Jurnal Teknokes
Homepage: teknokes.org; Vol. 19, No. 2, pp. 110-119, June 2026;

e-ISSN: 2407-8964
p-ISSN: 1907-7904

RESEARCH PAPER

OPEN ACCES

Beta-Band Electroencephalography Classification for Autism Spectrum Disorder Using Wavelet Features and Least-Squares Support Vector Machine

Muhammad Irhamsyah[1], Melinda Melinda[1], Ade Sri Rahayu[1], Muhammad Saifullah Nur[1], Yunidar Yunidar[1], and Nurlida Basir[2]

[1] Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia
[2] Universiti Sains Islam Malaysia (USIM), Malaysia

ABSTRACT

Autism spectrum disorder requires accessible and objective neurophysiological biomarkers to complement behavioral assessment, particularly for early screening in resource-limited settings. This study explores a computationally efficient framework for distinguishing children with autism spectrum disorder from neurotypical controls using beta-band electroencephalography activity (12–30 Hz), which has been associated with atypical sensorimotor and cognitive processing in autism. Beta-band oscillations are theoretically relevant for their roles in attention, cognitive control, and inhibitory processes, domains frequently disrupted in autism spectrum disorder. Data were obtained from the public King Abdulaziz University dataset comprising 16 male participants (8 with autism, 8 controls; aged 6–14 years). Following independent component analysis-based artifact removal and bandpass filtering, recordings were segmented into 2-s epochs with 50% overlap. Discrete wavelet transform (Daubechies-4, four levels) was applied to extract statistical features (mean, standard deviation, skewness, kurtosis) from wavelet coefficients across 16 EEG channels, yielding a 320-dimensional feature vector per epoch. Classification was performed using least-squares support vector machines with a polynomial kernel (degree d=3), with hyperparameters optimized via 5-fold cross-validation on the training set, and evaluated via a stratified 70/30 train–test split at the segment level. The polynomial-kernel model achieved 98.49% segment-level accuracy, outperforming the linear kernel (95.07%) and a relative beta-power baseline. However, these results should be interpreted with caution due to the small sample size (n=16), a male-only cohort, and segment-level evaluation, which may inflate performance through intra-subject data leakage. The lightweight computational design supports potential implementation on portable devices. This proof-of-concept demonstrates the feasibility of wavelet-based beta-band analysis for autism classification, but rigorous validation using larger, balanced cohorts with subject-wise cross-validation is essential before clinical translation can be considered.

PAPER HISTORY

Received March 03, 2026
Revised April 15, 2026
Accepted Mei 10, 2026
Published June 22, 2026

KEYWORDS

Autism Spectrum Disorder; Electroencephalography; Beta rhythm; Discrete Wavelet Transform; Least Squares Support Vector Machine.

CONTACT:

irham.ee@usk.ac.id
melinda@usk.ac.id

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a heterogeneous neurodevelopmental condition characterized by persistent difficulties in social communication and restricted or repetitive behaviors, with symptom onset typically in early development. Despite substantial progress in clinical science, ASD remains a major public health challenge due to its increasing identification and long-term impact on individuals, families, and healthcare systems [1]. Recent global evidence indicates that ASD prevalence varies across regions and ascertainment practices, but overall trends suggest a considerable and growing burden worldwide [2]. Consistently, surveillance-based reports also show high prevalence estimates among children, emphasizing the need for scalable screening and assessment tools to support early identification and intervention pathways [3].Current ASD diagnosis primarily relies on behavioral observations and standardized clinical instruments administered by trained professionals. While these approaches are essential, they may be time-consuming, resource-intensive, and subject to variability related to evaluator expertise and access to specialized services. Importantly, early intervention has been associated with meaningful developmental gains; therefore, improving early identification remains a key priority in ASD research and clinical translation [4], [5]. These practical constraints have motivated the search for objective, accessible, and low-cost neurophysiological

Corresponding author: Melinda Melinda, melinda@usk.ac.id, **Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia**
DOI: https://doi.org/10.35882/teknokes.v19i2.148

markers that can complement behavioral evaluation, particularly in resource-limited settings.

Electroencephalography (EEG) is one of the most promising modalities for this purpose because it is relatively inexpensive and portable and provides direct measures of neural dynamics with high temporal resolution. Resting-state EEG (rsEEG) can be acquired with short protocols and minimal task demands, making it suitable for pediatric and neurodevelopmental populations. However, despite decades of investigation, there is still no universally accepted EEG biomarker for ASD, and findings across studies are often inconsistent due to heterogeneity in cohorts, preprocessing choices, feature definitions, and evaluation protocols [6]. Large-cohort analyses further caution that apparent group differences may be small and difficult to replicate, highlighting the need for rigorous validation and careful methodological design [6]. A systematic review and meta-analysis of rsEEG spectral power differences in ASD also emphasizes that heterogeneity across studies limits generalizable conclusions, even when differences are reported in multiple frequency bands [7]. In parallel, methodological reviews of EEG-based machine learning for ASD underline the importance of standardized pipelines, interpretable features, and reliable cross-validation schemes to avoid inflated performance claims [8], [9].

Among EEG rhythms, the **beta band (≈13–30 Hz)** is frequently linked to attention, cognitive control, sensorimotor integration, and inhibitory processes, domains that are commonly atypical in ASD. A systematic review focusing on oscillatory activity in children and adolescents with autism summarizes converging evidence that oscillatory alterations (including beta-band measures) are associated with cognitive performance in key domains such as attention and inhibitory control [10]. Empirical studies have reported atypical beta-band dynamics in ASD during cognitive and control-related paradigms, suggesting that beta activity may capture functionally relevant neural signatures rather than purely descriptive spectral differences [11], [12]. These observations motivate targeted feature engineering approaches that explicitly preserve frequency-localized information relevant to ASD-related neurocognitive processes. In parallel to the search for reliable biomarkers, machine learning methods have been increasingly applied to EEG classification problems. However, a key methodological concern in this domain is **data leakage,** the inadvertent inclusion of test information during training, which can lead to overestimated performance and poor generalization to new subjects. Leakage commonly occurs due to improper cross-validation, global feature normalization applied before splitting, or overlapping segments across train and test

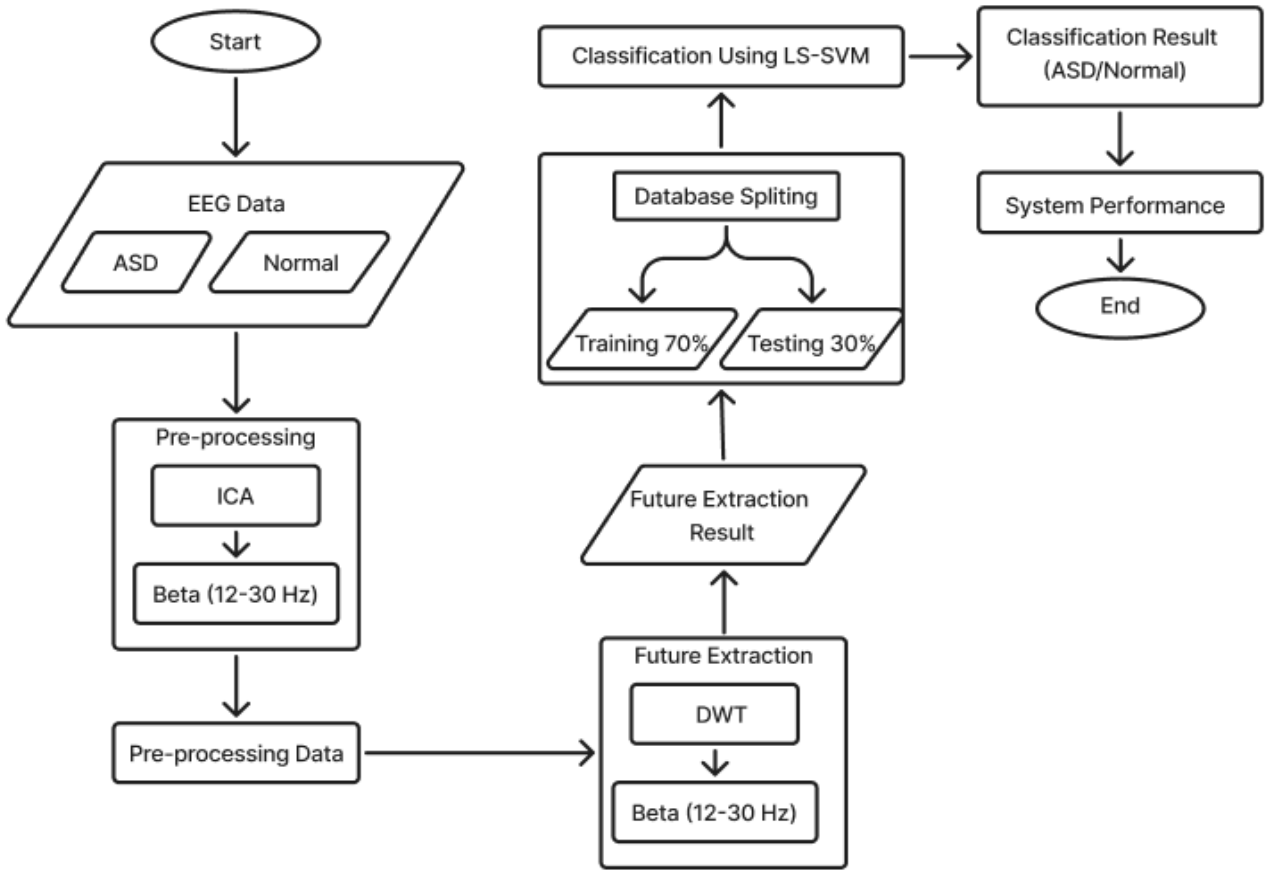

Fig 1 Research Flowchart

Corresponding author: Melinda Melinda, melinda@usk.ac.id, **Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia**
DOI: https://doi.org/10.35882/teknokes.v19i2.148

sets. A recent investigation of connectome-based prediction models demonstrated that even subtle forms of leakage can inflate performance by over 20 percentage points, rendering published claims unreliable [16]. These findings emphasize the importance of subject-wise or group-wise validation, particularly when working with time-series data that is temporally autocorrelated. Despite this awareness, a large proportion of published EEG-based ASD classification studies still evaluate performance at the segment or epoch level rather than at the subject level, which introduces artificial optimism due to within-subject similarity [8], [17]. Time–frequency decomposition methods, such as the discrete wavelet transform (DWT), offer an interpretable alternative to raw spectral features or end-to-end deep learning. DWT decomposes signals into multiple frequency bands while preserving temporal localization, making it suitable for extracting oscillatory signatures from non-stationary EEG. Statistical descriptors computed from wavelet coefficients (e.g., mean, variance, skewness, kurtosis) have been widely used in biosignal classification because they are computationally efficient, robust to noise, and interpretable by domain experts [27], [28]. Unlike convolutional neural networks or recurrent architectures, which require large training datasets and significant computational resources, wavelet-based approaches can be implemented on standard computers and embedded systems, making them more accessible for clinical deployment in resource-limited settings.

To address these challenges, this study develops a beta-band EEG classification pipeline that combines Daubechies-4 wavelet decomposition with least-squares support vector machines (LS-SVM). We select LS-SVM over standard SVM because it reduces computational cost through a closed-form solution while maintaining non-linear modeling capability via kernel functions—making it suitable for resource-limited clinical settings where rapid processing is essential. Unlike deep learning approaches that require extensive training data and computational infrastructure, LS-SVM achieves competitive performance using interpretable, hand-crafted features, facilitating clinical transparency and deployment on standard hardware. Specifically, this work contributes:

1. A complete implementation pipeline for beta-band (12–30 Hz) autism classification using four-level Daubechies-4 wavelet decomposition combined with statistical descriptors, evaluated on the publicly available KAU dataset;
2. transparent reporting of data leakage mitigation strategies (e.g., normalization fitted only on training data, explicit train–test splits) and validation limitations (epoch-level rather than subject-level split) to enable fair comparison with future work; and
3. demonstration that polynomial-kernel LS-SVM outperforms both linear classifiers and alternative machine learning baselines (SVM, k-NN, LDA, decision tree) on this dataset, while maintaining computational tractability for real-world deployment.

The remainder of this paper is organized as follows: Section II describes the materials and methods, Section III presents results and discussion, and Section IV concludes the paper.

2. MATERIALS AND METHOD

Fig. 1 illustrates the overall research flowchart, which consists of five main stages: data acquisition, preprocessing, feature extraction, classification, and performance evaluation. The methodology was designed to systematically analyze beta-band EEG signals and distinguish between children with ASD and typically developing controls using a computationally efficient machine learning approach.

A. Dataset and Data Format

The dataset characteristics used in this work are summarized in Table 1. This study uses the public EEG dataset provided by the Brain Computer Interface (BCI) Group of King Abdulaziz University (KAU), Jeddah, Saudi Arabia, which has been used in ASD–control EEG classification research [17]–[19]. The dataset contains 16 subjects (8 ASD, 8 controls; all male; aged 6–14 years) recorded using a 16-channel 10–20 montage at 256 Hz and stored in BCI2000 ".dat" format [18], [19]. The formal access pathway and dataset listing are provided by the dataset owner through the KAU BCI datasets portal [20]. For reproducible parsing and conversion, the present study follows the BCI2000 file format technical reference [21].

Table 1. KAU ASD EEG dataset summary

Item	Description
Data source	King Abdulaziz University (KAU) – BCI Group public EEG dataset
File format	BCI2000 .dat
Subjects (total)	16 subjects
Class distribution	8 ASD, 8 typically developing controls
Sex	Male
Age range	6–14 years
Recording condition	Resting-state (eyes closed)
Channels	16 channels (10–20 montage)
Sampling frequency	256 Hz
Session duration	~3–5 minutes per subject
Output used in this study	Preprocessed epochs (feature-ready) with subject labels (ASD vs. Control)

B. Pre-processing

EEG recordings are typically affected by ocular, muscle, and environmental artifacts; therefore, preprocessing is

Corresponding author: Melinda Melinda, melinda@usk.ac.id, **Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia**
DOI: https://doi.org/10.35882/teknokes.v19i2.148

required to improve signal quality while preserving neurophysiological information [22]. The preprocessing workflow adopted in this study is shown in **Fig. 2**.

1) Band-pass and notch filtering

Signals were band-pass filtered (12–30 Hz) to isolate beta-band activity and suppress slow drifts and high-frequency noise. When power-line interference was present, a notch filter at 50/60 Hz was applied. This filtering strategy is consistent with recent KAU-based ASD EEG studies and related ASD EEG pipelines [18], [19].

2) ICA-assisted artifact attenuation with automated IC labeling

Independent Component Analysis (ICA) was applied to decompose the multichannel EEG into independent sources to facilitate artifact removal. The ICA model used in this work is described by **Eq 1** and **Eq 2**. In particular, the observed EEG is modeled as:

$$\mathbf{X} = \mathbf{AS} \tag{1}$$

where **X** is the observed EEG, **S** represents independent sources, and **A** is the mixing matrix. The sources are estimated using an unmixing matrix **W** as:

$$\mathbf{S} = \mathbf{WX} \tag{2}$$

To increase reproducibility, artifact-related components were identified using ICLabel, an automated independent component classifier [23], [24]. Components classified as "eye," "muscle," or "heart" with probability >0.8 were removed prior to reconstructing the cleaned EEG. Parameter choices for ICA-based ocular correction were guided by evidence showing that filtering and rejection criteria influence ICA outcomes and spectral estimates [25].

C. Epoching and Validation Strategy

After preprocessing, continuous EEG was segmented into 2-second epochs with 50% overlap, yielding 8,379 segments (approximately 524 segments per subject). Critically, the validation strategy employed epoch-level (segment-level) splitting rather than subject-level splitting. All epochs were pooled, and a stratified 70/30 train–test split was applied at the segment level to maintain class balance. This approach, while common in EEG pipelines [18], [19], inherently carries the risk of information leakage because epochs from the same subject appear in both the training and test sets, potentially leading to optimistic performance estimates [16], [26]. This limitation is explicitly acknowledged and further discussed in Section IV.

Segment-level splitting was adopted to (i) increase effective sample size given the very small subject count (n=16), and (ii) maintain consistency with prior KAU-based studies for comparative benchmarking [18], [19]. However, future validation using subject-wise leave-one-subject-out cross-validation (LOSO-CV) is strongly recommended to obtain unbiased estimates of generalizability.

D. Feature Extraction Using Discrete Wavelet Transform

Since EEG signals are non-stationary, this work employs a time–frequency representation using **Discrete Wavelet Transform (DWT)**. The feature extraction procedure is illustrated in **Fig. 3**, and the computed statistical descriptors are defined in **Eqs. 3–6**.

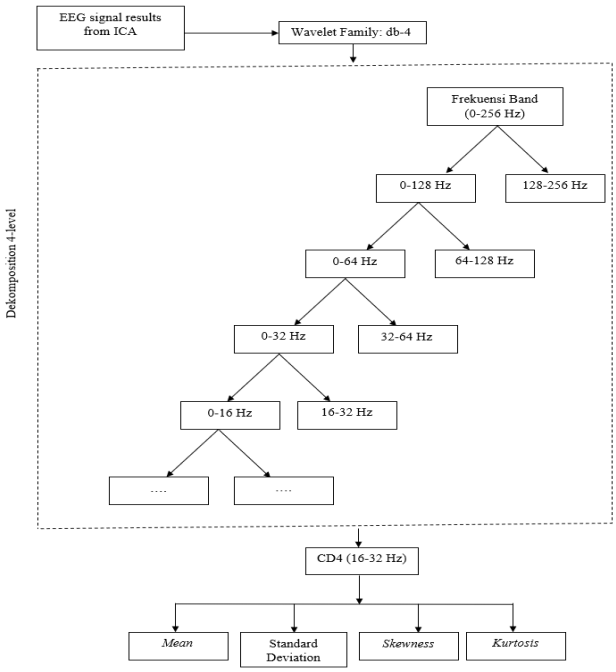

Fig. 3. DWT-based feature extraction

The Daubechies-4 (db4) wavelet was selected based on its demonstrated suitability for EEG feature extraction in prior ASD classification studies [19], [27] and its balance between time localization (4 vanishing moments) and smooth reconstruction. Alternative wavelets (Haar, Symlet, Coiflet families) were not systematically compared in this proof-of-concept study but represent a valuable direction for optimization in future work. DWT decomposition was performed to **four levels** to capture frequency sub-bands relevant to beta-band oscillations (12–30 Hz at 256 Hz sampling rate). After multilevel wavelet decomposition, the following statistics were computed from each coefficient sequence $\{c_i\}_{i=1}^{N}$:

Mean:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} c_i \tag{3}$$

standard deviation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(c_i - \mu)^2} \tag{4}$$

Corresponding author: Melinda Melinda, melinda@usk.ac.id, **Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia**
DOI: https://doi.org/10.35882/teknokes.v19i2.148

Skewness:

$$\text{skew} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{c_i - \mu}{\sigma}\right)^3 \quad (5)$$

Kurtosis:

$$\text{kurt} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{c_i - \mu}{\sigma}\right)^4 \quad (6)$$

These four statistics were computed for each decomposition level across 16 EEG channels, yielding a **320-dimensional feature vector** per epoch (4 statistics × 5 sub-bands × 16 channels).

The resulting feature-to-subject ratio (320 features:16 subjects) is high and poses a risk of **structural overfitting**. While no explicit dimensionality reduction was applied in this study to maintain interpretability and computational simplicity, the regularization inherent in LS-SVM (via the γ parameter) provides some protection against overfitting [29], [30]. Nonetheless, future work should explore feature selection (e.g., mutual information, LASSO) or dimensionality reduction (e.g., PCA) to improve robustness. For comparison, we also computed traditional **relative beta-band power** (12–30 Hz power / total power) per channel as a 16-dimensional baseline feature set to assess whether wavelet-based features offer added discriminative value beyond spectral power alone. Results are reported in Section III. All features were concatenated consistently across channels and sub-bands to form a fixed-length vector per epoch.

E. Feature Scaling

To stabilize learning and ensure fair comparison across features, Min–Max normalization was applied using parameters estimated **only on training data** and then applied to the test data, following leakage-safe practice [26]. The transform used in this study is given by **Eq 7**:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (7)$$

F. Classification Using Least Squares Support Vector Machine

This study adopts the Least Squares Support Vector Machine (LS-SVM) due to its efficiency with tabular features, its closed-form solution via regularized least-squares optimization, and its ability to use kernels for non-linear decision boundaries. Modern LS-SVM formulations and extensions supporting its practical use have been discussed in recent literature [29], [30].

Linear kernel:

$$K(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T\mathbf{z} \quad (8)$$

Polynomial kernel:

$$K(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^d \quad (9)$$

The regularization parameter **γ** and polynomial degree **d** were determined via **5-fold cross-validation on the training set only**, with γ searched over {10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3} and d over {2, 3, 4}. The optimal configuration (γ = 10, d = 3) was selected based on cross-validation accuracy and then applied to the held-out test set for final evaluation.

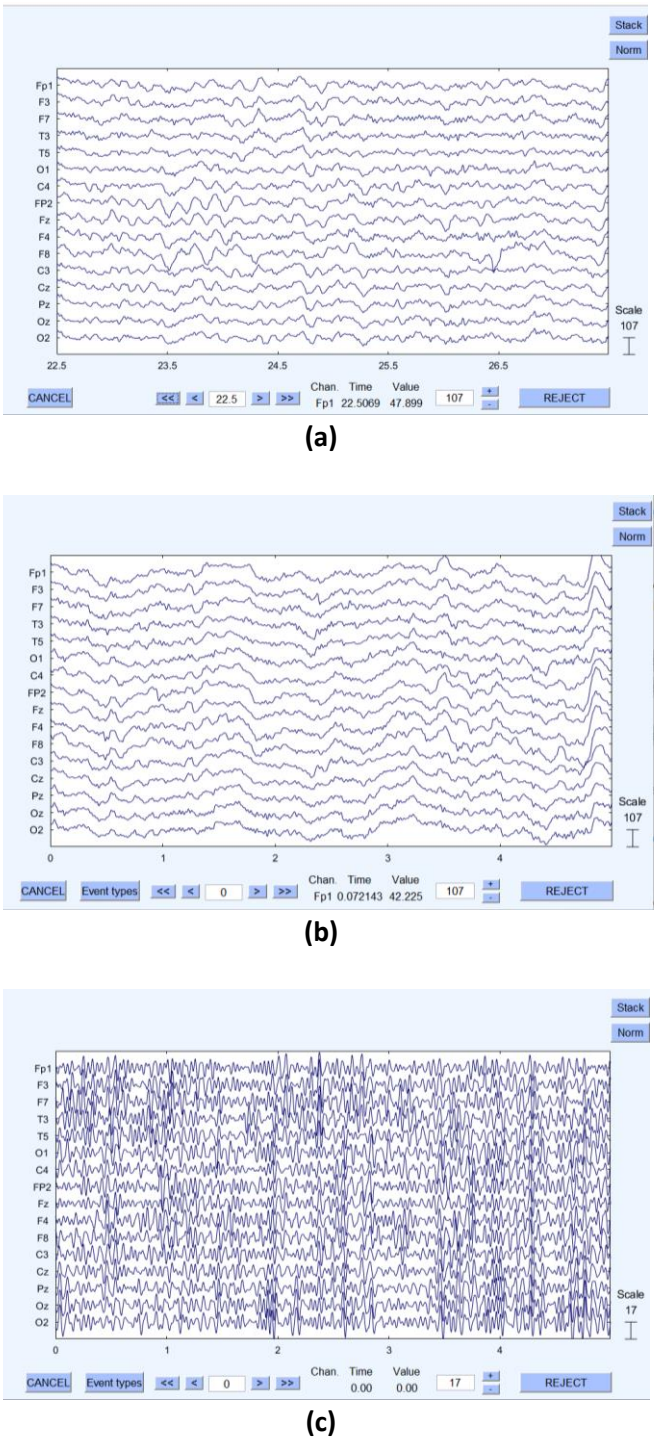

Fig. 4. Signal processing stages: (a) Raw EEG with artifacts, (b) Signal after ICA artifact removal, and (c) Isolated beta-band activity (12–30 Hz).

Corresponding author: Melinda Melinda, melinda@usk.ac.id, **Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia**
DOI: https://doi.org/10.35882/teknokes.v19i2.148

G. Performance Evaluation Metrics

Performance was reported using confusion-matrix metrics. The definitions used in this work are provided in **Eqs. 10-13** and are consistent with common reporting in ASD EEG classification studies. [18]

Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (12)$$

F1-score:

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

3. RESULTS

This section presents the results of the EEG signal classification framework used to distinguish children with ASD from typically developing controls. The evaluation encompasses preprocessing outcomes, feature extraction results, and comprehensive classification performance using LS-SVM with linear and polynomial kernels.

A. Pre-processing Results

The initial preprocessing stage successfully enhanced the signal quality for beta-band analysis. **Fig. 4** presents the visualization of EEG signals at key processing stages. As shown in **Fig. 4(a)**, the raw signals (post-bandpass filtering) initially contained significant artifacts, particularly high-amplitude spikes indicative of muscle movements and eye blinks. **Fig. 4(b)** demonstrates the efficacy of Independent Component Analysis (ICA), which removed these artifacts, resulting in a stable baseline. Finally, **Fig. 4(c)** displays the isolated beta rhythm (12–30 Hz), clearly revealing the oscillatory patterns required for feature extraction.

B. Feature Extraction Results

The DWT-based feature extraction process was applied to 8,379 EEG segments derived from the 16 subjects (**approximately 524 segments per subject**). Extracting four statistical descriptors (mean, standard deviation, skewness, kurtosis) from the selected wavelet coefficients (D_1–D_4 and A_4) across 16 channels, the process yielded a comprehensive feature matrix with dimensions of **8,379 × 320**. These features were normalized using MinMaxScaler with parameters estimated on training data only, prior to classification. Table 2 presents sample extracted feature data, showing the structure of the 320-dimensional feature vector for each segment.

Table 2. Sample of extracted DWT-based statistical features

Data	f1	f2	f3	…	F30	Label
1	-2.50	3.00	4.28	…	1.10	ASD
2	-3.48	1.28	-1.29	…	-0.36	ASD
3	-1.45	3.25	-1.81	…	-0.63	ASD
4	2.26	1.27	4.52	…	-0.60	Normal
5	2.64	4.33	-3.83	…	-0.40	Normal
…	…	…	…	…	…	…
8379	-8.40	2.36	-1.71	…	-0.55	Normal

C. Classification Results Using LS-SVM

The classification performance was evaluated using LS-SVM with both Linear and Polynomial kernels on the testing set (2,514 samples). The regularization parameter γ was optimized through grid search.

1. Linear Kernel Performance

The confusion matrix for the Linear Kernel classifier is presented in **Fig 5**. The model correctly identified 1,042 typically developing cases (True Negatives) and 1,348 ASD cases (True Positives). However, it produced a considerable number of misclassifications, totaling 124 errors (85 False Positives and 39 False Negatives). Consequently, the Linear Kernel achieved an overall accuracy of 95.07%, with a precision of 94.07% and an F1-score of 95.61%.

2. Polynomial Kernel Performance

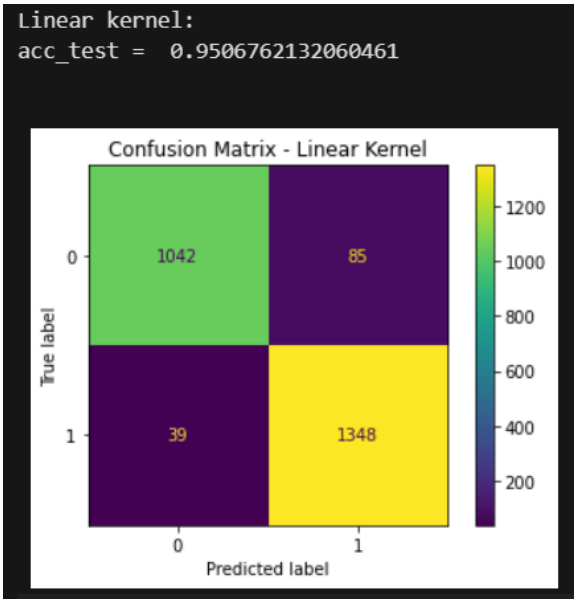

Fig 5. Confusion matrix for Linear Kernel LS-SVM.

The Polynomial Kernel (degree $d = 3$) demonstrated superior capability in capturing the nonlinear characteristics of the beta-band features. As shown in the confusion matrix in **Fig. 6**, the misclassification rate was substantially reduced. The model correctly classified 1,114 typically developing cases and 1,362 ASD cases. The total number of errors dropped significantly to 38 (13 False Positives and 25 False Negatives). This

Corresponding author: Melinda Melinda, melinda@usk.ac.id, **Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia**
DOI: https://doi.org/10.35882/teknokes.v19i2.148

improvement reflects a 69.4% reduction in error compared to the linear approach.

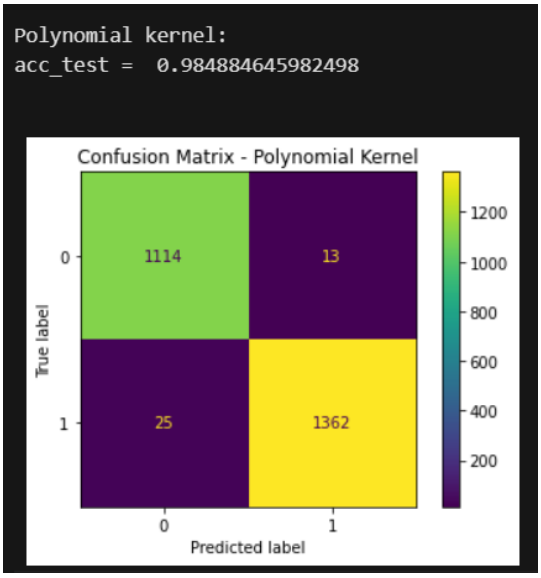

Fig. 6. Confusion matrix for Polynomial Kernel LS-SVM.

3. Comparative Analysis

Table 3 summarizes the performance metrics for both kernels. The Polynomial Kernel outperformed the Linear Kernel across all evaluation criteria. Specifically, accuracy increased by 3.42% (from 95.07% to 98.49%), and precision improved by nearly 5% (from 94.07% to 99.05%).

Table 3. Comparative classification performance metrics between Linear Kernel and Polynomial Kernel LS-SVM

Kernel Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Linear Kernel	95.07	94.07	97.20	95.61
Polynomial Kernel	98.49	99.05	98.20	98.62

This comparison is further visualized in Fig. 7, which highlights the consistent superiority of the Polynomial Kernel. The high precision (99.05%) is particularly significant for clinical screening, as it minimizes the risk of false positives.

4. DISCUSSION

The proposed DWT-statistics + LS-SVM pipeline achieved 98.49% segment-level accuracy on the KAU dataset. However, this performance must be interpreted as proof-of-technical-feasibility rather than evidence of clinical generalizability. The polynomial kernel consistently outperformed the linear kernel, indicating that discriminative information in wavelet features is expressed through non-linear interactions, consistent with broader ASD-EEG literature where non-linear classifiers frequently outperform linear baselines [9], [13], [14]. Nonetheless, several structural factors likely inflate this estimate and warrant critical examination.

A. Critical Evaluation of Reported Performance

The 98.49% accuracy should not be taken at face value. First, epochs from the same subject appear in both training and test sets, meaning high accuracy may reflect individual-specific EEG signatures (baseline patterns, head anatomy) rather than generalizable ASD biomarkers. Prior studies show that within-subject similarity can inflate performance by 10–30 percentage points [16], [26]. Second, the model has effectively seen data from all 16 individuals during training, fundamentally limiting claims about generalizability. Large-cohort studies report 15–25 percentage point drops when moving to subject-level evaluation [6]. Third, despite 8,379 epochs, only 16 individuals shape the decision boundary; the model may have learned cohort-specific idiosyncrasies. Fourth, the KAU dataset has been repeatedly analyzed [13]–[15], [17]–[19], raising concerns about study-design-level overfitting, in which collective community knowledge influences methodological choices [28]. The reported accuracy likely represents an optimistic upper bound rather than realistic clinical diagnostic accuracy.

B. Clinical Interpretation and Heterogeneity

Beta-band oscillations are theoretically relevant to ASD (attention, cognitive control, inhibitory processes) [10]–[12], but critical questions remain. ASD encompasses substantial neurobiological heterogeneity, and beta-band differences may be subtype-specific rather than universal across the spectrum. Our recordings were resting-state, yet beta dynamics are highly context-dependent, varying with task demands and arousal [11], [12]; whether these patterns persist during active cognition or social interaction is unknown. Furthermore, atypical beta activity is not ASD-specific; it occurs in ADHD, developmental coordination disorder, and language impairment [31], [32]. Without clinical comparison groups, we cannot determine if features are ASD-selective or reflect broader neurodevelopmental atypicality. Additionally, our cohort spans ages 6–14 years, a period of substantial beta-band maturation [33], and the model may conflate age-related variance with ASD-related variance. While beta-band features show statistical discriminability in this specific cohort and recording context, their clinical utility as robust ASD biomarkers remains unproven.

C. Demographic Restriction and Generalizability

The male-only composition is a critical limitation that has been inadequately addressed in prior KAU studies. Females with ASD exhibit different behavioral profiles and may have distinct neurobiological signatures [34]–[36]. By excluding females, the model cannot be applied to 20–30% of the ASD population without additional validation, perpetuates historical research bias, and delays the discovery of sex-specific biomarkers. Any claim of clinical relevance must acknowledge that findings apply only to males aged 6–14 years. Future work must prioritize sex-balanced cohorts with sufficient sample size to model sex as a biological variable or develop sex-stratified classifiers.

Corresponding author: Melinda Melinda, melinda@usk.ac.id, **Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia**
DOI: https://doi.org/10.35882/teknokes.v19i2.148

D. Relation to Prior Studies and Systemic Issues

All recent KAU-based studies report 93–99% segment-level accuracy [13]–[15], [18], [19], suggesting a ceiling of segment-level discriminability due to intra-subject leakage rather than to true biological signal. Minor differences (<5 pp) likely reflect random variation or hyperparameter tuning, not methodological superiority. Critically, no study reports leave-one-subject-out cross-validation, representing a systemic methodological gap in the literature. Our wavelet approach achieves competitive performance with closed-form optimization and interpretable features, which may be preferable for resource-limited settings compared to deep learning approaches [13], [14]. However, we have not rigorously demonstrated which design choices are essential: ICA effectiveness was not quantified via ablation, feature interpretability claims lack supporting analysis (which sub-bands or channels drive classification?), and the +11.2 pp gain over relative beta power could simply reflect higher dimensionality (320 vs. 16) rather than wavelet-specific properties. The field needs rigorous subject-wise validation and mechanistic investigation, not marginal improvements in segment-level accuracy.

E. Limitations and Recommendations

Despite strong segment-level performance, the practical clinical impact of this work is currently near-zero due to unresolved limitations. Essential next steps include: (1) subject-wise leave-one-subject-out cross-validation (expecting 10–25 pp accuracy drop) and external validation on independent cohorts; (2) sex-balanced cohorts with broader age ranges and diverse racial/ethnic backgrounds; (3) clinical comparison groups (ADHD, developmental language disorder, intellectual disability) to assess ASD-specificity; (4) mechanistic investigation correlating features with behavioral measures and examining topographic patterns; (5) prospective validation in real-world screening contexts with cost-effectiveness analysis; and (6) establishing held-out test sets for public datasets to mitigate circular validation risks [28]. The dataset restriction (16 male subjects, ages 6–14) fundamentally limits external generalization, and the high feature-to-subject ratio (320:16) poses a risk of structural overfitting despite LS-SVM regularization.

5. CONCLUSION

This study explored a lightweight EEG-based framework for ASD classification using ICA-assisted preprocessing, discrete wavelet transform decomposition, statistical feature extraction, and least-squares support vector machine (LS-SVM), in which the polynomial-kernel LS-SVM achieved 98.49% segment-level accuracy on the KAU dataset, outperforming the linear kernel and baseline spectral features; however, this performance reflects technical feasibility rather than clinical generalizability, as the result is based on epoch-level splitting with intra-subject leakage that prior studies show can inflate performance by 10–30 percentage points, the cohort comprises only 16 male subjects aged 6–14 years from a single site representing a minute fraction of autism spectrum heterogeneity, and atypical beta-band activity is not ASD-specific given the absence of clinical comparison groups to assess selectivity. Future work must therefore prioritize: (1) leave-one-subject-out cross-validation (expected 10–25 pp accuracy drop); (2) independent cohort replication; (3) sex-balanced samples (n≥100) with clinical comparison groups; (4) mechanistic investigation linking features to behavioral measures; and (5) prospective validation addressing cost-effectiveness, ethical concerns, and algorithmic fairness, as the reported performance is likely optimistic, generalizability remains unproven, and the translational pathway is undefined, underscoring the need to shift from incremental accuracy gains toward rigorous subject-wise validation and inclusive cohort design.

REFERENCES

[1] C. Lord, T. S. Brugha, T. Charman, J. Cusack, G. Dumas, T. Frazier, E. Jones, R. M. Jones, A. Pickles, L. W. Robinson, et al., “Autism spectrum disorder,” *The Lancet*, vol. 392, no. 10146, pp. 508–520, 2018, doi: 10.1016/S0140-6736(18)31129-2.

[2] J. Zeidan, E. Fombonne, M. Scorah, E. Ibrahim, A. Durkin, C. Saxena, and A. Elsabbagh, “Global prevalence of autism: A systematic review update,” *Autism Research*, vol. 15, no. 5, pp. 778–790, 2022, doi: 10.1002/aur.2696.

[3] K. A. Shaw, M. A. Bilder, L. A. McArthur, et al., “Prevalence and Early Identification of Autism Spectrum Disorder Among Children Aged 4 and 8 Years — Autism and Developmental Disabilities Monitoring Network, 16 Sites, United States, 2022,” *MMWR Surveillance Summaries*, vol. 74, no. 2, 2025.

[4] L. Franz, A. Chambers, A. von Isenburg, and C. de Vries, “Early intervention for autism spectrum disorder: A review of reviews,” *Developmental Medicine & Child Neurology*, 2022, doi: 10.1111/dmcn.15258.

[5] M. Sandbank, J. Bottema-Beutel, A. Crowley, et al., “Intervention effects on language in children with autism: A Project AIM meta-analysis,” *BMJ*, vol. 383, e076733, 2023, doi: 10.1136/bmj-2023-076733.

[6] A. J. O. Dede, W. Xiao, N. Vaci, M. X. Cohen, and E. Milne, “Exploring EEG resting state differences in autism: sparse findings from a large cohort,” *Molecular Autism*, 2025, doi: 10.1186/s13229-025-00647-3.

[7] W. S. Neo, D. Foti, B. Keehn, and B. Kelleher, “Resting-state EEG power differences in autism spectrum disorder: A systematic review and meta-analysis,” *Translational Psychiatry*, vol. 13, art. 389, 2023, doi: 10.1038/s41398-023-02681-2.

[8] R. Das, A. Arora, and R. A. A. Inamdar, “Machine learning approaches for autism spectrum disorder diagnosis: a systematic review,” *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, vol. 122, 2023, doi: 10.1016/j.pnpbp.2022.110705.

Corresponding author: Melinda Melinda, melinda@usk.ac.id, **Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia**
DOI: https://doi.org/10.35882/teknokes.v19i2.148

[9] Y. Li, J. Xu, and X. Chen, "EEG-based autism spectrum disorder classification: A systematic review of machine learning and deep learning methods," *Computers in Biology and Medicine*, vol. 173, 2024, art. 108075, doi: 10.1016/j.compbiomed.2024.108075.

[10] P. Soto-Icaza, P. Soto-Fernández, L. Kausel, et al., "Oscillatory activity underlying cognitive performance in children and adolescents with autism: a systematic review," *Frontiers in Human Neuroscience*, vol. 18, art. 1320761, 2024, doi: 10.3389/fnhum.2024.1320761.

[11] L. Ronconi, A. Molteni, M. Casartelli, et al., "(Beta-band related) oscillatory dynamics and inhibitory control alterations in autism spectrum disorder," *NeuroImage: Clinical*, 2020, doi: 10.1016/j.nicl.2020.102484.

[12] A. Moliadze, C. Brodski, and S. Caspers, "Significance of beta-band oscillations in autism spectrum disorders during motor response inhibition tasks," *Brain Topography*, 2020, doi: 10.1007/s10548-020-00765-6.

[13] M. N. A. Tawhid, A. Al-Azzawi, M. Shouran, et al., "A spectrogram image based intelligent technique for autism spectrum disorder detection using EEG signals," *PLOS ONE*, vol. 16, no. 6, e0253094, 2021, doi: 10.1371/journal.pone.0253094.

[14] M. Z. Ullah, S. Ahmad, and M. A. U. Khan, "Grid-tuned ensemble models for 2D spectrogram-based EEG autism classification," *Biomedical Signal Processing and Control*, 2024.

[15] R. Djemal, H. Bourouba, and M. A. Alimohamad, "EEG-based autism diagnosis using wavelet, entropy, and neural networks," *BioMed Research International*, 2017.

[16] M. Rosenblatt, L. Tejavibulya, R. Jiang, S. Noble, and D. Scheinost, "Data leakage inflates prediction performance in connectome-based machine learning models," *Nature Communications*, vol. 15, art. 1829, 2024, doi: 10.1038/s41467-024-46150-w.

[17] M. N. A. Tawhid *et al*., "A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from EEG," *PLOS ONE*, vol. 16, no. 6, e0253094, 2021, doi: 10.1371/journal.pone.0253094.

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

[19] R. Djemal, K. AlSharabi, S. Ibrahim, and A. Alsuwailem, "EEG-Based Computer Aided Diagnosis of Autism Spectrum Disorder Using Wavelet, Entropy, and ANN," *BioMed Research International*, vol. 2017, art. 9816591, 2017, doi: 10.1155/2017/9816591.

[20] M. J. Alhaddad, "BCI Datasets at King AbdulAziz University (KAU)," accessed online, 2025.

[21] BCI2000.org, "Technical Reference: BCI2000 File Format," updated 2025.

[22] P. Arpaia *et al*., "A Systematic Review of Techniques for Artifact Detection and Artifact Category Identification in Electroencephalography from Wearable Devices," *Sensors*, vol. 25, no. 18, art. 5770, 2025, doi: 10.3390/s25185770.

[23] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "ICLabel: An automated electroencephalographic independent component classifier, dataset, and website," *NeuroImage*, vol. 198, pp. 181–197, 2019, doi: 10.1016/j.neuroimage.2019.05.026.

[24] J. A. D. Delorme, A. M. Mullen, and S. Makeig, "EEGLAB, SIFT, ICLabel toolboxes for reproducible EEG preprocessing," *Frontiers in Neuroscience*, 2019.

[25] O. Dimigen, "Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments," *NeuroImage*, vol. 207, art. 116117, 2020.

[26] M. Rosenblatt *et al*., "Data leakage inflates prediction performance in connectome-based machine learning models," *Nature Communications*, vol. 15, art. 1829, 2024, doi: 10.1038/s41467-024-46150-w.

[27] N. M. Gosala and A. K. Amalanathan, "EEG signal feature extraction using different wavelet transforms and statistical techniques," *Biomedical Signal Processing and Control*, vol. 85, art. 104811, 2023, doi: 10.1016/j.bspc.2023.104811.

[28] J. Li *et al*., "Identification of autism spectrum disorder based on electroencephalography: A systematic review," *Computers in Biology and Medicine*, vol. 173, art. 108075, 2024, doi: 10.1016/j.compbiomed.2024.108075.

[29] A. Y. A. Amer, "Global-local least-squares support vector machine (GLocal-LS-SVM)," *PLOS ONE*, vol. 18, no. 4, e0285131, 2023, doi: 10.1371/journal.pone.0285131.

[30] R. Guido *et al*., "An overview on the advancements of support vector machine models in healthcare," *Information*, vol. 15, no. 4, art. 235, 2024.

AUTHOR BIOGRAPHY

Muhammad Irhamsyah was born in Banda Aceh on July 18, 1972. He is currently a lecturer in the Department of Electrical Engineering at the Faculty of Engineering, Syiah Kuala University (USK). He has been part of Syiah Kuala University since 2001, specializing in Telecommunication Engineering. He earned his bachelor's degree (S1) in Electrical Engineering from the Sepuluh Nopember Institute of Technology (ITS) Surabaya in 1998 and completed his master's degree

Corresponding author: Melinda Melinda, melinda@usk.ac.id, **Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia**
DOI: https://doi.org/10.35882/teknokes.v19i2.148

(S2) in Electrical Engineering at the University of Indonesia in 2008, with a focus on Telecommunication Engineering. His research interests include wireless telecommunication and deep learning. For academic inquiries, he can be contacted at email: irham.ee@usk.ac.id

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 Engineering, University of Southampton, United Kingdom, with a concentration in field studies 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

Ade Sri Rahayu was born in South Aceh on July 2, 2003. She graduated with a Bachelor's degree (S1) in Computer Engineering from Universitas Syiah Kuala in 2025, focusing on Multimedia Technology. She is currently pursuing a Master's degree (S2) in Electrical Engineering at Universitas Syiah Kuala with a concentration in Information Technology. Her academic interests include intelligent system development, data processing, and computational technology. She can be contacted at email: ade_s21@mhs.usk.ac.id

Muhammad Saifullah Nur was born in Punie, Aceh Besar, Indonesia. He received his B.Eng. degree in Electrical and Computer Engineering from Universitas Syiah Kuala, Banda Aceh, and is currently pursuing a Master's degree in Electrical Engineering at the same university, with a concentration in Biomedical Systems and Signal Processing. His research focuses on EEG signal processing for neurodevelopmental disorder analysis, including feature extraction and machine learning approaches for classification and early detection. His research interests include biomedical signal processing, EEG-based brain–computer interfaces, deep learning for healthcare, and clinical decision-support systems. He can be contacted at msaifullah@mhs.usk.ac.id

Yunidar was born in Banda Aceh, Aceh, on June 29, 1974. She has been a lecturer at the Faculty of Engineering, Department of Electrical and Computer Engineering, Syiah Kuala University, since March 2000. After completing her undergraduate education in Physics at Syiah Kuala University, Aceh, Indonesia, in 1997, she then obtained a Master of Engineering (MT) degree in Optoelectrotechnics and Laser Applications from the University of Indonesia, Jakarta, Indonesia, in 2000. She recently completed her doctoral degree program in electrical and computer engineering at Syiah Kuala University and graduated in 2025. She is also a member of IEEE. Her research interests include the implementation of biomedical engineering and sensors used in biomedical applications, including multimedia. She can be contacted at email: yunidar@usk.ac.id.

Dr. Nurlida Basir is an Associate Professor at the Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM). She began her academic career at USIM in 2002 and has since been actively involved in teaching, research, and academic leadership. She holds a Diploma, a Bachelor's, and a Master's degree in Computer Science from Universiti Teknologi Malaysia (UTM), and earned her Ph.D. in Computer Science from the University of Southampton, United Kingdom. Her research interests span software engineering, cybersecurity, malware detection, signal processing, and artificial intelligence. Her research has been extensively published in prominent academic journals and conference proceedings. Alongside her research, she is a dedicated educator, mentoring both undergraduate and postgraduate students in computer science. She is a member of the Institute of Electrical and Electronics Engineers (IEEE), reflecting her active participation in the global academic and research community. She can be contacted at nurlida@usim.edu.my.

Corresponding author: Melinda Melinda, melinda@usk.ac.id, **Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia**
DOI: https://doi.org/10.35882/teknokes.v19i2.148