Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 1, January 2026, pp: 409-429 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Heavy–Light Soft-Vote Fusion of EEG Heatmaps for Autism Spectrum Disorder Detection

Melinda Melinda[1], Syahrul Gazali[2,3], Yuwaldi Away[1], Aufa Rafiki[1], W.K. Wong[4], Muliyadi Muliyadi[5], and Siti Rusdiana[6]

[1] Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, Aceh, Indonesia.
[2] Department of Neurology, Faculty of Medicine, Universitas Syiah Kuala, Banda Aceh, Aceh, Indonesia.
[3] Department of Neurology, Faculty of Medicine, RSUD Dr. Zainoel Abidin, Banda Aceh, Aceh, Indonesia.
[4] Department of Electrical and Computer Engineering, Faculty of Engineering and Sciences, Curtin University Malaysia, Miri, Sarawak, Malaysia.
[5] Doctoral Programs of Engineering, Postgraduate School, Universitas Syiah Kuala, Banda Aceh, Aceh, Indonesia.
[6] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh, Aceh, Indonesia.

Corresponding author: Melinda Melinda (e-mail: melinda@usk.ac.id), **Author(s) Email**: Syahrul Gazali (e-mail: syahrulsps@usk.ac.id, Yuwaldi Away (e-mail: yuwaldi@usk.ac.id), Aufa Rafiki (e-mail: aufa35@mhs.usk.ac.id), W.K. Wong (e-mail: WeiKitt.w@curtin.edu.my), Muliyadi Muliyadi (e-mail: muliyadi@pnl.ac.id), Siti Rusdiana (e-mail: siti.rusdiana@usk.ac.id)

Abstract **Autism spectrum disorder is a neurodevelopmental condition that affects social communication and behaviour, and diagnosis still relies on subjective behavioural assessment. Electroencephalography provides a noninvasive view of brain activity but is noisy and often analysed with handcrafted features or evaluation schemes that risk data leakage. This study proposes a deep learning pipeline that combines wavelet denoising, EEG-to-image encoding, and heavy-light decision fusion for autism detection from EEG. Sixteen-channel EEG from children and adolescents with autism and typically developing peers in the KAU dataset is denoised using discrete wavelet transform shrinkage, segmented into fixed 4 second windows, and rendered as pseudo colour heatmaps. These images are used to fine-tune five ImageNet pretrained architectures under a unified training protocol with 5-fold cross-validation. Heavy-light fusion combines one heavyweight backbone and one lightweight backbone through weighted soft voting on class posterior probabilities. The strongest single model, ConvNeXt Tiny, attains about 97.25 percent accuracy and 97.10 percent F1 score at the window level. The best heavy light pair, ConvNeXt plus ShuffleNet, reaches about 99.56 percent accuracy and 99.53 percent F1, with sensitivity and specificity in the 99 percent range. Fusion mainly reduces missed ASD windows without increasing false alarms, indicating complementary error patterns between heavy and light models. These findings show that the proposed denoise encode classify pipeline with heavy light fusion yields more robust autism EEG classification than individual backbones and can support EEG-based decision support in autism screening.**

Keywords **Electroencephalography; Autism Spectrum Disorder; Wavelet Denoising; Heatmap; Deep Learning; Decision Fusion.**

I. Introduction

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental condition that affects social communication, behaviour, and sensory processing across the lifespan. Clinical diagnosis still relies primarily on behavioural assessment, which is time-consuming and inherently subjective, often leading to inter-clinician variability and delayed identification, particularly during early childhood when symptoms may be subtle or atypical. These limitations are further exacerbated in resource-limited settings with restricted access to specialised clinicians, motivating a strong interest in objective, physiology-based markers that can support earlier, more consistent, and scalable detection and follow-up. Electroencephalography (EEG) offers a noninvasive, high-temporal-resolution window into brain dynamics and has become a key modality for biomarker research in psychiatry and developmental neuroscience [1], [2]. Recent reviews show that EEG-derived measures, including oscillatory power, complexity, and connectivity, can distinguish ASD from typically developing cohorts and can be used as input to machine learning systems for computer-aided diagnosis [2], [3], [4], [5].

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

Raw EEG signals, however, have a low signal-to-noise ratio and are highly nonstationary. Muscle activity, eye movements, and environmental interference can obscure subtle neurophysiological patterns relevant to classification. Robust preprocessing is therefore essential before modelling [6]. Wavelet-based denoising with the discrete wavelet transform (DWT) provides joint time and frequency localisation and a multiresolution decomposition that matches both fast transients and slow rhythms. Survey papers and application studies report that wavelet shrinkage can effectively reduce ocular and muscle artefacts and power line interference while preserving clinically important morphology [7], [8], [9], [10]. In ASD EEG studies, wavelet domain pipelines such as DWT and stationary wavelet transform combined with linear discriminants or support vector machines have already demonstrated reliable case control separation, especially in paediatric cohorts [11], [12], [13].

Alongside denoising, evaluation protocols have a major impact on reported performance. Several deep learning papers on psychiatric EEG datasets have highlighted that segment-wise cross-validation can easily cause data leakage when windows from the same subject are split across training and test folds, leading to overly optimistic performance estimates [14], [15]. This concern is particularly pronounced in studies based on small subject cohorts, where limited sample diversity poses additional challenges for model generalizability. Recent work also shows that choices in filtering, referencing, and artefact handling can substantially change decoding accuracy, underscoring the need for transparent, leakage-aware preprocessing and subject-wise validation in EEG-based classification [16].

To exploit mature computer vision models, an emerging line of work converts preprocessed EEG into two-dimensional image representations, such as topographic maps, spatio spectral feature images, and connectivity maps [3], [17], [18], [19], [20]. These EEG to image encodings allow convolutional neural networks and transformer-based architectures to capture local patterns within channels and global relationships across channels in a single structured input, and they have achieved strong performance in several neuropsychiatric and neurodegenerative applications [17], [18], [19], [20]. Among these representations, channel-by-time heatmaps preserve the temporal evolution of each EEG channel within fixed windows while maintaining a consistent inter-channel ordering, enabling joint modelling of temporal dynamics and cross-channel relationships. Compared with topographic maps or spatio-spectral and connectivity images, this representation avoids additional feature engineering and provides a more direct and interpretable spatial–temporal structure for window-based ASD EEG analysis [21].

In parallel, modern computer vision backbones have evolved along two directions. Heavyweight networks such as ConvNeXt and transformer-based models offer high representational capacity at the cost of memory and computation, while lightweight architectures such as EfficientNet B0, GoogLeNet, and ShuffleNet are designed for deployment on constrained devices [22], [23], [24], [25]. Ensemble and decision fusion strategies that combine heterogeneous learners at the probability level have been shown to improve accuracy and stability in biomedical signal and medical image classification without major changes to the underlying models [26], [27], [28], [29], [30], [31], [32]. Building on this, combining one heavyweight and one lightweight model is hypothesised to exploit complementary characteristics, where the heavy model captures complex spatial–temporal patterns while the light model enhances robustness and efficiency, yielding improved performance without the full cost of heavy-only ensembles [26], [28].

Despite these advances, several limitations remain in current ASD EEG classification studies. Many wavelet-based pipelines rely on handcrafted features and a single shallow classifier, without assessing how different modern deep backbones behave on the same preprocessed data [11], [12], [13]. Deep learning approaches that operate on spectrograms or connectivity maps often use segment-level validation, which risks subject leakage and may overestimate performance on small datasets [4], [14], [15], [32]. Furthermore, the combined effect of wavelet denoising, non-spectral channel by time heatmap encoding, and heavy light model fusion under strictly subject-wise cross-validation on the widely used KAU ASD EEG dataset has not yet been systematically evaluated.

Based on the above observations, this study aims to directly address key methodological limitations in existing ASD EEG classification research, including the reliance on handcrafted features and shallow classifiers, the use of spectral or connectivity-driven representations that require additional modelling assumptions, and the risk of subject-level data leakage due to segment-wise evaluation. To this end, we develop and evaluate a leakage-aware EEG-to-heatmap pipeline that integrates wavelet-based denoising, non-spectral channel-by-time encoding, modern deep vision backbones, and heavy–light decision fusion under strictly subject-wise cross-validation. The main contributions of this work can be summarised as follows:

1) We propose an end-to-end pipeline that converts raw multi-channel EEG into denoised, fixed-length segments and subsequently into non-spectral channel-by-time heatmaps, thereby avoiding

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

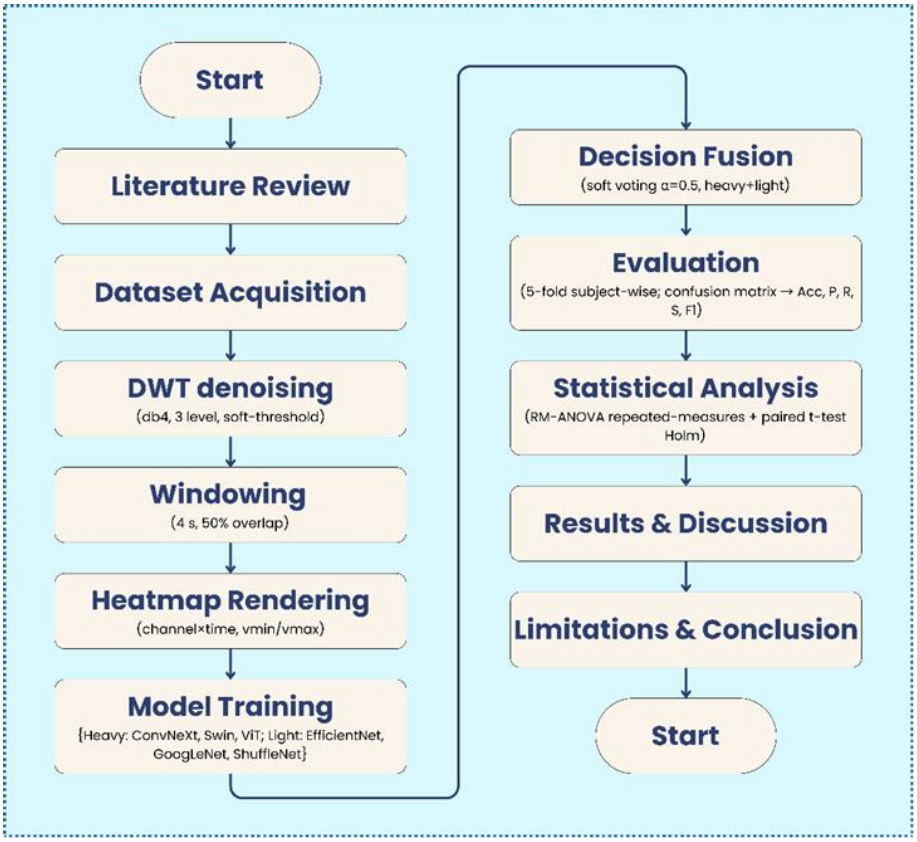


Fig. 1. Overview of the proposed EEG to heatmap ASD classification framework.

handcrafted feature extraction and frequency-domain assumptions commonly used in spectrogram- or connectivity-based approaches, while preserving interpretable spatial–temporal patterns across channels.

2) We design a heavy–light decision-level fusion scheme that explicitly combines models with complementary capacities, addressing the limitations of single shallow classifiers and homogeneous ensembles by leveraging both high-capacity representation learning and robustness from lightweight architectures under a unified inference framework.
3) We adopt a strictly subject-wise cross-validation protocol with leakage-aware preprocessing to directly mitigate the risk of data leakage associated with segment-level validation, and we report clinically relevant performance metrics together with confusion matrices and statistical significance tests.
4) We provide empirical evidence that the proposed combination of wavelet-based denoising, non-spectral EEG-to-heatmap encoding, and heavy–light fusion yields more robust ASD EEG classification than individual backbones, while achieving a favourable trade-off between classification performance and computational cost.

An overview of the proposed EEG to heatmap ASD classification framework is illustrated in Fig. 1. Raw 16-channel EEG from the KAU dataset is denoised with DWT, segmented into overlapping 4-second windows, converted into channel-by-time heatmaps, and then processed by heavy and light vision backbones whose outputs are combined through decision-level fusion. The rest of this paper is organised as follows. Section 2 describes the dataset, EEG preprocessing, and heatmap generation. Section 3 details the selected heavy- and light-vision backbones and the decision-fusion strategy. Section 4 presents the experimental setup and results. Section 5 discusses the findings, practical implications, and limitations. Section 6 concludes the paper and outlines directions for future research.

II. Materials and Methods

A. Dataset

This study analyzes a public EEG dataset provided by King Abdulaziz University (KAU), Jeddah, Saudi Arabia, comprising recordings from 16 subjects, 8 children with ASD and 8 typically developing controls, distributed as 16 recordings by 16 channels in BCI2000

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

.dat format [33]. Each sample contains the standard 10 20 montage (Fp1, F3, F7, T3, T5, O1, C4, Fp2, Fz, F4, F8, C3, Cz, Pz, Oz, O2). We retained the full 16-channel layout to bilaterally sample frontal, temporal, central, parietal, and occipital regions that are frequently implicated in ASD, consistent with prior EEG findings of atypical frontal and fronto-posterior coherence and altered spectral power profiles, including reduced alpha, relative increases in theta and beta, and reports of elevated high-frequency activity [34], [35], [36]. The same KAU cohort has also been used in recent wavelet-based ASD studies, for example stationary wavelet transform combined with Fisher linear discriminant analysis, which facilitates methodological comparability with our protocol [13].

All personal identifiers are absent in the public release. Group membership is provided at the subject level (ASD versus control) without per-event clinical markers or ASD subtyping. The cohort includes eight ASD participants (five males and three females, ages 6 to 20 years, total 4,104.2 seconds of EEG) and eight controls (all males, ages 9 to 13 years, total 4,534.9 seconds). Group labels were assigned by the KAU Hospital clinical team and are used here as ground truth. The public description does not specify the diagnostic instruments that were used, such as ADOS, ADI-R, or DSM 5, which we acknowledge as a limitation of the source data [33].

Recordings were acquired in a relaxed state using Ag/AgCl electrodes with a g.tec EEG device and a USB amplifier under BCI2000, sampled at 256 Hz, with an acquisition bandpass of 0.1 to 60 Hz and a 60 Hz notch filter. The original study reports that annotations were performed by qualified clinical staff following KAU Hospital standard diagnostic procedures for ASD, that written informed consent was obtained from all participants or their legal guardians prior to acquisition, and that all data were fully anonymised before public release, under approval from the KAU Ethics Committee. Dataset access requests can be directed to the data custodian listed in the original publication. The present work involves only secondary analysis, with no new data collection or direct contact with human participants, and therefore adheres to the ethical framework described in [33].

B. Discrete Wavelet Transform (DWT)

Electroencephalography (EEG) signals are characteristically low signal-to-noise ratio and nonstationary, so robust denoising is essential to preserve diagnostically relevant structure before modelling [6]. In this work, we adopt discrete wavelet transform (DWT) denoising, which provides joint time-frequency localisation and a natural multiresolution analysis that matches transient EEG bursts and slow rhythms. Compared with fixed resolution Fourier filtering, DWT separates transient components into scale-localised detail coefficients while concentrating slower neural oscillations into coarse approximations, enabling selective suppression of artefacts without smearing neurophysiological content. Contemporary reviews and application studies report that wavelet-based shrinkage can effectively suppress ocular and muscle artefacts and line noise while maintaining the morphology of neural rhythms, and they highlight wavelet-based denoising as a staple in modern EEG pipelines [7], [8], [9], [10].

Prior to wavelet-based denoising, raw EEG signals were subjected to standard band-pass filtering to suppress baseline drift and high-frequency noise outside the physiological EEG range. No additional notch filtering was applied, as power-line interference was not prominent in the recorded data. Discrete wavelet transform (DWT) shrinkage was then applied to further reduce noise, followed by fixed-length window segmentation. For a discrete-time EEG channel, the three-level DWT decomposes the signal into an approximation and detail component, as expressed in Eq. (1).

$$\boldsymbol{x[n]} = \boldsymbol{A_J[n]} + \sum_{j=1}^{J} \boldsymbol{D_j[n]} \quad (1)$$

where $D_j[n]$ captures oscillatory activity in a dyadic sub-band and $A_J[n]$ contains the slower baseline trend [29]. Let $d_{j,k}$ denote the wavelet coefficient at level j and time index k. Soft thresholding updates each coefficient according to Eq. (2).

$$\tilde{d}_{j,k} = sign\left(d_{j,k}\right) \max\left(\left|d_{j,k}\right| - \lambda_j, 0\right) \quad (2)$$

with λ_j chosen either by the universal threshold or by the BayesShrink rule described below. In this way, coefficients with small magnitude, which are likely dominated by noise, are shrunk towards zero, whereas large coefficients that carry neural information are preserved. We adopt an orthogonal $db4$ mother wavelet owing to its compact support and smooth Daubechies-style regularity, which promotes sparse representations of EEG bursts and edges while avoiding excessive ringing. Comparative studies on EEG denoising consistently report $db4$ and related families among the strongest choices for balancing fidelity and smoothness, providing a favourable accuracy to complexity trade-off [37].

DWT shrinkage was performed using a db4 wavelet with three decomposition levels. Three levels were selected to adequately capture the dominant EEG frequency bands while avoiding over-decomposition that may distort physiologically relevant signal components [7], [8], [9], [10]. For each 16-channel segment, a three-level Mallat decomposition is applied, yielding approximation and detail subbands $\{A_3, D_3, D_2, D_1\}$. Under a sampling rate of 256 Hz and ideal half-band splitting, these levels roughly correspond to very high frequency activity dominated

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

by muscle artefacts in D_1, higher beta and low gamma activity in D_2, alpha and low beta rhythms in D_3, and slower delta to theta activity in A_3. Exact passbands depend on the analysis filters, but this octave structure is standard in EEG wavelet analyses [29]. The three-level filter bank and the flow from the raw signal to the approximation and detail subbands are illustrated schematically in Fig. 2. Noise attenuation operates by soft-thresholding the detail coefficients at each level, a well-established approach that shrinks small-magnitude coefficients dominated by noise more than large coefficients that likely carry signal. In classical wavelet shrinkage, the universal threshold τ is defined as given in Eq. (3) [38].

$$\lambda = \sigma\sqrt{2\log N} \quad (3)$$

where N is the number of samples in the subband and σ is a robust noise estimate computed from the median absolute deviation of the finest scale details divided by 0.6745. BayesShrink refines this idea by adapting λ per subband using a generalised Gaussian prior, often improving the bias-variance trade-off [39]. In our implementation, σ is estimated from D_1 and soft thresholds are applied separately at each level. When BayesShrink is enabled, ssubband-specifichresholds are used; otherwise, a fixed universal threshold is applied across D_1 to D_3. The inverse DWT reconstructs denoised channels from A_3 and the thresholded details $\{D_3, D_2, D_1\}$. Qualitatively, DWT denoising reduces high-frequency fluctuations while preserving slower neural rhythms; this can be seen by comparing the raw and denoised 16-channel windows in Fig. 3(a) and Fig. 3(b), plotted with identical amplitude scales. Practically, artefacts with strong high-frequency content (muscle, abrupt motion) are attenuated mainly in D_1 and D_2, while slow ocular drifts and baseline wander contribute less to thresholded details after the acquisition band limits, so DWT complements the fixed front-end filters by adaptively suppressing bursty contaminants in a scale-aware manner. EEG workflows that integrate wavelet denoising in this way have been shown to improve downstream classification and regression (for example, depth of anaesthesia indices) without compromising temporal precision [40].

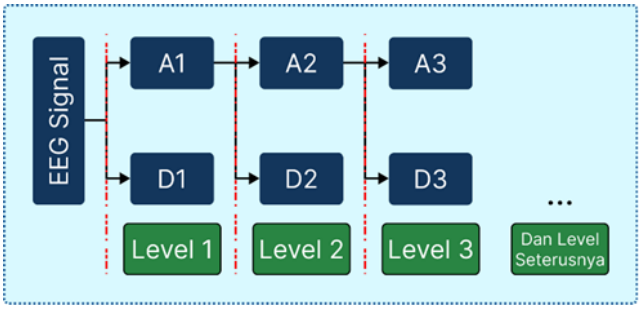
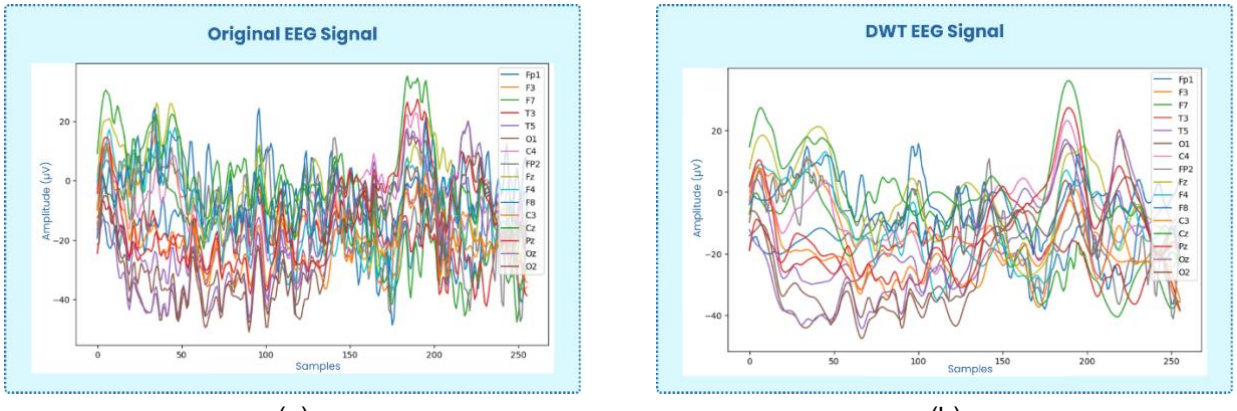


Fig. 2. DWT process overview. A three level Mallat tree decomposes each channel into approximation and detail coefficients.

C. Windowing

After denoising with DWT, each 16-channel EEG recording at 256 Hz is segmented into fixed 4-second windows with 50 percent overlap, corresponding to 1,024 samples per window and a hop of 512 samples. The choice of a 4-second window provides a balance between temporal resolution and contextual coverage, allowing multiple cycles of dominant EEG rhythms (e.g., theta, alpha, and beta bands) to be captured within each segment while maintaining quasi-stationarity. A 50% overlap is employed to increase the effective number of training samples and ensure smooth transitions between adjacent windows, while limiting excessive redundancy between segments. This window length balances temporal context, capturing multiple cycles of low-frequency rhythms, with a sufficient number of training instances, a trade-off that is commonly adopted in recent EEG classification pipelines. Fig. 4 illustrates the windowing process, where each window overlaps the



(a) (b)

Fig. 3. EEG before and after DWT denoising. (a) Raw 16 channel window; (b) DWT denoised window after soft thresholding.

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

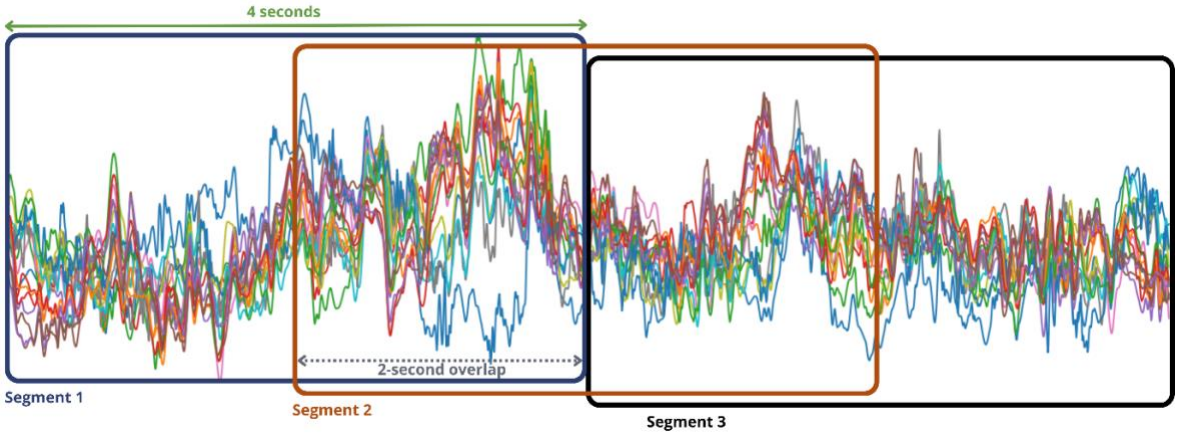


Fig. 4. Windowing of EEG signal with 50 percent overlap.

next by 50 percent to maintain continuity while increasing the number of training samples. A recent Q1 study explicitly selected 4-second windows with 50 percent overlap to ensure continuity while maintaining adequate sample counts for deep models [41]. Controlled comparisons also show that overlap-based segmentation can increase accuracy relative to non-overlapped baselines, improving training efficiency and robustness in EEG applications [42].

Let $x_c[n]$ denote the denoised discrete-time EEG signal of channel c sampled at $F_s = 256$ Hz. We set the window length to $L = 4F_s = 1{,}024$ samples and the hop size to $H = \frac{L}{2} = 512$ samples for a 50 percent overlap. For a recording of length N_s samples from the subject s, the m-th window covers indices defined in Eq. (4), with the total number of windows determined by Eq. (5).

$$n = mH, mH + 1, \dots, mH + L - 1, \quad (4)$$

$$M_s = \left\lfloor \frac{N_s - L}{H} \right\rfloor + 1 \quad (5)$$

The corresponding multi-channel window is represented as a matrix ${X_s}^{(m)} \in \mathbb{R}^{C \times L}$ with $C = 16$ channels, as defined in Eq. (6).

$${X_s}^{(m)}(c, i) = x_c[mH + i], c = 1, \dots, 16, i = 0, \dots, L - 1. \quad (6)$$

This construction yields a fixed-size tensor for each window that can be directly passed to the subsequent heatmap rendering and deep learning stages. For each subject, we construct a windowed tensor by stacking all channels over each 4-second interval. Any trailing fragment shorter than 4 seconds is discarded to maintain consistent window shapes. In practice, segmentation is implemented by array slicing on the denoised time series with a stride of 512 samples, which efficiently realises overlapping windows with minimal computational overhead. To assess generalisation performance, we use 5-fold cross-validation. The data are partitioned into five folds, and in each run, four folds are used for training and the remaining fold for testing. This practice avoids using the same samples for both training and evaluation, which can lead to overly optimistic performance estimates if not controlled properly [15].

D. Heatmap Rendering

Given a denoised EEG segment $\boldsymbol{X} \in R^{C \times T}$ with $C = 16$ channels and $L = 1{,}024$ time samples, we render a pseudo colour heatmap by mapping time samples to the horizontal axis and channels to the vertical axis, then applying a standardised colour scale. This direct amplitude-to-colour approach follows recent ASD EEG work that converts windowed EEG into heatmaps for vision backbones [21]. Fig. 5 illustrates the signal-to-heatmap pipeline.

Channels are arranged in rows according to the same 10 20 montage described in Section II A, so that the ordering preserves coarse bilateral and topographic structure across frontal, temporal, central, parietal, and occipital regions [29], [37], [38]. Specifically, the 16 EEG channels are arranged along the vertical axis in a fixed and reproducible order following the standard 10–20 electrode naming convention (e.g., Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, P4, P8), and this ordering is kept consistent across all subjects and windows. Although no spatial interpolation is performed, this anatomically informed ordering preserves coarse anterior–posterior and left–right relationships that can be exploited by convolutional neural networks. Each row linearly samples the corresponding channel waveform across the 4-second window, and columns correspond to uniformly spaced time bins. This yields a fixed-channel-by-time representation suitable for downstream vision backbones such as ConvNeXt, Swin, EfficientNet B0, GoogLeNet, and ShuffleNet [17]. To ensure comparability across images and avoid information leakage, colour limits v_{min} and v_{max} are computed only from the training fold, either globally or per channel as specified a priori, and then applied unchanged to validation and test windows in that fold. Given an entry $X(c, t)$, the amplitude is first clipped to the predefined

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

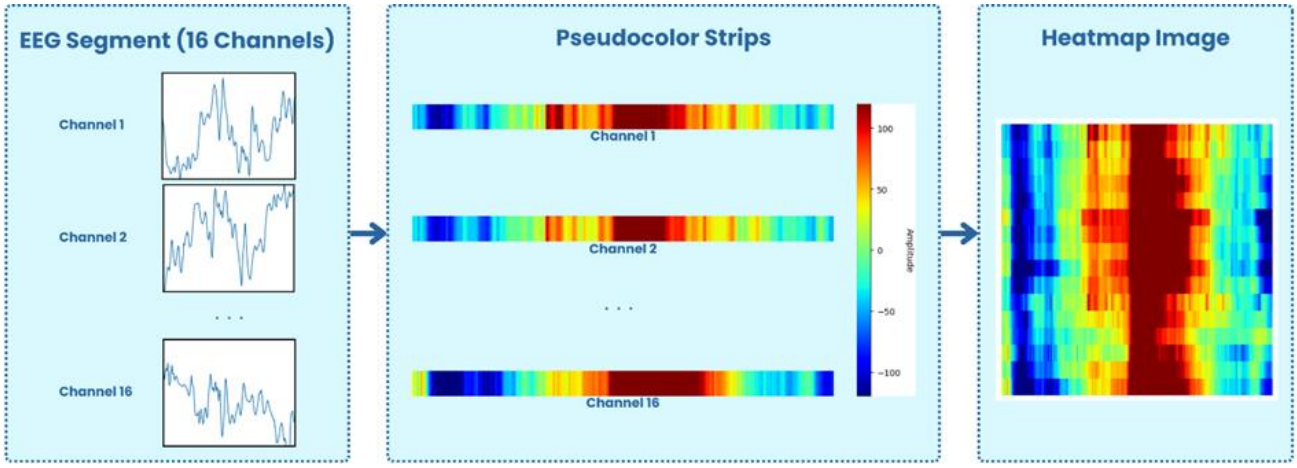


Fig. 5. Signal to heatmap pipeline. DWT denoised EEG windows are mapped to channel by time matrices, normalised using training derived colour limits, and rendered as pseudo colour heatmaps that preserve the channel order across rows and the temporal evolution along columns.

range according to Eq. (7) and subsequently normalised to the interval $[0,1]$ using Eq. (8).

$$\hat{X}(c,t) = \min(\max(X(c,t), v_{min}), v_{max}), \quad (7)$$

$$\tilde{X}(c,t) = \frac{\hat{X}(c,t) - v_{min}}{v_{max} - v_{min}} \quad (8)$$

The normalised matrix $\tilde{X} \in [0,1]^{C \times L}$ is finally mapped to image intensities via a perceptually uniform colormap $g: [0,1] \rightarrow [0,1]^3$, so that each pixel has an RGB triplet $I(c,t) = g(\tilde{X}(c,t))$. A perceptually ordered pseudo-colour mapping was applied using the jet colormap implemented in Matplotlib, where low amplitudes are encoded in blue tones and high amplitudes in yellow–red tones. This choice builds on our prior ASD EEG heatmap work and facilitates the visual interpretation of amplitude variations across channels and time [21]. Fold-wise standardisation in terms of v_{min} and v_{max} stabilises learning and prevents target information from seeping into preprocessing, and consistent EEG to image encodings have been reported to improve deep model reliability [17], [18].

As a concrete example, Fig. 5 shows the conversion of a single 4-second, 16-channel EEG window into a channel-by-time heatmap. The waveform panel ("DWT EEG Signal") displays the denoised multichannel segment arranged in 10 20 orders, while the "Heatmap Image" panel shows the corresponding pseudo colour representation obtained by mapping each channel to a row with global colour limits estimated from training data only. Columns encode time samples, and rows encode channels, producing a fixed canvas that maintains inter-channel relationships while exposing temporal patterns in a form that is convenient for image models [17], [18].

For compatibility with standard computer vision backbones pre-trained on ImageNet, the single-channel heatmap is replicated across three channels to form an RGB image and resized to the backbone input resolution (for example, 224 by 224 pixels) using antialiased interpolation [20], [22], [24]. We apply conservative augmentations that preserve the temporal and spectral structure of the signal, such as slight brightness and contrast jitter and small cutout, and avoid strong geometric warps that could distort time or frequency content. This aligns with recent literature showing that image like EEG encodings, including heatmaps, topographic maps, and connectivity maps, are effective inputs for convolutional and transformer-based classifiers [17], [18], [43], [44]. After rendering, the RGB heatmaps serve as inputs for the heavy and light vision backbones described in the following section.

E. Deep Learning Architectures

The proposed framework uses five pretrained deep learning architectures that are adapted to the channel-by-time EEG heatmaps: one heavyweight convolutional model (ConvNeXt Tiny), one heavyweight transformer-style model (Swin Transformer Tiny), and three lightweight convolutional networks (EfficientNet B0, GoogLeNet, and ShuffleNetV2 0.5x). Rather than proposing a new architecture from scratch, the present work focuses on how these complementary inductive biases behave on a unified EEG to heatmap pipeline and how heavy light decision fusion can exploit their differences under strict cross-validation [22], [23], [24], [25], [26].

All architectures receive the same type of input: $224 \, x \, 224$ RGB images obtained from 4-second windows as described before. Conceptually, the horizontal axis encodes time and the vertical axis encodes ordered EEG channels, so that convolutional kernels or local attention windows can jointly capture intra-channel temporal patterns and inter-channel interactions. By keeping the input representation fixed and normalising the classifier heads, we attribute

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

Table 1. CNN backbones and classifier heads used in this study.

Layer Number	Layer Type	Size Output			
		EfficientNet-B0	GoogLeNet	ShuffleNetV2	ConvNeXt
0	Input layer	$224 \times 224 \times 3$			
1	State-of-the-art convolution layers	$7 \times 7 \times 1280$	$7 \times 7 \times 1024$	$7 \times 7 \times 1024$	$7 \times 7 \times 768$
2	Global average pooling (GAP)	1280	1024	1024	768
3	Dense layer (256, ReLU)	256			
4	Fully connected layer	2			

differences in performance primarily to the architectures and their interaction with the EEG-specific image structure, rather than to arbitrary changes in downstream layers.

1. Convolutional backbones and inductive bias for EEG heatmaps

We adopt four convolutional families to span a range of model capacities and computational costs.

a) EfficientNet B0 uses depthwise separable MBConv blocks with squeeze and excitation and a compound scaling rule. Small 3×3 kernels in early layers primarily capture local temporal changes within channels and short-range correlations between neighbouring rows in the heatmap, while deeper layers aggregate progressively larger temporal contexts. Prior work has reported strong performance of EfficientNet B0 in EEG based seizure detection and other biomedical signal classification tasks [25].

b) GoogLeNet (Inception v1) employs multi branch Inception modules where 1×1, 3×3, and 5×5 filters operate in parallel. On channel by time heatmaps, these branches can be interpreted as capturing short temporal edges (1×1 and 3×3) and broader temporal structures spanning multiple cycles of low frequency oscillations (5×5), while the across row dimension implicitly encodes interactions between nearby electrodes. The multi scale design is thus well aligned with the multiscale nature of EEG rhythms and artefacts [22], [23].

c) ShuffleNetV2 0.5x is a highly efficient architecture built from channel split, channel shuffle, and depthwise separable convolutions. The channel shuffle operation encourages information mixing across feature groups, which is important when each feature channel corresponds to combinations of EEG channels and time. Because of its small parameter count and low multiply accumulate operations, ShuffleNetV2 is an attractive candidate for resource constrained ASD screening devices that require near real time inference [24].

d) ConvNeXt Tiny modernises the ResNet style design with large kernel depthwise convolutions (7×7), inverted bottlenecks, and simplified stage wise downsampling [25], [26]. Large depthwise kernels along the time axis allow ConvNeXt to integrate information over longer temporal horizons at relatively shallow depths, which can be beneficial for detecting slower oscillatory trends and cross-channel patterns. In preliminary experiments, ConvNeXt consistently ranked among the strongest heavy backbones on ASD heatmaps in terms of F1 score, underscoring its role as a primary heavy model in the fusion stage.

In all convolutional cases, we retain the ImageNet pretrained stem and block structure and replace the original classification layer with a lightweight EEG-specific head. The terminal feature map is processed by global average pooling, followed by a 256 unit fully connected layer with ReLU activation and dropout (dropout rate in $\{0.2, 0.3\}$ selected once on the validation folds and then fixed across models), and finally by a fully connected layer with two output logits (ASD and TD). This shared head design, summarised in Table 1 and Fig. 6, standardises classifier capacity across architectures and facilitates fair comparison and fusion.

2. Swin Transformer Tiny for structured EEG images

To complement the convolutional encoders, we include Swin Transformer Tiny as a hierarchical transformer-style backbone for EEG heatmaps [18], [22], [23]. The modified Swin-based classifier used in this work is summarised in Fig. 7. Swin first divides the 224×224 RGB image into non-overlapping patches (for example 4×4 pixels), projects each patch into a low-dimensional embedding, and then applies multi-head self-attention within fixed-size windows (for example 7×7 patches). Between stages, windows are shifted and patches are merged, creating a pyramid of feature maps with decreasing spatial resolution and increasing channel depth. On channel-by-time heatmaps, local attention windows cover short temporal spans and small groups of neighbouring channels, allowing Swin to adaptively weight interactions between electrodes across time

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

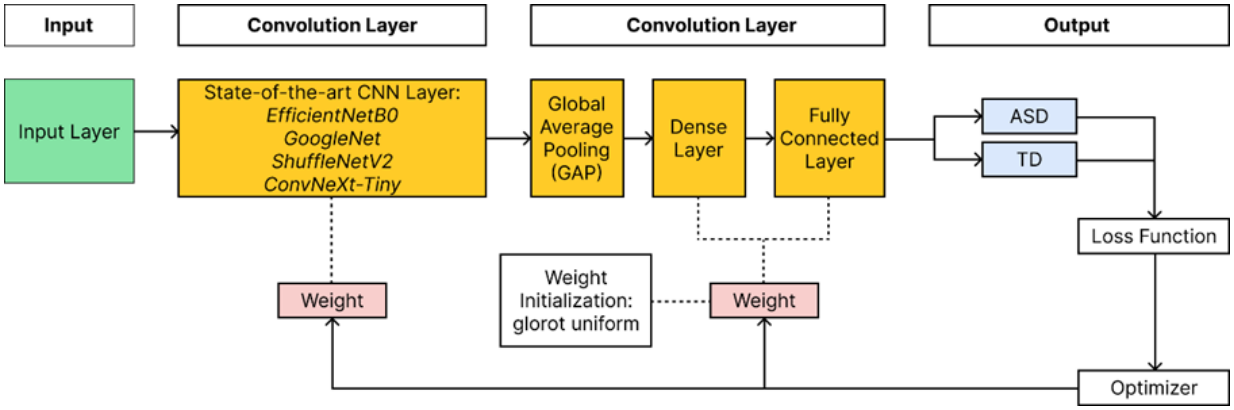


Fig. 6. Modified convolutional neural network (CNN) architectures used in this study.

without the rigid locality of convolutional kernels. The shifted window mechanism then enables information flow across adjacent temporal blocks and channel bands while keeping computational cost subquadratic in the number of patches. The final stage feature map is pooled to obtain a 768 dimensional representation, which is fed to a fully connected layer with two output units. As with the convolutional models, we start from ImageNet pretrained weights and fine-tune the entire network on the ASD EEG heatmaps.

3. Training protocol and hyperparameter selection

To address concerns about fairness and replicability in comparative deep learning studies, all five architectures are trained under a unified protocol with minimal, explicitly documented hyperparameter tuning [14], [15], [29]. For each cross-validation fold:

a) Inputs are 224×224 RGB heatmaps standardised using fold-specific training statistics.
b) We use the AdamW optimizer with an initial learning rate in $\{1e-4, 3e-4\}$ and weight decay $1e-4$. A cosine learning rate schedule with linear warmup over the first five epochs is applied. The learning rate setting is selected once on validation folds for a reference architecture, then reused unchanged for the others.
c) Mini batches contain 32 heatmaps, drawn by sampling windows uniformly from the training subjects. To reduce class imbalance at the window level, ASD and TD windows are oversampled so that mini-batches are approximately balanced.
d) Training is run for up to 80 epochs with early stopping based on validation F1 score, with a patience of 10 epochs. The model state that achieves the highest validation F1 is used for evaluation on the held out test fold.
e) Random seeds for weight initialisation and data shuffling are fixed and reported so that the experiments are reproducible.

No model-specific tricks such as custom learning rate schedules, auxiliary heads, or aggressive data augmentation are employed. This design choice intentionally trades off some potential peak performance for transparency k comparability: all architectures are exposed to the same optimisation regime, data preprocessing, and stopping criteria, enabling interpretation of differences in ASD detection performance in terms of architectural biases rather than idiosyncratic training recipes.

F. Decision Level Fusion

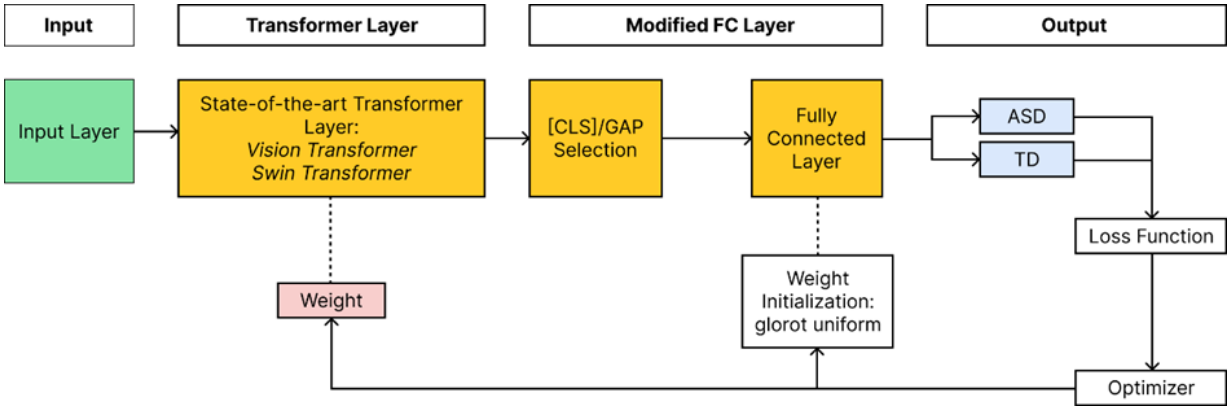


Fig. 7. Modified transformer-based architecture used in this study.

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

To balance accuracy and efficiency while exploiting complementary inductive biases, we integrate one heavyweight architecture with one lightweight architecture at the decision level using weighted soft voting, a strategy that has been widely adopted in biomedical ensemble learning [26], [27], [28]. Let $P_k^H(x)$ denote the posterior probability for class $k \in \{ASD, TD\}$ produced by a heavyweight model given an input heatmap x, and let $P_k^L(x)$ denote the corresponding posterior produced by a lightweight model. The fused posterior probability is then obtained via weighted soft voting, as defined in Eq. (9), where the fusion weight $\alpha \in [0,1]$ controls the relative contribution of the heavyweight and lightweight branches. The final class label is determined by selecting the class with the maximum fused posterior probability, as given in Eq. (10).

$$P_k^F(x) = \alpha P_k^H(x) + (1-\alpha) P_k^L(x), \quad (9)$$

$$\hat{y}(x) = \arg\max_k P_k^F(x). \quad (10)$$

Here, the heavyweight branch is either ConvNeXt Tiny or Swin Transformer Tiny, and the lightweight branch is one of EfficientNet B0, GoogLeNet, or ShuffleNetV2 0.5x, reflecting the heavy versus light families discussed before [22], [23], [24], [25], [26]. Within each cross-validation fold, the fusion weight α is selected on the validation split only, using a coarse grid $\alpha \in \{0.3, 0.5, 0.7\}$ to maximise the mean F1 score, in line with common practice for tuning ensemble weights on held-out data [26], [27]. The selected α is then fixed for the corresponding test split, which prevents information from the test subjects from influencing the fusion rule and avoids evaluation bias [14], [15]. Fig. 8 illustrates the special case $\alpha = 0.5$, where heavy and light posteriors are combined with equal weight.

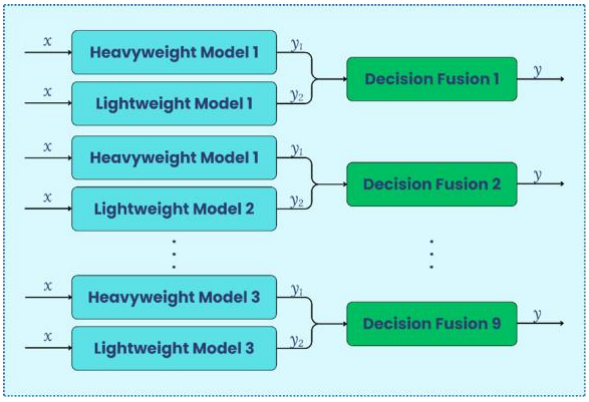


Fig. 8. Decision level fusion between heavy and light models. Heavy and light posteriors are merged by weighted soft voting with fusion weight α to obtain the final ASD versus TD decision.

In this study, we evaluate six heavy-light pairs using the same preprocessing and training pipeline as the single architectures, namely ConvNeXt Tiny + EfficientNet B0, ConvNeXt Tiny + GoogLeNet, ConvNeXt Tiny + ShuffleNetV2 0.5x, Swin Tiny + EfficientNet B0, Swin Tiny + GoogLeNet, and Swin Tiny + ShuffleNetV2 0.5x. These combinations are chosen to probe how pairing different inductive biases and capacities affects ASD detection on the same EEG to heatmap representation [22], [23], [24], [25]. For each pair, fused posteriors are computed for all windows in the test fold and converted to hard labels via $\hat{y}(x)$. We construct confusion matrices on the held out test data and derive Accuracy, Precision, Recall or Sensitivity, Specificity, and F1 score, following standard evaluation practice in EEG based classification and medical image analysis [29], [41], [45]. Metrics are then averaged across folds. Ties in hard decisions are broken by the larger fused posterior value. Unless otherwise stated, we do not apply post hoc calibration on test predictions; when calibration is explored, it is fitted on validation outputs only, consistent with recommended procedures for probability calibration in biomedical models [30], [45].

Decision-level fusion in probability space has been shown to improve robustness and accuracy in various biomedical signal and image classification tasks, including EEG, cardiac sounds, and multimodal medical imaging [26], [27], [28], [30]. Studies on medical imaging and multimodal decision fusion report that weighted soft voting between heterogeneous learners often yields better performance than any single model without substantially increasing model complexity [31], [32], [44], [45]. In the context of EEG-based ASD detection, pairing a heavyweight architecture with a lightweight encoder in this manner aims to retain high sensitivity on ASD windows while controlling false positives and maintaining an acceptable computational footprint for potential deployment [24], [25], [27].

G. Cross Validation and Evaluation Metrics

We evaluate all architectures using K-fold cross-validation with K = 5 and class stratification between ASD and TD, a protocol widely adopted in EEG-based diagnostic studies to obtain stable performance estimates from limited cohorts [14], [15], [43]. The overall scheme is illustrated in Fig. 9, where each fold is used once as the test set, while the remaining folds form the training and validation pools. The 16 available subjects are first partitioned into five folds, with each fold containing three or four subjects, while preserving the overall ASD-to-TD ratio. In each run, three folds are used for training, one for validation, and one for testing; the roles of the folds are rotated until every subject appears in the test set exactly once. Splits are defined at the subject level, and all windows belonging to a given subject are assigned to the same fold, which avoids identity leakage between training, validation, and test data and addresses concerns raised in recent work on data leakage in psychiatric EEG deep learning [14], [16], [17].

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

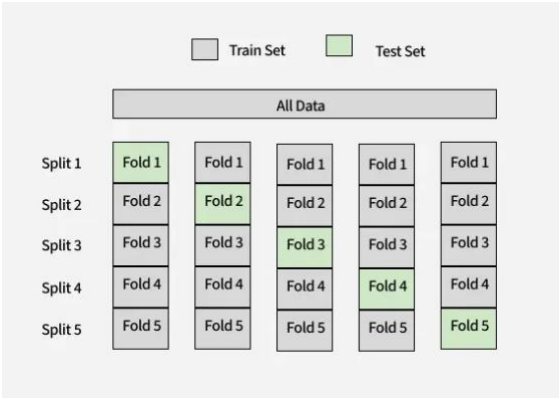


Fig. 9. 5-fold cross validation scheme. All data are partitioned into five disjoint folds.

For each fold, all data-dependent preprocessing parameters are estimated from the training portion only and then applied unchanged to the test. This includes the wavelet thresholding configuration, normalisation factors, and the colour limits used to render EEG heatmaps, consistent with recommendations for leakage-aware EEG preprocessing [6], [18]. All deep learning architectures ingest identically preprocessed $224\ x\ 224$ RGB heatmaps with the same augmentation policy. Optimiser, batch size, maximum number of epochs, and regularisation settings are fixed a priori and shared across architectures so that performance differences can be attributed primarily to representational capacity and inductive bias rather than to model-specific optimisation tricks [29], [45], [46]. This experimental design follows recent comparative studies in biomedical signal classification and multimodal fusion, where unified training pipelines are emphasised to ensure fair model comparison [28], [44], [47], [48].

With ASD treated as the positive class, performance is evaluated using standard metrics derived from the confusion matrix of each test fold. Let true positives (TP) denote the number of ASD windows correctly classified as ASD, true negatives (TN) denote the number of TD windows correctly classified as TD, false positives (FP) correspond to TD windows incorrectly classified as ASD, and false negatives (FN) correspond to ASD windows incorrectly classified as TD. Based on these quantities, Accuracy and F1-score are computed using Eq. (11) and Eq. (12), while Precision, Recall (Sensitivity), and Specificity are defined in Eq. (13), Eq. (14), and Eq. (15).

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP + TN + FP + FN}} \tag{11}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

To better characterise clinical relevance, we also report Precision, Recall, or Sensitivity, and Specificity, which distinguish between false positives and false negatives and are standard in diagnostic performance assessment:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP + FP}} \tag{13}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP + FN}} \tag{14}$$

$$Specificity = \frac{\mathrm{TN}}{\mathrm{TN + FP}} \tag{15}$$

These metrics are computed for each architecture and each fusion pair on every test fold. The final reported performance is given as the mean plus or minus standard deviation across the five folds, as is common in EEG and medical imaging benchmarks [29], [41], [43]. In the results section, we additionally analyse confusion matrices and perform paired statistical comparisons across folds, for example, using repeated measures analysis of variance and adjusted pairwise tests, to assess whether observed differences between single architectures and heavy-light fusions are likely to be due to chance [44], [45], [46], [47], [48].

III. Result

A. Single Model Performance

Table 2 summarises the 5-fold cross-validation performance of the five individual architectures on the EEG heatmaps. For each model, we report the mean and standard deviation of Accuracy, Precision, Recall, Sensitivity, Specificity, and F1 score across the test folds, with ASD treated as the positive class. These figures serve as the baseline against which the heavy light fusion schemes in the next subsection are evaluated. Among the single models, ConvNeXt Tiny attains the best overall performance, with an accuracy of around 97.25 ± 0.19 percent and an F1 score of 97.10 ±

Table 2. 5-Fold Test Performance of Single Backbones (mean ± SD).

Model	Model Size (MB)	Param (M)	FLOPs	Acc (%)	P (%)	R (%)	S (%)	F1 (%)
EfficientNet	16	~5.3	~0.39	94.90 ± 0.34	96.81 ± 0.29	92.30 ± 0.64	97.25 ± 0.25	94.50 ± 0.38
GoogleNet	22.1	~6.8	~1.5	95.09 ± 0.30	**97.85 ± 0.21**	91.67 ± 0.62	98.18 ± 0.19	94.66 ± 0.34
ShuffleNet	**1.5**	**~1.4**	**~0.14**	94.48 ± 0.18	94.52 ± 0.18	93.82 ± 0.47	95.08 ± 0.19	94.17 ± 0.20
SwinT	107.8	~28.3	~4.5	96.34 ± 0.13	96.35 ± 0.20	95.93 ± 0.22	96.72 ± 0.18	96.14 ± 0.14
ConvNeXt	108.7	~28.6	~4.64	**97.25 ± 0.19**	97.39 ± 0.28	**96.81 ± 0.17**	**97.65 ± 0.25**	**97.10 ± 0.20**

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

0.20 percent. Swin Transformer Tiny forms the second tier, with an accuracy of 96.34 ± 0.13 percent and an F1 score of 96.14 ± 0.14 percent. In both cases, Recall on ASD windows are close to or above 95 percent, indicating that the heavyweight architectures tend to miss fewer ASD segments while maintaining high Specificity on TD windows. The lightweight architectures show a slightly different profile. EfficientNet B0 and GoogLeNet cluster in the mid 94 percent range for F1 score: EfficientNet B0 reaches 94.50 ± 0.38 percent, with a relatively balanced Precision and Recall, whereas GoogLeNet achieves the highest Precision among the light models (97.85 ± 0.21 percent) at the cost of lower Recall (91.67 ± 0.62 percent). ShuffleNetV2 0.5x yields the lowest F1 score of 94.17 ± 0.20 percent, driven mainly by modestly reduced Recall, but its Accuracy and F1 remain within about half a percentage point of the other lightweight models.

Table 2 also lists model sizes, highlighting the trade-off between performance and computational footprint. ConvNeXt Tiny and Swin Tiny are the largest models, with sizes around 108.7 MB and 107.8 MB, respectively, whereas ShuffleNetV2 0.5x is by far the most compact at 1.5 MB. EfficientNet B0 (16 MB) and GoogLeNet (22.1 MB) occupy intermediate positions. The fact that EfficientNet B0 and GoogLeNet approach the F1 scores of Swin Tiny, and lie within roughly 2 percentage points of ConvNeXt Tiny, suggests that the proposed DWT plus heatmap preprocessing pipeline provides a strong and consistent input representation across architectures, rather than relying solely on large model capacity to extract useful features. In addition to classification performance, Table 2 also reports quantitative measures of computational footprint in terms of the number of parameters and theoretical FLOPs for each single backbone. As expected, lightweight models such as ShuffleNetV2 and EfficientNet-B0 exhibit substantially lower parameter counts (≈1–5 M) and FLOPs (<0.5 GFLOPs), whereas heavyweight architectures (ConvNeXt Tiny and Swin Transformer Tiny) operate in the range of ≈28 M parameters and ≈4.5 GFLOPs. These results provide a quantitative basis for assessing the accuracy–efficiency trade-off, showing that the proposed EEG-to-heatmap representation enables competitive performance even for architectures with markedly reduced computational complexity.

Overall, the single model results establish a clear hierarchy and reveal complementary strengths. Heavyweight architectures provide the highest F1 scores and the best sensitivity to ASD windows, but at a higher computational cost, while lightweight architectures offer competitive performance with much smaller footprints. This motivates the decision-level fusion experiments in the next subsection, where we investigate whether pairing heavy and light models can further improve robustness and F1 score without incurring the full cost of deploying multiple independent systems.

B. Heavy Light Fusion Performance

Table 3 summarises the 5-fold cross-validation performance of the six heavy-light fusion pairs. For each pair, we report the same metrics as for the single architectures, namely Accuracy, Precision, Recall or Sensitivity, Specificity, and F1 score, averaged over the test folds with ASD as the positive class. The best heavy light combination is ConvNeXt + ShuffleNet, which reaches 99.56 ± 0.05% Accuracy and 99.53 ± 0.05% F1 across folds. This represents an absolute gain of about 2.4 percentage points in F1 compared with the ConvNeXt single model, indicating that soft voting recovers a nontrivial number of misclassified windows. Importantly, this improvement does not come at the cost of Precision: for ConvNeXt-based fusions, Precision remains at or above 99%, while Recall and Specificity both move into the high 99% range. Among the Swin-based pairs, Swin + ShuffleNet yields the strongest result, with Accuracy and F1 around 98.8% and 98.7%, respectively, lifting the transformer baseline into a regime comparable with the best single convolutional backbone. Beyond performance gains, Table 3 further extends this analysis by reporting the total parameter count and FLOPs of each heavy–light fusion model, computed as the sum of the two constituent backbones. This explicitly quantifies the additional computational cost introduced by decision-level fusion. Notably, the best-performing fusion (ConvNeXt + ShuffleNet) increases computational complexity only marginally compared with ConvNeXt Tiny alone (≈4.64 vs. ≈4.50

Table 3. 5-Fold Test Performance of Heavy–Light Fusion (α = 0.5), (mean ± SD).

Model	Model Size (MB)	Param (M)	FLOPs	Acc (%)	P (%)	R (%)	S (%)	F1 (%)
ConvNeXt + EffNet	124.7	~33.9	~4.89	99.19 ± 0.08	99.17 ± 0.13	99.12 ± 0.13	99.25 ± 0.12	99.14 ± 0.09
ConvNeXt + GoogleNet	130.8	~35.4	~6.00	99.21 ± 0.10	99.31 ± 0.11	99.02 ± 0.17	99.38 ± 0.10	99.17 ± 0.10
ConvNeXt + ShuffleNet	**110.2**	~30.0	**~4.64**	**99.56 ± 0.05**	**99.46 ± 0.20**	**99.61 ± 0.13**	**99.51 ± 0.19**	**99.53 ± 0.05**
Swin + EffNet	123.8	~33.6	~4.89	98.70 ± 0.10	98.58 ± 0.20	98.68 ± 0.13	98.71 ± 0.18	98.63 ± 0.10
Swin + GoogleNet	129.9	~35.1	~6.00	98.60 ± 0.14	98.58 ± 0.20	98.48 ± 0.20	98.71 ± 0.19	98.53 ± 0.15
Swin + ShuffleNet	109.3	**~29.7**	**~4.64**	98.77 ± 0.21	98.63 ± 0.22	98.77 ± 0.24	98.76 ± 0.20	98.70 ± 0.22

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

GFLOPs), while yielding a substantial improvement in F1-score of approximately 2.4 percentage points. This demonstrates that the observed performance gains are achieved at a relatively low computational overhead.

To assess whether the performance improvement from heavy–light fusion is statistically significant, we conducted paired statistical tests across the five cross-validation folds comparing the strongest single backbone (ConvNeXt Tiny) and the best fusion model (ConvNeXt + ShuffleNet). A paired t-test revealed that the fusion model significantly outperforms the single model in both Accuracy ($p < 1 \times 10^{-5}$) and F1-score ($p < 1 \times 10^{-5}$). A complementary Wilcoxon signed-rank test showed consistent directional improvements across all folds, although statistical significance was not reached due to the small number of folds. These results indicate that the observed gains from decision-level fusion are systematic rather than attributable to chance.

A consistent pattern in Table 3 is that every heavy light pair improves over its lightweight component on F1 and Recall. Even when a fusion does not strictly surpass the strongest heavy architecture across all metrics,

it typically narrows the gap while retaining a footprint much closer to that of the light model. Standard deviations across folds remain small (on the order of a few tenths of a percentage point), suggesting that the gains from fusion are stable rather than driven by a particular split. From a deployment perspective, the model size column in Table 3 highlights the practicality of the design. Among ConvNeXt fusions, the top-scoring ConvNeXt + ShuffleNet is also the most compact, at about 110.2 MB, compared with 124.7 MB for ConvNeXt + EfficientNet and 130.8 MB for ConvNeXt + GoogLeNet. Decision level aggregation itself adds only a constant time average over two posterior vectors, so inference cost is dominated by the two forward passes. The resulting operating point of "one heavy + one light" therefore offers a favourable accuracy versus footprint trade-off compared with either a single heavy model or heavier ensembles.

Overall, the fusion results indicate that combining heavy and light models at the probability level is not redundant with simply choosing the best single backbone. In several configurations, heavy light pairs establish new best values for F1 and Recall, and even when improvements are modest, they are achieved with negligible extra training effort and minimal additional runtime cost. This motivates a more detailed examination of confusion matrices and error patterns in the next subsection to clarify how fusion alters false negative and false positive profiles relative to the best single model.

C. Confusion Matrices

We analyse confusion matrices for the best single backbone (ConvNeXt Tiny) and the best heavy-light pair (ConvNeXt + ShuffleNet), as shown in Fig. 10, to clarify how fusion modifies the classifiers' error profiles. The figure displays confusion matrices aggregated across the five test folds for both models. As illustrated in Fig. 3, DWT-based denoising substantially suppresses high-frequency artefacts while preserving the underlying oscillatory structure of the EEG signals. Since the proposed heatmap representation is directly constructed from these denoised signals, improvements in signal quality yield cleaner, more structured time–channel images for downstream deep learning. This preprocessing step, therefore, provides a more stable and informative input for the classifiers, which is reflected in the error patterns analysed below.

For the ConvNeXt Tiny baseline, the pooled confusion matrix shows that the model correctly classifies 1,974 ASD windows and 2,203 TD windows, while 66 ASD

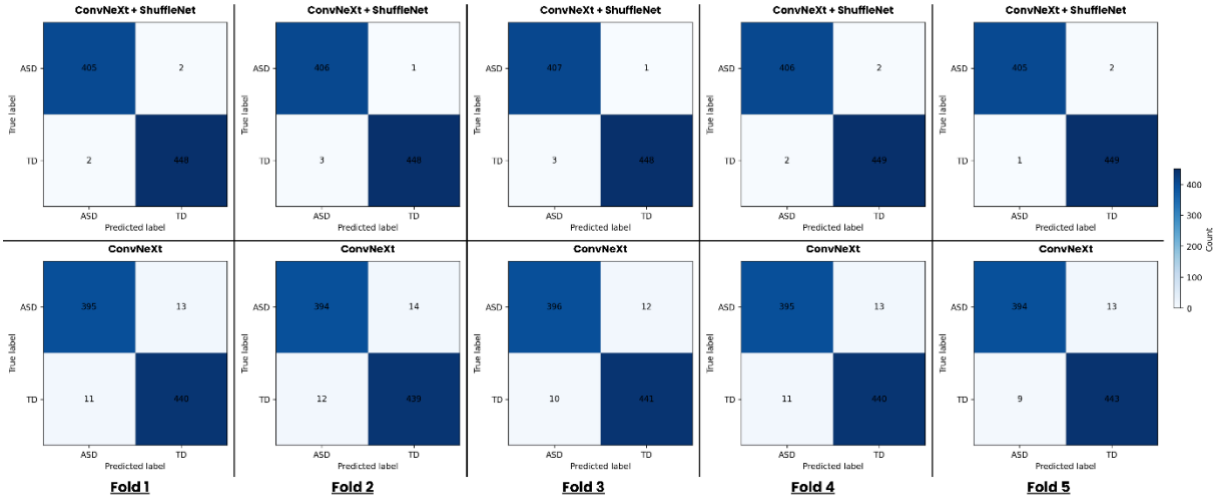


Fig. 10. Per-fold confusion matrices (5-fold). Top row: ConvNeXt + ShuffleNet (α = 0.5); bottom row: ConvNeXt.

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

windows are incorrectly labelled as TD (false negatives) and 53 TD windows are incorrectly labelled as ASD (false positives). These counts correspond to an Accuracy of 97.23 percent, a precision of 97.39 percent, a recall of 96.76 percent, a specificity of 97.65 percent, and an F1 score of 97.07 percent, consistent with the mean values reported in Table 2. The dominant residual error mode is therefore ASD windows misclassified as TD, which is undesirable in a screening setting because it implies missed ASD segments. When ConvNeXt is fused with ShuffleNet, the pooled confusion matrix changes in two important ways. First, the number of correctly detected ASD windows increases to 2,027, while ASD windows misclassified as TD drop from 66 to 8. Second, the number of correctly identified TD windows rises to 2,242, with TD windows misclassified as ASD decreasing from 53 to 11. In metric form, the fused model achieves an accuracy of 99.56 percent, a precision of 99.46 percent, a recall of 99.61 percent, a specificity of 99.51 percent, and an F1 score of 99.53 percent, in line with Table 3. The large reduction in false negatives, accompanied by a simultaneous reduction in false positives, explains the observed gains in both sensitivity and specificity. Across the six heavy light combinations, a similar qualitative pattern is observed in the corresponding confusion matrices. Every fusion reduces the number of false negatives relative to its lightweight component and typically narrows the gap to the corresponding heavyweight architecture. In Swin-based pairs, the absolute reduction in errors is smaller than in ConvNeXt-based fusions, but still sufficient to lift the transformer baseline into the performance regime of the best single convolutional backbone. These observations suggest that the heavy and light models make partially complementary errors on EEG heatmaps, and that probability level fusion leverages this complementarity rather than simply replicating the behaviour of the stronger single model. Taken together, the confusion matrix analysis confirms that the improvements reported in Table 2 and Table 3 are primarily driven by a reduction in missed ASD windows, with only a small residual change in false alarms. This error profile is consistent with the intended use of the models as decision support tools for ASD screening, where prioritising high Recall while maintaining high Specificity is more valuable than maximising overall accuracy alone.

IV. Discussion

The results in this study show that the proposed DWT-based preprocessing and heatmap encoding pipeline, combined with modern deep learning architectures, yields high performance for ASD detection on the KAU EEG dataset. Under 5-fold cross-validation, the best single backbone, ConvNeXt Tiny, achieved an Accuracy of about 97.25 ± 0.19 percent and an F1 score of about 97.10 ± 0.20 percent, with a recall on ASD windows close to or above 95 percent. The best heavy-light fusion, ConvNeXt + ShuffleNet, further improved performance to approximately 99.56 ± 0.05 percent Accuracy and 99.53 ± 0.05 percent F1, with both sensitivity and specificity in the 99 percent range. These gains were obtained under the same preprocessing and training protocol, indicating that the combination of wavelet denoising, windowed heatmap representation, and probability level fusion can significantly improve ASD EEG classification beyond a single strong backbone.

Importantly, the inclusion of explicit computational metrics in Table 2 and Table 3 allows a more objective interpretation of the accuracy–efficiency trade-off. While heavyweight backbones provide the strongest single-model performance, the proposed heavy–light fusion strategy achieves superior accuracy with only a modest increase in computational footprint. In particular, pairing a high-capacity backbone with an ultra-light model such as ShuffleNet preserves most of the efficiency advantages of lightweight architectures, while substantially reducing missed ASD windows. This balance is especially relevant for practical deployment scenarios, where computational resources may be constrained.

Although the reported metrics are computed at the window level, clinical diagnosis is ultimately made at the subject level. In practice, window-level predictions can be aggregated into a subject-level decision using simple, interpretable rules, such as majority voting or averaging across all windows for a subject. Given the high consistency of window-level predictions observed in this study, such aggregation is expected to stabilise subject-level decisions and reduce the impact of sporadic misclassified windows. A systematic evaluation of subject-level aggregation strategies is an important direction for future work toward clinical deployment.

The single-model results in Table 2 suggest a clear hierarchy consistent with the architectures' representational capacity. Heavyweight models, ConvNeXt Tiny and Swin Transformer Tiny, consistently occupy the upper end of the performance range, with F1 scores in the mid- to high-90s and relatively high Recall on ASD windows. This indicates that deeper hierarchies with larger effective receptive fields can capture more complex spatial-temporal patterns in the EEG heatmaps, echoing previous observations in medical imaging and physiological signal classification [21], [22], [24], [25], [39]. Lightweight architectures, particularly EfficientNet B0 and GoogLeNet, achieve F1 scores around 94 to 95 percent, only a few percentage points below Swin and within roughly 2 percentage points of ConvNeXt, while using far fewer parameters. ShuffleNetV2 0.5x

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

maintains a slightly lower F1 score but remains competitive given its extremely small footprint. This pattern suggests that the proposed DWT plus heatmap preprocessing pipeline provides a robust input representation that all architectures can exploit, rather than requiring highly specialised models to extract useful EEG features.

The heavy light fusion results in Table 3 refine this picture by showing that combining one heavy and one light architecture is not redundant with simply picking the best single model. For the strongest pair, ConvNeXt + ShuffleNet, fusion improves F1 by about 2.4 percentage points relative to ConvNeXt alone and raises Accuracy, Recall, and Specificity into the high 99 percent range. Swin-based pairs show similar trends, with Swin + ShuffleNet lifting the transformer baseline into a regime comparable with the best single convolutional backbone. Importantly, these improvements are achieved without individually tuning the architectures for each fusion and with only a simple weighted soft voting scheme. This behaviour is consistent with ensemble learning theory and prior biomedical studies where probability level fusion between heterogeneous classifiers has been shown to reduce variance and correct complementary errors [25], [26], [27], [42], [44], [46].

The confusion matrix analysis in Section III-C helps explain how these numerical gains arise. For ConvNeXt Tiny, the pooled confusion matrix shows that residual errors are dominated by ASD windows misclassified as TD, leading to a small but non negligible number of false negatives. When ConvNeXt is fused with ShuffleNet, the number of correctly detected ASD windows increases and both false negatives and false positives are substantially reduced. As a result, F1, Recall, and Specificity all increase simultaneously, rather than trading sensitivity against specificity. Similar, though smaller, improvements are observed across the other heavy light combinations. These patterns suggest that the light models do not simply replicate the decisions of the heavy architectures; instead, they introduce alternative decision boundaries that help recover difficult ASD windows and filter out spurious alarms on TD windows, especially under the DWT denoised and heatmap encoded input.

Although a systematic window-level error attribution is beyond the scope of this study, the confusion matrix trends indicate that heavy and light models exhibit partially complementary error patterns. Heavy backbones emphasise global temporal coherence, whereas lightweight models preserve sensitivity to local variations. Their combination at the decision level, therefore, enables correction of missed ASD windows that would otherwise persist in a single-model setting.

From a signal-processing perspective, the strong performance of all models supports the use of discrete wavelet transform denoising as an effective front-end for ASD EEG classification. Wavelet shrinkage, as implemented here, reduces high frequency artefacts and power line noise while preserving slower oscillatory components that are known to be relevant in ASD and other neuropsychiatric conditions. In addition, mapping multichannel windows into channel by time heatmaps allows general purpose vision architectures to exploit both local temporal patterns and inter channel relationships. This is in line with studies that convert EEG into topographic maps, spectrograms, or connectivity images for deep learning-based diagnosis. The present work shows that even a relatively simple amplitude based heatmap, without explicit spectral or connectivity features, can support high performance when combined with a consistent denoising and cross validation protocol.

Compared with alternative EEG-to-image encodings such as topographic maps, spatio-spectral representations, or functional connectivity images, the proposed channel-by-time heatmap retains the raw temporal dynamics of each channel without imposing additional modelling assumptions. Spectral or connectivity-based representations require predefined frequency decompositions or pairwise interactions, which may obscure transient temporal patterns or introduce bias under limited data conditions. In contrast, the channel-by-time encoding preserves fine-grained temporal structure across all channels and allows convolutional and transformer-based vision backbones to learn relevant temporal and inter-channel relationships directly from the data. The strong performance observed in this study suggests that, when combined with robust denoising and leakage-aware validation, this simple non-spectral representation provides a favourable balance between expressiveness, data efficiency, and model generality.

Compared with earlier ASD EEG work, the proposed framework extends wavelet-based pipelines that relied on handcrafted features and shallow classifiers, and complements recent deep learning approaches based on spectrograms or connectivity maps. Previous studies using DWT or stationary wavelet transform with Fisher linear discriminant analysis, support vector machines, or similar classifiers reported promising separation between ASD and control groups on the same KAU cohort and related datasets. Other work has focused on learned representations from spectro temporal images or functional connectivity matrices, often using segment level validation schemes that risk data leakage. In contrast, this study uses wavelet denoising, fixed length overlapping windows, and leakage aware cross validation with explicit control over model capacity and

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

training conditions, and shows that modern CNN and transformer architectures, together with heavy light fusion, can achieve very high F1 scores on the KAU dataset while offering a clear analysis of accuracy versus computational footprint.

To place these findings in the context of recent ASD EEG work on the same cohort, Table 4 summarises representative pipelines and their reported accuracies on the KAU dataset. Early approaches that combined Butterworth filtering, independent component analysis, and k nearest neighbours reached about 85.4 percent accuracy [49], providing an initial machine learning baseline. Subsequent methods based on continuous wavelet transform features with support vector machines and spectrogram or short time Fourier transform representations with classical classifiers reported accuracies around 95 to 95.25 percent [12], [50]. The stationary wavelet-transform plus Fisher linear discriminant analysis framework in [13] achieved about 95 percent accuracy on the KAU data under a subject wise split, demonstrating the effectiveness of wavelet features with shallow models. In comparison, the proposed DWT plus heatmap pipeline combined with heavy light fusion attains approximately 99.56 percent accuracy under subject wise 5-fold cross validation, placing our results at the upper end of reported performance on this dataset.

Importantly, several prior studies summarised in Table 4 rely on segment-level evaluation schemes or do not explicitly address control for subject overlap between the training and test sets, which may lead to optimistic performance estimates. In the present study, although evaluation is conducted at the window level, we explicitly acknowledge this limitation and apply a consistent cross-validation protocol with leakage-aware preprocessing, where normalisation parameters are derived exclusively from training folds. This design improves transparency and reduces indirect information leakage during preprocessing. Nevertheless, a fully subject-wise validation protocol represents an important direction for future work to further assess cross-subject generalization. While direct numerical comparisons should be interpreted with care due to differences in feature design and resampling strategies, the aggregated evidence in Table 4 suggests that integrating DWT denoising, channel-by-time heatmap encoding, and probability level fusion yields a competitive and often stronger alternative to existing wavelet-based and EEG-to-image approaches on the KAU cohort.

Although the proposed framework achieves very high window-level accuracy under subject-wise cross-validation, the relatively small, single-site nature of the KAU dataset (16 subjects) remains an important limitation for generalizability. With limited subject diversity, performance estimates may exhibit higher variance when applied to larger or more heterogeneous cohorts, and part of the observed accuracy may reflect dataset-specific characteristics related to acquisition protocols or site-dependent factors. While leakage-aware preprocessing and subject-wise validation mitigate trivial forms of overfitting, they cannot substitute for validation on independent external datasets. Accordingly, the reported results should be interpreted as evidence of strong within-dataset discrimination rather than definitive clinical generalization. External validation on multi-site cohorts using different EEG systems and acquisition protocols is, therefore, a critical direction for future work.

V. Conclusion

This paper presented an EEG to heatmap pipeline for autism spectrum disorder classification that combines discrete wavelet transform denoising, fixed-length overlapping windows, and channel-by-time pseudo colour images processed by modern deep learning architectures. Five ImageNet pretrained models, ConvNeXt Tiny, Swin Transformer Tiny, EfficientNet B0, GoogLeNet, and ShuffleNetV2 0.5x, were fine-tuned under a unified training protocol and evaluated with 5-fold cross-validation on the KAU ASD EEG dataset. A heavy-light decision fusion scheme was then used to combine one heavyweight and one lightweight backbone at the probability level. The experimental results showed that ConvNeXt Tiny provided the strongest single model baseline, with an accuracy of around 97.25 percent and an F1 score of around 97.10 percent at the window level, while Swin Transformer Tiny formed a close second tier. Lightweight architectures, especially EfficientNet B0 and

Table 4. Comparison of ASD EEG classification performance on the KAU dataset.

Ref	Methods	Accuracy (%)
[12]	CWT features → SVM	95
[13]	SWT (Levels 3/4/6) → FLDA	95
[49]	Butterworth → ICA → KNN	85.4
[50]	Spectrogram/STFT features → classical ML	95.25
This Study	**DWT → ConvNeXt + ShuffleNet**	**99.56**

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

GoogLeNet, achieved F1 scores in the mid-94 percent range with substantially smaller model sizes, indicating that the proposed DWT plus heatmap representation offers a robust and architecture-agnostic input. The best heavy-light pair, ConvNeXt plus ShuffleNet, achieved approximately 99.56 percent Accuracy and 99.53 percent F1, with both sensitivity and specificity in the 99 percent range, and confusion matrix analysis confirmed that most of this gain comes from reducing missed ASD windows without increasing false alarms.

Although these findings are promising, the experiments were conducted on a relatively small, single-site dataset with binary ASD versus TD labels, and the analysis focused on window-level decisions. Future work will therefore focus on validating the proposed pipeline on larger, more diverse ASD EEG datasets, including external test sets, and on extending the image representation to incorporate complementary information such as time frequency structure or connectivity measures. It would also be valuable to design simple subject-level decision rules and lightweight implementations of the best heavy-light pair, so that the present results can move closer to practical decision support for ASD screening and follow-up in real clinical settings. In addition, the generalizability of the optimal heavy-light backbone pairing and fusion weights identified on the KAU dataset should be systematically examined across datasets with different recording paradigms, age groups, and EEG montages. Practical considerations for real-time deployment, including inference latency under overlapping window processing and probability level fusion, also warrant further investigation. Finally, transfer learning from larger general EEG corpora may provide a promising strategy to improve robustness and reduce dataset dependency when adapting the proposed framework to new clinical settings.

Acknowledgment

The authors would like to thank the Directorate of Research and Community Service and the Directorate General of Research and Development at the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia for their support through the Fundamental Research Program – Regular Scheme (FY 2025). We also acknowledge the Institute for Research and Community Service (LPPM), Universitas Syiah Kuala, for its administrative assistance and institutional support in managing the grant and ensuring project compliance.

Funding

This work was funded by the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, under the Fundamental Research Program – Regular Scheme, Fiscal Year 2025, Contract No. 113/C3/DT.05.00/PL/2025.

Data Availability

The dataset used in this study is a third-party dataset obtained from King Abdulaziz University (KAU) Hospital, Jeddah, Saudi Arabia, and is available from the data owner upon reasonable request and permission.

Author Contribution

Melinda conceptualised and designed the study, supervised the research, and led manuscript preparation. Syahrul Gazali and Muliyadi Muliyadi coordinated EEG acquisition, clinical labelling, and dataset organisation. Yuwaldi Away contributed to study conceptualisation, supported data curation, and assisted in interpreting the engineering aspects of the results. Aufa Rafiki implemented the algorithms and experimental pipeline, performed the main data analysis and visualisation, and drafted the initial manuscript. W.K. Wong advised on the modelling strategy and deep learning architectures and critically revised the technical content. Siti Rusdiana conducted the statistical analysis and contributed to the interpretation of the quantitative results and manuscript refinement. All authors reviewed and approved the final version of the manuscript and agreed to be responsible for all aspects of the work.

Declarations

Ethical Approval

This study involves secondary analysis of anonymised EEG data collected previously at King Abdulaziz University (KAU) Hospital, Jeddah, Saudi Arabia, as described in [33]. The original data acquisition was approved by the KAU Ethics Committee, and written informed consent was obtained from all participants or their legal guardians prior to recording. No new data collection or direct contact with human participants was conducted by the present authors, and, in accordance with institutional and national guidelines, additional ethical approval at the authors' institutions was not required.

Consent for Publication Participants.

Consent for data use and publication was obtained by the original investigators at KAU Hospital as part of the initial study, and all EEG recordings were fully anonymised before being shared with the authors. No identifiable personal information is included in this manuscript, and no additional consent for publication was required for this secondary analysis.

Competing Interests

The authors declare no competing interests.

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

References

[1] S. Yun, "Advances, challenges, and prospects of electroencephalography-based biomarkers for psychiatric disorders: a narrative review," *Journal of Yeungnam Medical Science*, vol. 41, no. 4, pp. 261–268, Oct. 2024, doi: 10.12701/jyms.2024.00668.

[2] J. Shan *et al.*, "A scoping review of physiological biomarkers in autism," *Front Neurosci*, vol. 17, 2023, doi: 10.3389/fnins.2023.1269880.

[3] J. Li *et al.*, "Identification of autism spectrum disorder based on electroencephalography: A systematic review," *Comput Biol Med*, vol. 170, p. 108075, 2024, doi: https://doi.org/10.1016/j.compbiomed.2024.108075.

[4] Y. Xu, Z. Yu, Y. Li, Y. Liu, Y. Li, and Y. Wang, "Autism spectrum disorder diagnosis with EEG signals using time series maps of brain functional connectivity and a combined CNN–LSTM model," *Comput Methods Programs Biomed*, vol. 250, p. 108196, 2024, doi: https://doi.org/10.1016/j.cmpb.2024.108196.

[5] J. Rogala *et al.*, "Enhancing autism spectrum disorder classification in children through the integration of traditional statistics and classical machine learning techniques in EEG analysis," *Sci Rep*, vol. 13, no. 1, p. 21748, 2023, doi: 10.1038/s41598-023-49048-7.

[6] S. Phadikar, N. Sinha, R. Ghosh, and E. Ghaderpour, "Automatic Muscle Artifacts Identification and Removal from Single-Channel EEG Using Wavelet Transform with Meta-Heuristically Optimized Non-Local Means Filter," *Sensors*, vol. 22, no. 8, 2022, doi: 10.3390/s22082948.

[7] M. Grobbelaar *et al.*, "A Survey on Denoising Techniques of Electroencephalogram Signals Using Wavelet Transform," Sep. 01, 2022, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/signals3030035.

[8] I. H. Elshekhidris, M. B. Mohamedamien, and A. Fragoon, "Wavelet Transforms for EEG Signal Denoising and Decomposition," 2023.

[9] A. Chaddad, Y. Wu, R. Kateb, and A. Bouridane, "Electroencephalography Signal Processing: A Comprehensive Review and Analysis of Methods and Techniques," Jul. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/s23146434.

[10] S. N. S. S. Daud and R. Sudirman, "Wavelet Based Filters for Artifact Elimination in Electroencephalography Signal: A Review," *Ann Biomed Eng*, vol. 50, no. 10, pp. 1271–1291, 2022, doi: 10.1007/s10439-022-03053-5.

[11] M. Melinda, M. Oktiana, Y. Yunidar, N. H. Nabila, and I. K. A. Enriko, "Classification of EEG Signal using Independent Component Analysis and Discrete Wavelet Transform based on Linear Discriminant Analysis," *International Journal on Informatics Visualization (JOIV)*, vol. 7, no. 3, pp. 830–838, Sep. 2023.

[12] M. Melinda, F. H. Juwono, I. K. A. Enriko, M. Oktiana, S. Mulyani, and K. Saddami, "Application Of Continuous Wavelet Transform and Support Vector Machine for Autism Spectrum Disorder Electroencephalography Signal Classification," *Radioelectronic and Computer Systems*, no. 3(107), pp. 73–90, 2023, doi: 10.32620/reks.2023.3.07.

[13] F. Fahmi, M. Melinda, P. D. Purnamasari, E. Elizar, and A. Rafiki, "Recognition of EEG Features in Autism Disorder Using SWT and Fisher Linear Discriminant Analysis," *Diagnostics*, vol. 15, no. 18, p. 2291, Sep. 2025, doi: 10.3390/diagnostics15182291.

[14] H. T. Lee, H. R. Cheon, S. H. Lee, M. Shim, and H. J. Hwang, "Risk of data leakage in estimating the diagnostic performance of a deep-learning-based computer-aided system for psychiatric disorders," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-43542-8.

[15] G. Brookshire *et al.*, "Data leakage in deep learning studies of translational EEG," *Front Neurosci*, vol. Volume 18-2024, 2024, doi: 10.3389/fnins.2024.1373515.

[16] R. Kessler, A. Enge, and M. A. Skeide, "How EEG preprocessing shapes decoding performance," *Commun Biol*, vol. 8, no. 1, Dec. 2025, doi: 10.1038/s42003-025-08464-3.

[17] N. S. Amer and S. B. Belhaouari, "Exploring new horizons in neuroscience disease detection through innovative visual signal analysis," *Sci Rep*, vol. 14, no. 1, p. 4217, 2024, doi: 10.1038/s41598-024-54416-y.

[18] M. A. Bravo-Ortiz *et al.*, "SpectroCVT-Net: A convolutional vision transformer architecture and channel attention for classifying Alzheimer's disease using spectrograms," *Comput Biol Med*, vol. 181, p. 109022, 2024, doi: https://doi.org/10.1016/j.compbiomed.2024.109022.

[19] E. Vafaei, F. Nowshiravan Rahatabad, S. K. Setarehdan, and P. Azadfallah, "Extracting a Novel Emotional EEG Topographic Map Based on a Stacked Autoencoder Network," *J Healthc Eng*, vol. 2023, 2023, doi: 10.1155/2023/9223599.

[20] N. Bajaj and J. Requena Carrión, "Deep Representation of EEG Signals Using Spatio-Spectral Feature Images," *Applied Sciences (Switzerland)*, vol. 13, no. 17, Sep. 2023, doi: 10.3390/app13179825.

[21] A. Rafiki *et al.*, "Implementation of Vision Transformer for Early Detection of Autism Based

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

on EEG Signal Heatmap Visualization," *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 1, pp. 102–112, 2025.

[22] J. Li, J. Chen, Y. Tang, C. Wang, B. A. Landman, and S. K. Zhou, "Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives," *Med Image Anal*, vol. 85, p. 102762, 2023, doi: https://doi.org/10.1016/j.media.2023.102762.

[23] Z. Li, R. Zhang, Y. Zeng, L. Tong, R. Lu, and B. Yan, "MST-net: A multi-scale swin transformer network for EEG-based cognitive load assessment," *Brain Res Bull*, vol. 206, p. 110834, 2024, doi: https://doi.org/10.1016/j.brainresbull.2023.110834.

[24] Y. Mehmood and U. I. Bajwa, "Brain tumor grade classification using the ConvNext architecture," *Digit Health*, vol. 10, p. 20552076241284920, 2024, doi: 10.1177/20552076241284920.

[25] Y. A. Saadoon, M. Khalil, and D. Battikh, "Predicting Epileptic Seizures Using EfficientNet-B0 and SVMs: A Deep Learning Methodology for EEG Analysis," *Bioengineering*, vol. 12, no. 2, 2025, doi: 10.3390/bioengineering12020109.

[26] S. K. Prabhakar, J. J. Lee, and D.-O. Won, "Ensemble Fusion Models Using Various Strategies and Machine Learning for EEG Classification," *Bioengineering*, vol. 11, no. 10, 2024, doi: 10.3390/bioengineering11100986.

[27] A. Karim, S. Ryu, and I. cheol Jeong, "Ensemble learning for biomedical signal classification: a high-accuracy framework using spectrograms from percussion and palpation," *Sci Rep*, vol. 15, no. 1, p. 21592, 2025, doi: 10.1038/s41598-025-05027-8.

[28] M. Salvi *et al.*, "Multi-modality approaches for medical support systems: A systematic review of the last decade," *Information Fusion*, vol. 103, Mar. 2024, doi: 10.1016/j.inffus.2023.102134.

[29] T. Wu, X. Kong, Y. Zhong, and L. Chen, "Automatic detection of abnormal EEG signals using multiscale features with ensemble learning," *Front Hum Neurosci*, vol. Volume 16-2022, 2022, doi: 10.3389/fnhum.2022.943258.

[30] K. Munadi *et al.*, "A Deep Learning Method for Early Detection of Diabetic Foot Using Decision Fusion and Thermal Images," *Applied Sciences (Switzerland)*, vol. 12, no. 15, Aug. 2022, doi: 10.3390/app12157524.

[31] Md. H. R. Rabbani and S. Md. R. Islam, "Deep learning networks based decision fusion model of EEG and fNIRS for classification of cognitive tasks," *Cogn Neurodyn*, vol. 18, no. 4, pp. 1489–1506, 2024, doi: 10.1007/s11571-023-09986-4.

[32] M. Zakir Ullah and D. Yu, "Grid-tuned ensemble models for 2D spectrogram-based autism classification," *Biomed Signal Process Control*, vol. 93, p. 106151, 2024, doi: https://doi.org/10.1016/j.bspc.2024.106151.

[33] M. J. Alhaddad *et al.*, "Diagnosis Autism by Fisher Linear Discriminant Analysis FLDA via EEG," 2012.

[34] M. Murias, S. J. Webb, J. Greenson, and G. Dawson, "Resting State Cortical Connectivity Reflected in EEG Coherence in Individuals With Autism," *Biol Psychiatry*, vol. 62, no. 3, pp. 270–273, 2007, doi: https://doi.org/10.1016/j.biopsych.2006.11.012.

[35] E. V. Orekhova *et al.*, "Excess of High Frequency Electroencephalogram Oscillations in Boys with Autism," *Biol Psychiatry*, vol. 62, no. 9, pp. 1022–1029, Nov. 2007, doi: 10.1016/j.biopsych.2006.12.029.

[36] R. Coben, A. R. Clarke, W. Hudspeth, and R. J. Barry, "EEG power and coherence in autistic spectrum disorder," *Clinical Neurophysiology*, vol. 119, no. 5, pp. 1002–1009, May 2008, doi: 10.1016/j.clinph.2008.01.013.

[37] Y. Xia, K. Li, D. Li, J. Nan, and R. Lu, "An Improved VMD and Wavelet Hybrid Denoising Model for Wearable SSVEP-BCI," 2024. [Online]. Available: www.ijacsa.thesai.org

[38] D. L. Donoho, "De-Noising by Soft-Thresholding," 1995.

[39] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000, doi: 10.1109/83.862633.

[40] Y. Huang, P. Wen, B. Song, and Y. Li, "Real-Time Depth of Anaesthesia Assessment Based on Hybrid Statistical Features of EEG," *Sensors*, vol. 22, no. 16, 2022, doi: 10.3390/s22166099.

[41] W. Liu, K. Jia, and Z. Wang, "Graph-based EEG approach for depression prediction: integrating time-frequency complexity and spatial topology," *Front Neurosci*, vol. Volume 18-2024, 2024, doi: 10.3389/fnins.2024.1367212.

[42] L. Cao *et al.*, "A Novel Deep Learning Method Based on an Overlapping Time Window Strategy for Brain–Computer Interface-Based Stroke Rehabilitation," *Brain Sci*, vol. 12, no. 11, Nov. 2022, doi: 10.3390/brainsci12111502.

[43] S. Y. Ke *et al.*, "Classification of autism spectrum disorder using electroencephalography in Chinese children: a cross-sectional retrospective study.," *Front Neurosci*, vol. 18, p. 1330556, 2024, doi: 10.3389/fnins.2024.1330556.

[44] T. Xu, Y. Zhou, Z. Hou, and W. Zhang, "Decode Brain System: A Dynamic Adaptive Convolutional Quorum Voting Approach for Variable-Length

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

EEG Data," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/6929546.
[45] J. Duan, J. Xiong, Y. Li, and W. Ding, "Deep learning based multimodal biomedical data fusion: An overview and comparative review," *Information Fusion*, vol. 112, p. 102536, 2024, doi: https://doi.org/10.1016/j.inffus.2024.102536.
[46] Y. Dong *et al.*, "Subject-Independent EEG Classification of Motor Imagery Based on Dual-Branch Feature Fusion," *Brain Sci*, vol. 13, no. 7, 2023, doi: 10.3390/brainsci13071109.
[47] I. Jemal, L. Abou-Abbas, K. Henni, A. Mitiche, and N. Mezghani, "Domain adaptation for EEG-based, cross-subject epileptic seizure prediction," *Front Neuroinform*, vol. Volume 18-2024, 2024, doi: 10.3389/fninf.2024.1303380.
[48] A. M. Alghamdi, M. U. Ashraf, A. A. Bahaddad, K. A. Almarhabi, W. A. Al Shehri, and A. Daraz, "Cross-subject EEG signals-based emotion recognition using contrastive learning," *Sci Rep*, vol. 15, no. 1, p. 28295, 2025, doi: 10.1038/s41598-025-13289-5.
[49] L. Shi *et al.*, "TFSNet: A Time–Frequency Synergy Network Based on EEG Signals for Autism Spectrum Disorder Classification," *Brain Sci*, vol. 15, no. 7, Jul. 2025, doi: 10.3390/brainsci15070684.
[50] M. N. A. Tawhid, S. Siuly, and H. Wang, "Diagnosis of autism spectrum disorder from EEG using a time-frequency spectrogram image-based approach," *Electron Lett*, vol. 56, no. 25, pp. 1372–1375, Dec. 2020, doi: 10.1049/el.2020.2646.

Author Biography

Melinda was born in Bireuen, Aceh, on June 10, 1979. She received a B.Eng degree from the Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh in 2002. She completed her master's degree at the Faculty of Electrical Department, University of Southampton, United Kingdom, with a concentration in field study of Radio Frequency Communication Systems in 2009. She has already completed her Doctoral degree at the Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, in February 2018. She has been with the Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala since 2002. She is also a member of IEEE. Her research interests include multimedia signal processing and fluctuation processing. She can be contacted at email: melinda@usk.ac.id.

Syahrul Gazali was born in Banda Aceh in 1962. He received the M.D. degree from Andalas University, Padang, in 1988, and the Ph.D. degree from Gadjah Mada University, Yogyakarta, in 2013. He joined the Faculty of Medicine, Syiah Kuala University, Banda Aceh, as a Lecturer in 1989. He completed the Neurology specialist training with the University of Indonesia, Jakarta, in 1997, and earned the Stroke Consultant credential from the Indonesian Collegium of Neurology in 2012. In 2021, he obtained Neurovascular Subspecialist Certification and was promoted to a professor of Neurology. His leadership roles include Assistant Dean, two terms as Dean, and Director of RSUD Dr. Zainoel Abidin. He has been the Chair of Indonesian Neurology Collegium since 2023, and the Director of human resources, education, and research with the Mahar Mardjono National Brain Center Hospital, Jakarta, since August 2024. His research focuses on cerebrovascular disease, with numerous publications and invited talks at national and international forums, including the World Stroke Congress and the World Congress of Neurology.

Yuwaldi Away was born in South Aceh, Indonesia, in 1964. He received the degree in electrical-computer engineering from the Sepuluh Nopember Institute of Technology (ITS), Indonesia, in 1988, the M.Sc. degree from Bandung Institute of Technology (ITB), Indonesia, in 1993, and the Ph.D. degree in industrial computer from the National University of Malaysia (UKM), in 2000. Since 1990, he has been a Lecturer with the Department of Electrical Engineering, Faculty of Engineering, Syiah Kuala University, Indonesia. From 1996 to 2000, he was a Research Assistant and Lecturer at the National University of Malaysia, and from 2001 to 2004. Since 2007, he has been a Professor and the Head of the Research Group for Automation and Robotics Studies at Universitas Syiah Kuala. His research interests include a combination of theory and practice, including microprocessor-based systems, simulation, automation, and optimization.

Aufa Rafiki was born on April 20, 2003, in Banda Aceh. He received his Bachelor's degree in 2025 from the Department of Electrical and Computer Engineering at Universitas Syiah Kuala, where he focused on multimedia technology and EEG signal analysis. He is currently pursuing a Master's degree in Electrical Engineering at Universitas Syiah Kuala, with research interests centred on biomedical signal processing and deep learning for EEG-based applications. During his

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377

undergraduate studies, he served as a teaching assistant and programming laboratory assistant, gaining experience in both teaching and practical implementation. He is committed to strengthening his expertise in theory and applied research to advance technological development in his field. He can be contacted at aufa35@mhs.usk.ac.id.

W. K. Wong is a highly experienced professional engineer (P.Eng) with a strong background in the telecommunications and building services industries prior to involvement in academia. He is currently the Director of an M&E consultancy firm and serves as an Associate Professor in the Department of Electrical and Computer Engineering at Curtin University Malaysia. Dr. Wong received his PhD and Master's degrees from Universiti Malaysia Sabah in 2008 and 2016, respectively. He is a registered member of the Board of Engineers Malaysia and a member of IEEE. At Curtin Malaysia, he leads the IoT Research Group, where his research focuses on biometrics, bioinformatics, sensor technology, applied machine learning, and applied optimization. Dr. Wong has published over 100 academic articles and actively contributes to the research community as a reviewer and editor for numerous reputable journals. He can be contacted via email at: WeiKitt.w@curtin.edu.my.

Mulyadi was born in Punteut, Lhokseumawe, on October 28, 1976. He earned his Bachelor's degree in Applied Science from Institut Teknologi Sepuluh Nopember, Surabaya, majoring in Electrical Engineering - Information Technology. He then pursued his Postgraduate studies in Electrical Engineering at Universitas Sumatera Utara, Medan. Currently, he serves as a faculty member at Politeknik Negeri Lhokseumawe (PNL). Throughout his career, he has held various professional roles at PNL, including teaching staff, practitioner, and expert consultant for both government and non-government institutions. His research interests focus on Electrical Engineering and Information Engineering, and he actively engages in research and community service. He can be contacted at mulyadi@pnl.ac.id.

Siti Rusdiana was born in Banda Aceh, Indonesia, on 10 September 1963. She received the B.S. degree in mathematics science from the Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, in 1988, the M.Eng. degree in mathematics science from Osaka University, Osaka, Japan, in 1998, and the Ph.D. degree in applied mathematics from Universitas Sumatera Utara, Medan, Indonesia, in 2013. She was the Chairperson of the Department of Mathematics, Universitas Syiah Kuala, Banda Aceh, Indonesia (1998 – 2004). From 2013 to 2020, she was the Head of the Laboratory for Dynamic and Optimization Applications, Department of Mathematics, Universitas Syiah Kuala, where she has been an Associate Professor with the Department of Mathematics since 2014. She is also a member of IEEE. Her research interests in applied mathematics include optimization, operations research, and data science. She can be contacted at email: siti.rusdiana@usk.ac.id.

Manuscript Received 12 November 2025; Revised 15 January 2026; Accepted 20 January 2026; Available online 30 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1377