

Brain-Computer Interface for Emotion Recognition Based on Electroencephalography

Ziyan Fan *

College of Chemistry, University of California, Berkeley, USA

* Corresponding Author Email: zfan0114@berkeley.edu

Abstract. Human emotions are erratic, and so are their brain signals throughout the day. Brain-computer interface devices must handle high signal throughputs in the form of electroencephalography from different areas of the human brain for us to learn more about these emotions. Researchers can recognize emotions from these signals after data acquisition, preprocessing, feature extraction, and feature selection to the classifier. Efficient machine learning algorithms for the process are imperative to quickly provide emotional feedback to a device. Models of such devices are already applicable in medical fields, gaming, and more. In the meantime, ethical considerations arise. Most people are concerned about their privacy infringements. This paper aims to assist researchers in swiftly comprehending the basic theory of emotion recognition while also offering insights into the future development of this highly interdisciplinary technology. Furthermore, it underscores the need for a delicate equilibrium between technological progress and human values.

Keywords: Emotion recognition; EEG; BCI.

1. Introduction

Our daily existence depends on emotion. It is pertinent to cognitive processes like decision-making, language comprehension, and selective attention [1]. Researchers put out numerous approaches to identify emotions in people. Speech, gestures, and facial expressions are used to identify emotions. Non-physiological signals are simple to use and do not require sophisticated signal reception equipment. However, people can disguise their genuine emotions and skew the results of an experiment. Another class of methods could be more explicit. It involves receiving physiological signals such as brain waves. Brain waves convey in different forms. Specific characteristics of brain waves like electroencephalography (EEG) can represent emotions. An EEG device covers a patient's scalp with electrodes to detect the electrical activities of the cortical pyramidal neurons in the cerebral cortex. The detective measurement derives from the membrane potential changes induced by neurotransmitters binding to receptors on the postsynaptic membrane. EEG generally consists of 5 bands. The Delta (δ) waves are the slowest waves (0.5–4 Hz). Often, they are observed during deep sleep. Theta (θ) waves (4–8 Hz) are present during sleep induction, drowsiness, and daydreaming. They are associated with memory and emotional processing in the temporal lobe. Alpha (α) waves (8 - 13 Hz) appear during a relaxed state from the temporal, parietal, and occipital lobes. The most frequent times to detect beta waves (13–30 Hz) are when a person is awake and actively thinking. They are predominant in the frontal lobe and are associated with mental concentration and focus.

Additionally, some are seen in the parietal region during sensorimotor tasks or in the occipital lobe during visual tasks and eye movements. Gamma (γ) waves (30 - 100 Hz and beyond) are the fastest frequencies. They are linked to the frontal lobe's perception, learning, memory, and information processing during cognitive tasks.

This paper summarizes the paradigmatic procedure of emotion recognition in Brain-Computer Interface (BCI). Using brain waves, a BCI device establishes a direct communication channel between the human brain and external devices. The signal acquisition step obtains and records signals in EEG form. Preprocessing filters out unnecessary edge signals. The rest of the steps are to find signal patterns and classify and recognize emotions from them. BCIs have potential for use in various fields, including improved human-computer interaction, neuroscientific research, and assistive technology for persons with disabilities.

2. Emotion Extraction in Brain Signals

2.1. Signal Acquisition

There are three categories of BCI devices: intrusive, semi-intrusive, and non-intrusive BCI [2]. Intrusive BCIs involve directly implanting sensors, electrodes, or other devices into the brain tissue or neural pathways. In this case, the neural signal acquired presents high quality, but his method's invasiveness raises safety and ethical concerns that make their application exorbitant. Semi-intrusive BCIs, as their name suggests, balance invasiveness and non-invasiveness. They involve placing sensors or electrodes on or within the body without directly penetrating the brain tissue. Lastly, non-intrusive BCIs often involve external sensors placed on the scalp. Though their collection of brain signals could be better, non-intrusive BCI's simplicity and affordability prevail in the market. Thus, this discussion uses non-invasive wearables such as EEG helmets and headsets for signal reception. Their electrodes are usually made of metal. Nonpolarized electrodes such as Ag/AgCl are preferable due to their compatibility with neurophysiological applications. These electrodes maintain a stable interface with the skin, leading to accurate signal acquisition [3]. The signal median is the voltage-induced brain waves.

2.2. Emotion Recognition Model

Emotion recognition includes categorical and dimensional models [2]. The former classifies emotions into discrete categories. Each category corresponds to a specific emotion with unique neural and physiological signatures, such as happiness, sadness, anger, fear, disgust, and surprise. This system utilizes machine learning algorithms to train labeled datasets that associate specific patterns of brain activity with particular emotions. In a multidimensional space, the rear model can express complicated emotions. These dimensions can involve valence (positive to negative emotions), arousal (calm to excited), and possibly others like dominance or potency. This model captures the variation of emotions along these continuous dimensions. The valence-arousal plane in *Figure 1* depicts various primary emotional states. These axes are normalized, so their ranges are between -1 and 1.

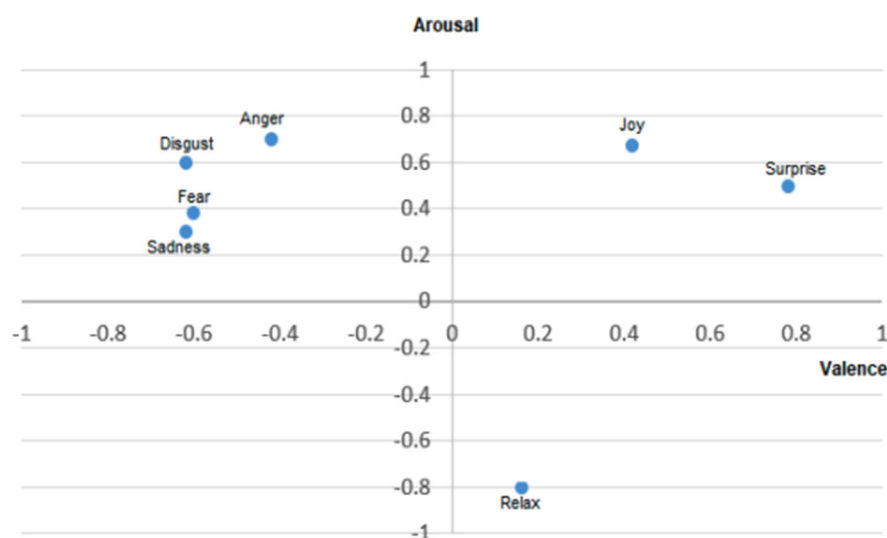


Figure 1. Emotion States in a Two-Dimensional Plane [4]

Ten healthy students from the South China University of Technology served as the control group in a BCI study of patients with disorder of consciousness (DOC), including those in comas, vegetative, minimally conscious, and more. The study also included eight DOC patients from a nearby hospital as the experimental group. The experiment achieved a high average online accuracy of 91.5 ± 6.34 percent [1]. The researchers focused on the discrete categories of emotions and emotions in several

continuous dimensions. Because there are considerable intersubject differences in the evaluations of valence and arousal using the Self-Assessment Manikin and the scale's meaning is arbitrary, this model was limited to two standard emotional states with a consistent label for each. Additionally, given their limited consciousness and disabilities, patients may feel more stress due to various emotional conditions. Remarkably, it makes sense that increasing the dimensions makes analysis and classification much more difficult.

2.3. Preprocessing

There are numerous methods to elicit emotions for experiments. These methods allow researchers to study the corresponding brain activity patterns and develop emotion recognition models. Emotion-eliciting stimuli present participants with various stimuli (images, videos, music, stories, and more) that elicit specific emotions. For example, sad scenes or music might induce sadness, while funny videos might induce happiness. In this method, external triggers are used to evoke emotional responses. Imagery, visualization, or audio instructions immerses participants in imagined situations associated with specific emotions. Virtual reality (VR) environments also create immersive environments that evoke specific emotions. Participants can interact with virtual scenarios that trigger emotional responses, providing a more ecologically valid way of inducing emotions. Social interaction scenarios involve designing scenarios where participants interact with virtual characters or real individuals to provoke social emotions like empathy, envy, or embarrassment. Individual variability and research goals are essential factors in selecting an induction method. Participants in the DOC experiment was shown forty 30-second clips, featuring very positive or negative situations from well-known Chinese films or crosstalk shows [1]. Each emotion category contained 20 clips. The video clips were displayed in random order, with a hint at the beginning of each clip about whether the clip was positive or negative. Ultimately, a smiling/crying cartoon face indicates positive or negative detection.

Another thing to worry about during preprocessing is the artifact. Signals with no neurological origin are called artifacts. Unwanted physiological signals that could seriously interfere with the EEG are referred to as subject-related artifacts. Due to the dipole in the eyes, ocular movement produces a sluggish signal (< 4 Hz) that corresponds to its mechanical movement. The changes in the dipole's distance to the electrodes create the signal mainly from the frontal and temporal areas.

Conversely, muscular artifacts produce frequency signals much higher than cerebral signals (> 13 Hz). Technical artifacts come from electronics. A low-frequency artifact (2 Hz) is produced by EEG data associated with wire or electrode movements on one electrode. Frequently, the signal has large amplitudes. The noise of the AC power line can be decreased by lowering electrode impedance and using shorter electrode wires.

Thus, denoising is momentous in preprocessing to make signal classification more accurate in the latter steps. Preprocessing mainly focuses on filtering, segmentation, artifact removal, and baseline correction [4, 5]. Low-pass filters remove high-frequency noise and attenuate muscle artifacts. High-pass filters eliminate slow drifts and some baseline shifts. The notch filter eliminates power-line interference (50 - 60 Hz) and its harmonics. Then, EEG data are segmented into shorter segments (epochs) aligned with stimulus presentation or experimental conditions, which allows analysis of specific time windows associated with emotional responses. Independent Component Analysis (ICA) and Principal Component Analysis (PCA) are employed to enhance the elimination of artifacts. ICA separates mixed EEG signals into independent components, while PCA can further enhance the separation of noise and relevant signals within those components. Their combination helps to obtain cleaner EEG data for emotion recognition by identifying and removing independent components representing eye movements and muscle artifacts.

Furthermore, baseline correction adjusts the baseline of each epoch to zero to ensure that baseline shifts do not affect the subsequent analysis. EEGLAB is also a popular preprocessing toolkit that would conveniently run these steps for researchers. A notch filter was used to remove the 50 Hz

power-line noise from the DOC reception data. Then, between 0.1 and 70 Hz, it was filtered using a tenth-order minimum-phase FIR bandpass filter [1].

2.4. Feature Extraction

Feature extraction converts preprocessed EEG signals consisting of complex time-series data from multiple electrodes into meaningful features that can be used to characterize different emotional states. Three primary sources of information can be used to extract features from EEG signals. The spectral information describes the power spectral density (PSD) or the band power in specific frequency bands, which reflect the energy distribution across different frequencies. It involves the frequency distribution of EEG signals. The temporal information describes how the relevant EEG signals vary with time. Features derived from this source of information include statistical measures (mean, variance), temporal properties (waveform length, zero-crossing rate), and temporal dynamics (changes in signal amplitude and frequency over time). Finally, the spatial information describes where the relevant signals come from. It pertains to the distribution of EEG signals across different electrode locations on the scalp. The scalp topologies, which represent the spatial distribution of activity in specific frequency bands or periods, are obtained. By combining these sources of information, researchers can create a comprehensive set of features that captures the dynamic, frequency-specific, and spatial aspects of EEG activity during emotional responses.

The traditional EEG feature analyses are generally split into the time, frequency, and time-frequency domains [4]. However, considering the EEG signal's non-linear characteristics, a non-linear dynamic features analytical approach can be used for more profound research. Time-domain methods include histogram analysis, Hjorth parameters, statistical characteristics, and more. They mainly start with the EEG signal's geometric features. Then, statistical analyses of its features are performed. The estimation of spectral power takes place in the frequency domain analysis, whereas time-frequency analysis relates to the localization of power in time and frequency. Methods in the frequency domain involve Fast Fourier Transform (FFT) and PSD. Discrete Fourier Transform (DFT) represents a signal regarding its frequency components, as in *Equation 1*.

$$X[k] = \sum_{n=0}^{N-1} x[n] \times e^{-i2\pi kn/N} \quad (1)$$

The FFT algorithm efficiently calculates the DFT by exploiting the symmetry properties of the complex exponential terms (divide and conquer). It reduces the runtime from N^2 to $N \log N$, making it much faster for large datasets. *Equation 3* is the simplified version of *Equation 2*. It provides a recursive method for computing the DFT by constructing a and b, then recursively computing their $N/2$ -point DFTs A and B. See its algorithm below.

$$X[k] = [\sum_{m=0}^{\frac{N}{2}-1} x(2m) e^{-i\frac{2\pi}{N/2}mk}] + e^{i\frac{2\pi}{N}k} [\sum_{m=0}^{\frac{N}{2}-1} x(2m+1) e^{-i\frac{2\pi}{N/2}mk}] \quad (2)$$

Define: $a(n) = [x(0), x(2), x(4), \dots, x(N-2)]$, $b(n) = [x(1), x(3), x(5), \dots, x(N-1)]$

$$X[k] = A[k] + e^{-i\frac{2\pi}{N}k} B[k], \quad k = 0, 1, 2, \dots, N-1 \quad (3)$$

FFT is a fundamental tool frequently applied in calculating the Short-Time Fourier Transform (STFT). STFT is valuable in feature extraction because it allows for time-frequency analysis of signals, making it well-suited for non-stationary signals and applications where dynamic changes in frequency content are significant. *Equation 4* is the general formula of continuous-time STFT.

$$STFT\{x(t)\}(\tau, \omega) \equiv X(\tau, \omega) = \int_{-\infty}^{\infty} x(t) w(t - \tau) e^{-i\omega t} dt \quad (4)$$

The PSD describes the power distribution across different frequency components within a signal [6]. Therefore, its unit is energy per frequency. Additionally, it is beneficial to express it in logarithmic decibels per hertz to clarify graphs. The following procedure calculates the PSD values: 1. Find the sequence length as a sample number (N); 2. Run the FFT algorithm on the EEG signal; 3. The PSD is the absolute squared value of the FFT output; 4. The frequency range becomes from zero

in steps of $2\pi/N$ up to $2\pi - 2\pi/N$; 5. Plot $frequency/\pi$ on the x-axis and $10 \times \log(PSD)$ on the y-axis. Analyzing the signal in the time or frequency domain cannot extract the feature information at present. Thus, time-frequency analysis comes in handy for comprehensive analysis. Since PSD and differential entropy (DE) are equivalent for a fixed length in DOC research, the logarithmic PSD feature is the same as the DE feature. In Equation 5, $|x(m, f_k)|$ denotes the absolute amplitude of the original signal after STFT [1].

$$DE = \log(|x(m, f_k)|^2) \quad (5)$$

Wavelet transform is a standard method in the time-frequency domain [7]. The general procedure starts with taking a wavelet and comparing it to the section at the start of the original signal. Next, the correlation between the wavelet and the signal is calculated inside the support of the window. Then, the wavelet is shifted right for new correlation computations until the whole signal is covered. We can create the most straightforward wavelet (Morlet Wavelet) by multiplying the sine wave by a Gaussian curve on the premise that they have the same time points and sampling rates. Online datasets store other essential mother wavelets. Wavelet combinations effectively create a set of bandpass filters. Each version of the wavelet function is sensitive to a specific frequency band, and the combination of these functions covers a range of frequencies. Figure 2 shows that $x[n]$ is the raw EEG data. Each step is a representation of two digital filters: a high pass filter ($g[n]$) and a low pass filter ($h[n]$). The component of the EEG data with the dominant frequency determines how many levels the wavelet decomposes into. Overall, wavelet transforms divide EEG signal into small segments. Each segment taken with a frequency band can be considered a stationary signal. The discrete wavelet transforms (DWT) divides signals at different frequency bands by composing the signal into detail coefficients, passing through low-pass and high-pass filters.

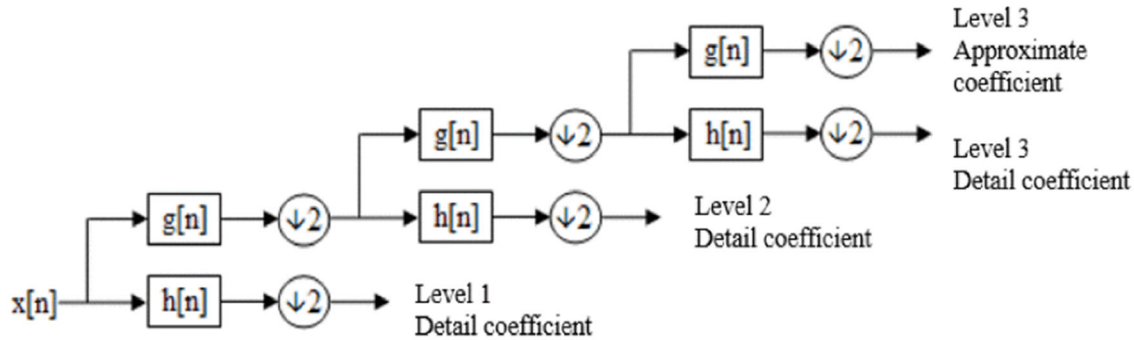


Figure 2. Implementation of decomposition of DWT [8]

For instance, a multichannel EEG signals study creates windows for each EEG channel that are 4 seconds long and overlap each other by 2 seconds [9, 10]. Then, using db4 DWT, all the high-frequency components, including the gamma, beta, alpha, and theta bands, were extracted from each window's data after being decomposed four times. Wavelet transform's time-frequency localization, adaptability, and ability to handle non-stationary signals make it a valuable tool for improving the accuracy and effectiveness of emotion recognition systems.

2.5. Optimal Feature Selection

Optimal feature selection refers to selecting a subset of features from a more extensive set of potential features in such a way that it improves the performance of a machine learning model.

2.6. Classification Algorithm

The K-Nearest Neighbors (KNN) algorithm is a simple and initiative classifier [11]. "K" represents the number of nearest neighbors to consider when predicting a new data point. A larger K reduces noise but may result in a more biased prediction. KNN uses a distance metric. If we have a dataset of EEG recordings from various individuals, each labeled with the corresponding emotional state, and each EEG recording consists of multiple features extracted from different frequency bands and periods. A training set and a test set are first created from the dataset. EEG recording is a data point whose corresponding emotional state label is the target variable. A normalization of EEG features keeps them on a similar scale. Find the optimal K value and calculate the Euclidean distance from this data point to all data points in the training set for each EEG recording in the test set. Locate the K training points that are the shortest distance from the test point. For classification, determine the most frequent emotional state among the K nearest neighbors. Assign this emotional state as the predicted emotion for the test data point.

Finally, to evaluate the KNN model's correctness, compare the predicted emotions and the actual emotions in the test set. The multichannel EEG signals researchers divided the total 37120 EEG samples from 32 subjects who each watched 40 videos into ten parts, of which nine parts were for training and 1 part for testing. The value of K was set to 3 [9]. Subsequently, valence and arousal accuracy are 92.28 ± 0.62 and 92.24 ± 0.33 on the DEAP dataset. KNN is simple to implement but can be sensitive to noisy or irrelevant features. EEG datasets are used for emotional recognition [11]. SEED and DEEP datasets are the most popular.

On the other hand, the Support Vector Machine (SVM) algorithm is more robust and can handle non-linear data with high dimensionality. The DOC experiment utilized the LIBSVM toolbox to establish the SVM classifier. As a result, three out of eight participants achieved significant online emotion recognition accuracy. It demonstrates that the suggested BCI system might be a promising instrument for DOC-based emotional state detection in patients [1]. The Adversarial Learning based on the subject-independent deep neural network (DNN) model achieves lower individual variation and higher average accuracy than SVM. Researchers used the randomization function $R: S \rightarrow \hat{S}$ to create incorrect labels with the goal of learning adversarially. The next step is to optimize the following two functions [8]. The DNN is made to anticipate the real and fake subject labels.

$$L_{confusing} = -\log P(R(S_s)|DNN(X_s)) \quad (6)$$

$$L_{emotion} = -\log P(Y_s|DNN(X_s)) \quad (7)$$

The following dual loss function is optimized by combining the adversarial loss and emotion classification loss.

$$L_{adversarial} = \alpha L_{emotion} + \beta L_{confusing} \quad (8)$$

Overall, adversarial learning in emotion based BCIs leverages the power of adversarial networks to enhance emotion recognition accuracy, improves generalization to different datasets, and increases robustness against adversarial attacks. It is an emerging approach that holds promise in advancing the field of emotion based BCIs.

3. Real World Applications

Emotion-based BCIs offer diverse real-world applications that bridge the gap between human emotions and technology [12]. These interfaces have the potential to enhance user experiences in various domains. They can revolutionize human-computer interaction by tailoring interfaces to users' emotional states. For instance, BCIs can monitor cognitive load and mental fatigue, providing insights into users' mental states during tasks like driving or operating machinery. In healthcare and mental health monitoring, emotion based BCIs can aid in monitoring mental health conditions like depression.

Moreover, they provide objective insight into emotional responses, helping therapists and clinicians tailor treatment plans more effectively.

4. Ethical Considerations and Future Directions

Emotion-based BCIs offer unlimited possibilities; however, privacy concerns emerge as BCIs can directly access users' intimate emotional states [12]. In addition, interpreting emotions could lead to unintentional biases or misinterpretations, influencing decision-making. In the future, ethical frameworks must evolve to address the challenges of informed consent, data ownership, and transparency. Research should explore ways to ensure fairness, mitigate biases, and establish universal patterns across diverse populations. Emotion-based BCIs can reshape human-technology interaction and enrich mental health interventions, provided ethical considerations guide their development and utilization towards a more empathetic and equitable technological landscape.

5. Conclusion

In this comprehensive exploration of emotion recognition through BCIs, we have delved into the intricate processes that forge strong connections between individuals and their affective states. In the article, we primarily reviewed signal acquisition, emotion recognition model, preprocessing, feature extraction, feature selection, classification, and, towards the end, real-world applications and ethical considerations. They have unveiled both the potential and challenges of this transformative field. BCIs have unlocked the ability to decode emotions from brain signals, offering profound insights into our innermost experiences. This process, from raw signals to informative features, is a testament to the power of signal processing and analytical techniques. Utilizing non-invasive wearables, such as EEG helmets, has facilitated capturing emotional responses, transforming them into meaningful features that characterize emotional states. Researchers have significantly progressed in developing low-cost BCI devices with increasingly better usability in recent years. Due to time constraints, 12 of 25 representative papers were selected to generalize contemporary BCI studies. The classification algorithms, ranging from the simplicity of KNN to the complexity of deep neural networks, have demonstrated their efficacy in translating EEG features into actionable emotional insights. As these algorithms evolve, they promise to contribute to a more nuanced understanding of the complex landscape of human emotions. The real-world applications of emotion-based BCIs are also far-reaching. From revolutionizing human-computer interaction to enhancing mental health monitoring and enriching the lives of individuals with disabilities, BCIs hold the potential to reshape diverse domains. However, the research forward must be guided by ethical considerations. As we navigate this uncharted territory, collaboration between researchers, ethicists, and policymakers is paramount. Overall, the path to emotion recognition through BCIs is a journey that goes beyond the confines of scientific exploration, extending into the realm of ethics, privacy, and human values. By embracing this multidisciplinary approach, we pave the way for a future where technology understands our emotions and respects and enhances our humanity.

References

- [1] Huang, Haiyun, et al. An EEG-Based Brain-Computer Interface for Emotion Recognition and Its Application in Patients with Disorder of Consciousness. *IEEE Transactions on Affective Computing*, 2021, 12(4): 832–842.
- [2] Panoulas, Konstantinos J., et al. Brain-Computer Interface (BCI): Types, Processing Perspectives and Applications. *Multimedia Services in Intelligent Environments: Integrated Systems*, edited by George A. Tsihrintzis and Lakhmi C. Jain, Springer, 2010, pp. 299–321.
- [3] Usakli, Ali Bulent. Improvement of EEG Signal Acquisition: An Electrical Aspect for State of the Art of Front End. *Computational Intelligence and Neuroscience*, 2010, 630–649.
- [4] Torres, Edgar P., et al. EEG-Based BCI Emotion Recognition: A Survey. *Sensors*, 2020, 20(18): 50–83.

- [5] Maiorana, Emanuele, et al. EEG Signal Preprocessing for Biometric Recognition. *Machine Vision and Applications*, 2016, 27(8): 1351–1360.
- [6] Gupta, Akshansh, et al. On the Utility of Power Spectral Techniques With Feature Selection Techniques for Effective Mental Task Classification in Noninvasive BCI. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, 51(5): 3080–3092.
- [7] Islam, Md. Rabiul, and Mohiuddin Ahmad. Wavelet Analysis Based Classification of Emotion from EEG Signal. *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019, 1–6.
- [8] Hwang, Sunhee, et al. Subject-Independent EEG-Based Emotion Recognition Using Adversarial Learning. *2020 8th International Winter Conference on Brain-Computer Interface (BCI)*, 2020, 1–4.
- [9] Li, Mi, et al. Emotion Recognition from Multichannel EEG Signals Using K-Nearest Neighbor Classification. *Technology and Health Care*, 2018, 26(S1): 509–519.
- [10] Parvin, Hamid, Hosein Alizadeh, and Behrouz Minaei-Bidgoli. MKNN: Modified k-nearest neighbor. *Proceedings of the world congress on engineering and computer science*. 2008, 1.
- [11] Houssein, Essam H., et al. Human Emotion Recognition from EEG-Based Brain–Computer Interface Using Machine Learning: A Comprehensive Review. *Neural Computing and Applications*, 2022, 34(15): 12527–12557.
- [12] Fontanillo Lopez, Cesar Augusto, et al. Beyond Technologies of Electroencephalography-Based Brain-Computer Interfaces: A Systematic Review From Commercial and Ethical Aspects. *Frontiers in Neuroscience*, 2020, 14, 611130.