



# Effects of Sampling Length and Overlap Ratio on EEG Mental Arithmetic Task Performance: A Comparative Study

Samet ORAN\* , Esen YILDIRIM 

Adana Alparslan Turkes Science and Technology University, Artificial Intelligence Engineering Department, 01250, Adana, Türkiye

## Highlights

- This paper focuses on a comparison study for EEG mental tasks with Good/Bad selection.
- A combinational approach is proposed for feature extraction and selection in the study.
- A significantly enhanced and more effective classification accuracy was achieved.

## Article Info

Received: 01 Jan 2024  
Accepted: 19 Mar 2024

## Keywords

Classification algorithms  
Cognitive science  
Electroencephalography  
Feature extraction  
Wavelet Transform

## Abstract

Cognitive tasks have become quite popular in recent years. Understanding this sort of neurological research, its real-world applications, and how it may be improved in future studies are crucial. For this purpose, our study compares the classification accuracies for various segment lengths and overlap ratios for EEG recordings collected from 36 healthy volunteers during mental arithmetic tasks. EEG features are extracted from brain signals using the wavelet spectrum and the sample length and the overlap ratio of the sliding Windows are used as parameters. Feature selection was conducted using Correlation-Based and ReliefF feature selections. Subsequently, for classification results, Support Vector Machine, Random Forest, C4.5 Algorithm and k-Nearest Neighbor algorithms were employed, with the outcomes supported by the F1-score and Matthew's correlation coefficient. Therefore, the reliability of the obtained results has been ensured. In the comparisons obtained, the best average results for Accuracy, F1-score, and Matthew's correlation coefficient were found to be 0.990, 0.987, and 0.975 respectively, when applying the ReliefF feature selection method with the Support Vector Machine classifier.

## 1. INTRODUCTION

Researchers have been interested in human cognitive activity and signal processing for decades. Signal processing techniques are used for investigating the principles of the working brain, the effects of various diseases on the brain, and neural pathways during cognitive processes ever since [1-3]. Furthermore, brain-computer interfaces (BCIs) are developed with the purpose of helping people with disabilities due to neuromuscular disorders [4,5]. The purpose of a BCI system is to analyze the acquired brain signals and send the required command to a device for the desired action. For such systems, brain signals should be captured during mental tasks, which point to the activities of the brain such as imagining counting, raising a hand, or calculating. During the mental effort, physiological recordings are collected for further processing. EEG is one of the most frequently used modalities, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) and provides higher temporal resolution.

The related studies utilizing different datasets and the dataset we used in this study are shown in Table 1 and Table 2, respectively. Table 1 demonstrates various techniques for mental tasks that belong to different datasets. Sharma et al. (2021) [6] proposed an efficient mental arithmetic task load characterization approach using EEG signals collected from 30 healthy subjects and Bayesian structured k-Nearest Neighbor (BO-k-NN) classifier. Yavuz & Aydemir (2020) [7], a two-class mental arithmetic-based EEG+NIRS dataset, which was collected from 29 participants, was used. Higuchi fractal dimension-based features were used to extract the feature oxygenated hemoglobin and deoxygenated hemoglobin signals. Ergün & Aydemir (2020) [8], applied the dataset which is used in [7]. They aimed to

apply novel fusion preprocessing and the features were extracted by the auto-regressive model to detect mental work. Edla et al. (2018) [9] proposed an experiment to predict the concentration and meditation of data established on 40 subjects. Lim et al. (2015) [10], presented and performed an experiment for identifying the mental workload associated with no-task, visual task, auditory task, and multitasking performance with 12 volunteers.

**Table 1. An Overview of Relevant Works with the Different Datasets from the Literature**

Study References	# of Participant	# of Channels	Mental Task Types	Signals	Feature Extraction	Validation Methods	Best Accuracy
Sharma et al. [6]	30	2	Mental Arithmetic	EEG	Sample Entropy	k-fold cross validation	BO-k-NN – 96.0%
Yavuz & Aydemir [7]	29	30	Mental Arithmetic	(EEG+NIRS)	Higuchi Fractal Dimension	k-fold cross validation	LDA – 94.0%
Ergün & Aydemir [8]	29	30	Mental Arithmetic	EEG	Autoregressive Model	k-fold cross validation	k-NN – 99.7%
Edla et al. [9]	40	21	Concentration and Meditation	EEG	Mean, SD and Min-Max Amplitude	NM*	RF – 75.0%
Lim et al. [10]	12	14	None, Visual, Auditory, Multitask	EEG	Fourier Transform and Higuchi	k-fold cross validation	SVM – 90.3%

NM: not mentioned; k-NN: k nearest neighbor; SVM: support vector machine; RF: random forest; LDA: linear discrimination analysis; NIRS: near-infrared spectroscopy; BO-k-NN: bayesian optimized k-nearest neighbor. #: number.

**Table 2. An Overview of Relevant Works with the Same Dataset from the Literature**

Study References	Feature Extraction	Feature Selection	Validation Methods	Classifier(s)	Classification Output	Best Accuracy
Fatimah et al. [11]	FDM	Kruskal-Wallis	6-fold cross validation	k-NN, SVM, LR	Before / After	SVM – 98.6%
Ahammed & Ahmed [12]	MMSE	NM*	NM*	ANOVA, t-test, SVM	Good / Bad	SVM – 87.5%
Rahman et al. [13]	DWT	NM*	Gaussian PDF	k-NN, SVM	Good / Bad	SVM – 72.2%
O'Reilly & Chanmittakul [14]	PSD	L1 regularization	6-fold cross validation	SVM, LR, MLP, RF, GNB, DT	Resting / Counting	SVM – 88.9%
Mridha et al. [15]	TF, Mean, Entropy, Shannon	NM*	10-fold cross validation	k-NN, SVM, GB, LR, RF, GNB	Resting / Counting	RF – 99.8%
Babu et al. [16]	TF, WT	NM*	70% Train / 30% Test	LDA, ANN, RSC	Good / Bad	RSC – 87.5%
Malviya & Mal [17]	DWT	CNN	10-fold cross validation	BLSTM	Good / Bad	BLSTM – 98.10%
Saini et al. [18]	SWT	Two-sample t-test	5-fold cross validation	k-NN, SVM, RF, GDA, NB, LR, DT	Resting / Counting	SVM – 98.0% (Average)
Bergil et al. [19]	WT	NM*	10-fold cross validation	LR, SVM, LDA, k-NN	Good / Bad	kNN – 97.22%
Baygin et al. [20]	FRLP	INCA	10-fold cross validation,	SVM	Good / Bad	SVM – 97.88%
<b>This study</b>	<b>WT</b>	<b>CFS, ReliefF</b>	<b>10-fold cross validation</b>	<b>k-NN, RF, J48, SVM</b>	<b>Good / Bad</b>	<b>SVM – 99.00%</b>

NM: not mentioned; FDM: fourier decomposition method; MMSE: multivariate multiscale entropy; TF: temporal features; WT: wavelet transform; DWT: discrete wavelet transform; PSD: power spectral density; RSC: random subspace classifier; ANN: artificial neural network; SWT: stationary wavelet transform; LR: logistic regression; LOSO: leave-one-subject-out; CFS: correlation-based feature selection; INCA: iterative neighborhood component analysis; ANOVA: analysis of variance; GB: gradient booster; CNN: convolutional neural network; BLSTM: bidirectional long short-term memory; GNB: gaussian naive bayes; MLP: multilayer perceptron; DT: decision tree; GDA: gaussian discriminant analysis; NB: naive-bayes; #: number.

Table 2 shows distinct methods for mental arithmetic tasks for EEG signal classification using the dataset we use in this study [21]. Fatimah et al. (2020) [11] presented a single lead EEG signal. The Fourier Decomposition method was applied for feature extraction. Additionally, the Kruskal-Wallis [22] method was used for feature selection. The selected features were classified into “before” and “after” mental states. The state results were obtained by using machine learning algorithms that are SVM, kNN and LR. Ahammed & Ahmed (2020) [12] explained a novel nonlinear complexity analysis method MMSE that was utilized for detecting mental stress. EEG signals were analyzed in the complexity domain for resting, mental counting, and good-bad counting, with the results classified using SVM. Rahman et al. (2021) [13] has

focused on statistical modeling of recorded EEG signals where Gaussian distribution is used to statistically model the selection of the signals from an arithmetic task. The signals in the dataset are divided into a number of groups using DWT. The major goal of this study is to establish a model that can assess the quality of arithmetic task signals. O'Reilly & Chanmittakul (2021) [14] used heart rate and EEG spectral power data collected from individuals doing a mental arithmetic activity to categorize their cognitive state. PSD was used to extract features and L1 regularization was utilized to select features. To categorize resting and active states with six-fold cross-validation, the following six machine learning models were used: SVM, LR, MLP, RF, GNB, and DT. Mridha et al. (2021) [15] proposed using temporal features like energy, Shannon energy, entropy, and temporal energy in combination with various classifiers to determine the relaxation state of people when they are executing mental activities like arithmetic operations. Babu et al. (2022) [16] focused on using EEG signals to analyze cognitive activity in students performing mathematical tasks. Feature extraction involved deriving the temporal features (minimum, maximum, mean, kurtosis, and skewness) and frequency features (wavelet). These features were then classified using LDA, ANN, and RSC. The dataset was validated using a 70-30% dataset split. Malviya & Mal (2022) [17] introduces a hybrid deep learning model (CNN-BLSTM) based on DWT for detecting stress levels using EEG signals. EEG signals are denoised using DWT and feature selection is performed automatically using CNN. Classification is conducted with BLSTM. The proposed model employed 10-fold cross-validation. Saini et al. (2022) [18] has aimed EEG signal decomposition using SWT into sub-bands, computation of sub-band energy features, and using seven classifiers' steps were followed. The categorization of mental/mental and baseline (resting)/mental (counting) tasks using the suggested technique with the SVM classifier has the greatest average subject-dependent classification accuracy. Bergil et al. (2023) [19] focused on classifying arithmetic task performance using EEG and ECG signals. The feature extraction method involved wavelet transform, creating a feature set from the energy of the wavelet. The study used logistic LR, SVM, LDA and k-NN methods for classification. Baygin et al. (2023) [20] presented a novel EEG signal classification model. This model incorporates a multilevel feature extraction method using a Four Rhombuses Lattice Pattern and triangle pooling. Feature selection is achieved through INCA. The classification is done using an SVM with LOSO and 10-fold cross-validation methods.

In this paper, we used an open-access dataset [21] containing EEG recordings, collected from 36 subjects, during mental counting tasks to predict if the mental arithmetic task was successful or not. We investigated the mental arithmetic task performance of the participants by classifying the counting quality (good/bad) of the test. Besides, we compared the effect of the various sample lengths and overlap ratio of segments on classification results based on the features obtained by Wavelet Transform features. Wavelet Transform was used for extracting the power on four frequency bands namely: delta ( $\delta$ , 0.5–4Hz), theta ( $\theta$ , 4–8Hz), alpha ( $\alpha$ , 8–12Hz), and beta ( $\beta$ , 12–30Hz). In order to select the relevant features, Correlation-Based Feature Selection and ReliefF were applied. The same comparison was also applied to the features selected by the Correlation-Based Feature Selection and ReliefF algorithms. The objective is to ascertain more efficacious outcomes from machine learning classifiers such as k-NN, RF, J48 and SVM by manipulating parameters such as sample length and overlap ratio. Finally, to bolster the accuracy of the classification results, a comparison was made using the F1-score and Matthew's Correlation Coefficient (MCC).

## 2. MATERIAL METHOD

In this section, we present the dataset, organization of EEG, pre-processing, the proposed solution method, and classification. The principal part focused on in this title is to extract more efficient features from Wavelet Transform by altering the data sample length and Sliding Window technique into the system to get a more efficient result from the features. Afterward, with the Correlation-Based Feature Selection and ReliefF algorithms, the best features from the Whole-Set database. Finally, inferences were made from the obtained accuracy rates. The structure of the paper is summarized as follows (see Figure 1): In the first stage, a description of the EEG dataset used in this study and its recording structure is provided. The second stage involves the application of preprocessing procedures. In the third stage, techniques for feature extraction and selection are examined. The fourth stage entails classification using machine learning algorithms, and the final stage confirms the accuracy using evaluation metrics.

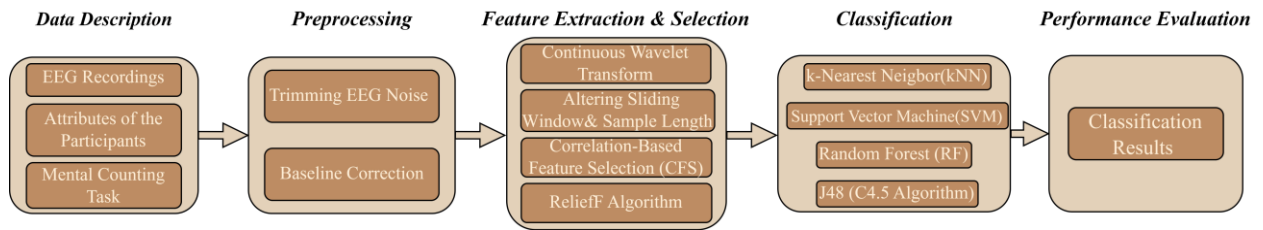


Figure 1. General view of the flowchart in the study

2.1. Data Description

In this section, the dataset [21] inspected in the study is examined under three subheadings: EEG Recordings, Attributes of Participants, and Mental Counting Task.

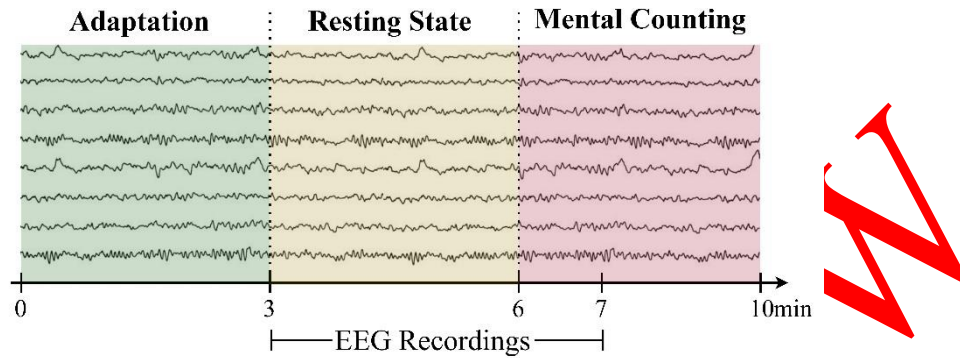
**EEG Recordings:** In this section EEG relations with the mental activities of these subjects during cognitive tasks were examined. In the preprocessing stage of the records, the 50Hz power line was eliminated by the band-stop filtering method. The sample rate of the EEG recordings is 500Hz and recordings are filtered with a low-pass filter with a cut-off frequency of 45Hz, and a high-pass filter with a cut-off frequency of 0.5Hz was used. On the other hand, using the Independent Component Analysis (ICA) filtering, the effect of blinking, muscle, and heart movements were removed. The electrodes (Fp1, Fp2, F3, F4, Fz, F7, F8, C3, C4, Cz, P3, P4, Pz, O1, O2, T3, T4, T5, T6) were located in scalp sites in accordance with to the International 10/20 scheme [23].

**Attributes of the Participants:** The open-access dataset [21] was collected from 36 (27 women and 9 men) volunteers who are nearly age-matched (between 18-26) male and female subjects. Participants were qualified to partake in the study given that they exhibited no clinical manifestations of any learning disabilities. A binary classification structure was also established to measure the quality of the task performance. The “Count quality” is a parameter that is labeled on the serial subtraction as bad (B) or good (G). Numbers of good and bad labels are obtained after the final score calculation. In Table 3, the age, gender, and count qualities of the participants are shown.

Table 3. Summary of The Participated Subjects

Name	Age	Gender	# of Subtractions	Count Quality	Name	Age	Gender	# of Subtractions	Count Quality
Sub1	21	F	9.7	B	Sub19	17	F	20	G
Sub2	18	F	29.35	G	Sub20	22	M	7.06	B
Sub3	19	F	12.88	G	Sub21	17	F	15.41	G
Sub4	17	F	31	G	Sub22	19	F	4.47	B
Sub5	17	F	8.6	B	Sub23	20	F	1	B
Sub6	16	F	20.71	G	Sub24	16	F	27.47	G
Sub7	18	M	4.35	B	Sub25	17	M	14.76	G
Sub8	18	F	13.38	G	Sub26	17	M	30.53	G
Sub9	26	M	18.24	G	Sub27	17	F	13.59	G
Sub10	16	F	7	B	Sub28	19	F	34.59	G
Sub11	17	F	1	B	Sub29	19	F	27	G
Sub12	18	F	26	G	Sub30	19	M	16.59	G
Sub13	17	F	26.36	G	Sub31	17	M	10	B
Sub14	24	M	34	G	Sub32	19	F	19.88	G
Sub15	17	F	9	B	Sub33	20	F	13	G
Sub16	17	F	22.18	G	Sub34	17	M	21.47	G
Sub17	17	F	11.59	G	Sub35	18	F	31	G
Sub18	17	F	28.7	G	Sub36	17	F	12.18	G

**Mental Counting Task:** Each volunteer's recording process takes 10 minutes. The first 3 minutes of the recording are reserved for adaptation to the experiment, the next 3 minutes for resting with closed eyes, and the last 4 minutes for the mental task. The task involved sequential subtraction of a two-digit number from a four-digit number (for instance  $1997 - 24 = 1973$ ). During the EEG recording, participants performed a mental subtraction task for a duration of 4 minutes in an acoustically isolated room. However, EEG task recordings in the dataset include the first 60 seconds out of the 4 minutes. Figure 2 illustrates the overall flow of the experiment.



*Figure 2. Structure of EEG data acquisition throughout the experiment*

During the 4-minute period, the total operations were completed, and their correctness was measured for every participant, leading to a derived mental arithmetic rating. A task was considered accurately performed if the resulting score matched a multiple of the given two-digit subtrahend. Additionally, if a participant's reported result was within 20% of the expected value, they were determined to have effectively engaged in the task. The individual challenge posed by the task to participants was determined by the rate of operations carried out and the nature of the numbers introduced.

## 2.2. Preprocessing

In this section, the extraction of features from the dataset in a healthy manner is examined under two subtopics: Trimming EEG Noise and Baseline Correction.

**Trimming EEG Noise:** The EEG recordings are indicated in two files which are 180 seconds of resting state and 60 seconds of mental tasks. However, the file lengths appear 182 seconds for resting and 62 seconds for arithmetic mental task time. When data is analyzed, the last two seconds of the recording are the noise originating from the EEG device therefore, this part was trimmed from the recorded data resulting in 180-seconds resting and 60-seconds counting time.

**Baseline Correction:** In this dataset, there is a 180-second resting period just before the 60-second task recording. The resting signal activity should be removed from the task record in order to get a more accurate result. In this study, average voltage values calculated in the resting state were subtracted from corresponding channels in the mental counting state for baseline correction.

## 2.3. Feature Extraction

In this section, the process by which features extracted through the Wavelet Transform method are formed is detailed, followed by an explanation of how the Sliding Window technique is applied to EEG signals with parameters of sample length and overlap ratio, and finally, the feature selection algorithms Correlation-based Feature Selection and ReliefF are explicated.

**Wavelet Transform:** Wavelet Transform has been used in EEG analysis frequently [24,25] to acquire the distribution of power among frequency bands within the data array. Wavelet coefficients give information about the correlation between the selected wavelet and the recorded EEG data array [26]. In this study, Morlet was used as the mother wavelet to determine the spectrum for time-frequency analysis. Power values

in 4 frequency bands (delta, theta, alpha, and beta) are obtained for each EEG channel to be used as features. The formula used to extract the EEG features is shown in Equation (1)

$$F_i^j = \frac{1}{N} \int_{f \in f_i} \int_{\tau} |S_i(f, \tau)|^2 d\tau df, \quad (1)$$

$$i \in (\delta, \theta, \alpha, \beta), \quad \text{and} \quad j = 1, \dots, 19$$

where is the extracted feature of  $j^{\text{th}}$  channel and  $i^{\text{th}}$  sub-band and  $S_i(f, t)$  is the wavelet spectrum of the  $j^{\text{th}}$  channel from the EEG segment. Finally, the number of features for a particular subject obtained was 76 (19 channels x 4 bands). During the feature extraction process, both the segment lengths and the overlap ratios between segments were used as parameters. Segment lengths were tested as 0.5, 1, 2, 3, and 4-seconds with 0%, 25%, 50%, and 75% overlap ratios. As a result, a total of 16 different evaluations were experimented with.

**Sliding Window:** The sliding window technique is performed for investigating the patterns in EEG data. EEG studies frequently use the sliding window technique for exploring brain signals through recordings divided into segments of a few seconds (or hundreds of milliseconds) with a possible overlap between segments [27-29]. In this study, the length of the segments is used as a parameter for the analysis. The EEG dataset has 36 subjects, and each subject has 60 seconds long EEG signals. In order to analyze the effect of segment length on classification accuracy, we experimented with segment lengths of 0.5, 1, 2, and 4 seconds. Besides, 0, 25, 50, and 75% overlap ratios are tested. Figure 3 shows the illustration of all overlap ratios for a 500ms long segment as an example. In this technique, the inclusion of windows (segments) by sliding them over each other is referred to as overlap, and the percentage to which the windows are included is termed the overlap ratio.

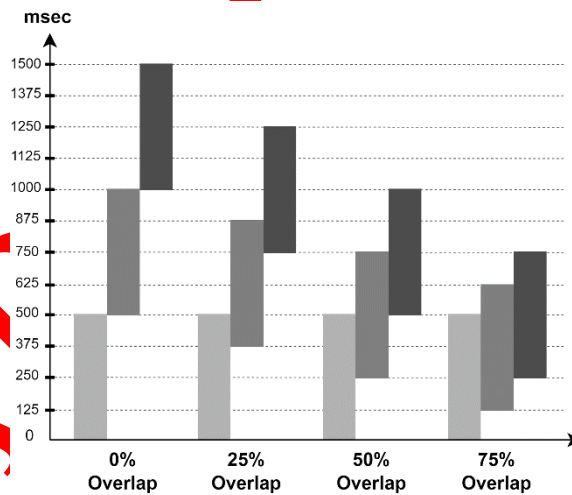


Figure 3. Overlap-ratios for a 500ms segment

In Figure 3's illustration, the section indicating 0% overlap demonstrates the windows as being completely separated, with the first window (0-500ms), the second window (500-1000ms), and the third window (1000-1500ms) illustrated accordingly. The instances where these three windows approach each other are sequentially shown as 25%, 50%, and 75% overlap. The use of overlap ratio increases the number of samples in the dataset, thereby facilitating more extensive training with the available data. Additionally, it is utilized to prevent data losses caused by the non-stationary nature of EEG signals.

**Correlation-based Feature Selection (CFS):** After the feature extraction step, we applied Correlation-Based Feature Selection (CFS) [30] to identify the best features for mental task classification. The aim of the CFS method is to eliminate irrelevant and redundant features. CFS is applied to the extracted feature set and a subset of features is selected by finding the ones that have the highest relation (correlation) to the

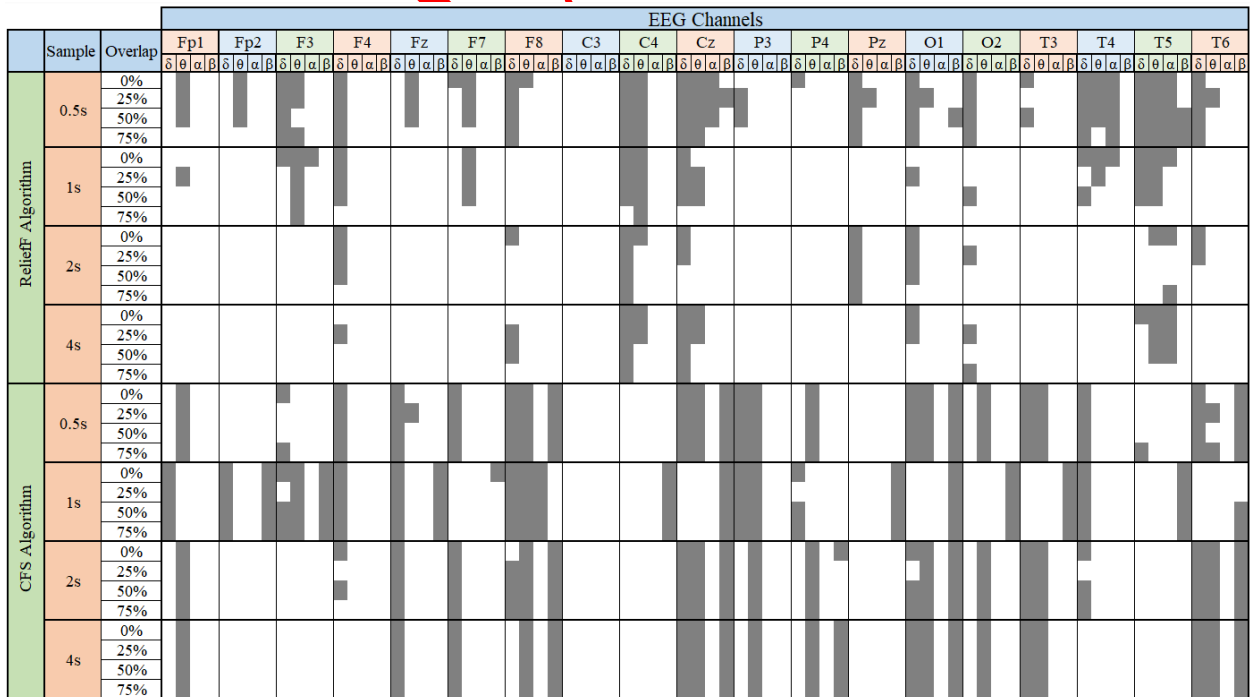
class but the lowest relation to each other [31,32]. Decreasing the number of features decreases the training and testing durations while preserving comparable, even better for some cases, classification accuracies.

**ReliefF Algorithm:** The Relief algorithm, formulated by Kira and Rendell in 1992 [33], employs a filter-method strategy for feature selection, particularly emphasizing its keen sensitivity to interactions among features. ReliefF is an enhanced version of the Relief statistical model. The ReliefF method performs feature selection by taking a sample from the dataset and constructing a model based on the proximity of the selected sample to other samples within its class and its distance from samples in different classes. The algorithm is employed for determining the significance of features within a dataset, operating through a formula that updates their weights. This process involves randomly selecting an instance from the dataset, followed by identifying k nearest neighbors within the same class (hits) and k nearest neighbors from different classes (misses). The weight update for each feature is contingent upon the difference in feature values between the chosen instance and these neighbors. Specifically, if a feature demonstrates similarity in values with hits and disparity with misses, its weight is augmented. The formula for weight adjustment is shown in Equation (2)

$$W[i] := W[i] - \frac{diff(hit)}{k} + \sum_{C \neq C_x} \frac{P(C)}{1 - P(C_x)} * \frac{diff(x, miss_C)}{k} \quad (2)$$

where  $W[i]$  denotes the weight of the  $i^{th}$  feature, diff represents the difference in feature values between the selected instance and its neighbors, and  $P(C)$  signifies the predetermined probability of class  $C$ . In each iteration,  $x$  is a randomly selected sample from the dataset. It updates the weights for each feature by comparing the features of the closest neighbors from the same class (hit) and different classes (miss). This methodology is iteratively applied across numerous instances, thus ascertaining the efficacy of features in discriminating between classes.

Figure 4 illustrates which features were selected (white) or discarded (grey) for the analyses of all sampling length/overlap ratios. Each applied analysis is labeled on the left side, while the columns represent the frequency bands (delta, theta, alpha, beta) of 19 channels. The average number of selected features is 54.06 ( $\pm 1.87$ ) for CFS and 64.18 ( $\pm 8.83$ ) for ReliefF.



**Figure 4.** Visualization of selected (white) and discarded (grey) features for CFS and ReliefF Algorithms

## 2.4. Classification

In this study, k-Nearest Neighborhood (k-NN), Random Forest (RF), C4.5 Algorithm (J48) and Support Vector Machine (SVM) algorithms, which are frequently used in analyzing EEG signals [34-39], were exercised in the classification process. In machine learning algorithms, there are adjustable parameters that allow for the oversight of the model training process, termed hyperparameters. To ascertain optimal parameters, the Grid Search technique has been employed. Grid Search can be described as an automated version of manual hyperparameter optimization. It tests every possible combination of hyperparameters. In this study, optimal hyperparameters were obtained using the GridSearchCV [40] implementation from the Python library, Scikit-Learn.

k-NN algorithm depends on distance for classification and the characteristics represented come from multifarious scales, normalizing the training data increases its accuracy significantly. In this study, kNN has three fundamental parameters which are "k", "distance metrics" and "weighting function". The k values between 1 to 50 (1,3,5,7...49) and Minkowski, Manhattan, and Euclidean, as values of the distance metric, are utilized. The weighting function was configured with "distance" and "uniform" as the selected values. The RF is a powerful supervised learning method that builds and merges decision trees to comprise a forest. RF yields an outcome that is the class chosen by the majority of trees. J48 is an algorithm in the WEKA program, an extension of the ID3 algorithm [36]. It is an algorithm that generates a decision tree created by C4.5 (an extension of ID3). The RF and J48 classifiers encompass three fundamental parameters: "random states", "n estimators", and "max depth". The values for "random states" are 0, 50, and 100; the values for "n estimators" are 64, 128, 256, and 512; and the values for "max depth" are 4, 8, 16, and 32. The SVM algorithm draws a line on a plane to separate placed points. It aims for this line to be at a maximum distance from the points of both classes. The SVM is characterized by three essential parameters: "gamma", "C", and "kernels". The gamma values explored were 1, 0.1, 0.01, 0.001, and 0.0001, while the C values were set at 0.1, 1, 10, 100, and 1000. The kernel parameter was set using "linear", "polynomial", and "radial basis function" as the values. For the aforementioned machine learning algorithms, a Grid Search was conducted over the used parameters. The names of these parameters and their values that yield the best results are presented in Table 4.

**Table 4.** The Optimal Hyperparameters of Classifiers

Classifiers	Parameters	Optimal Values
k-NN	k neighbors	5
	Distance metric	Euclidean
	Weighting function	Distance
RF & J48	n estimators	256
	Maximum depth	32
	Random state	0
SVM	Gamma	1
	C	10
	Kernel	Radial basis function

**Performance Evaluation:** Various methods are used to assess performances such as sensitivity (True Positive rate), specificity (True Negative rate), and precision (Positive Predictive value). In this study, classification accuracy, F1-score and Matthew's Correlation Coefficient (MCC) criteria are preferred in performance evaluation. In imbalanced datasets, as with this dataset, accuracy results can prove to be deceptive. Under such circumstances, The F1-score and MCC metrics have been examined [41]. The F1-score acts as the harmonic mean of precision and recall, taking into consideration both false positives and false negatives. This provides a more balanced and realistic evaluation of the true performance of the model. For the same reasons, in situations where the simple accuracy rate might be misleading, the MCC metric is as illuminating as the F1-score. MCC considers all combinations of correct and incorrect classifications, yielding values between -1 (complete disagreement) and +1 (perfect agreement). A value of 0 indicates performance equivalent to random classification. Therefore, it is employed to assess classification performance in a more comprehensive and balanced manner. Table 5 describes the accuracy, F1-score and MCC formulas.



**Table 5.** The Formulas of Evaluation Metrics

<b>Accuracy</b>	$\frac{N_{TP}+N_{TN}}{N_{TN}+N_{FP}+N_{FN}+N_{TP}}$
<b>F1-score</b>	$\frac{N_{TP}}{N_{TP}+\frac{1}{2}(N_{FP}+N_{FN})}$
<b>MCC</b>	$\frac{(N_{TP}+N_{TN}) - (N_{FP}+N_{FN})}{\sqrt{(N_{TP}+N_{FP})(N_{TP}+N_{FN})(N_{TN}+N_{FP})(N_{TN}+N_{FN})}}$
<ul style="list-style-type: none"> <li>• <math>N_{TP}</math> (True Positive) = # of the successful task correctly classified.</li> <li>• <math>N_{TN}</math> (True Negative) = # of the unsuccessful task correctly classified.</li> <li>• <math>N_{FP}</math> (False Positive) = # of the unsuccessful task incorrectly classified as successful.</li> <li>• <math>N_{FN}</math> (False Negative) = # of the successful task incorrectly classified as unsuccessful.</li> </ul>	

For performance evaluation, a leave-one-out cross validation (LOOCV) method was applied. In this method, each sample in the dataset is sequentially selected as a test instance while all other samples are used as the training set. This process is repeated for every sample in the dataset. LOOCV is beneficial for our dataset, which is not particularly large, as it allows for the evaluation of the model's performance using the maximum training data [42]. Moreover, this method is employed to assess the model's sensitivity to the data and its generalization capability.

### 3. RESULTS AND DISCUSSIONS

In this section, using machine learning algorithms, combinational results have been obtained by applying the determinative performance criteria of "sample length values" and "overlap ratios" parameters to features derived from the Wavelet Transform and to features selected through feature selection algorithms. Tables 6, 7 and 8 show subject-independent results for 36 subjects evaluated with the leave-one-out cross-validation method. In terms of content, Table 6 refers to the classification using the "Whole feature set" created from all subjects. Tables 7 and 8 present the classification results of the features obtained from the "CFS" and "ReliefF" algorithms, respectively. In Figure 5, the results of the aforementioned tables have been graphically represented, facilitating further interpretation. For Tables 6, 7 and 8, at first glance, an effective increase is observed in all classifier accuracy values as the sample length of the data records increases. Similarly, a noticeable change appeared in all classifier accuracy values of increasing overlap ratio within the same sample length. Hence, based on the data in the three tables, increasing the sample length and overlap ratio values has a positive effect on the outcomes.

Another observation regarding these parameters is that, while there's a direct proportionality between the success rate and increases in overlap ratio and sample length, it can be stated for each graph in Figure 5 that the influence of these parameters decreases with the increase in overlap ratios when moving from a 2-second sample length to 4 seconds. On the other hand, most of the poorest accuracies are recorded for 0.5-second sample length and 0% (25% for a few cases) overlap ratio value. In the tables for Whole-Set, CFS, and ReliefF, the values for accuracy, F1-score, and MCC across all classifiers have been observed to be closed. Consequently, the close alignment of the MCC and F1-score results has prevented occasional inaccuracies in the accuracy values. Furthermore, upon examining the average results of the evaluation metrics for all classifiers, the performance values are ranked as "Accuracy > F1-score > MCC".

**Table 6. The Classification Results of The Whole-Set**

Whole-Set Classification Results													
Sample	Overlap	KNN			SVM			J48			RF		
		Acc	F1score	MCC	Acc	F1score	MCC	Acc	F1score	MCC	Acc	F1score	MCC
0.5s	0%	0.930	0.909	0.822	0.957	0.945	0.892	0.906	0.872	0.759	0.908	0.876	0.766
	25%	0.938	0.921	0.844	0.965	0.956	0.912	0.905	0.871	0.759	0.912	0.880	0.775
	50%	0.950	0.936	0.873	0.977	0.971	0.943	0.921	0.894	0.800	0.927	0.903	0.815
	75%	0.976	0.970	0.939	0.986	0.983	0.965	0.933	0.911	0.832	0.940	0.920	0.848
1s	0%	0.973	0.966	0.933	0.983	0.978	0.957	0.938	0.918	0.843	0.940	0.921	0.849
	25%	0.979	0.973	0.946	0.990	0.987	0.975	0.949	0.932	0.871	0.953	0.939	0.882
	50%	0.986	0.983	0.965	0.996	0.995	0.990	0.959	0.946	0.896	0.965	0.955	0.912
	75%	0.995	0.994	0.989	0.998	0.998	0.995	0.973	0.966	0.933	0.979	0.974	0.948
2s	0%	0.988	0.985	0.970	0.992	0.989	0.979	0.962	0.951	0.904	0.967	0.957	0.916
	25%	0.990	0.987	0.975	0.994	0.992	0.984	0.964	0.954	0.911	0.973	0.966	0.932
	50%	0.995	0.994	0.987	0.998	0.998	0.995	0.979	0.973	0.947	0.980	0.974	0.949
	75%	0.999	0.999	0.998	1.000	1.000	0.999	0.990	0.988	0.976	0.993	0.991	0.983
4s	0%	0.985	0.981	0.964	0.994	0.993	0.986	0.963	0.952	0.907	0.961	0.950	0.903
	25%	0.996	0.994	0.989	0.997	0.996	0.993	0.974	0.966	0.935	0.974	0.966	0.935
	50%	0.998	0.998	0.995	0.999	0.999	0.998	0.983	0.978	0.957	0.985	0.981	0.962
	75%	1.000	0.999	0.999	1.000	0.999	0.999	0.994	0.992	0.984	0.996	0.994	0.989
<b>Average</b>		<b>0.980</b>	<b>0.974</b>	<b>0.949</b>	<b>0.989</b>	<b>0.986</b>	<b>0.973</b>	<b>0.956</b>	<b>0.942</b>	<b>0.888</b>	<b>0.959</b>	<b>0.947</b>	<b>0.898</b>

**Table 7. The Classification Results of the CFS-Set**

Feature-Set After CFS Classification Results													
Sample	Overlap	KNN			SVM			J48			RF		
		Acc	F1score	MCC	Acc	F1score	MCC	Acc	F1score	MCC	Acc	F1score	MCC
0.5s	0%	0.933	0.913	0.829	0.963	0.953	0.906	0.906	0.873	0.760	0.907	0.875	0.762
	25%	0.937	0.919	0.840	0.961	0.950	0.902	0.904	0.869	0.754	0.909	0.876	0.767
	50%	0.949	0.935	0.872	0.975	0.969	0.938	0.918	0.890	0.792	0.925	0.900	0.809
	75%	0.973	0.966	0.933	0.983	0.979	0.958	0.931	0.908	0.826	0.938	0.919	0.845
1s	0%	0.965	0.956	0.912	0.979	0.974	0.948	0.930	0.907	0.823	0.938	0.918	0.844
	25%	0.974	0.967	0.934	0.984	0.980	0.960	0.945	0.928	0.862	0.949	0.934	0.873
	50%	0.981	0.977	0.954	0.993	0.991	0.982	0.951	0.935	0.876	0.957	0.944	0.892
	75%	0.995	0.994	0.989	0.996	0.995	0.991	0.969	0.960	0.923	0.974	0.967	0.935
2s	0%	0.988	0.985	0.970	0.992	0.990	0.979	0.956	0.944	0.891	0.966	0.956	0.914
	25%	0.991	0.988	0.977	0.994	0.992	0.984	0.966	0.956	0.914	0.970	0.962	0.925
	50%	0.995	0.994	0.988	0.998	0.998	0.995	0.980	0.974	0.949	0.978	0.972	0.945
	75%	0.999	0.999	0.997	1.000	1.000	0.999	0.990	0.987	0.975	0.992	0.990	0.980
4s	0%	0.987	0.984	0.968	0.998	0.998	0.995	0.965	0.955	0.912	0.965	0.955	0.912
	25%	0.994	0.993	0.986	0.997	0.996	0.993	0.972	0.964	0.931	0.981	0.976	0.953
	50%	0.998	0.998	0.995	0.999	0.999	0.998	0.984	0.980	0.960	0.985	0.981	0.962
	75%	1.000	0.999	0.999	1.000	0.999	0.999	0.996	0.994	0.989	0.995	0.993	0.986
<b>Average</b>		<b>0.979</b>	<b>0.973</b>	<b>0.946</b>	<b>0.988</b>	<b>0.985</b>	<b>0.970</b>	<b>0.954</b>	<b>0.939</b>	<b>0.883</b>	<b>0.958</b>	<b>0.945</b>	<b>0.894</b>

When the results in Table 6 are analyzed, the best average values are obtained as “Accuracy = 0.989, F1-score = 0.986 and MCC = 0.973” with SVM classifier. Upon examining Figure 5, it can be observed that, excluding the segment for 0.5-second sample length, there is virtually no rate of performance increase for other sample lengths in the Whole-Set / SVM graph. In Table 6, when the average MCC values are compared with other classifiers, they are ranked as follows: "SVM = 0.973, kNN = 0.949, RF = 0.898, J48 = 0.888". An analysis of the results in Table 7 reveals that the optimal average values are "Accuracy = 0.988, F1-score = 0.985, and MCC = 0.970" using SVM classifier. Figure 5 suggests that, aside from the data for the 1-second sample length, there is negligible performance improvement across other sample lengths in the CFS-Set / SVM representation. Furthermore, based on Table 7, when evaluating average

**Table 8.** The Classification Results of the ReliefF-Set

Feature-Set After ReliefF Classification Results													
Sample	Overlap	KNN			SVM			J48			RF		
		Acc	F1score	MCC	Acc	F1score	MCC	Acc	F1score	MCC	Acc	F1score	MCC
0.5s	0%	0.939	0.923	0.846	0.961	0.950	0.901	0.905	0.872	0.758	0.913	0.884	0.778
	25%	0.949	0.935	0.871	0.966	0.957	0.915	0.912	0.881	0.777	0.916	0.887	0.787
	50%	0.956	0.945	0.890	0.975	0.968	0.937	0.921	0.894	0.800	0.930	0.906	0.822
	75%	0.979	0.974	0.948	0.984	0.980	0.960	0.935	0.914	0.836	0.941	0.922	0.851
1s	0%	0.979	0.974	0.948	0.988	0.984	0.969	0.938	0.917	0.842	0.943	0.924	0.855
	25%	0.982	0.978	0.956	0.989	0.987	0.974	0.950	0.935	0.874	0.955	0.941	0.886
	50%	0.989	0.987	0.974	0.996	0.996	0.991	0.961	0.950	0.903	0.965	0.955	0.913
	75%	0.996	0.995	0.990	0.998	0.998	0.995	0.975	0.968	0.938	0.980	0.974	0.949
2s	0%	0.995	0.994	0.989	0.993	0.991	0.982	0.964	0.953	0.909	0.970	0.962	0.926
	25%	0.993	0.991	0.982	0.996	0.995	0.989	0.969	0.960	0.921	0.972	0.964	0.929
	50%	0.996	0.995	0.991	0.999	0.998	0.996	0.978	0.972	0.945	0.980	0.975	0.951
	75%	0.999	0.999	0.998	1.000	1.000	0.999	0.991	0.989	0.979	0.993	0.991	0.983
4s	0%	0.993	0.991	0.982	0.998	0.998	0.995	0.957	0.945	0.893	0.967	0.957	0.917
	25%	0.996	0.994	0.989	0.997	0.996	0.993	0.974	0.967	0.935	0.981	0.976	0.953
	50%	0.999	0.999	0.998	0.999	0.999	0.998	0.982	0.977	0.955	0.984	0.979	0.960
	75%	1.000	0.999	0.999	1.000	0.999	0.999	0.995	0.993	0.986	0.996	0.995	0.990
<b>Average</b>		<b>0.984</b>	<b>0.980</b>	<b>0.959</b>	<b>0.990</b>	<b>0.987</b>	<b>0.975</b>	<b>0.957</b>	<b>0.943</b>	<b>0.891</b>	<b>0.962</b>	<b>0.950</b>	<b>0.903</b>

MCC values among different classifiers, the sequence is: "SVM = 0.970, kNN = 0.946, RF = 0.894, J48 = 0.883". A review of the findings in Table 8 indicates that the best average metrics achieved are "Accuracy = 0.990, F1-score = 0.987, and MCC = 0.975" when employing the SVM classifier. Observations from Figure 5 imply that other than for the 1-second sample length, there's a minimal increase in performance for other sample durations in the ReliefF-Set / SVM chart. Additionally, according to Table 8, in terms of average MCC scores among various classifiers, they are ranked as: "SVM = 0.975, kNN = 0.959, RF = 0.903, J48 = 0.891".

In Figure 5, based on the obtained results, the most successful classifier was SVM, achieving an F1-score of 0.987 and an MCC of 0.975 when using the Feature-Set After ReliefF. From the results gathered, RF emerged as the low-performing classifier with an F1-score of 0.945 and an MCC of 0.894 when applied to the Feature-Set After CFS. Upon examining Figure 5 for the Feature-Set After ReliefF and MCC metric across all classifiers, a notable increase in results is observed when transitioning from a 0.5-second sample length to a 1-second sample length, amounting to approximately 8.5%. On the other hand, for SVM and kNN, the rate is around 3.5% when moving from a 1-second to a 2-second sample length, while for RF and J48, this average is about 10%. Similarly, for RF and J48, there's a 7.5% performance increase rate when shifting from a 2-second to a 4-second sample length whereas kNN and SVM demonstrate an average increment of just 0.8%. Therefore, the increase in sample length and overlap ratio has most positively impacted the RF and J48 classifiers.

One of the primary issues to be discussed in the above results is the optimal sample length and overlap ratio. The results presented in the three tables demonstrate that elevating the values of sample length and overlap ratio yields beneficial effects on the results. Such a tendency underscores the importance of refining these parameters to augment the performance of the model significantly. Generally, a 50% overlap and a 1-second sample length are deemed sufficient in EEG studies, but this study compared outcomes by providing various values for sample length and overlap ratio. Although the increase in overlap ratio consistently yields positive effects, the computational demand escalates with the size of the data due to the increased volume generated by the overlapping process. Therefore, determining the optimal value based on the dataset in use is of importance.

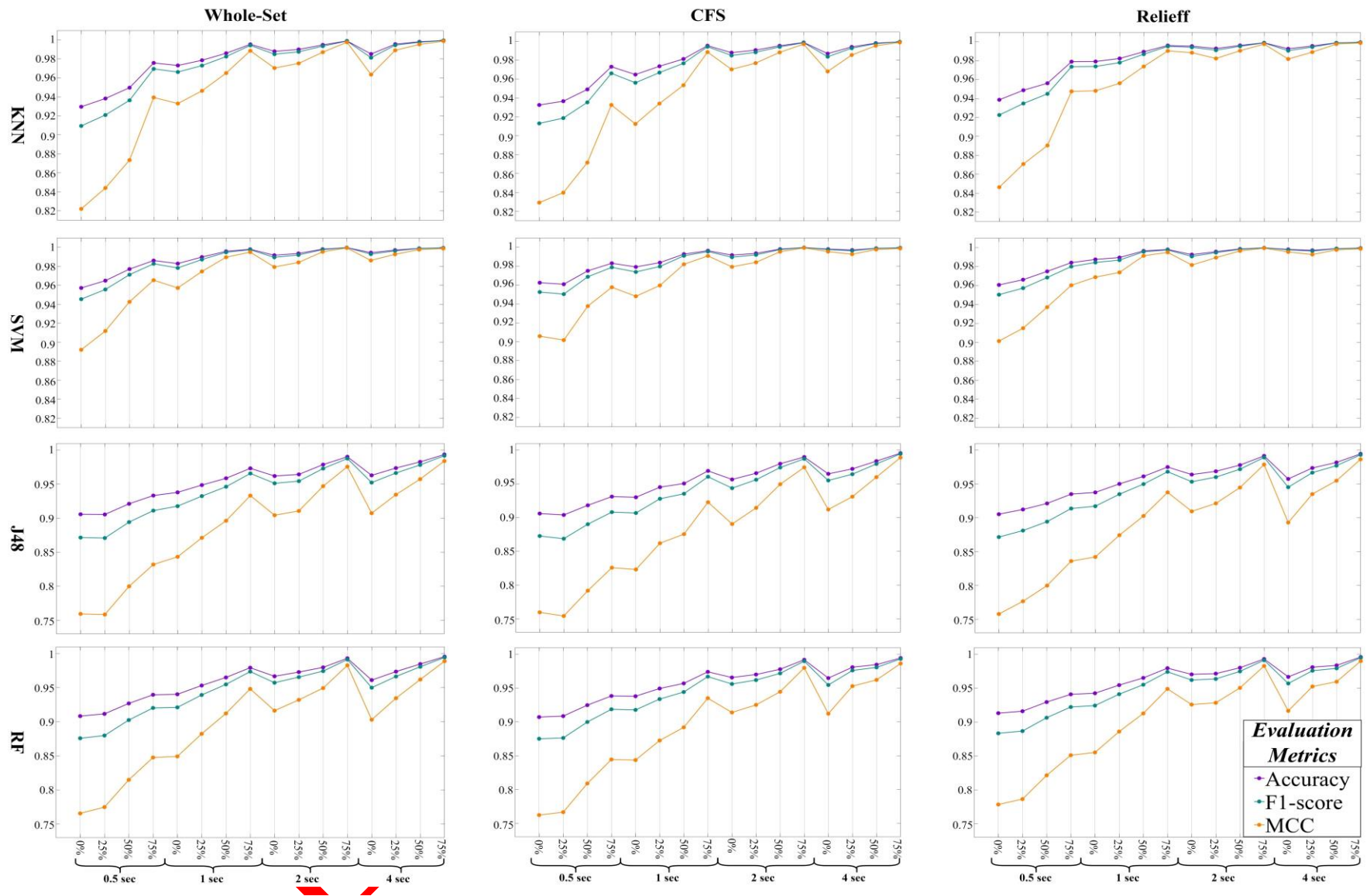


Figure 5. The results of machine learning algorithms (kNN, SVM, J48, and RF) using wavelet features (Whole-Set, CFS, and ReliefF) are depicted graphically in terms of Accuracy, F1-score, and MCC metrics

Another point of interest in the study is that the most traditional feature selection method, CFS, has yielded slightly lower results from the features already obtained. Conversely, the ReliefF method has shown a marginally positive impact on the results. The most significant inference that can be drawn from this is that each EEG channel in the dataset is capable of capturing a sufficient and significant amount of information.

Lastly, an aspect that requires discussion pertains to the classifiers. The kNN and SVM algorithms have shown successful outcomes with a small margin of difference, demonstrating correlated performance patterns. On the other hand, when considering the decision tree algorithms, RF and J48, although they exhibit some internal correlation, they have not achieved as successful results as kNN and SVM. Finally, the SVM achieved the highest average success rates (Confusion matrix: TP = 1140, FN = 21, FP = 20, TN = 2986), recording an accuracy of 0.990, an MCC of 0.975 and an F1-score of 0.987. In contrast, the lowest success rates were found in the J48, which attained with MCC of 0.883 and an F1-score of 0.939.

As can be seen from Table 2, upon examining articles in the literature that utilize the same dataset, it is observed that studies classifying Resting / Counting (including Before/After study) have achieved results close to 100%. However, it is already known that Resting and Mental states exhibit distinct differences, hence achieving high accuracies in such classifications is expected. On the other hand, as shown in Table 2, studies that made Good / Bad distinctions reveal accuracy values ranging between 72.20% and 98.10%. When comparing our results to those obtained in studies focusing on Good / Bad, our best outcome was achieved using SVM with an accuracy of 99.0%, suggesting that this study is promising.

#### **4. CONCLUSION**

EEG cognitive tasks are pivotal in neuroscience for their capacity to non-invasively record electrical activity in the brain, providing real-time insight into the neural underpinnings of cognitive processes. This area is crucial for advancing our understanding of the brain-behavior relationship and enabling the development of targeted interventions. In order to contribute to this research, we used an open-access dataset “Electroencephalograms during Mental Arithmetic Task Performance” [21] in this study. A mental task analysis is explained for precise classification of whether the mental arithmetic task was successful or not (Good/Bad selection). The proposed method is leaned on Wavelet Transform obtaining energy sub-bands as EEG features. The purpose is to demonstrate the results obtained from machine learning classifiers by altering the parameters of sample length and overlap ratio, utilizing EEG features that are Whole-Set, CFS-Set, and ReliefF-Set. Additionally, we utilized the grid search technique to ascertain the optimal values of hyperparameters for the machine learning classifiers. The F1-score and MCC metrics have been employed to reinforce the accuracy results. The leave-one-out cross-validation method was employed to reduce the model's bias towards training data and to provide more insight into the model's generalization capabilities. When the results are examined, as mentioned in the previous section, the highest values are obtained using ReliefF features with an SVM classifier, averaging across all participants, resulting in an accuracy of 0.990, an F1-score of 0.987, and an MCC of 0.975.

#### **CONFLICTS OF INTEREST**

No conflict of interest was declared by the authors.

#### **REFERENCES**

- [1] Sörnmo, L., and Laguna, P., “Bioelectrical Signal Processing In Cardiac And Neurological Applications”, Elsevier Academic Press, 8: 25-53, (2005).
- [2] Baillet, S., Mosher, J. C., and Leahy, R. M., “Electromagnetic brain mapping”, IEEE, 18(6): 14–30, (2001).

- [3] Fukushima, K., Fukushima, J., Warabi, T., and Barnes, G. R., "Cognitive processes involved in smooth pursuit eye movements: behavioral evidence, neural substrate and clinical correlation", *Frontiers in Systems Neuroscience*, 7(4), (2013).
- [4] McFarland D. J., and Wolpaw, J. R., "EEG-based brain-computer interfaces", *current opinion in Biomedical Engineering*, 4: 194-200, (2017).
- [5] Chaudhary, U., Birbaumer, N., and Ramos-Murguialday, A., "Brain-computer interfaces for communication and rehabilitation", *Nature Reviews Neurology*, 12(9): 513-525, (2016).
- [6] Sharma, L. D., Chhabra, H., Chauhan, U., Saraswat R. K., and Sunkaria, R. K., "Mental Arithmetic Task Load Recognition Using EEG Signal and Bayesian Optimized K-Nearest Neighbor", *International Journal of Information Technology*, 13(6): 2363–2369, (2021).
- [7] Yavuz, E., and Aydemir, O., "Classification of Mental Arithmetic Based Hybrid EEG+Nirs Signals", 28th Signal Processing and Communications Applications Conference (SIU). IEEE, Gaziantep, Turkey 1–4, (2020).
- [8] Ergün, E., and Aydemir, O., "A New Evolutionary Preprocessing Approach for Classification of Mental Arithmetic Based EEG Signals", *Cognitive Neurodynamics*, 14(5): 609–617, (2020).
- [9] Edla, D. R., Mangalorekar, K., Dhavalikar, G., and Dodia, S., "Classification of EEG Data for Human Mental State Analysis Using Random Forest Classifier", *Procedia Computer Science*, 132: 1523–1532, (2018).
- [10] Lim, W., Sourina, O., Liu, Y., and Wang, L., "EEG-Based Mental Workload Recognition Related to Multitasking", 10th International Conference on Information, Communications and Signal Processing (ICICS). IEEE, Singapore, 1–4, (2015).
- [11] Fatimah, B., Javali, A., Ansar, H., Harshitha, B., and Kumar, H., "Mental Arithmetic Task Classification Using Fourier Decomposition Method", *International Conference on Communication and Signal Processing (ICCSP)*. IEEE, Chennai, India, 46–50, July (2020).
- [12] Ahammed, K., and Ahmed, M. U., "Quantification of mental stress using complexity analysis of EEG signals", *Biomedical Engineering: Applications, Basis and Communications*, 32(2): 2050011, (2020).
- [13] Rahman, S. M. Z., Tawana, I., Mostafiz, H. R., Chowdhury T. T., and Shahnaz, C., "Arithmetic Mental Task of EEG Signal Classification Using Statistical Modeling of The Dwt Coefficient", *International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*. IEEE, Rajshahi, Bangladesh, 1–4, (2021).
- [14] O'Reilly J. A., and Chanmittakul, W., "L1 Regularization-Based Selection of EEG Spectral Power and Ecg Features for Classification of Cognitive State", 9th International Electrical Engineering Congress (iEECON). IEEE, Pattaya, Thailand, 365–368, (2021).
- [15] Mridha, K., Kumar, D., Shukla, M., and Jani, M., "Temporal Features and Machine Learning Approaches to Study Brain Activity with EEG And Ecg", *International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. IEEE, Greater Noida, India, 409–414, (2021).
- [16] Babu, T. A., Gadde, S., Ravi, S., Rao, K. V., Mamillu, Y., and Krishna, D., "Analysis of Mental Task Ability in Students based on Electroencephalography Signals", In 2022 IEEE International

- Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), 1: 274-278, IEEE, (2022).
- [17] Malviya, L., and Mal, S., "A novel technique for stress detection from EEG signal using hybrid deep learning model", *Neural Computing and Applications*, 34(22): 19819-19830, (2022).
- [18] Saini, M., Satija U., and Upadhayay, M. D., "Discriminatory Features Based on Wavelet Energy for Effective Analysis of Electroencephalogram During Mental Tasks", *Circuits, Systems, and Signal Processing*, 1–29, (2022)
- [19] Bergil, E., Oral C., and Ergül, E. U., "Classification of arithmetic mental task performances using EEG and ECG signals", *The Journal of Supercomputing*, 1-13, (2023).
- [20] Baygin, N., Aydemir, E., Barua, P. D., Baygin, M., Dogan, S., Tuncer, T., Tan R. S., and Acharya, U. R., "Automated mental arithmetic performance detection using quantum pattern-and triangle pooling techniques with EEG signals", *Expert Systems with Applications*, 227, 120306, (2023).
- [21] Zyma, I., Tukaev, S., Seleznov, I., Kiyono, K., Popov, A., Chernykh, M., and Shpenkov, O., "Electroencephalograms During Mental Arithmetic Task Performance", *Data*, 4(1): 14, (2019).
- [22] Vargha, A., and Delaney, H. D., "The Kruskal-Wallis Test and Stochastic Homogeneity", *Journal of Educational and Behavioral Statistics*, 23(2): 170–192, (1998).
- [23] Homan, R. W., Herman J., and Purdy, P., "Cerebral Location of International 10–20 System Electrode Placement", *Electroencephalography and Clinical Neurophysiology*, 66(4): 376–382, (1987).
- [24] Adeli, H., Zhou, Z., and Dadmehr, N., "Analysis of EEG Records in An Epileptic Patient Using Wavelet Transform", *Journal of Neuroscience Methods*, 123(1): 69–87, (2003).
- [25] Hazarika, N., Chen, J. Z., Tsoi, A. C., and Sergejew, A., "Classification of EEG Signals Using the Wavelet Transform", *Signal Processing*, 59: 1, 61–72, (1997).
- [26] Pathak, R. S., "The Wavelet Transform", *Springer Science & Business Media*, 4: 109-128, 2009.
- [27] Tzamourta, K. D., Astrakas, L. G., Gianni, A. M., Tzallas, A. T., Giannakeas, N., Paliokas, I., Tsalikakis, D. G., and Tsipouras, M. G., "Evaluation of window size in classification of epileptic short-term EEG signals using a Brain Computer Interface software", *Engineering, Technology & Applied Science Research*, 8(4): 3093-3097, (2018).
- [28] Gaur, R., Gupta, H., Chowdhury, A., McCreddie, K., Pachori, R. B., and Wang, H., "A sliding window common spatial pattern for enhancing motor imagery classification in EEG-bci", *IEEE Transactions on Instrumentation and Measurement*, 70: 1–9, (2021).
- [29] Li, Y., Wei, H. L., Billings, S. A., and Sarrigiannis, P. G., "Identification of Nonlinear Time-Varying Systems Using An Online Sliding-Window and Common Model Structure Selection (Cmss) Approach with Applications to EEG", *International Journal of Systems Science*, 47: 11, 2671–2681, (2016).
- [30] Hall, M. A., "Correlation-based feature selection for machine learning", Ph.D. dissertation, The University of Waikato, (1999).
- [31] Jin, J., Miao, Y., Daly, I., Zuo, C., Hu, D., and Cichocki, A., "Correlation-based channel selection and regularized feature optimization for MI-based BCI", *Neural Networks*, 118: 262-270, (2019).

- [32] Şen, B., Peker, M., Çavuşoğlu, A., and Çelebi, F. V., “A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms”, *Journal of Medical Systems*, 38: 1-21, (2014).
- [33] Kira, K., and Rendell, A. L., “The feature selection problem: Traditional methods and a new algorithm”, In *Proceedings of The Tenth National Conference On Artificial Intelligence*, 129-134, (1992).
- [34] Zhang, S., Cheng, D., Deng, Z., Zong, M., and Deng, X., “A Novel Knn Algorithm with Data-Driven K Parameter Computation”, *Pattern Recognition Letters*, 109: 44–54, (2018).
- [35] Edla, D. R., Mangalorekar, K., Dhavalikar, G., and Dodia, S., “Classification of EEG Data for Human Mental State Analysis Using Random Forest Classifier”, *Procedia Computer Science*, 132: 1523-1532, (2018).
- [36] Wang, S., Li, Y., Wen, P., and Zhu, G., “Analyzing EEG signals using graph entropy-based principle component analysis and J48 decision tree”, In *Proceedings of the 6th International Conference on Signal Processing Systems (ICSPPS 2014)*, University of Southern Queensland, (2014).
- [37] Du, K. L., and Swamy, M., “Support vector machines”, in *Neural Networks and Statistical Learning*, Springer, 469–524, (2014).
- [38] Sidaoui, B., and Sadouni, K., “Epilepsy Seizure Prediction from EEG Signal Using Machine Learning Techniques”, *Advances in Electrical & Computer Engineering*, 23: 2, (2023).
- [39] Stancic, I., Veic, L., Music, J., and Grujic, T., “Classification of Low-Resolution Flying Objects in Videos Using the Machine Learning Approach”, *Advances in Electrical & Computer Engineering*, 22: 2, (2022).
- [40] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., Louppe, G., “Scikit-learn: Machine Learning in Python”, *JMLR* 12: 2825-2830, (2011).
- [41] Chicco, D., and Jurman, G., “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation”, *BMC Genomics*, 21: 1, 1-13, (2020).
- [42] Berrar, D., “Cross-Validation”, 542-545, (2019).