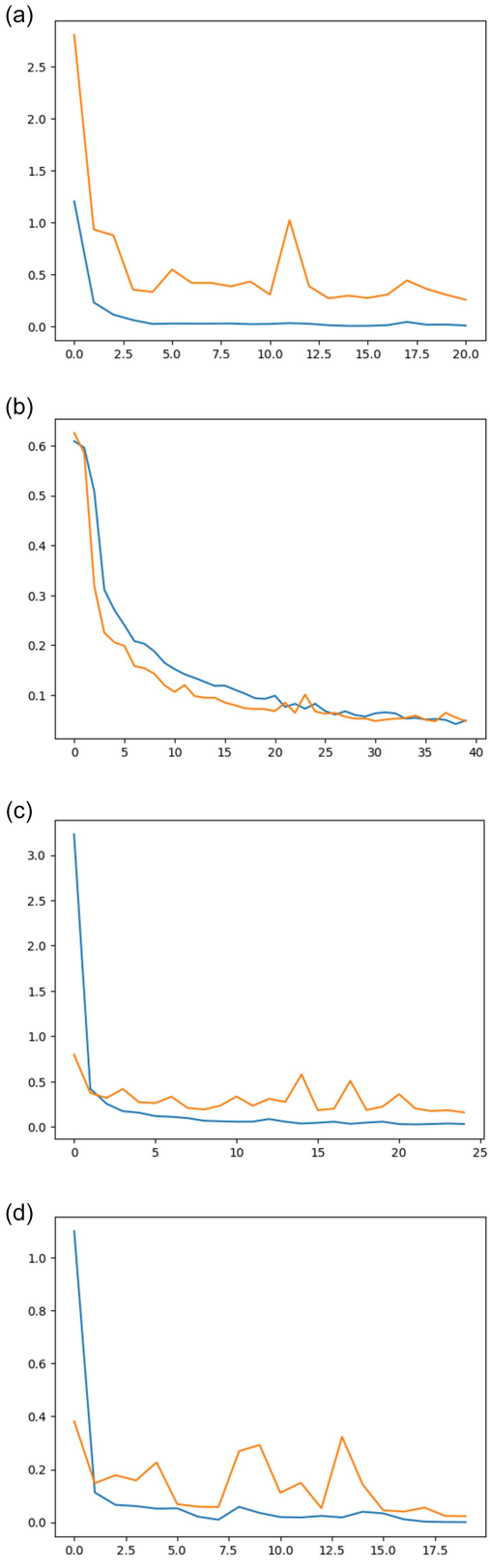
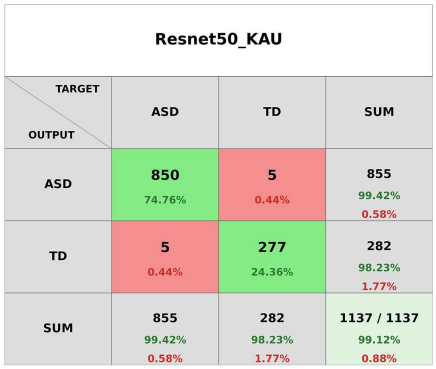
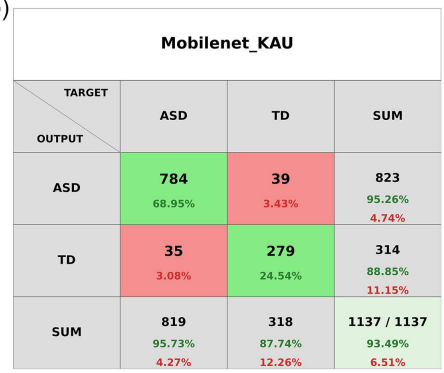


Figure 6
Confusion matrices classifying the as per ASD and TD on the KAU dataset (Dataset I) by (a) ResNet50, (b) MobileNet, (c) EfficientNet, and (d) custom architecture

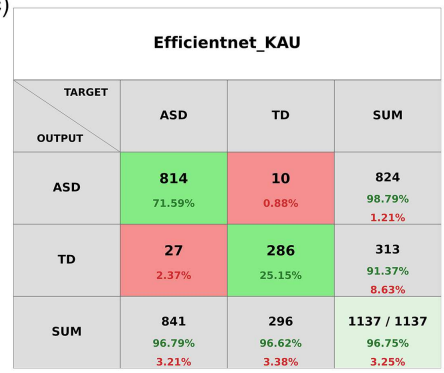
(a)



(b)



(c)



(d)

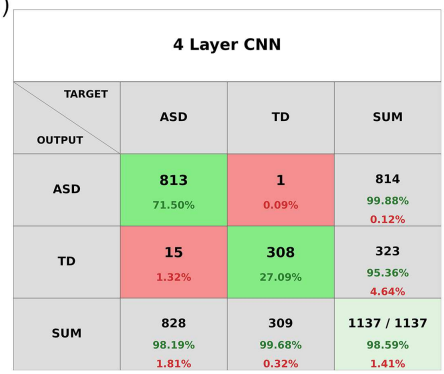


Table 2
The difference in performance metrics between the models on Dataset I

Model	Test accuracy(%)	Class	F1-score	Precision	Recall	Specificity(%)
4-layer CNN	98.59	0	0.99	0.98	1.00	95.36
		1	0.97	1.00	0.95	
Efficientnet_b0	96.74	0	0.98	0.97	0.99	91.37
		1	0.94	0.97	0.91	
Mobilenet_v3	93.49	0	0.95	0.96	0.95	88.85
		1	0.88	0.88	0.89	
Resnet50	98.94	0	0.99	0.99	0.99	98.23
		1	0.98	0.98	0.98	

Table 3
The difference in performance metrics for all the models when applied to Dataset II

Model	Test accuracy(%)	Class	F1-score	Precision	Recall	Specificity(%)
4-layer CNN	90.88	0	0.91	0.88	0.94	87.54
		1	0.91	0.94	0.88	
Efficientnet_b0	83.41	0	0.83	0.81	0.84	82.46
		1	0.84	0.85	0.82	
Mobilenet_v3	83.16	0	0.81	0.86	0.77	74.24
		1	0.85	0.81	0.89	
Resnet50	95.23	0	0.95	0.97	0.93	97.43
		1	0.96	0.94	0.97	

Figure 7
Loss curves comparing training and validation losses when the models (a) custom architecture, (b) MobileNet, (c) EfficientNet, and (d) ResNet50 were trained on Dataset II

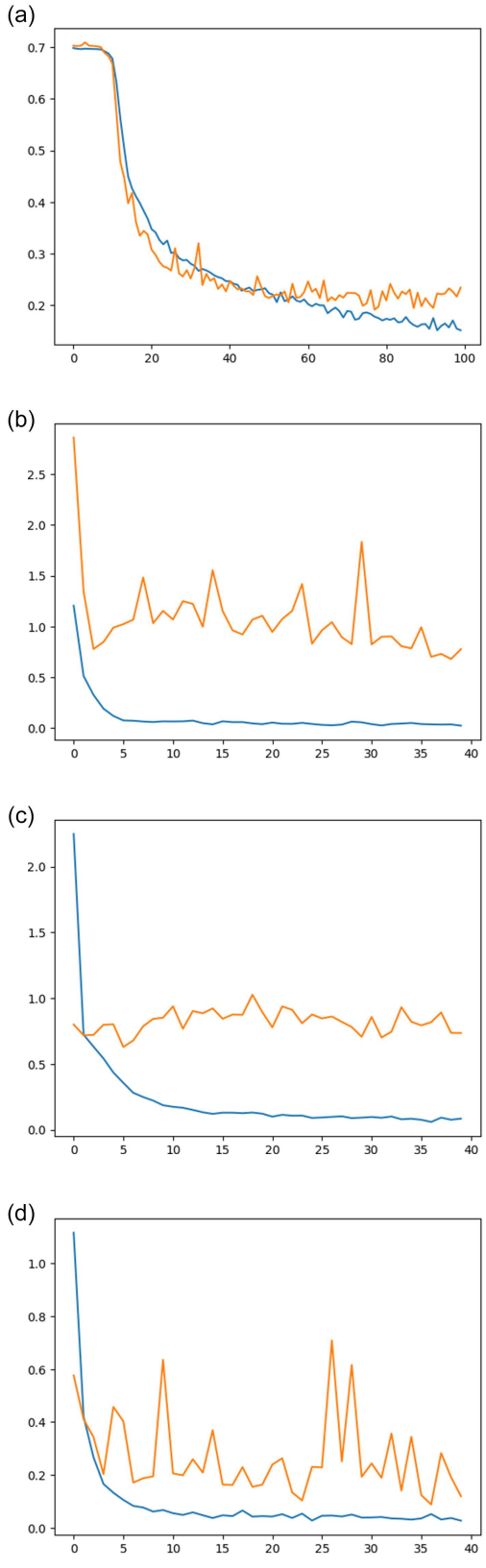


Figure 8
Confusion matrices classifying the as per ASD and TD on Dataset II by (a) ResNet50, (b) MobileNet, (c) EfficientNet, and (d) custom architecture

(a)

Resnet50_Dataset2

TARGET / OUTPUT	ASD	TD	SUM
ASD	852 44.15%	66 3.42%	918 92.81% 7.19%
TD	26 1.35%	986 51.09%	1012 97.43% 2.57%
SUM	878 97.04% 2.96%	1052 93.73% 6.27%	1930 / 1930 95.23% 4.77%

(b)

Mobilenet_dataset2

TARGET / OUTPUT	ASD	TD	SUM
ASD	662 34.30%	251 13.01%	913 72.51% 27.49%
TD	262 13.58%	755 39.12%	1017 74.24% 25.76%
SUM	924 71.65% 28.35%	1006 75.05% 24.95%	1930 / 1930 73.42% 26.58%

(c)

4LayerCNN_Dataset2

TARGET / OUTPUT	ASD	TD	SUM
ASD	897 46.48%	54 2.80%	951 94.32% 5.68%
TD	122 6.32%	857 44.40%	979 87.54% 12.46%
SUM	1019 88.03% 11.97%	911 94.07% 5.93%	1930 / 1930 90.88% 9.12%

(d)

EffocientNet_Dataset2

TARGET / OUTPUT	ASD	TD	SUM
ASD	773 40.05%	142 7.36%	915 84.48% 15.52%
TD	178 9.22%	837 43.37%	1015 82.46% 17.54%
SUM	951 81.28% 18.72%	979 85.50% 14.50%	1930 / 1930 83.42% 16.58%

seen in Figure 6. MobileNet and EfficientNet gave relatively low accuracies as compared to when trained on Dataset I. ResNet50 performed the best, giving an accuracy of 95.23%.

As can be observed, the 4-layer lightweight CNN as suggested by Ari et al. [17] and Resnet50 showed comparable results for the KAU dataset, with ResNet50 achieving 98.94% test accuracy. However, when applied to the second dataset, ResNet50 shows a considerable advantage with 95.23%. The accuracy achieved in the second dataset using different orders and assortments of EEG channels validates the process using EEG waves converted to spectrogram images for the diagnosis of ASD as the models underwent no changes between the two datasets.

4.2. Explainable AI (XAI) results

To interpret the spatial features driving the model's predictions, Smooth Grad-CAM++ was applied to generate class-specific activation maps. For Dataset I (KAU), the generated heatmaps demonstrate high gradient activations predominantly localized within the regions corresponding to the FP1, FP2, F7, and F3 EEG channels. This localization is neurodevelopmentally significant: these channels map to the frontal and prefrontal cortices, regions fundamentally associated with executive functioning, social cognition, and emotional regulation domains characteristically impaired in individuals with ASD. While it remains unclear if all these channels or if any of them specifically provide more insight than the others, it is apparent that their significance for classification is notable.

As illustrated from Figures 9 and 10, the model focuses on specific channels/areas for finding patterns to check whether the patient has ASD or not.

Dataset II also shows certain regions having higher gradient concentrations than the others; however, due to the lack of channel-specific data, we are unable to find greater contributions.

4.3. Discussion

This study aims to advance research on the computer-aided diagnosis of ASD. The study explores methods of using EEG for diagnosis.

Images are obtained from the EEG signals utilizing the STFT. This enables the utilization of inductive biases inherent in CNNs, optimizing image-based analysis. The inductive biases of CNNs, particularly their ability to capture local spatial hierarchies and translation invariance, make them well-suited for image-based analysis of EEG data. By leveraging these biases, CNNs can detect subtle patterns within the transformed EEG images that are critical for distinguishing between ASD and typical development. This architecture is particularly advantageous when handling complex, high-dimensional medical data like EEG signals, which would be challenging for traditional classifiers. Two datasets were used in this study. The first dataset used is the KAU dataset, which has been used by many studies for classification purposes. The process was validated, achieving an accuracy of approximately 99%. We have also used the second dataset from the ACE. This helps us verify if the proposed method can be generalized.

We have implemented XAI so we can gain greater insights into the decision-making process used by the model. It can be

Figure 9
Examples of the gradient heatmap in cases of (a)–(b) ASD patients and (c)–(d) TD patients for Dataset I

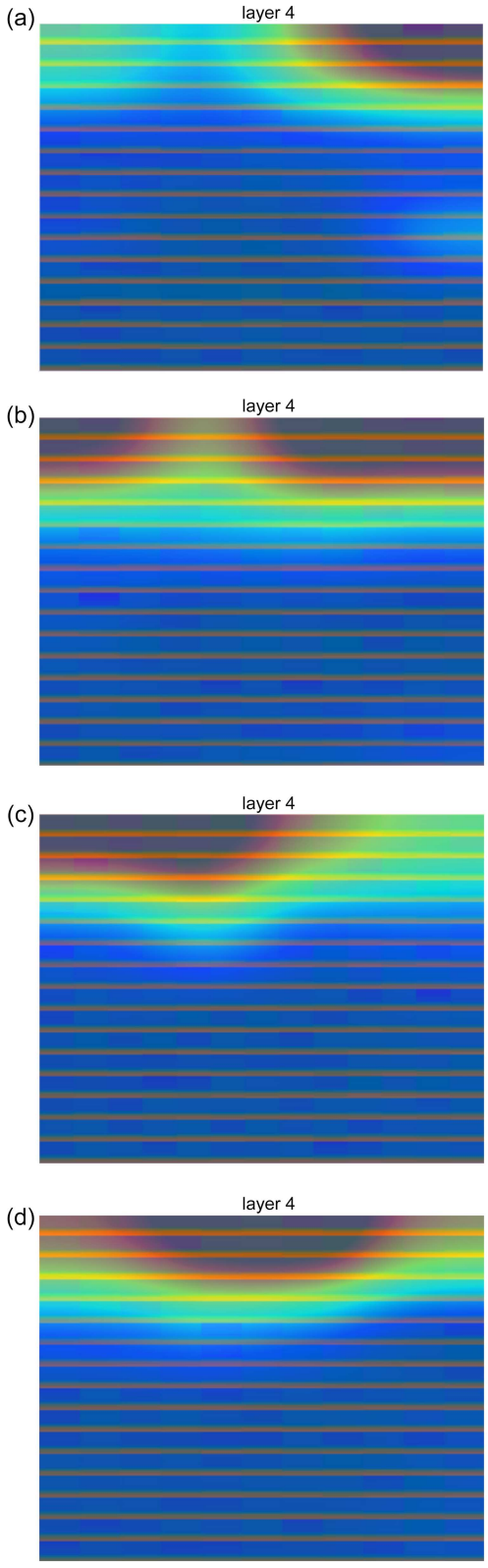
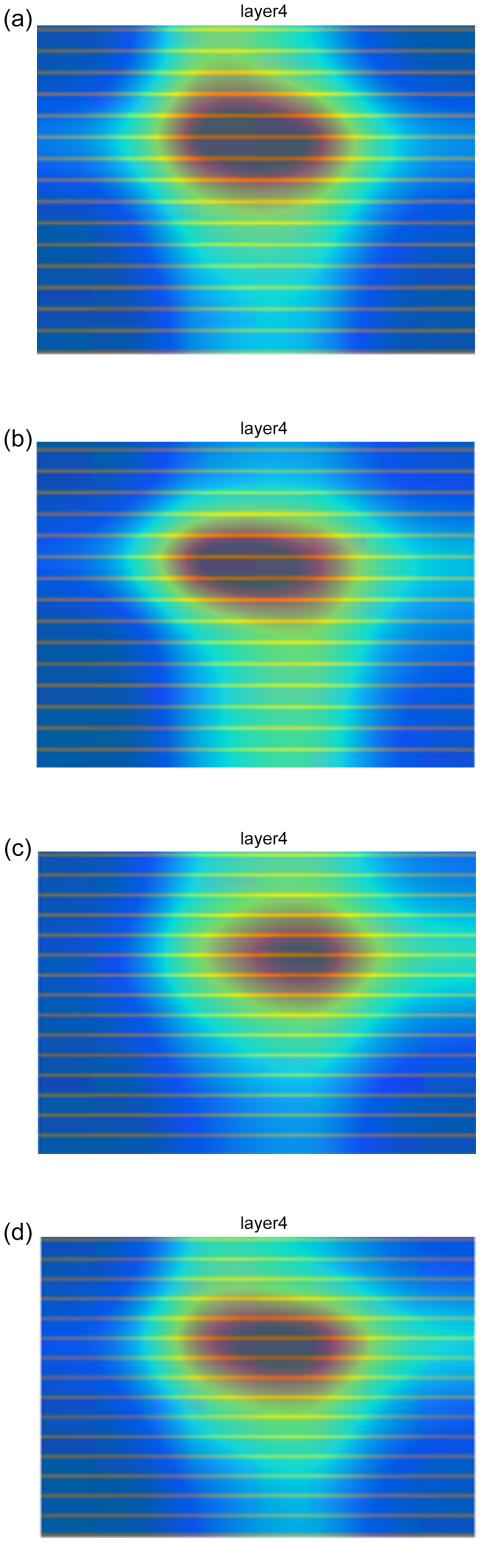


Figure 10
Examples of the gradient heatmap in cases of (a)–(b) ASD patients and (c)–(d) TD patients in Dataset II



observed that the model uses certain specific channels for the diagnosis of autism. In the KAU dataset, the regions corresponding to the FP1, FP2, F7, and F3 channels show the highest Grad-Cam activations in the class decision. These activations might be caused by positive or negative kernels. Hence, it is possible that certain features implying ASD are present in these channels or the absence of certain features in these channels implies ASD. We observed certain specific regions and channels contributed to the decision even in the second dataset; however, the exact channel names could not be extrapolated.

A critical limitation of earlier studies and of Dataset I (KAU dataset) in isolation is the reliance on extremely small sample sizes and severe class imbalances (16 subjects: 12 ASD, 4 TD). While our models achieved near-perfect accuracy on Dataset I, such high metrics on small cohorts must be interpreted with caution due to the risk of deep neural networks memorizing subject-specific noise rather than true pathological features. To directly address this limitation and rigorously validate the clinical viability of our approach, we deliberately evaluated the exact same STFT-CNN pipeline on Dataset II (ACE dataset), a significantly larger and balanced independent cohort (280 subjects). The strong performance on this second dataset achieving an F1-score of 0.96 using ResNet50 empirically demonstrates the broad generalizability of the proposed technique. It confirms that the extraction of spectral features via STFT, coupled with CNN-based classification, captures robust, transferable neurophysiological biomarkers of ASD across disparate clinical populations, rather than relying on dataset-specific artifacts.

Using the knowledge that certain channels hold greater weightage in differences between ASD and TD participants, future works can conduct an exact analysis of which channels show the most differences. The specific contributions of channels FP1, FP2, F7, and F3 in the diagnosis of ASD align with established research on the role of the frontal cortex in social cognition and executive functioning, both of which are often impaired in individuals with ASD. A detailed analysis of the activity in these regions could provide deeper insights into the neurophysiological underpinnings of ASD, linking specific brain regions to behavioral and cognitive symptoms associated with the disorder. Further studies could find the behavioral and intellectual implications of said differences. This entire system can be utilized to develop a real-world CAD system.

A limitation in the present study is that the XAI cannot pinpoint the exact channels containing the information required for diagnosis. Another limitation of the current study is the absence of channel-specific information in Dataset II, which hindered a more granular analysis of which brain regions contributed most to the diagnostic decision. Future research could address this gap by incorporating datasets with more detailed channel data, allowing for a more comprehensive understanding of the neurophysiological differences between ASD and TD populations.

Looking forward, integrating advanced architectural frameworks could further elevate this diagnostic pipeline. While Smooth Grad-CAM++ provides vital spatial explainability, future iterations could leverage medical image segmentation techniques to achieve precise, pixel-wise localization of abnormal spectral events within the time–frequency domain. Furthermore, to mitigate the overfitting observed in high-density datasets like ACE, incorporating fuzzy ensemble techniques and FL could stabilize model convergence and enable privacy-preserving training across multiple pediatric clinics. Finally, transitioning from static classification to continuous monitoring via sliding windows and sequential modeling (e.g., Bi-LSTMs), as demonstrated in robust EEG seizure prediction pipelines, could better capture the temporal evolution of ASD biomarkers and minimize false positives in real-world clinical settings.

5. Conclusion

ASD affects approximately 1 in 100 children worldwide, making early and accurate diagnosis essential for timely intervention. Traditional ASD assessments rely on behavioral observations, which are inherently subjective and time-consuming, while manual EEG interpretation is impeded by complex signal dynamics and noise. In this work, we overcome these limitations by developing a fully automated, noninvasive diagnostic pipeline that combines time–frequency EEG imaging with deep convolutional classifiers and XAI. We first transformed resting-state EEG into spectrogram images using the STFT and then trained multiple CNN architectures—a custom 4-layer CNN, ResNet50, and EfficientNet—on two independent cohorts (KAU and ACE) to ensure robustness. Our best model achieved 98.94% accuracy on the KAU dataset and maintained above 95% accuracy on the larger ACE dataset, demonstrating strong generalizability. By integrating Smooth Grad-CAM++, we produced high-resolution activation maps that not only highlighted critical spectral–temporal features but also identified the EEG channels most predictive of ASD, thereby enhancing clinical interpretability and trust. Looking forward, refining XAI methods to localize contributions at the electrode level could yield deeper neurophysiological insights and support individualized intervention strategies. Further, correlating these spectro-temporal biomarkers with behavioral phenotypes will be an important step toward personalized, data-driven care for children with ASD.

Ethical Statement

The data used in this study were obtained from publicly available and fully de-identified datasets (KAU dataset and ACE datasets). No new experiments or data collection involving human participants were conducted by the authors. The datasets were originally collected by their respective providers under approved institutional ethical protocols. The data used in this study are fully anonymized and publicly authorized for research use. Therefore, ethical approval and informed consent were not required for the present study, as the analysis was conducted solely on secondary, de-identified data.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The Dataset I (KAU dataset) that supports the findings of this study is openly available at https://www.earticle.net/Article/A207042, Reference number [23]. The Dataset II (ACE dataset) that supports the findings of this study is openly available in the NIH National Database for Autism Research (NDA) at https://nda.nih.gov/edit_collection.html?id=2021.

Author Contribution Statement

Andrew Jeyabose: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Arav Chadda:** Methodology, Software, Writing – original draft, Writing – review & editing, Visualization. **Venkatesh Bhandage:** Conceptualization, Software, Formal analysis, Resources, Data curation, Visualization, Supervision.

References

[1] Sobieski, M., Sobieska, A., Sekułowicz, M., & Bujnowska-Fedak, M. M. (2022). Tools for early screening of autism spectrum disorders in primary health care–A scoping review. *BMC Primary Care*, *23*(1), 46. https://doi.org/10.1186/s12875-022-01645-7

[2] Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M. S., Saxena, S., . . . , & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, *15*(5), 778–790. https://doi.org/10.1002/aur.2696

[3] Garic, D., McKinstry, R. C., Rutsohn, J., Slomowitz, R., Wolff, J., MacIntyre, L. C., . . . , & Shen, M. D. (2023). Enlarged perivascular spaces in infancy and autism diagnosis, cerebrospinal fluid volume, and later sleep problems. *JAMA Network Open*, *6*(12), e2348341. https://doi.org/10.1001/jamanetworkopen.2023.48341

[4] Bogéa Ribeiro, L., & da Silva Filho, M. (2023). Systematic review on EEG analysis to diagnose and treat autism by evaluating functional connectivity and spectral power. *Neuropsychiatric Disease and Treatment*, *2023*(19), 415–424. https://doi.org/10.2147/NDT.S394363

[5] Zhang, H., Zhou, Q.-Q., Chen, H., Hu, X.-Q., Li, W.-G., Bai, Y., . . . , & Li, X.-L. (2023). The applied principles of EEG analysis methods in neuroscience and clinical neurology. *Military Medical Research*, *10*(1), 67. https://doi.org/10.1186/s40779-023-00502-7

[6] Hussein, S. A., Bayoumi, A. E. R. S., & Soliman, A. M. (2023). Automated detection of human mental disorder. *Journal of Electrical Systems and Information Technology*, *10*(1), 9. https://doi.org/10.1186/s43067-023-00076-3

[7] Mienye, I. D., & Swart, T. G. (2024). A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information*, *15*(12), 755. https://doi.org/10.3390/info15120755

[8] Wadhera, T., & Kakkar, D. (2021). Social cognition and functional brain network in autism spectrum disorder: Insights from EEG graph-theoretic measures. *Biomedical Signal Processing and Control*, *67*, 102556. https://doi.org/10.1016/j.bspc.2021.102556

[9] Baygin, M., Dogan, S., Tuncer, T., Barua, P. D., Faust, O., Arunkumar, N., . . . , & Acharya, U. R. (2021). Automated ASD detection using hybrid deep lightweight features extracted from EEG signals. *Computers in Biology and Medicine*, *134*, 104548. https://doi.org/10.1016/j.compbiomed.2021.104548

[10] Ghosh-Dastidar, S., Adeli, H., & Dadmehr, N. (2007). Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection. *IEEE Transactions on Biomedical Engineering*, *54*(9), 1545–1551. https://doi.org/10.1109/TBME.2007.891945

[11] Ecker, C., Rocha-Rego, V., Johnston, P., Mourao-Miranda, J., Marquand, A., Daly, E. M., . . . , & Murphy, D. G. (2010). Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach. *NeuroImage*, *49*(1), 44–56. https://doi.org/10.1016/J.NEUROIMAGE.2009.08.024

[12] Ibrahim, S., Djemal, R., & Alsuwailem, A. (2018). Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis. *Biocybernetics and Biomedical Engineering*, *38*(1), 16–26. https://doi.org/10.1016/j.bbe.2017.08.006

[13] Harun, N. F., Hamzah, N., Zaini, N., Sani, M. M., Norhazman, H., & Yassin, I. M. (2018). EEG classification analysis for diagnosing autism spectrum disorder based on emotions. *Journal of Telecommunication, Electronic and Computer Engineering*, *10*(1-2), 87–93. https://jtec.utem.edu.my/jtec/article/view/3326

[14] Din, Mohi ud., Q, & Jayanthy, A. K. (2022). Wavelet scattering transform and deep learning networks based autism spectrum disorder identification using EEG signals. *Traitement du Signal*, *39*(6), 2069–2076. https://doi.org/10.18280/ts.390619

[15] Liao, M., Duan, H., & Wang, G. (2022). Application of machine learning techniques to detect the children with autism spectrum disorder. *Journal of Healthcare Engineering*, *2022*(1), 9340027. https://doi.org/10.1155/2022/9340027

[16] Han, J., Jiang, G., Ouyang, G., & Li, X. (2022). A multimodal approach for identifying autism spectrum disorders in children. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 2003–2011. https://doi.org/10.1109/TNSRE.2022.3192431

[17] Ari, B., Sobahi, N., Alçin, Ö. F., Sengur, A., & Acharya, U. R. (2022). Accurate detection of autism using Douglas-Peucker algorithm, sparse coding based feature mapping and convolutional neural network techniques with EEG signals. *Computers in Biology and Medicine*, *143*, 105311. https://doi.org/10.1016/j.compbiomed.2022.105311

[18] Tawhid, M. N. A., Siuly, S., Wang, H., Whittaker, F., Wang, K., & Zhang, Y. (2021). A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from EEG. *PLOS One*, *16*(6), e0253094. https://doi.org/10.1371/journal.pone.0253094

[19] Tawhid, M. N. A., Siuly, S., Wang, K., & Wang, H. (2023). Automatic and efficient framework for identifying multiple neurological disorders from EEG signals. *IEEE Transactions on Technology and Society*, *4*(1), 76–86. https://doi.org/10.1109/tts.2023.3239526

[20] Gao, Y., Jiang, Y., Peng, Y., Yuan, F., Zhang, X., & Wang, J. (2025). Medical image segmentation: A comprehensive review of deep learning-based methods. *Tomography*, *11*(5), 52. https://doi.org/10.3390/tomography11050052

[21] Wang, D., Li, E., Wang, Y., Liu, Z., Sun, A., Wei, W., . . . , & Zhang, X. (2025). From data to diagnosis: An innovative approach to epilepsy prediction with CGTNet incorporating spatio-temporal features. *PLOS One*, *20*(12), e0337007. https://doi.org/10.1371/journal.pone.0337007

[22] Jiang, W., Zhang, Y., Han, H., Liu, X., Gwak, J., Gu, W., . . . , & Shankar, A. (2025). Fuzzy ensemble-based federated learning for EEG-based emotion recognition in Internet of Medical Things. *Journal of Industrial Information Integration*, *44*, 100789. https://doi.org/10.1016/j.jii.2025.100789

[23] Alhaddad, M. J., Kamel, M. I., Malibary, H. M., Alsaggaf, E. A., Thabit, K., Dahlwi, F., & Hadi, A. A. (2012). Diagnosis autism by Fisher linear discriminant analysis FLDA via EEG. *International Journal of Bio-Science and Bio-Technology*, *4*(2), 45–54. https://www.earticle.net/Article/A207042

[24] Jack, A., McQuaid, G. A., & Gupta, A. R. (2024). Neurogenetics of autism spectrum conditions in individuals assigned female at birth. In L. Mazzone, M. Siracusano, & K. A. Pelphrey (Eds.), *Autism spectrum disorder: Understanding the female phenotype* (pp. 49–79). Springer International Publishing. https://doi/10.1007/978-3-031-62072-0_5

[25] Lord, C., Rutter, M., & le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685. https://doi.org/10.1007/BF02172145

[26] Manjur, S. M., Diaz, L. R. M., Lee, I. O., Skuse, D. H., Thompson, D. A., Marmolejos-Ramos, F., . . . , & Posada-Quintero, H. F. (2025). Detecting autism spectrum disorder and attention deficit hyperactivity disorder using multimodal time-frequency analysis with machine learning using the electroretinogram from two flash strengths. *Journal of Autism and Developmental Disorders*, *55*(4), 1365–1378. https://doi.org/10.1007/s10803-024-06290-w

[27] Elliott, S. N., Anthony, C. J., Lei, P.-W., & DiPerna, J. C. (2022). Parents' assessment of students' social emotional learning competencies: The SSIS SEL brief scales-parent version. *Family Relations*, *71*(3), 1102–1121. https://doi.org/10.1111/fare.12615

[28] Brusa, C., Buchignani, B., Cutri, C., Coratti, G., Clark, E., Johnson, E., & Baranello, G. (2026). Expressive language and social communication abilities in children with spinal muscular atrophy type 1. *Developmental Medicine & Child Neurology*, *68*(5), 696–705. https://doi.org/10.1111/dmcn.16461

[29] Volkmar, F. R (Ed.). (2021). *Encyclopedia of autism spectrum disorders*. Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-91280-6

[30] Mehrabi, A., Sreenivasan, N., Gunawardana, U., & Gargiulo, G. (2026). Hybrid spike-encoded spiking neural networks for real-time EEG seizure detection: A comparative benchmark. *Biomimetics*, *11*(1), 75. https://doi.org/10.3390/biomimetics11010075

[31] Bitar, A., Rosales, R., & Paulitsch, M. (2023). Gradient-based feature-attribution explainability methods for spiking neural networks. *Frontiers in Neuroscience*, *17*, 1153999. https://doi.org/10.3389/fnins.2023.1153999

[32] Agarwal, D., Berbís, M. Á., Luna, A., Lipari, V., Ballester, J. B., & de la Torre-Díez, I. (2023). Automated medical diagnosis of Alzheimer's disease using an efficient net convolutional neural network. *Journal of Medical Systems*, *47*(1), 57. https://doi.org/10.1007/s10916-023-01941-4

[33] Shafiq, M., & Gu, Z. (2022). Deep residual learning for image recognition: A survey. *Applied Sciences*, *12*(18), 8972. https://doi.org/10.3390/app12188972

[34] Rybczak, M., & Kozakiewicz, K. (2024). Deep machine learning of MobileNet, efficient, and inception models. *Algorithms*, *17*(3), 96. https://doi.org/10.3390/a17030096

[35] Ennab, M., & Mcheick, H. (2025). Advancing AI interpretability in medical imaging: A comparative analysis of pixel-level interpretability and Grad-CAM models. *Machine Learning and Knowledge Extraction*, *7*(1), 12. https://doi.org/10.3390/make7010012

[36] Saleem, H., Shahid, A. R., & Raza, B. (2021). Visual interpretability in 3D brain tumor segmentation network. *Computers in Biology and Medicine*, *133*, 104410. https://doi.org/10.1016/J.COMPBIOMED.2021.104410

[37] Singh, R. K., Pandey, R., & Babu, R. N. (2021). COVIDScreen: Explainable deep learning framework for differential diagnosis of COVID-19 using chest X-rays. *Neural Computing and Applications*, *33*(14), 8871–8892. https://doi.org/10.1007/s00521-020-05636-6

How to Cite: Jeyabose, A., Chadda, A., & Bhandage, V. (2026). Robust and Interpretable Deep Learning on EEG Spectrograms for Autism Spectrum Disorder Detection. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE62027780