ITM Web of Conferences **67**, 01007 (2024) https://doi.org/10.1051/itmconf/20246701007
ICMSA 2024

Extraversion prediction from EEG coherence during a face-to-face interaction task using machine learning techniques

Nur Syahirah Roslan[1*], *Nur Lyana Shahfiqa* Albashah[1], and *Ibrahima* Faye[2,3]

[1]Department of Information Systems, Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, 31900 Kampar, Perak, Malaysia.
[2]Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak, Malaysia
[3]Centre for Intelligent Signal & Imaging Research (CISIR), Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak, Malaysia

Abstract. Researchers have begun investigating personality assessments using brain-imaging techniques, such as electroencephalography (EEG). However, previous studies usually utilised EEG power, resting state, and video stimulus in the extraversion classification study, which could be the factors contributing to insufficient accuracy. Thus, this study proposes to classify extraversion using EEG coherence during a face-to-face interaction task. A total of 32 healthy male individuals were selected for this study based on their scores on the Big Five Inventory (BFI) and the Eysenck Personality Inventory (EPI). Sixteen of the individuals were identified as extraverts, whereas the remaining sixteen were identified as introverts. The study employed the Kruskal-Wallis H test to identify the high-ranking features. For the extraversion classification, optimizable KNN and SVM were utilised, along with leave-one-out cross-validation. The findings indicated that employing 1624 EEG coherence features yielded an accuracy of less than 80%. However, when applying feature selection, the accuracy increased up to 84.4%. Hence, we believe the study offers valuable insights for extraversion classification.

1 Introduction

The extraversion personality trait is considered one of the Big Five personality theories, which represents individuals' diverse inclinations for showing spontaneity and sociability, particularly in unfamiliar social situations [1]. Persons with elevated levels of extraversion, commonly called extraverts (or occasionally extroverts), demonstrate chatty and highly gregarious behaviours. Conversely, those who have lower levels of extraversion, commonly called introverts, show less sociability and favour solitude in most situations. Due to their contrasting social preferences, extraverts and introverts have divergent impacts on several elements of life, including happiness or subjective well-being, mental health, and

*Corresponding author: syahirahr@utar.edu.my

occupation. In terms of happiness, extraverts were discovered to be happier than introverts, whether they live alone or live together with others in social or non-social environments [2, 3]. In terms of mental health, extraversion has been reported to have an inverse association with several mental disorders, such as social anxiety, personality disorder, depression, and schizophrenia [4]. One possible explanation for this deduction is that introverts may have a tendency to overthink things [5]. Overthinking is widely recognised as a contributing factor to mental anguish since it can result in feelings of worthlessness, hopelessness, and low self-esteem. Regarding occupation, extraverts may find occupations with positions such as presenter, guide, director, or inspirer (e.g., sales managers, event planners, or attorneys) more fitting [6]. This is likely due to extraverts' inclination towards social surroundings, their preference for interpersonal communication and influence, as well as their desire to be acknowledged and rewarded. Careers that include specialist positions, such as researchers or writers, may be well-suited for introverts [6]. It is likely that these professions entail minimal social engagement and a significant amount of solitary work. In light of the various benefits linked to comprehending an individual's personality traits, it is believed to be beneficial for developing highly accurate personality assessments that can subsequently be utilised by many parties.

In the era of modern technology, researchers from various domains like psychology, neuroscience, engineering, and computer science have started identifying human personality traits based on brain imaging modalities and machine learning [7-9]. The complexity of manipulating brain imaging modalities makes them more challenging to manipulate than personality assessment questionnaires. Furthermore, integrating machine learning might enhance the accuracy and speed of the assessment. The involvement of brain imaging modalities and machine learning has the capability to address other limitations of questionnaires, such as unanswered questions, variations in question interpretation, and survey fatigue. Given these advantages, it is deemed valuable to pursue the development of human personality recognition by brain imaging methods, as it can yield numerous advantages across various fields. One of the brain imaging modalities that have been used to identify human personality traits is electroencephalography (EEG). EEG is an electrophysiological method that involves the placement of tiny electrodes on the scalp to record the electrical activity generated by the human brain. EEG is the most cost-effective and user-friendly brain imaging modality compared to other modalities, such as Functional Magnetic Resonance Imaging (fMRI), Functional Near-infrared Spectroscopy (fNIRS), and Positron Emission Tomography (PET). The method also possesses a very high temporal resolution, enabling the tracking of brain responses in milliseconds. Due to its exceptional temporal resolution, this technique has been employed in real-time recordings for both clinical and non-clinical investigations. For example, EEG has been used in investigating patients with COVID-19 [10], stress assessment [11], and personality recognition [12].

In EEG studies, some researchers classified extraversion based on resting-state tasks, which included eyes open and closed [7, 13]. In 2015, researchers indicated that personality could not be discerned from the analysis of EEG power derived from resting states [7]. The study suggested that the utilisation of EEG power can be a contributing factor to the unfavourable results observed [7]. The study hypothesised that important information about extraversion may not be indicated by the power features. Their suggestion claimed that time-domain or other EEG features might encompass noteworthy extraversion-related information, hence enhancing classification accuracy. Another factor that may have led to insufficient classification performance is a lack of stimulation during resting states, as suggested in research conducted in 2020 [7, 13]. The findings suggested that external stimuli are necessary to elicit individual personality traits, indicating that resting states alone are insufficient for identifying personality. The suggestion to utilise external stimuli seems to be highly effective, as evidenced by extraversion classification accuracy exceeding 95% [14].

However, the study used a public speaking task that was deemed difficult and impractical to include in an EEG-based extraversion assessment, mainly due to the need for a large audience during the assessment. In addition, the study used a K-Nearest Neighbour (KNN) with a value of $k = 1$. According to some researchers, the use of $k = 1$ may lead to overfitting and has poor tolerance to noise [15, 16].

Other than public speaking tasks, emotional video stimuli have been used as external stimuli to classify extraversion [12, 17-19]. For instance, a study conducted in 2018 used emotional video clips to classify extraversion by using EEG power and Support Vector Machine (SVM) [17]. The study achieved less than 80% accuracy, which is likely because of the use of EEG features or the experimental task that inadequately reflects the essential information of extraversion. Furthermore, a study carried out in 2020 also used emotional video stimuli and SVM for classifying extraversion [12]. Specifically, the study employed short video experiments sourced from the AMIGOS dataset [20]. Moreover, the study distinguished itself by employing EEG coherence instead of EEG power for the purpose of improving extraversion classification. As a result, the study obtained an accuracy of 83.8% [12]. The study recommends that researchers should not solely rely on emotional factors but also consider behavioural and cognitive factors to enhance the accuracy of classification. This is because personality is closely related to an individual's thoughts, emotions, and behaviours.

Overall, previous EEG studies have demonstrated that EEG features and experimental tasks significantly contribute to improving the accuracy of extraversion classification [12, 14]. Some EEG features and experimental tasks may not provide sufficient strength to evoke the personality traits of an individual. Moreover, the study with the highest accuracy utilised challenging and impractical tasks for real-world assessment, which presents a challenge when a large number of individuals wish to be assessed [14].

Therefore, in the present study, we suggest a face-to-face interaction task, which is a simpler social-based experiment than the public speaking task, as a way to improve classification performance and offer a more practicable assessment. In lieu of EEG power, we classified extraversion using EEG coherence in conjunction with the optimizable KNN and SVM. The choice of KNN and SVM was influenced by EEG studies conducted in 2020, which showed remarkable accuracy rates of 96.4% [14] and 83.8% [12], respectively. These findings suggest that KNN and SVM exhibit promising potential for classifying extraversion. Moreover, it is anticipated that the incorporation of the face-to-face interaction task, EEG coherence, and optimizable machine learning into the present study could result in higher extraversion classification accuracy and provide numerous insightful observations that can be applied to future studies. This paper is divided into five parts, with the first part being the introduction. The paper provides an explanation of the materials and methods used, which could be found in Section 2 and Section 3, respectively. The findings and analysis of the extraversion classification utilising optimizable KNN and SVM, as well as EEG coherence during the face-to-face interaction task, are outlined in Section 4. Lastly, the conclusion of the study is presented in Section 5.

2 Materials

2.1 Participants

In this research, the Eysenck Personality Inventory (EPI) and the Big Five Inventory were used to evaluate the extraversion scores of the possible participants. Two personality tests were employed to rigorously screen the subjects, in which participants who obtained contradictory results were automatically excluded from the study. Participants were also disqualified if they obtained a score of exactly 50% on either of the tests. Thus, of the ninetyone male students from Universiti Teknologi PETRONAS who volunteered to participate, only 41 individuals fulfilled the necessary criteria. However, only 32 out of the initial 41 individuals came to participate in the EEG experiment, which consists of 16 extraverts and 16 introverts.

The participants in the study were aged between 18 and 23 years (mean = 19.53, standard deviation = 1.22). They were right-handed, had normal or corrected-to-normal vision, had no hearing impairment, had no personal or family history of cognitive disorders, were not taking any drugs or medication, and were not experiencing chronic mental stress or adverse psychological states. Each participant provided informed consent and received compensation for their involvement in the experiment conducted for this study.

This research was approved by the Medical Research Ethics Committee of the University of Kuala Lumpur, Royal College of Medicine, Perak, Malaysia.

2.2 Experimental task

The study employed a face-to-face interaction task involving a participant and an inquirer (see Fig. 1). Before performing the task, the participants were instructed not to move their bodies while EEG signals were being recorded. This is due to the movement sensitivity of the EEG. In order to ensure that the participants maintained minimal movement during the experiment, a practice question was provided to give them a clear understanding of what to do and what not to do.

Following the practice session, the actual face-to-face interaction task commenced, which included four distinct questions (see Fig. 2). Throughout the task, a questioner asked the participant one question at a time. The participants were mandated to spontaneously respond to each question within a time limit of two minutes. Upon reaching the two-minute mark, the researcher will promptly stop the participants from further responding to the question. At the end of the task, four different EEG datasets were obtained, representing four different questions in the face-to-face interaction task. To maintain the authenticity of the obtained EEG datasets, all experimental sessions were conducted in a partially soundproofed EEG experiment room, and the participants were prohibited from sharing any aspects of the experimental task with any party.

All questions in the face-to-face interaction task were easy to understand and not related to any academic subjects that require previous knowledge to answer. This is to encourage participants to provide genuine and unrestricted replies to the questions, as our emphasis is on the participants' personality traits rather than their answers to the questions.

Fig. 1. An illustration of the face-to-face interaction task, where a participant is seated in front of a questioner.

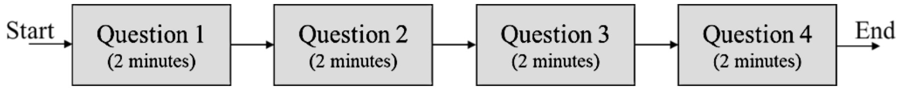

Fig. 2. The protocol for the face-to-face interaction task.

3 Methods

In the study, EEG signals were recorded at a sampling rate of 512 Hz by using an eego™ sports amplifier with a 32-channel waveguard EEG cap. During the recordings, the impedance of the electrodes was kept below 10 kΩ. After the acquisition of EEG data, the raw EEG signals underwent several procedures, such as EEG signal preprocessing, feature extraction (EEG coherence), feature selection, and extraversion classification using the whole EEG features and the high-ranked EEG features, as illustrated in Fig. 3.

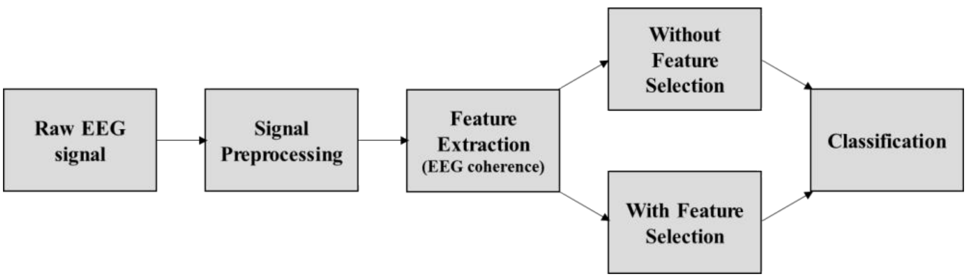

Fig. 3. The flowchart of extraversion classification from EEG data.

3.1 Signal preprocessing

The raw EEG signals obtained from the experiments were pre-processed using BESA Research 6.0. At first, the EEG signals underwent bandpass filtering with a cut-off frequency of 0.53–48 Hz to remove low-frequency noises like sweat artefacts (0.25–0.50 Hz) and high-frequency noises like power line noises (e.g., 50 Hz in Malaysia). Subsequently, other artefacts, including eye blinks, horizontal eye movements, and muscle activity, were visually examined and automatically corrected using Berg's technique [21], as implemented in the BESA software. The cleaned EEG data was then used for the extraction of EEG coherence features.

3.2 Feature extraction: EEG coherence

EEG coherence is one of the methods for measuring functional connectivity between two brain regions. It can be defined mathematically as the ratio of the normalised cross-power spectrum to the auto-power spectrum. The Fast Fourier Transform (FFT) algorithm was applied to 512 samples with 50% overlap between successive 2-second segments (1024 points) of the EEG signals to compute the power spectrum. Coherence was then calculated as [22]:

$$C_{ab}^2(f) = \frac{|P_{ab}(f)|^2}{(P_{aa}(f)P_{bb}(f))} \tag{1}$$

where f represents the frequency, P_{ab} represents the cross-power spectrum between the signals of two brain regions, a and b, P_{aa} represents the auto-power spectrum for the signals of brain region a, and P_{bb} is the auto-power spectrum for the signals of brain region b. The coherence values of 0 suggesting a lack of functional connectivity between the two brain signals, while values close to 1 indicating a high level of connectivity.

For the feature extraction, only the first 15 seconds of every EEG dataset obtained during the face-to-face interaction task were used. We assume that the initial 15 seconds are enough to classify extraversion since introverts struggle to start conversations, particularly with unfamiliar individuals [23, 24]. In addition, only 29 out of the 32 electrodes were utilised for feature extraction. The electrodes, often referred to as brain regions, included Fp1, Fpz, Fp2, F3, F4, F7, F8, Fz, FC5, FC1, FC2, FC6, Cz, T7, T8, P7, P8, C3, C4, CP5, CP1, CP2, CP6, P3, Pz, P4, Poz, O1, and O2. Since EEG coherence quantifies the connectivity between two specific brain regions, a total of 406 electrode pairs were identified from the initial set of 29 electrodes. The number of these pairing combinations, denoted as $\mathbb{N}$, was computed using the formula $\mathbb{N} = (\mathrm{n}(\mathrm{n}-1))/2$, where n represents the number of electrodes. The EEG coherence was determined for the delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta bands (13–30 Hz) of each participant. Each EEG dataset involves a total of 406 pairings. The study opted not to include the gamma band (30–48 Hz) due to the possibility that it represented muscular activity in the neck and face during the face-to-face interaction task [25]. At the end, the average coherence of each dataset was computed for each band to represent the EEG coherence during the face-to-face interaction task.

3.3 Feature selection: Kruskal-Wallis H test

In the study, extraversion classification was carried out using all EEG coherence features (1624 features), including 406 delta coherence features, 406 theta coherence features, 406 alpha coherence features, and 406 beta coherence features. To enhance the accuracy of extraversion classification, high-ranked EEG features were determined to eliminate unnecessary and redundant features. To select the high-ranked features, the Kruskal-Wallis H test was used in the study. It is basically a rank-based, nonparametric test for two or more groups of independent data. It can be used as an alternative to the one-way ANOVA that could be used to rank all the involved features.

The test statistic for the Kruskal-Wallis H test can be calculated using

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{m} \frac{R_i^2}{n_i} - 3(N+1) \tag{2}$$

where N represents the total number of observations in all samples combined, m represents the number of observations in one group, R_i represents the sum of rank in the group i, and n_i represents the number of observations in group i.

3.4 Classification

To classify extraversion, optimizable KNN and SVM implemented in the Classification Learner App included in MATLAB were used. By using Classification Learner, Bayesian Optimisation was also applied to find the optimal hyperparameter for KNN and SVM.

3.4.1 K-Nearest Neighbours (KNN)

Within the realm of machine learning, KNN is one of the most straightforward and efficient classification algorithms. The core principle of KNN is to perform the classification through a majority vote among its neighbouring data points. When confronted with new data requiring prediction, the k closest neighbours to that data point must be identified to facilitate prediction through reference to the majority vote among these neighbours. Normally, selecting $k = 1$ represents the simplest scenario, while for binary classification, it is advisable to opt for an odd value of k to avoid tie situations.

The KNN algorithm can generally be delineated into several steps:

1) Compute the distances between the testing data and all the training data.
2) Arrange the distances in ascending order.
3) Select k points with the smallest distances.
4) Determine the frequency of occurrence for each class among the selected k points.
5) Assign the class with the highest frequency among the selected k points as the predicted classification for the test data.

The KNN algorithm comprises two crucial phases: first, determining k close neighbours, and second, defining class types using these close neighbours. The training data space S is represented by the equations:

$$S = \{X_1, X_2, \dots, X_n\} \tag{3}$$

which has n samples, and each sample X_i is defined by f features as

$$X_i = \left(x_{i1}, x_{i2}, \dots, x_{if}\right) \tag{4}$$

The dataset encompasses α different classes. To identify the data class of X', the initial step involves calculating its distances from all data points in space S. Subsequently, K data points are selected from space S based on their proximity to X', forming its nearest neighborhood. To identify the data class of X', the initial step involves measuring its distance from all the data points in space S. Subsequently, K data points are selected from space S based on their proximity to X', forming its nearest neighborhood. The class type of all K data is identified, and thus, X' is assigned to the class that has undergone the highest number of iterations among all K data. There are several criteria for assessing the distance between X' and X_i, with the Euclidean distance criterion being the most used. The Euclidean distance between X' and X_i, using the feature dimension f, is determined by the following equation:

$$d = \sqrt{(x_{i1} - x_1')^2 + \cdots + \left(x_{if} - x_f'\right)^2} \tag{5}$$

In this study, we employed City block, Chebyshev, Correlation, Cosine, Hamming, Jaccard, Mahalanobis, Minkowski, and Spearman distance within the Bayesian Optimisation method, in conjunction with additional parameters (see Table 1), to identify the optimal hyperparameter for the KNN classifier.

3.4.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) stands as a renowned supervised machine learning technique widely applied across various analyses, including neuroimaging [26]. Its conceptual simplicity and adaptability have led to its extensive utilization in numerous classification studies. SVM exhibits the capacity to offer balanced predictive performance, even when confronted with studies characterized by small sample sizes. In essence, SVM operates by delineating classes through a hyperplane with the maximum margin. This margin is gauged as the perpendicular distance from the hyperplane to the nearest data points, commonly referred to as support vectors. Optimal margin attainment occurs when this distance reaches its maximum extent, thereby delineating the optimal hyperplane.

In cases where the input space, such as EEG signals, demonstrates non-linearity, SVM still proves viable through the application of kernels to enable linear modeling. Supposed $\psi: \chi \rightarrow Z$ represents a non-linear transformation from the input, space χ, to the feature space $\mathbb{Z}$, where the problem can be linearly separable. Let $\mathbb{Z} = \psi(x)$ and $z^T z' = S(x, x')$, then the $x_n^T x_m$ term in the dual form can be written as $S(x, x')$. Now, by solving the previous dual form using the kernel function, we can find the optimal a_n^* and thereafter, calculate the optimal weight vector $w^* = \sum_{z_n} a_n^* y_n z_n$,where z_n is the support vector and $\mathrm{y_n}$ is the correct output of the SVM for the nth training. Therefore, the decision function becomes as follows, where $b = y_m - \sum_{a_n>0} a_n y_n S(x_n, x_m)$ is the bias term.

$$f(x) = sign\left(\sum_{a_n>0} a_n y_n S(x_n, x) + b\right) \tag{6}$$

In this study, we employed Linear, Quadratic, Cubic, and Gaussian kernels within the Bayesian Optimisation method, in conjunction with additional parameters (see Table 1), to identify the optimal hyperparameter for the SVM classifier.

3.4.3 Bayesian optimisation

In the study, Bayesian Optimisation was employed to identify the optimal set of hyperparameters capable of improving classification performance. Renowned for its efficacy in the realm of machine learning and optimisation, Bayesian Optimisation stands as a potent technique for navigating the hyperparameter space of a model. The decision to employ Bayesian Optimisation stemmed from its systematic and efficient methodology, especially in cases where evaluating the objective function proves costly or when dealing with intricate, high-dimensional search spaces. Through a process of iterative exploration and exploitation of pertinent information within the objective function, Bayesian Optimisation can swiftly converge towards nearly optimal solutions.

The hyperparameter search range involved in the study is listed in Table 1.

Table 1. Hyperparameter search range for KNN and SVM.

KNN	**SVM**
Number of neighbours: 1 – 16	Box constraint level: 0.001 – 1000
Distance metric: City block, Chebyshev, Correlation, Cosine, Euclidean, Hamming, Jaccard, Mahalanobis, Minkowski (Cubic), Spearman	Kernel function: Linear, Quadratic, Cubic, Gaussian.
Distance weight: Equal, Inverse, Squared inverse	Kernel scale: 0.001 – 1000
Standardize data: true, false	Standardize data: true, false

3.4.4 Leave-one-out cross-validation (LOOCV)

Since a limited sample size was involved in the study, leave-one-out cross-validation (LOOCV) was employed for classifying extraversion based on optimizable KNN and SVM. The use of LOOCV is to prevent overfitting issues from occurring. LOOCV can be described as a specific instance of k-fold cross-validation, in which k equals n, the total size of data involved in the study. Since $k = n$, each data point in the entire dataset will be the test data (see Fig. 4).

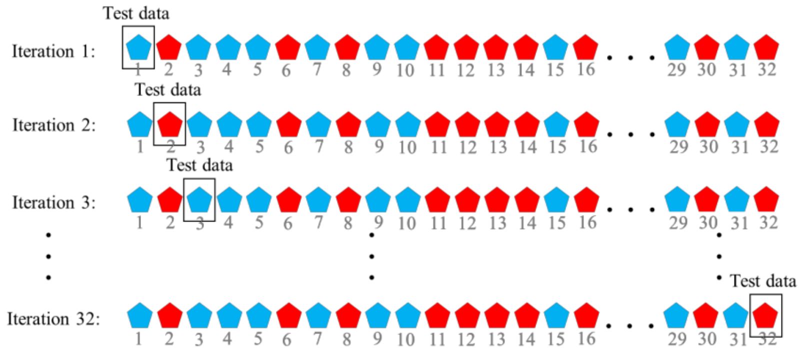

Fig. 4. The illustration of leave-one-out cross validation (blue pentagon represents introverts and red represents extraverts).

3.5 Evaluation

In the study, a confusion matrix that involves true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) was used to evaluate the classification performance of a machine learning algorithm (see Table 2). The core idea of the confusion matrix is to summarise correct and incorrect predictions with count values, which are broken down by class. Then, the values can be used in evaluating the performance of the classifiers as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

Table 2. Confusion matrix for evaluating classifiers' performances.

		Actual values	
		Positive (Class 1)	Negative (Class 2)
Predicted values	Positive (Class 1)	TP	FP
	Negative (Class 2)	FN	TN

4 Results and discussion

As stated in Section 3, the study classified extraversion by utilising all features (1624 features) and the top-ranked features of EEG coherence during the face-to-face interaction task. The top-ranked features were selected using the Kruskal-Wallis H test. In addition, a thorough investigation of extraversion classification was performed by utilising the entire features of each band, including the bands' highest-ranked features. A thorough investigation is expected to offer valuable insights into the comparative accuracy of different bands, shedding light on their correlation with extraversion. The extraversion classification in the study was conducted using optimizable KNN and SVM algorithms, including leave-one-out cross-validation.

As shown in Table 3, when utilising the whole features of EEG coherence (1624 features) of the face-to-face interaction task, the accuracies achieved were 75% and 71.9% using optimised KNN and SVM, respectively. However, by exclusively utilising EEG coherence from specified frequency bands, the accuracy was enhanced to 78.1% (see Table 3). This was accomplished by employing all theta coherence along with optimised KNN. The use of optimised KNN achieved an accuracy of 78.1% by utilising the optimal hyperparameters, which consisted of Spearman as the distance metric, the number of neighbours equal to 16, and equal distance weight (see Fig. 5). In addition, the optimised SVM achieved the lowest accuracy of 50%, obtained using 406 features of EEG delta coherence. The model with the lowest accuracy (50%) was a linear SVM (see Fig. 6). Overall, based on Table 2, it can be inferred that the inadequate accuracy of the extraversion categorization can be attributed to the extensive number of features, some of which may be irrelevant and redundant. Consequently, feature selection is employed to determine the essential features while eliminating irrelevant and redundant ones.

Table 3. Classification accuracy based on the whole EEG coherence features of the face-to-face interaction task using optimized KNN and SVM.

	Number of features	Accuracy (%)	
		KNN	SVM
All features	1624	75.0	71.9
All delta coherence	406	56.2	50.0
All theta coherence	406	78.1	71.9
All alpha coherence	406	62.5	68.8
All beta coherence	406	75.0	71.9

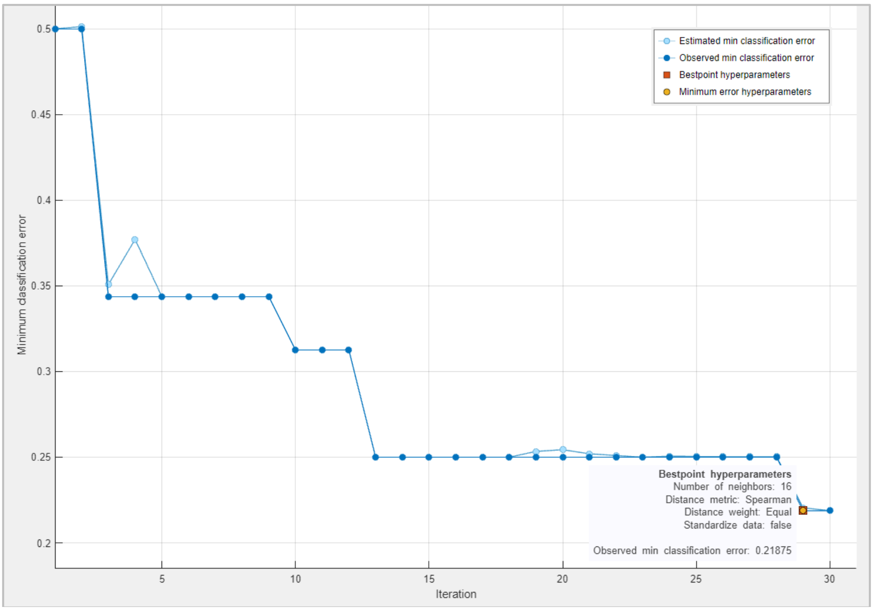

Fig. 5. The minimum error classification plot for the model with the highest accuracy (KNN) obtained using all features of EEG theta coherence (406 features).

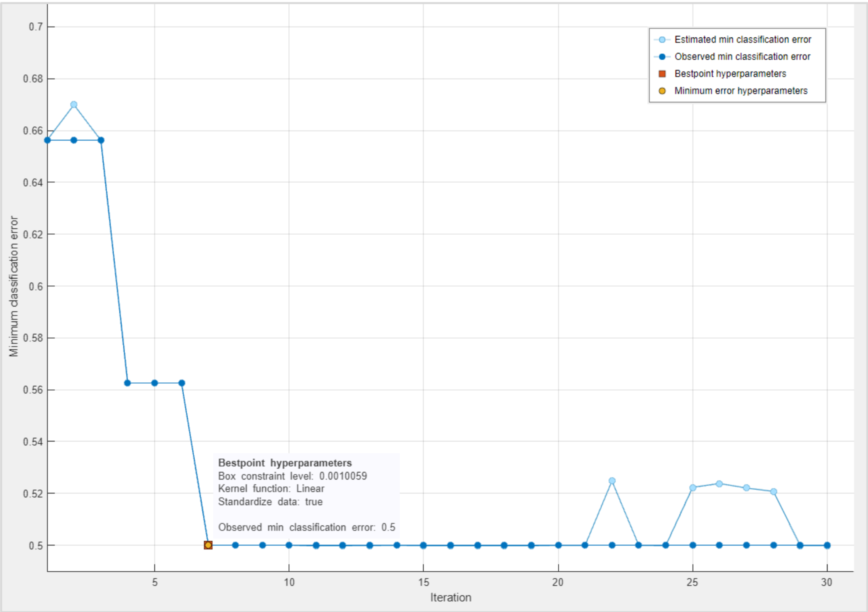

Fig. 6. The minimum error classification plot for the model with the lowest accuracy (SVM) obtained using all features of EEG delta coherence (406 features).

Table 4 demonstrates the accuracy of extraversion classification utilizing high-ranked EEG coherence features of the face-to-face interaction task, based on optimized KNN and SVM. By utilising a subset of the top 10 features from a pool of 1624 features, the optimised KNN and SVM algorithms attained accuracies of 75.0% and 68.8%, respectively, as seen in Table 3. The results indicated that there was a minimal difference when comparing it to classification using 1624 features, likely due to the existence of redundant or irrelevant features within the 10 selected features. Both optimised KNN and SVM increased the classification accuracy to 84.4% by removing the top 9 features and choosing only the single best feature out of a total of 1624 (see Table 4). The optimised KNN achieved an accuracy of 84.4% by utilising Chebyshev as the distance metric, three neighbours, and equal distance weight (see Fig. 7). While the optimised SVM achieved 84.4% using quadratic as the kernel function (see Fig. 8). In addition, a bar graph illustrating the accuracy of extraversion classification was generated to visually illustrate the difference between classification with and without feature selection (see Fig. 9). The increase is observed in all features of EEG coherence and EEG theta coherence. For instance, extraversion classification accuracy is below 80% when all features of EEG coherence are employed, but it exceeds 80% when only the highest feature out of 1624 is employed. This result suggests that feature selection should be utilised as it has the potential to enhance classification accuracy. The next part will present a more comprehensive extraversion classification using the top 10 features of EEG coherence for various frequency bands, ultimately narrowing it down to the top 1.

Table 4. Extraversion classification accuracy based on high-ranked EEG coherence features of the face-to-face interaction task using optimised KNN and SVM.

No. of features	All features (1624 features)		Delta Coh. (406 features)		Theta Coh. (406 features)		Alpha Coh. (406 features)		Beta Coh. (406 features)	
	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM
10	75.0	68.8	59.4	75.0	78.1	75.0	75.0	71.9	68.8	53.1
9	75.0	78.1	59.4	78.1	78.1	75.0	75.0	71.9	65.6	56.2
8	75.0	75.0	62.5	68.8	78.1	71.9	75.0	71.9	65.6	56.2
7	68.8	75.0	62.5	68.8	75.0	75.0	71.9	71.9	62.5	56.2
6	71.9	68.8	59.4	68.8	71.9	71.9	71.9	68.8	65.6	56.2
5	68.8	65.6	68.8	75.0	68.8	75.0	68.8	68.8	68.8	50.0
4	68.8	62.5	62.5	65.6	71.9	75.0	78.1	75.0	62.5	50.0
3	65.6	75.0	71.9	62.5	78.1	78.1	71.9	71.9	65.6	37.5
2	71.9	84.4	59.4	53.1	75.0	78.1	62.5	62.5	56.2	37.5
1	84.4	84.4	53.1	53.1	84.4	84.4	62.5	53.1	50.0	40.6
Average	72.5	73.8	61.9	66.9	75.9	75.9	71.3	68.8	63.1	49.4
SD	5.3	7.4	5.3	8.6	4.4	3.6	5.3	6.4	5.9	7.9

Note: Coh. = Coherence, SD = Standard deviation.

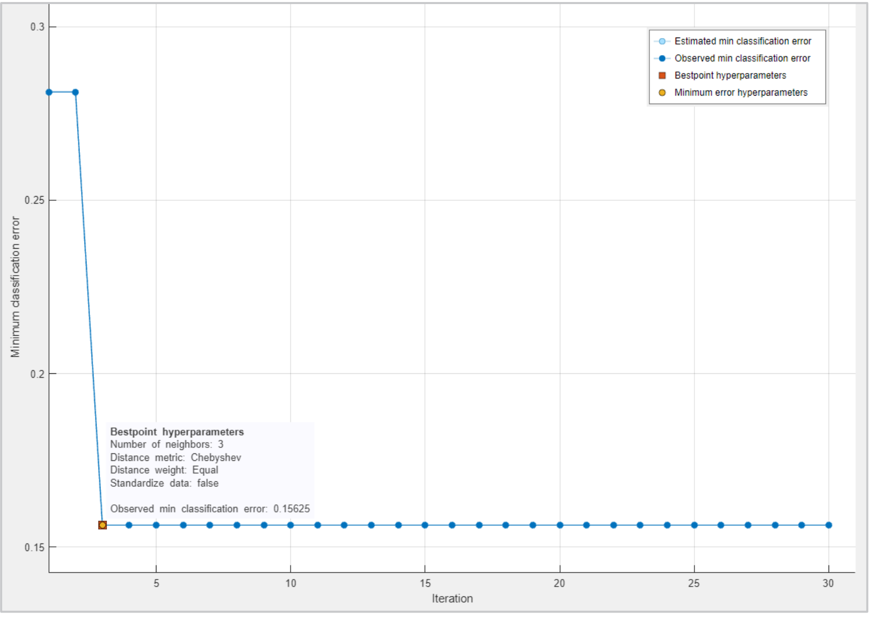

Fig. 7. The minimum error classification plot for the model with the highest accuracy (KNN) obtained using top 1 feature among 1624 features.

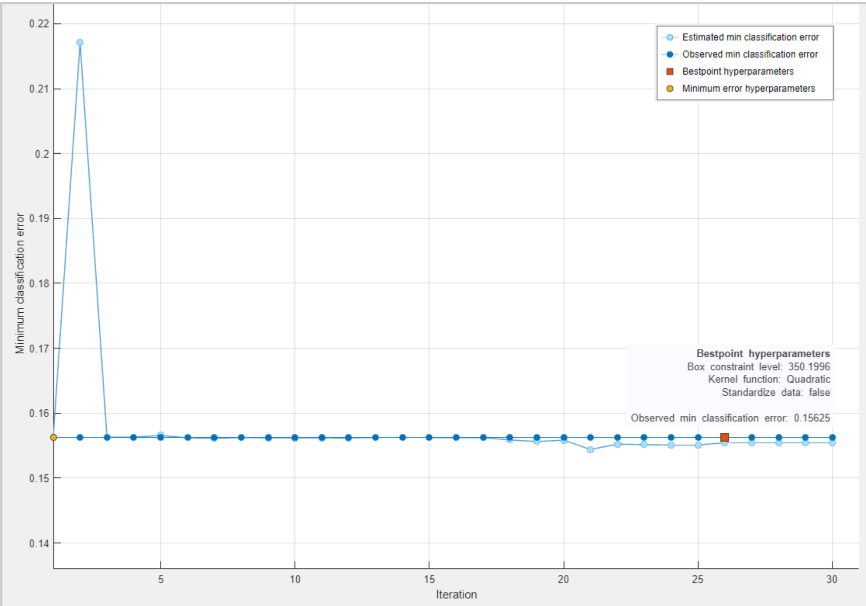

Fig. 8. The minimum error classification plot for the model with the highest accuracy (SVM) obtained using top 1 feature among 1624 features.

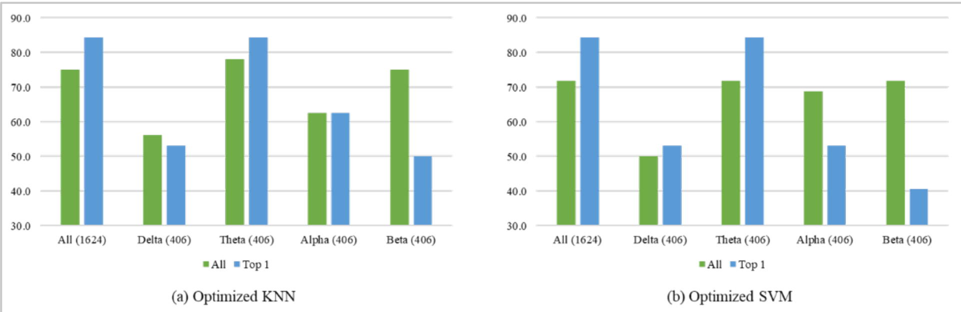

Fig. 9. Comparison of extraversion classification accuracy based on all features and top 1 feature.

When it comes to the different frequency bands, using the top 10 EEG coherence features of each band gave the highest accuracy of 78.1% with optimised KNN and the lowest accuracy of 53.1% with optimised SVM for classifying extraversion (see Fig. 10). The top 10 theta coherences yielded the highest accuracy, whereas the lowest accuracy was obtained using the top 10 beta coherences. Furthermore, by exclusively utilising the highest-ranked feature, theta coherence yielded the highest accuracy of 84.4% for both optimised classifiers. The optimised KNN achieved an accuracy of 84.4% by utilising the City Block distance metric, 16 neighbours, and equal distance weight (see Fig. 11). While the optimised SVM achieved 84.4% employing a linear kernel function (see Fig. 12). Compared to other frequency bands, theta coherence during the face-to-face interaction task produced the highest average accuracy when extraversion was classified using the highest-ranked features. It was also observed that beta coherence during yielded the lowest accuracy compared to other frequency bands, with an average accuracy of 49.4% for optimised SVM (see Fig. 10). Based on the results obtained, it can be inferred that theta coherence of the face-to-face interaction potentially offers additional insights into extraversion, which could account for its accuracy exceeding 80%.

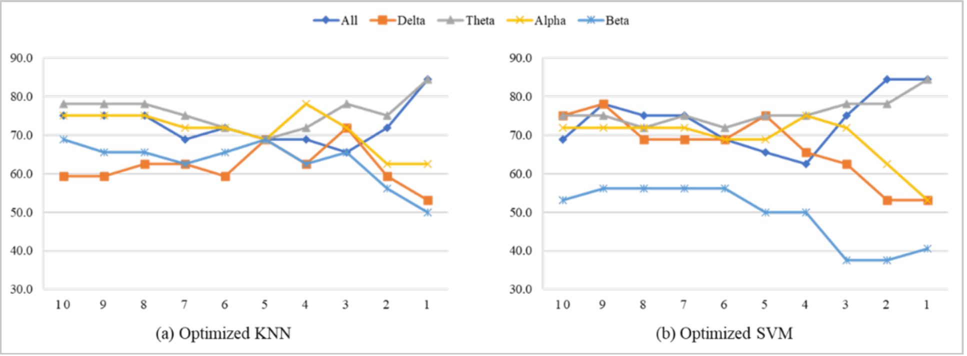

Fig. 10. The accuracy of extraversion classification using optimised KNN and SVM with respect to high-ranked EEG coherence features (top 10 to top 1 features).

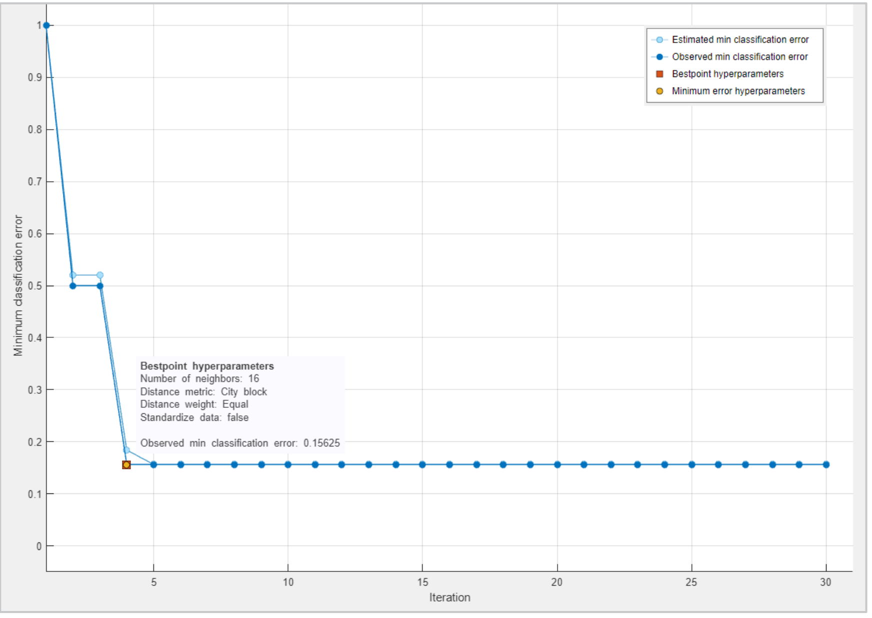

Fig. 11. The minimum error classification plot for the model with the highest accuracy (KNN) obtained using top 1 feature of theta coherence.

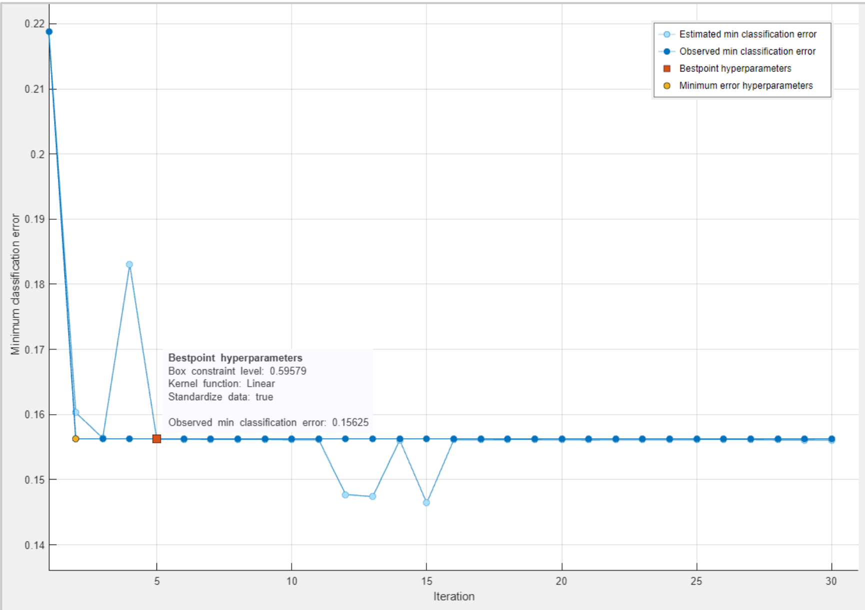

Fig. 12. The minimum error classification plot for the model with the highest accuracy (SVM) obtained using top 1 feature of theta coherence.

In all, the study achieved 84.4% accuracy for classifying extraversion, surpassing the results of several recent published studies [18, 19]. However, inherent limitations in the study necessitate further exploration. Firstly, the study exclusively focuses on EEG coherence during face-to-face interaction tasks. Researchers might find it useful to include more types of EEG features in the future to get a more complete picture of the best features for classifying extraversion, which would indirectly lead to new insights into the neural correlates of extraversion. Enhanced comprehension of the underlying brain mechanisms associated with personality traits could deepen our understanding of human behaviour and interpersonal dynamics, thus enriching theoretical frameworks in the field. Secondly, the study solely employed KNN and SVM for classification purposes, thereby neglecting the potential of alternative machine learning algorithms such as decision trees, random forests, and neural networks. Subsequent investigations that consider and compare these alternative approaches could augment the methodology section and offer supplementary insights. Lastly, the study's confinement to the binary extraversion classification represents an additional limitation. Future research endeavours could broaden the scope to encompass multiple categories of extraversion, enabling a more nuanced examination of extraversion traits. Despite these limitations, our findings primarily pave the way for the development of EEG-based biomarkers for personality traits, facilitating clinicians' comprehension of individual differences and the customization of treatment plans. These EEG-based biomarkers could also prove advantageous in tailoring interventions and educational programs to optimize learning outcomes and job performance based on individuals' personality traits. Hence, our study holds significant implications across various domains, including psychology, neuroscience, engineering, and computer science.

5 Conclusion

The main goal of the present study is to develop an optimised model for extraversion classification based on EEG coherence during face-to-face interaction tasks to improve extraversion classification accuracy. A total of 32 participants were employed in the study, consisting of 16 introverts and 16 extraverts. The type of EEG feature used in the study was EEG coherence, which measures the connectivity between two brain regions. For classifying extraversion, all features of the EEG coherence were used, as well as the top-ranked features that were selected using the Kruskal-Wallis H test. Two optimizable machine learning algorithms used were KNN and SVM. Bayesian optimisation was employed to find the best hyperparameters for KNN and SVM. To avoid overfitting, leave-one-out cross-validation was used.

This study showed that the accuracy of identifying extraversion using all features of EEG coherence is lower than the accuracy obtained using the single best characteristic out of 1624. This suggests that by eliminating redundant and insignificant features, accuracy could be improved while the quantity of EEG electrodes required could be reduced. For future research, we recommend exploring alternative EEG feature types and employing advanced machine learning algorithms to further improve the existing accuracy. We also recommend utilising fusion EEG features for the purpose of improving extraversion classification. Overall, we believe that the current study has the capability to offer several potential investigations in the development of EEG-based extraversion assessment.

The Ministry of Education of Malaysia, through the Higher Institution Centre of Excellence (HICoE) grant to the Centre for Intelligent Signal & Imaging Research (CISIR), provided financial support for this study. The decision to submit this article for publication and all aspects of study design, data acquisition, analysis and interpretation, and writing were all outside the purview of the funders.

References

1. L. Ellis, D.P. Farrington, A.W. Hoskin, in *Handbook of Crime Correlates (Second Edition),* (Academic Press, San Diego, 2019), 205 − 257 (2019)
2. E. Diener, E. Sandvik, W. Pavot, F. Fujita, J. Res. Pers., **26**, 205 − 215 (1992)
3. W. Pavot, E. Diener, F. Fujita, Pers. Individ. Differ., **11**(12), 1299 −1306 (1990)
4. D. Watson, K. Stanton, S. Khoo, S. Ellickson-Larew, S. M. Stasik-O'Brien, J. Res. Pers., **81**, 1-10 (2019)
5. M. Humpal, Ray Stanford Strong, West Coast Landscape Artist, vol. 28 (University of Oklahoma Press, 2017)
6. N. de Jong, B. Wisse, J. A. M. Heesink, K. I. van der Zee, Front. Psychol., **10**, 461663 (2019)
7. K. Korjus, A. Uusberg, H. Uusberg, N. Kuldkepp, K. Kreegipuu, J. Allik, R. Vicente, J. Aru, Front. Hum. Neurosci., **9**, (2015)
8. R. Subramanian, J. Wache, M.K. Abadi, R.L. Vieriu, S. Winkler, N. Sebe, IEEE Trans. Affect. Comput., **9**(2), 147 − 160 (2018)
9. H. Bhardwaj, P. Tomar, A. Sakalle, A. Bhardwaj, *Classification of extraversion and introversion personality trait using electroencephalogram signals*, in Artificial Intelligence and Sustainable Computing for Smart City: First International Conference, AIS2C2 2021, Greater Noida, India, March 22–23, 2021, Revised Selected Papers 1. (Springer, 2021), 31 − 39 (2021)
10. A.R. Antony, Z. Haneef, Seizure **83**, 234 − 241 (2020)
11. R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, H. Al-Nashash, Sensors **21**(15), 5043 (2021)

12. M.A. Klados, P. Konstantinidi, R. Dacosta-Aguayo, V.D. Kostaridou, A. Vinciarelli, M. Zervakis, Brain Sci., **10**(5), 278 (2020)
13. H. Baumgartl, S. Bayerlein, R. Buettner, In NeuroIS Retreat. (Springer, 2020), 259 − 265 (2020)
14. A.R. Butt, A. Arsalan, M. Majid, IEEE Sens. J., **20**(12), 6532 − 6541 (2020)
15. J. Derrac, S. García, F. Herrera, Inf. Sci., **260**, 98 − 119 (2014)
16. S. Taneja, C. Gupta, S. Aggarwal, V. Jindal, *MFZ-KNN - A modified fuzzy based K nearest neighbor algorithm*, in 2015 International Conference on Cognitive Computing and Information Processing(CCIP), 1 − 5, (2015)
17. G. Zhao, Y. Ge, B. Shen, X. Wei, H. Wang, IEEE Trans. Affect. Comput., **9**(3), 362 − 371 (2018)
18. F.Q. Annisa, E. Supriyanto, S. Taheri, *Personality dimensions classification with EEG analysis using support vector machine*, in 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 79 − 82 (2020)
19. R. Arya, J. Singh, *EEG Signals based Personality and Mood Recognition using Neural Network Approach*, in 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM). (IEEE), 217 − 220 (2022)
20. J.A.M. Correa, M.K. Abadi, N. Sebe, I. Patras, IEEE Trans. Affect. Comput., **12**(2), 479 − 493 (2018)
21. P. Berg, M. Scherg, Electroencephalogr. Clin. Neurophysiol., **90**(3), 229 − 241 (1994)
22. R.W. Thatcher, D. North, C. Biver, Clin. Neurophysiol. **116**(9), 2129 − 2141 (2005)
23. R. Mihaila, J. Inf. Syst. Oper. Manage., 460 (2016)
24. N. Silaban, T. Giawa, I. Hutabarat, Jurnal Littera: Fakultas Sastra Darma Agung **1**(1), 10 − 17 (2021)
25. J.F. Nottage, J. Horder, Neuropsychobiology **72**(3 − 4), 219 − 228 (2016)
26. D.A. Pisner, D.M. Schnyer, Support vector machine, in Machine Learning 101 − 121 (Academic press, 2020)