ISSN 2085-4579

Redundancy-Aware Feature Selection using mRMR and F-Test for EEG Emotion Classification

Ira Febrianti[1], Carens Chanda Claudhyta Hasan[2], Nadzifatu Chomtsa[3], Hanifa Khairunisa[4], Deandra Faysa Mardatila[5], Yuri Pamungkas[6*]
[1-6] Department of Medical Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
yuri@its.ac.id

Accepted on April 04th, 2026
Approved on June 21st, 2026

***Abstract*—Emotions play an essential role in human interaction, driving the development of reliable automatic emotion recognition systems. Electroencephalography (EEG) offers a noninvasive method to record neural activity related to emotional states; however, many existing studies focus on limited feature configurations or binary classification problems. This research examines the influence of feature dimensionality and classifier selection on three-class EEG-based emotion recognition involving positive, neutral, and negative categories. The primary contribution of this study is a systematic assessment of feature and classifier compatibility across 28 experimental scenarios within a unified evaluation framework. Using a publicly available EEG dataset containing statistical and spectral features, selection was conducted using F-test and Minimum Redundancy Maximum Relevance (mRMR) methods, isolating the top 5, 10, and 15 features alongside the complete set. Four classifiers (Random Forest, Support Vector Machine, K-Nearest Neighbors, and Neural Networks) were evaluated via a 70/30 hold-out validation scheme using accuracy, F1-score, and Area Under the Curve (AUC). Results indicate that Random Forest trained with the full feature set achieved the highest performance, reaching 99.53% accuracy and 0.9994 AUC. These findings suggest that ensemble-based models demonstrate greater robustness when handling high-dimensional EEG features in multi-class emotion recognition.**

***Index Terms*—EEG; Emotion Recognition; Machine Learning; Feature Selection**

I. INTRODUCTION

Emotions are fundamental to human social interaction and societal functioning because they support coordination, communication, and group cohesion that developed alongside the growth of human communities over evolutionary timescales [1]. Beyond internal experiences, emotions also operate interpersonally and are shaped by social context, while simultaneously shaping others' responses and decisions [2]. The emotions as social information model further explains that emotional expressions convey cues about the expresser's feelings, goals, and intentions, which then influence observers through affective responses and cognitive inferences [3]. These perspectives motivate the need for reliable emotion recognition methods, especially those that can capture emotional states objectively rather than relying solely on self-report or behavioral observation.

Electroencephalography has become a prominent approach for emotion recognition because it is non-invasive and provides real-time measurements of brain electrical activity related to central nervous system dynamics [4]. EEG-based emotion recognition is commonly formulated as a signal-processing and pattern-recognition pipeline that includes acquisition, preprocessing, feature extraction, feature selection, and classification [5]. However, a persistent research problem remains that many EEG-derived features are high-dimensional and redundant, while EEG signals themselves are noisy and non-stationary. This combination can degrade generalization performance, inflate computational cost, and complicate interpretability, particularly when models are trained on limited samples relative to the number of extracted features [6].

State-of-the-art studies have explored diverse feature representations and classifiers to improve performance. Multi-feature fusion in the time domain and composite representations based on wavelet analysis and entropy have reported accuracies above 70% in common emotion dimensions such as valence and arousal [7]. Other work has investigated spectral and time-frequency strategies including Gabor-based spectral descriptors, wavelet transforms, and general time-frequency analysis, as well as entropy-driven feature families [8], [9], [10]. In parallel, machine learning classifiers such as neural networks, support vector machines, and random forests have been widely adopted for EEG emotion classification and have achieved competitive results across datasets [11], [12], [13]. These efforts collectively show that both feature engineering and model choice strongly affect recognition accuracy, yet they also suggest that improvements are often constrained by the quality and compactness of the selected feature set.

Prior studies highlight the importance of feature selection and discriminative representations, especially in relation to frequency bands. A mutual information-based feature selection combined with kernel classifiers achieved around 73% accuracy for both arousal and valence, indicating that reducing irrelevant information can improve prediction [14]. Differential entropy features have been shown to outperform conventional energy spectrum descriptors, with reported accuracy above 84%, and gamma-band activity was found to be particularly associated with emotional variation [15]. Comparisons across multiple feature extraction families indicate that time-domain statistical characteristics can provide strong results, with accuracies around the high 70% range for valence, arousal, and dominance [16]. More recent evaluations comparing extraction, reduction, and classification approaches reported strong binary performance for kernel spectral regression with random forest, while again reaffirming gamma rhythm as informative for emotional changes [17]. Additional hybrid reduction schemes such as combining PCA with statistical screening have also reported mid-80% performance with SVM, and Hjorth-parameter-based pipelines with ANOVA-based selection have achieved accuracy in the mid-70% range using ensemble classifiers [18], [19]. Cross-dataset analyses further suggest that the best-performing features and classifiers can vary by dataset and protocol, although certain families such as fractal-dimension features can be consistently competitive when paired with appropriate classifiers [20].

Despite these advances, the research gap is that many pipelines either rely on large feature sets that are prone to redundancy or apply a single selection criterion that may not simultaneously maximize relevance to the target labels and minimize overlap among chosen features. This study addresses that gap by proposing a redundancy-aware feature selection strategy that integrates minimum redundancy maximum relevance with F-test screening for EEG emotion classification. The solution is designed to prioritize features that are both discriminative and non-redundant, thereby reducing dimensionality while preserving informative variability related to emotional states. We extract a broad set of statistical and mathematical descriptors including mean, minimum, maximum, covariance matrix representations, matrix logarithm features, correlation measures, and frequency-domain components via Fast Fourier Transform, and then apply the proposed selection strategy prior to classification.

The novelty of this work lies in combining redundancy-aware selection with statistical significance screening on a heterogeneous feature pool spanning time-domain statistics, matrix-based descriptors, and frequency components, aiming to yield a compact and informative subset for emotion prediction. The contribution of the research is to introduce and evaluate an integrated feature selection pipeline using minimum redundancy maximum relevance and F-test screening to improve EEG emotion classification performance while reducing feature redundancy and computational burden. The contribution of the research is also to provide an empirical analysis of how the selected non-redundant features impact classifier performance for emotion dimensions such as valence and arousal using common machine learning models, thereby supporting more robust and interpretable EEG-based affect recognition.

II. PRIOR WORKS

Table 1 provides a consolidated summary of representative studies published between 2018 and 2025 on EEG-based emotion recognition. It brings together recent contributions that reflect how the field has progressed through different combinations of feature engineering, feature selection, and machine learning classification, while reporting performance using commonly adopted metrics. This synthesis is intended to offer a clear snapshot of current research directions and typical performance ranges reported in recent literature.

TABLE I. SUMMARY OF PREVIOUS RESEARCH

Authors	Year	Method	Evaluation	Key finding
Vempati, et al. [21]	2023	mRMR feature selection plus ensemble ML (subspace KNN, SVM, ANN) on multichannel EEG rhythms	Accuracy, LOSOCV	mRMR-selected gamma rhythm features with ensemble KNN reached 93.5–99.8%, confirming mRMR effectiveness for EEG emotion classification.
Zhang, et al. [22]	2023	ML (RF, SVM, KNN, XGBoost) with DWT and PCA	Accuracy, Recall, Precision, F1, ROC	Higher-frequency bands, especially gamma, were most informative. Beta and gamma correlated with emotional dynamics, with frontal sites reflecting agitation.
Abdumalikov, et al. [23]	2024	ML (RF, LR, XGBoost, SVM) plus feature selection (ANOVA F-test/LASSO/GA) with Bayesian optimization	Accuracy, hold-out validation	RF with LASSO achieved 99.39% on EEG Emotion dataset, consistently outperforming baseline across three feature selection methods.
Hamzah, et al. [24]	2024	Deep learning with time, frequency, and time-frequency features	Not stated explicitly	Review indicates increasing use of deep learning for complex EEG patterns, with feature choice remaining critical.
Rios, et al. [25]	2024	ML with PCA and PSD, RF and Extra-Trees	Accuracy, F1-score	Real-time VAD estimation system. Extra-Trees reached ~94% overall accuracy, with optimized channels including frontal, temporal, parietal, and occipital sites.

Authors	Year	Method	Evaluation	Key finding
Wang, et al. [26]	2024	GAN-based deep model with self-attention and residual design	Emotion detection accuracy	GAN framework improved emotion detection accuracy and addressed training instability such as vanishing gradients.
Mouazen, et al. [27]	2025	Hybrid ML and DL models, CNNs, RNNs, transformers, autoencoders, SHAP	Peak accuracy range	Hybrid approaches reported peak accuracy around 85–95% and emphasized interpretability and real-time feasibility.

III. METHODOLOGY

The workflow of this study is organized as a sequential pipeline that transforms EEG recordings into final emotion classification outcomes. It begins with the available EEG data representing three emotional states, namely positive, neutral, and negative, and proceeds through standardized processing stages that ensure the data are suitable for machine learning analysis and performance evaluation. This structure follows common EEG-based emotion recognition practice, where each stage is designed to reduce noise and complexity while preserving emotion-related information needed for accurate prediction.

The preprocessing stage converts continuous EEG recordings into consistent and analyzable segments. In the original data collection protocol, the raw signals were resampled and downsampled to a uniform sampling rate, then segmented using a sliding-window approach to handle the non-stationary nature of EEG. Feature extraction was subsequently performed to generate a high-dimensional feature representation, yielding 2,549 attributes that summarize relevant temporal and spectral characteristics of brain activity across standard EEG frequency bands. This feature-based representation supports efficient learning while maintaining a broad coverage of signal properties relevant to emotional variation.

Dimensionality reduction in this study is achieved through feature selection rather than transformation-based methods. Two statistical strategies are applied to identify informative features and reduce redundancy. The F-test ranks features by how strongly they differ across emotion classes, while minimum redundancy maximum relevance selects features that are highly associated with the target labels and minimally overlapping in information content. This combination is intended to produce compact feature subsets that remain discriminative, improve generalization, and reduce computational burden during model training and testing.

The selected features are then used as inputs to multiple classifiers to evaluate how feature selection influences emotion recognition performance. The study employs support vector machines, k-nearest neighbors, random forests, and neural networks, providing a comparative perspective across commonly used machine learning paradigms for EEG classification tasks. Model performance is assessed using complementary metrics that include precision, recall, F1-score, area under the curve, true negative rate, and accuracy, enabling both class-wise discrimination assessment and overall correctness evaluation. This evaluation strategy supports a systematic comparison of classifier behavior under different feature selection settings and feature subset sizes.

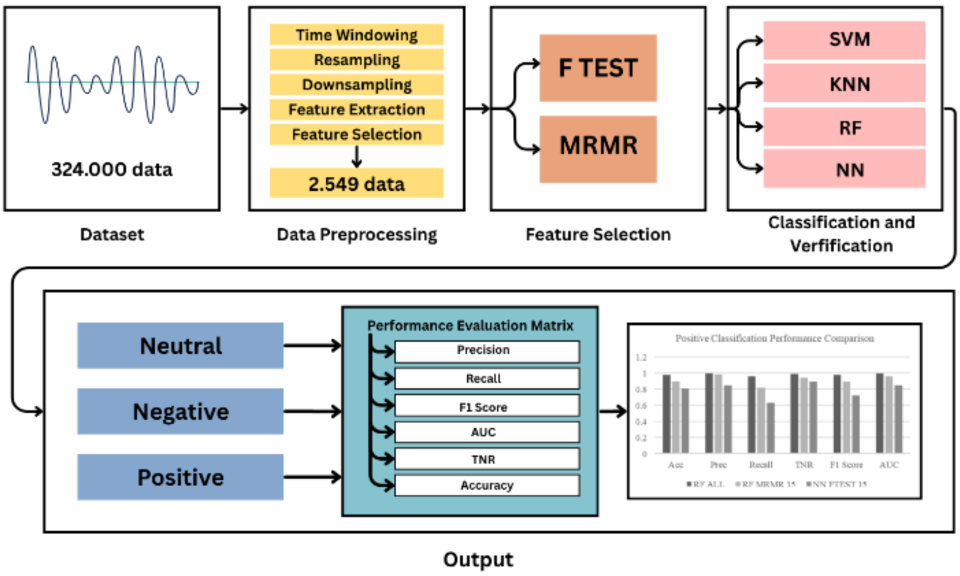

Fig. 1. Methodology of the research

A. Dataset

This study used a publicly available dataset entitled "Detecting Emotions Using EEG Waves", which was developed by Karn Suksomkit and released on Kaggle [28]. The dataset contains electroencephalogram recordings that are labeled into three emotion categories, namely positive, neutral, and negative. Each sample is represented by a set of features that were extracted from the underlying raw EEG signals, enabling supervised learning for EEG-based emotion classification. The feature set includes common time-domain descriptive statistics that summarize the distribution of EEG amplitudes. These features include the mean as an estimate of average signal level, as well as minimum and maximum values that capture extreme amplitudes within the signal segment. Such statistical descriptors provide compact summaries of overall signal behavior and are frequently used as baseline representations in emotion-related EEG modeling.

In addition to descriptive statistics, the dataset provides matrix-based and inter-channel relationship features. Covariance matrices are included to represent signal variance and cross-channel dependencies, reflecting how EEG channels co-vary during emotional processing [29]. The matrix logarithm is also provided as a transformation applied to covariance matrices to map them into a space that is often more suitable for learning algorithms [30]. Correlation coefficients further quantify linear relationships among channels, offering additional information about functional coupling patterns that may vary with emotional states [31]. The dataset also incorporates frequency-domain information derived using the Fast Fourier Transform, which converts EEG signals from the time domain into spectral components. These features aim to capture brainwave-related patterns in the frequency spectrum that are relevant to emotional processing [32]. By combining time-domain statistical descriptors, inter-channel measures, and spectral representations, the dataset supports comprehensive modeling of EEG dynamics for emotion recognition tasks.

B. Preprocessing

The EEG dataset used in this study was originally reported in earlier work that investigated emotion classification using a Muse headband device. Data were collected from two participants, one male and one female, aged 20 to 22 years. For each emotional condition, EEG activity was recorded for three minutes. In addition, six minutes of neutral resting-state data were obtained, resulting in recordings representing positive, negative, and neutral emotional states. EEG acquisition was conducted using a commercial Muse headband equipped with four dry electrodes placed at TP9, AF7, AF8, and TP10 in accordance with the international EEG placement system. Emotional states were elicited through audiovisual stimuli [33]. Negative emotions were induced using selected scenes from Marley and Me, Up, and My Girl, while positive emotions were elicited using La La Land, Slow Life, and Funny Dogs. The neutral condition was recorded without external stimuli and was performed before the emotional sessions to reduce potential carryover effects that could influence subsequent recordings [34].

The raw EEG signals were resampled and down-sampled to a uniform sampling rate of 150 Hz, producing a total of 324,000 data points [35]. To handle the non-stationary nature of continuous EEG streams, the recordings were segmented using a one-second sliding window with a 0.5-second overlap, generating shorter epochs suitable for analysis. From these segments, a total of 2,549 attributes were extracted, including statistical descriptors computed across multiple canonical frequency bands comprising delta, theta, alpha, beta, and gamma. Feature selection was subsequently applied using evaluators such as Information Gain and Symmetrical Uncertainty to identify the most informative predictors and improve computational efficiency [36]. To reduce electromyographic contamination, participants were instructed to avoid intentional movements during the recording sessions [37]. At the same time, natural behaviors such as blinking were neither explicitly encouraged nor suppressed in order to preserve ecological validity. Prior evidence indicates that blink dynamics are associated with cognitive demand and attentional processes, and thus may provide informative signals that contribute to emotion-related classification. Participants were also instructed to keep their eyes open during the tasks to maintain consistent recording conditions across sessions [38].

C. Feature Selection

Unlike studies that start from raw EEG signal processing, this research used a structured dataset obtained from Kaggle that already provides pre-extracted features derived from EEG recordings. The available attributes include both statistical descriptors and spectral characteristics, allowing the analysis to proceed directly at the feature level rather than at the continuous-signal level [39]. Because the dataset is feature-based, preprocessing in this study emphasized selecting informative variables and building classification models, rather than applying signal filtering, artifact correction, or other waveform-level operations [40]. All preprocessing and analysis were implemented in MATLAB 2024. Given the high dimensionality of the feature space, the study applied feature selection to reduce complexity and improve learning stability. Two statistical selection approaches were used, namely the F-test and minimum redundancy maximum relevance. The F-test was used to assess the discriminative association between individual features and the emotion labels, supporting the identification of features that vary significantly across classes [41].

Minimum redundancy maximum relevance was then employed to prioritize features that are strongly related to the emotion class labels while simultaneously minimizing redundancy among the selected features. This strategy aims to retain complementary information by reducing overlap between features that carry similar patterns, thereby improving the compactness and interpretability of the final feature subset used for classification [42]. The ranked feature list produced by the minimum redundancy maximum relevance procedure is shown and summarized in Figure 2 and Table 2.

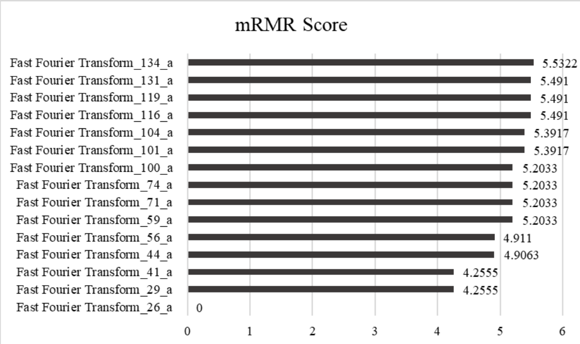

Fig. 2. Top 15 features selected using mRMR Method

TABLE II. TOP 15 FEATURES SELECTED USING MRMR METHOD

Rank	Features	mRMR Score
1	Fast Fourier Transform_746_b	5.5322
2	Mean_0_a	5.4910
3	Fast Fourier Transform_748_b	5.4910
4	Log Magnitude_72_a	5.4910
5	Covariance Matrix_50_a	5.3917
6	Log Magnitude_60_a	5.3917
7	Covariance Matrix_30_a	5.2033
8	Covariance Matrix_99_b	5.2033
9	Fast Fourier Transform_233_a	5.2033
10	Log Magnitude_4_b	5.2033
11	Correlate_59_b	4.9110
12	Moments_2_a	4.9063
13	Fast Fourier Transform_491_a	4.2555
14	Minimum Quantile_19_a	4.2555
15	Maximum Quantile_17_b	3.5684

After applying minimum redundancy maximum relevance, the F-test was used to further evaluate the statistical relevance of individual features across the emotion classes [43]. This step aimed to quantify how strongly each feature differs among class groups, providing an additional perspective for prioritizing discriminative attributes before classification. The F-test assesses the ratio between between-class variance and within-class variance for a given feature. Features that produce higher F-scores indicate larger differences among class means relative to the variability within each class, and are therefore considered to have stronger discriminative capability for emotion classification [44]. The resulting feature rankings obtained from the F-test are summarized in Table 3.

TABLE III. FEATURE RANKING RESULTS USING F-TEST METHOD

Rank	Features	F-test Score
1	Fast Fourier Transform_26_a	Inf
2	Fast Fourier Transform_29_a	Inf
3	Fast Fourier Transform_41_a	Inf
4	Fast Fourier Transform_44_a	Inf
5	Fast Fourier Transform_56_a	Inf
6	Fast Fourier Transform_59_a	Inf
7	Fast Fourier Transform_71_a	Inf
8	Fast Fourier Transform_74_a	Inf
9	Fast Fourier Transform_100_a	Inf
10	Fast Fourier Transform_101_a	Inf
11	Fast Fourier Transform_104_a	Inf
12	Fast Fourier Transform_116_a	Inf
13	Fast Fourier Transform_119_a	Inf
14	Fast Fourier Transform_131_a	Inf
15	Fast Fourier Transform_134_a	Inf

The results indicate that all selected features produced infinite F-scores, which suggests extremely strong separation among the emotion classes under the F-test criterion. In practical terms, an infinite F-score typically arises when the within-class variance for a feature approaches zero while between-class differences remain non-zero, leading to a ratio that diverges. This outcome implies that, for the evaluated samples, the corresponding features exhibit highly consistent values within each class while differing markedly across classes, and therefore appear to be highly discriminative according to the statistical test [45]. To examine how classification performance changes as the feature space becomes more compact or more expressive, subsets of the highest-ranked features were formed based on the F-test ordering. Specifically, three feature dimensions were evaluated by selecting the top 5, top 10, and top 15 ranked features for subsequent classification experiments. This staged selection supports a systematic comparison of model performance under different levels of dimensionality and helps assess whether a smaller subset can achieve comparable accuracy to larger feature sets, while potentially improving efficiency and reducing redundancy effects [46].

D. Classification Models

In this study, emotion classification was formulated as a single-label multi-class task in which each sample was assigned to exactly one category among positive, neutral, and negative emotional states. All classification procedures were implemented in MATLAB 2024 using the Classification Learner App, ensuring a consistent workflow for training, validation, and performance reporting across models. To examine how feature selection influences predictive performance, four widely used classifiers were evaluated, namely K-Nearest Neighbors, Support Vector Machine, Neural Network, and Random Forest. Each classifier was first trained using the complete feature set to establish a baseline. The same models were then re-trained and re-tested using reduced feature subsets produced by the F-test and minimum redundancy maximum relevance methods. For both selection strategies, three subset sizes were considered, consisting of 5, 10, and 15 features, to assess performance under different dimensionalities.

Across the full-feature baseline and the reduced-feature settings, a total of 28 experimental scenarios were conducted to enable a structured comparison of classifier behavior across alternative feature configurations. Model evaluation followed a hold-out validation strategy in which 70% of the samples were used for training and the remaining 30% were reserved for testing. All analyses were performed within the MATLAB environment, including confusion matrix generation and metric computation to support consistent and reproducible reporting. The primary performance measure was the area under the curve, computed on a per-class basis to reflect discrimination capability for each emotional category. Accuracy and F1-score were also reported to provide a complementary view of overall correctness and class-sensitive predictive balance, thereby enabling a more comprehensive interpretation of classification performance across models and feature selection conditions.

The following hyperparameters were applied to each classifier using the default presets in MATLAB 2024 Classification Learner. Random Forest (Bagged Trees): learner type = Decision Tree, number of learners = 30, maximum number of splits = 1492. Support Vector Machine (SVM): kernel = Linear, box constraint C = 1, kernel scale = automatic, multiclass coding = One-vs-One, with data standardization enabled. K-Nearest Neighbors (KNN): preset = Fine KNN, number of neighbors K = 1, distance metric = Euclidean, distance weight = equal, with data standardization enabled. Neural Network: preset = Wide Neural Network, hidden layer sizes = 100 (1 layer), activation = ReLU, iteration limit = 1000. These settings were held constant across all 28 experimental scenarios to ensure fair comparison.

E. Feature Extraction

In this study, feature extraction was not performed directly from raw EEG signals because the analysis relied on a publicly available Kaggle EEG dataset that already provides a rich set of pre-extracted features. The feature set is designed to represent multiple aspects of EEG activity under different emotional conditions, including statistical descriptors, inter-channel relationship measures, and frequency-domain characteristics. This approach allows the workflow to focus on feature selection and classification while still capturing essential temporal and spectral information relevant to emotion recognition [47].

Basic statistical descriptors such as mean, minimum, and maximum were included to summarize the amplitude characteristics of EEG segments. The mean reflects the average signal amplitude within a segment, while the minimum and maximum values represent the lowest and highest amplitudes, capturing the intensity and variability of neural activity. These features provide compact temporal summaries and are commonly used in EEG-based emotion recognition because they can reflect changes in baseline activity and amplitude fluctuations that occur across emotional states [48]. The mean of a signal segment can be expressed as

$$Mean = \frac{1}{N} \sum_{i=1}^{N} x_i \quad (1)$$

where x_i denotes the signal samples and N is the number of samples in the segment.

To represent coordinated activity across EEG channels, the dataset also includes covariance matrices. Covariance captures linear co-variation between channels and can reflect shared neural dynamics that may change under emotional stimulation. Because covariance matrices are positive semi-definite and are often treated as lying on a Riemannian manifold, applying the matrix logarithm maps them into a Euclidean space, which facilitates their use in conventional machine learning pipelines. The covariance between two channel signals x and y can be written as

$$\mathrm{Cov}(x, y) = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \quad (2)$$

where x_i and y_i are samples from two channels, N is the number of samples, and $\bar{x}$ and $\bar{y}$ are the corresponding channel means.

Correlation features were further used to quantify the similarity of signal patterns across channels in a normalized manner. Correlation can help identify synchronization or functional coupling, which may vary depending on emotional processing and stimulus engagement. The correlation coefficient is defined as

$$\mathrm{Corr}(x, y) = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y} \quad (3)$$

where σ_x and σ_y are the standard deviations of x and y, respectively. The coefficient ranges from -1 to 1, with values closer to ± 1 indicating stronger linear association.

In addition to time-domain and relational descriptors, frequency-domain features were derived using the Fast Fourier Transform to compute spectral representations efficiently [49]. FFT enables transformation of EEG signals from the time domain into the frequency domain so that energy or power distribution across frequency bins can be analyzed. Such spectral information is highly relevant to emotion recognition because canonical EEG bands including theta, alpha, beta, and gamma are linked to cognitive and affective processes such as relaxation, alertness, and emotional arousal [50]. The discrete Fourier transform of a length-N discrete signal $x(n)$ is given by

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi kn}{N}}, \quad k = 0,1,\dots,N-1 \quad (4)$$

where $x(n)$ is the time-domain EEG signal, k is the frequency index, N is the number of samples, and j denotes the imaginary unit.

F. Dimensionality Reduction

Rather than relying on transformation-based dimensionality reduction techniques such as Principal Component Analysis, this study adopted statistical feature selection to reduce dimensionality at the feature level. The main objective was to identify a compact subset of informative features from the high-dimensional EEG feature space while preserving interpretability, improving classification effectiveness, and reducing computational cost during model training and testing. Two complementary selection strategies were applied, namely the F-test and minimum redundancy maximum relevance. The F-test was used to quantify how strongly each feature differentiates the target emotion classes by evaluating the ratio between between-class variation and within-class variation. Features with larger F-scores were considered more discriminative and were therefore prioritized for downstream classification.

Minimum redundancy maximum relevance was used to address a common limitation of purely univariate ranking methods, which may select features that are individually strong yet highly overlapping in information content. This method selects features that are highly associated with the class labels while simultaneously minimizing redundancy among the chosen features, encouraging a feature subset that is both relevant and diverse in the information it carries. Through these selection procedures, the original feature space containing more than 2,500 features was reduced into several compact subsets. The top-ranked 5, 10, and 15 features were retained as inputs for classification experiments, enabling a systematic evaluation of model performance under different feature dimensionalities and supporting a clear assessment of the trade-off between compactness and predictive accuracy.

IV. Results

In the initial phase of this study, four classifiers were evaluated, comprising Random Forest, Neural Network, K-Nearest Neighbors, and Support Vector Machine, using feature subsets generated by the F-test and minimum redundancy maximum relevance methods. To assess the influence of dimensionality and to understand each classifier's sensitivity to feature reduction, experiments were conducted using the full feature set and reduced subsets consisting of the top 5, top 10, and top 15 ranked features. In total, 28 experimental scenarios were executed: four configurations used the full feature set (one per classifier, without feature selection), while the remaining 24 reflected combinations of two feature selection methods (F-test and mRMR), three subset sizes (top 5, 10, and 15 features), and four classifier types.

The results showed notable variability across configurations, indicating that classification performance depends strongly on the interaction between the learning algorithm and the selected feature representation. Among all evaluated settings, three configurations consistently produced high performance across the positive, neutral, and negative emotion categories. These top-performing settings were Random Forest trained on the full feature set, Random Forest trained using 15 features selected by minimum redundancy maximum relevance, and Neural Network trained using 15 features selected by the F-test. The corresponding performance summaries are presented in Table 4 and Figure 3 for the positive class, Table 5 and Figure 4 for the neutral class, and Table 6 and Figure 5 for the negative class.

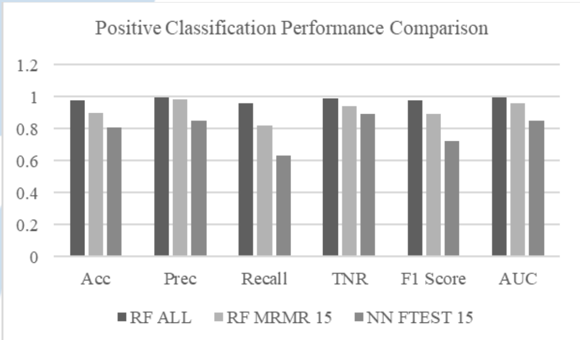

Fig. 3. Performance on Positive Emotion Classification

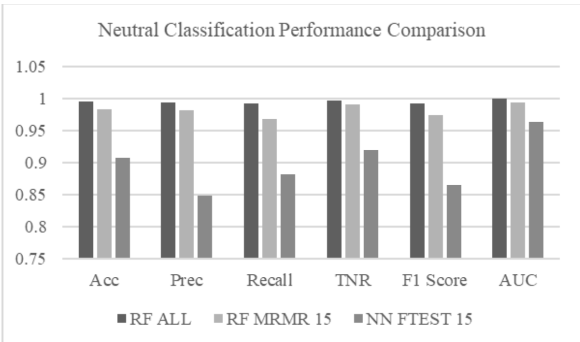

Fig. 4. Performance on Neutral Emotion Classification

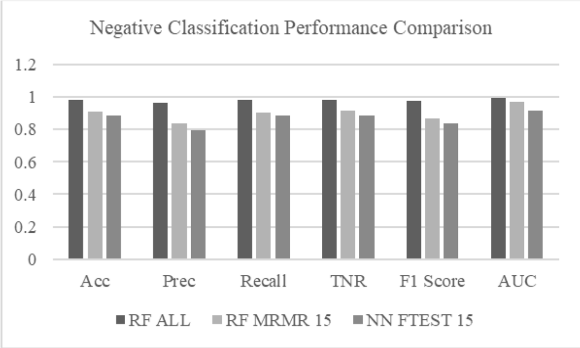

Fig. 5. Performance on Negative Emotion Classification

TABLE IV. PERFORMANCE COMPARISON OF CLASSIFIERS FOR POSITIVE EMOTION CLASSIFICATION

Metrics Evaluation	RF ALL	RF mRMR 15	NN FTEST 15
Accuracy	0.97790	0.90221	0.80643
Precision	0.99400	0.98178	0.84837
Recall	0.95766	0.82056	0.63105
TNR	0.98796	0.94283	0.89368
F1 Score	0.97549	0.89396	0.72375
AUC	0.99796	0.96020	0.85180

TABLE V. PERFORMANCE COMPARISON OF CLASSIFIERS FOR NEUTRAL EMOTION CLASSIFICATION

Metrics Evaluation	RF ALL	RF mRMR 15	NN FTEST 15
Accuracy	0.99531	0.98326	0.90757
Precision	0.99400	0.98178	0.84837
Recall	0.99202	0.96806	0.88224
TNR	0.99698	0.99093	0.92036
F1 Score	0.99301	0.97487	0.86497
AUC	0.99943	0.99460	0.96420

TABLE VI. PERFORMANCE COMPARISON OF CLASSIFIERS FOR NEGATIVE EMOTION CLASSIFICATION

Metrics Evaluation	RF ALL	RF mRMR 15	NN FTEST 15
Accuracy	0.98259	0.91092	0.88547
Precision	0.96443	0.83925	0.79385
Recall	0.98387	0.90524	0.88508
TNR	0.98195	0.91374	0.88566
F1 Score	0.97405	0.87100	0.83699
AUC	0.99643	0.97250	0.91930

For positive emotion classification (Table 4, Figure 3), RF ALL achieved the highest performance (Recall: 0.95766, F1-score: 0.97549, AUC: 0.99796). RF mRMR 15 yielded a Recall of 0.82056, F1-score of 0.89396, and AUC of 0.96020. NN FTEST 15 recorded the lowest performance among the three (Recall: 0.63105, F1-score: 0.72375, AUC: 0.85180).

For neutral emotion classification (Table 5, Figure 4), RF ALL achieved a Recall of 0.99202, F1-score of 0.99301, and AUC of 0.99943. RF mRMR 15 yielded a Recall of 0.96806, F1-score of 0.97487, and AUC of 0.99460. NN FTEST 15 recorded a Recall of 0.88224, F1-score of 0.86497, and AUC of 0.96420.

For negative emotion classification (Table 6, Figure 5), RF ALL achieved a Recall of 0.98387, F1-score of 0.97405, and AUC of 0.99643. RF mRMR 15 yielded a Recall of 0.90524, F1-score of 0.87100, and AUC of 0.97250. NN FTEST 15 recorded a Recall of 0.88508, F1-score of 0.83699, and AUC of 0.91930.

These findings demonstrate that classification performance varies notably across configurations, with Random Forest using the full feature set consistently achieving the highest results across all three emotion classes, while reduced-feature subsets and alternative classifiers showed varying degrees of performance decline depending on the selection method and subset size.

V. DISCUSSIONS

The experimental results demonstrate that EEG-based emotion recognition is strongly influenced by both the selected feature subset and the classifier used for learning. In this study, Random Forest, Neural Network, K-Nearest Neighbors, and Support Vector Machine were evaluated using two commonly applied feature selection strategies, namely the F-test and minimum redundancy maximum relevance. Each classifier was assessed using multiple feature dimensionalities, including the top 5, top 10, top 15, and the full feature set, to examine the trade-off between compact representations and predictive effectiveness for positive, neutral, and negative emotional states.

Across the 28 experimental scenarios, Random Forest trained using the full feature set, denoted as RF ALL, consistently achieved the best overall performance for all emotion classes. Its strongest outcomes were observed for the neutral class, where it reached an accuracy of 99.53 percent, an F1 score of 0.9930, and an AUC of 0.9994. Comparable results for the positive and negative classes further indicate that Random Forest remains robust in high-dimensional EEG feature spaces, likely because its ensemble structure can accommodate feature redundancy and implicitly exploit feature importance during split optimization.

When stronger dimensionality reduction was applied, performance differences became more apparent across classifiers. Neural Network trained with the top 15 features selected by the F-test, denoted as NN FTEST 15, showed noticeably lower performance, particularly for the positive class, where the F1 score and AUC declined to 0.72375 and 0.8518, respectively. This pattern suggests that some classifiers are more sensitive to reduced feature representations, especially when informative interactions among features are removed or when the remaining subset is insufficient to support stable decision boundaries. In such cases, Neural Network models may require architectural adjustment or additional regularization to better exploit limited input representations, particularly when the features are not learned end to end from the raw signal but provided as engineered descriptors.

At the same time, Random Forest trained using the top 15 features selected by minimum redundancy maximum relevance, denoted as RF mRMR 15, remained competitively strong. This indicates that redundancy-aware feature selection can preserve a large proportion of discriminative information when paired with an appropriate learner, while also reducing dimensionality and computational cost. The stable performance of Random Forest under both the full and reduced feature sets highlights its versatility and supports its suitability for settings where either maximum accuracy or a compact feature set is required.

The Support Vector Machine results showed notable sensitivity to feature dimensionality. Under the full feature set, SVM produced lower performance compared to configurations using reduced subsets selected by mRMR or F-test, which is consistent with the well-documented challenge of margin estimation in high-dimensional spaces where training samples are limited relative to feature count. The choice of kernel function and regularization parameter C are particularly important for SVM performance and should be carefully tuned for each feature configuration. K-Nearest Neighbors exhibited comparable sensitivity, performing more competitively under mRMR-reduced feature subsets where the non-redundant feature space supports more meaningful distance-based similarity computations. However, KNN was consistently outperformed by Random Forest across all configurations, likely due to KNN's susceptibility to irrelevant or overlapping features and its reliance on local neighborhood structures that may not generalize well in multi-class EEG emotion classification. These findings underline the importance of aligning feature selection strategy with the strengths and limitations of each classifier.

A broader comparison with prior work is summarized in Table 1, which reports recent studies on EEG-based emotion classification. Many prior approaches report moderate performance ranges and often focus on binary labeling or limited evaluation metrics. In contrast, the approach presented in this study achieves high three-class performance with strong discrimination measured by AUC and complementary reporting using precision, recall, and F1-score, providing a more complete assessment of classification behavior across emotion categories. This comparison supports the view that robust ensemble-based learning combined with well-chosen features can yield strong results even without end-to-end deep architectures.

Compared with previous studies summarized in Table 1, the proposed approach demonstrates several important advantages. First, this study evaluates a comprehensive combination of classifiers and feature selection strategies within a unified experimental framework consisting of 28 different scenarios, enabling a more systematic analysis of classifier–feature compatibility than most prior works, which commonly focus on a single model or limited feature configuration. Second, unlike studies that primarily emphasize binary emotion recognition, this work addresses a more challenging three-class classification problem involving positive, neutral, and negative emotional states while still achieving very high-performance metrics.

Another important advantage lies in the integration of redundancy-aware feature selection using mRMR combined with F-test evaluation. This approach not only reduces dimensionality but also preserves discriminative information effectively, as demonstrated by the strong performance of RF mRMR 15 despite using a substantially smaller feature subset. In comparison, several previous studies relied on larger feature spaces, deep architectures with higher computational requirements, or transformation-based reduction techniques that may reduce interpretability.

Furthermore, the proposed framework provides strong classification performance without requiring computationally expensive end-to-end deep learning architectures. The Random Forest model using the complete feature set achieved 99.53% accuracy and 0.9994 AUC for neutral emotion classification, outperforming or remaining competitive with many recent studies summarized in Table 1. These findings suggest that carefully optimized machine learning pipelines combined with effective feature selection strategies can achieve highly robust EEG emotion recognition while maintaining lower computational complexity and better interpretability.

The findings emphasize the importance of compatibility between the classifier and the selected feature representation. Feature selection can improve interpretability and efficiency, but its benefits depend on whether the classifier can effectively exploit the remaining information after dimensionality reduction. For high-resource scenarios, RF ALL provides the strongest option, whereas reduced-feature configurations such as RF mRMR 15 offer practical alternatives for resource-constrained deployment. Future work could investigate hybrid selection strategies, richer temporal and spectral representations, and models tailored for low-dimensional inputs, as well as validation on larger and more diverse datasets to improve generalizability and robustness.

Several limitations of this study merit acknowledgment. First, the EEG dataset was collected from only two participants using a consumer-grade Muse headband with four electrodes, which limits generalizability to broader populations and diverse acquisition settings. The small participant count and limited electrode coverage may not capture the full spatial distribution of emotion-related neural dynamics accessible through clinical-grade high-density EEG systems. Second, emotion elicitation relied on audiovisual stimuli with individually variable emotional responses, introducing subjectivity in label quality. Third, all classifiers were trained and evaluated within a single-dataset framework without cross-dataset validation, meaning transferability of findings to different EEG protocols or emotional paradigms remains untested. Fourth, the use of pre-extracted engineered features rather than end-to-end deep learning pipelines may constrain the ceiling of achievable performance. Future studies should address these limitations through larger and more diverse participant cohorts, higher-density EEG systems, cross-dataset generalization experiments, and deep learning architectures that can learn directly from raw EEG signals.

VI. CONCLUSION

This study examined how classifier choice and feature dimensionality influence EEG-based emotion recognition in a multi-class setting. Across the tested configurations, Random Forest trained using the full feature set consistently delivered the strongest performance for all emotion categories, achieving high accuracy and AUC values that exceeded those reported in several earlier studies. These findings indicate that ensemble learning methods can effectively exploit rich EEG feature representations and remain robust in high-dimensional feature spaces. Feature selection using the F-test and minimum redundancy maximum relevance substantially reduced the input dimensionality; however, the resulting impact on performance varied depending on the classifier. In particular, the Random Forest model using 15 mRMR-selected features maintained competitive accuracy and discrimination performance while using a significantly smaller feature subset, demonstrating that redundancy-aware feature selection can preserve important class-related information while improving computational efficiency. Nevertheless, the results also indicate that dimensionality reduction does not always guarantee superior classification performance, since the highest overall performance in this study was still achieved using the complete feature set.

In contrast, Neural Network models were more sensitive to feature reduction, particularly for the positive emotion class, suggesting that aggressive dimensionality reduction may remove informative feature interactions required to construct stable decision boundaries. Overall, these findings emphasize that EEG emotion recognition performance is influenced not only by the classifier or feature selection method individually, but also by the compatibility between the learning model and the selected feature representation. Random Forest appears particularly suitable for scenarios where classification accuracy is prioritized and computational resources allow the use of high-dimensional features, whereas reduced-feature configurations may provide practical advantages for real-time or resource-constrained implementations. This study also has several limitations, including the limited number of participants and the reliance on pre-extracted features from a public dataset, which may affect generalizability and restrict direct control over EEG preprocessing procedures. Future research should therefore investigate adaptive or hybrid feature selection approaches, more advanced deep learning architectures tailored to compact feature representations, and broader validation using larger and more diverse EEG datasets.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Department of Medical Technology, Institut Teknologi Sepuluh Nopember, for providing the facilities and institutional support that contributed to the completion of this research.

REFERENCES

[1] M. M. N. Bieńkiewicz *et al.*, "Bridging the gap between emotion and joint action," *Neurosci. Biobehav. Rev.*, vol. 131, pp. 806–833, Dec. 2021, doi: 10.1016/J.NEUBIOREV.2021.08.014.

[2] M. Tekke, "Beyond emotion: social awareness as a key to wellbeing and reduced violence tendency in university students," *Humanit. Soc. Sci. Commun.*, vol. 13, no. 1, 2026, doi: 10.1057/s41599-025-06475-3.

[3] L. O. H. Kroczek, A. Lingnau, V. Schwind, C. Wolff, and A. Mühlberger, "Observers predict actions from facial emotional expressions during real-time social interactions," *Behav. Brain Res.*, vol. 471, p. 115126, 2024, doi: 10.1016/j.bbr.2024.115126.

[4] A. Chaddad, Y. Wu, R. Kateb, and A. Bouridane, "Electroencephalography Signal Processing: A Comprehensive Review and Analysis of Methods and Techniques," *Sensors*, vol. 23, no. 14, p. 6434, 2023, doi: 10.3390/s23146434.

[5] E. Gkintoni, A. Aroutzidis, H. Antonopoulou, and C. Halkiopoulos, "From Neural Networks to Emotional Networks: A Systematic Review of EEG-Based Emotion Recognition in Cognitive Neuroscience and Real-World Applications," *Brain Sci.*, vol. 15, no. 3, p. 220, 2025, doi: 10.3390/brainsci15030220.

[6] A. Mehrabi, N. Sreenivasan, U. Gunawardana, and G. Gargiulo, "Hybrid Spike-Encoded Spiking Neural Networks for Real-Time EEG Seizure Detection: A Comparative Benchmark," *Biomimetics*, vol. 11, no. 1, p. 75, 2026, doi: 10.3390/biomimetics11010075.

[7] F. M. Heir, "A hybrid CNN and reinforcement learning framework for speaker identification using Mel-Spectrogram and continuous wavelet transform features," *Sci. Rep.*, 2026, doi: 10.1038/s41598-026-35858-y.

[8] A. Orhan, "A Comparative Study of Time–Frequency Representations for Bearing and Rotating Fault Diagnosis Using Vision Transformer," *Machines*, 2025, doi: 10.3390/machines13080737.

[9] N. Filimonova, "Determination of the Time-frequency Features for Impulse Components in EEG Signals," *Neuroinformatics*, 2025, doi: 10.1007/s12021-024-09698-y.

[10] R. Ahmed, "A Lightweight Hybrid Gabor Deep Learning Approach and its Application to Medical Image Classification," *Int. J. Comput. Vis.*, 2026, doi: 10.1007/s11263-025-02658-2.

[11] S. S. Al-Hadithy, "Emotion recognition in EEG Signals: Deep and machine learning approaches, challenges, and future directions," *Comput. Biol. Med.*, 2025, doi: 10.1016/j.compbiomed.2025.110713.

[12] M. Saeidi, "Neural Decoding of EEG Signals with Machine Learning: A Systematic Review," *Brain Sci.*, 2021, doi: 10.3390/brainsci11111525.

[13] A. Prakash, "Electroencephalogram-based emotion recognition: a comparative analysis of supervised machine learning algorithms," *Data Sci. Manag.*, 2025, doi:

10.1016/j.dsm.2024.12.004.

[14] H. Roubhi, "Mutual Information-based Feature Selection Strategy for Speech Emotion Recognition using Machine Learning Algorithms Combined with the Voting Rules Method," *ETASR*, 2025, doi: 10.48084/etasr.9066.

[15] Y. Xu, "Emotional recognition of EEG signals utilizing residual structure fusion in bi-directional LSTM," *Complex Intell. Syst.*, 2024, doi: 10.1007/s40747-024-01682-y.

[16] G. S. Kumar, "Classification of Human Emotional States Based on Valence-Arousal Scale using Electroencephalogram," *J. Med. Signals Sensors*, 2023, doi: 10.4103/jmss.jmss_169_21.

[17] J. Barrowclough, "Personalised Affective Classification Through Enhanced EEG Signal Analysis," *Appl. Artif. Intell.*, 2025, doi: 10.1080/08839514.2025.2450568.

[18] K. Slama, "Comprehensive review of machine learning and deep learning techniques for epileptic seizure detection and prediction based on neuroimaging modalities," *VCIBA*, 2025, doi: 10.1186/s42492-025-00208-8.

[19] N. Bhadra, "Multiclass classification of environmental chemical stimuli from unbalanced plant electrophysiological data," *PLoS One*, 2023, doi: 10.1371/journal.pone.0285321.

[20] M. H. Kabir, "Exploring Feature Selection and Classification Techniques to Improve the Performance of an Electroencephalography-Based Motor Imagery Brain–Computer Interface System," *Sensors*, 2024, doi: 10.3390/s24154989.

[21] R. Vempati and L. D. Sharma, "EEG rhythm based emotion recognition using multivariate decomposition and ensemble machine learning classifier," *J. Neurosci. Methods*, vol. 393, p. 109879, 2023, doi: https://doi.org/10.1016/j.jneumeth.2023.109879.

[22] J. Zhang, Z. Yin, and P. Chen, "EEG-based Affect Classification with Machine Learning Algorithms," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 11627–11632, 2023, doi: 10.1016/j.ifacol.2023.10.486.

[23] S. Abdumalikov, J. Kim, and Y. Yoon, "Performance Analysis and Improvement of Machine Learning with Various Feature Selection Methods for EEG-Based Emotion Classification," *Appl. Sci.*, vol. 14, no. 22, Nov. 2024, doi: 10.3390/app142210511.

[24] H. A. Hamzah and K. K. Abdalla, "EEG-based emotion recognition systems; comprehensive study," *Heliyon*, vol. 10, no. 10, p. e31485, 2024, doi: 10.1016/j.heliyon.2024.e31485.

[25] M. A. Blanco-Ríos *et al.*, "Real-time EEG-based emotion recognition for neurohumanities: perspectives from principal component analysis and tree-based algorithms," *Front. Hum. Neurosci.*, vol. 18, 2024, doi: 10.3389/fnhum.2024.1319574.

[26] L. I. Wang, J. Go, and X. Chen, "AN EEG-BASED EMOTION RECOGNITION MODEL USING AN INTERACTION DESIGN FRAMEWORK and DEEP LEARNING," *J. Mech. Med. Biol.*, vol. 24, no. 2, pp. 1–16, 2024, doi: 10.1142/S0219519424400220.

[27] B. Mouazen, A. Benali, N. T. Chebchoub, E. H. Abdelwahed, and G. De Marco, "Enhancing EEG-Based Emotion Detection with Hybrid Models: Insights from DEAP Dataset Applications," *Sensors*, vol. 25, no. 6, pp. 1–22, 2025, doi: 10.3390/s25061827.

[28] D. Gao, "A multimodal functional structure-based graph neural network for fatigue detection," *Brain Res. Bull.*, 2025, doi: 10.1016/j.brainresbull.2025.111420.

[29] A. Mellot, "Harmonizing and aligning M/EEG datasets with covariance-based techniques to enhance predictive regression modeling," *Imaging Neurosci.*, 2023, doi: 10.1162/imag_a_00040.

[30] F. Yu, "Multidimensional structural-functional coupling uncovers network dysregulation and predicts binge-eating severity in bulimia nervosa," *BMC Med.*, 2025, doi: 10.1186/s12916-025-04556-3.

[31] I. Stancin, "A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems," *Sensors*, 2021, doi: 10.3390/s21113786.

[32] M. Blanco-Ruiz, "Emotion Elicitation Under Audiovisual Stimuli Reception: Should Artificial Intelligence Consider the Gender Perspective?," *IJERPH*, 2020, doi: 10.3390/ijerph17228534.

[33] Y. Mashiki, "Affective dimensions of emotional memory shape corticospinal excitability in the upper and lower limbs in young males," *Neuroscience*, 2025, doi: 10.1016/j.neuroscience.2025.10.034.

[34] D. Mikhaylov, "Comparison of EEG Signal Spectral Characteristics Obtained with Consumer- and Research-Grade Devices," *Sensors*, 2024, doi: 10.3390/s24248108.

[35] A. Tukhtaev, "Feature Selection and Machine Learning Approaches for Detecting Sarcopenia Through Predictive Modeling," *Mathematics*, 2024, doi: 10.3390/math13010098.

[36] K. Bhadra, "Learning to operate an imagined speech Brain-Computer Interface involves the spatial and frequency tuning of neural activity," *Commun. Biol.*, 2025, doi: 10.1038/s42003-025-07464-7.

[37] S. M. Ceh, "Neurophysiological indicators of internal attention: An electroencephalography–eye-tracking coregistration study," *Brain Behav.*, 2020, doi: 10.1002/brb3.1790.

[38] A. Jabbar, "Spectral feature modeling with graph signal processing for brain connectivity in autism spectrum disorder," *Sci. Rep.*, 2025, doi: 10.1038/s41598-025-06489-6.

[39] B. Mdluli, "Signal Preprocessing, Decomposition and Feature Extraction Methods in EEG-Based BCIs," *Appl. Sci.*, 2025, doi: 10.3390/app152212075.

[40] M. H. T. Najaran, "A deep learning feature mapping algorithm for emotion detection via facial and audio signals," *Appl. Soft Comput.*, 2026, doi: 10.1016/j.asoc.2026.114998.

[41] M. Habibollahi, "How PCA helps multi-criteria decision making for feature selection: A feature fusion approach in bioinformatics and gene expression data," *Alexandria Eng. J.*, 2025, doi: 10.1016/j.aej.2025.09.028.

[42] I. K. Ihianle, "Minimising redundancy, maximising relevance: HRV feature selection for stress classification," *Expert Syst. Appl.*, 2024, doi: 10.1016/j.eswa.2023.122490.

[43] M. Bilucaglia, "Applying machine learning EEG signal classification to emotion-related brain anticipatory activity," *F1000Research*, 2021, doi: 10.12688/f1000research.22202.3.

[44] J. B. Nasejje, "Statistical approaches to identifying significant differences in predictive performance between machine learning and classical statistical models for survival data," *PLoS One*, 2022, doi: 10.1371/journal.pone.0279435.

[45] M. Malekipirbazari, "Performance comparison of feature selection and extraction methods with random instance selection," *Expert Syst. Appl.*, 2021, doi: 10.1016/j.eswa.2021.115072.

[46] Z. Wang, "Emotion recognition based on multimodal physiological electrical signals," *Front. Neurosci.*, 2025, doi: 10.3389/fnins.2025.1512799.

[47] F. Albasu, "Electroretinogram Analysis Using a Short-Time Fourier Transform and Machine Learning Techniques," *Bioengineering*, 2024, doi: 10.3390/bioengineering11090866.

[48] N. S. Suhaimi, "EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities," *Comput. Intell. Neurosci.*, 2020, doi:

10.1155/2020/8875426.

[49] M. Henry, “An ultra-precise Fast Fourier Transform,” *Measurement*, 2023, doi: 10.1016/j.measurement.2023.113372.

[50] L. Gong, M. Li, T. Zhang, and W. Chen, “EEG emotion recognition using attention-based convolutional transformer neural network,” *Biomed. Signal Process. Control.*, vol. 84, p. 104835, 2023, doi: 10.1016/j.bspc.2023.104835.