Preprints.org

Article

Inner Speech Recognition for Mutism and Speech Disorder Using Brain-Computer Interface

Mokhles M. Abdulghani , Wilbur L. Walters , Khalid H. Abed [*]

Article

Inner Speech Recognition for Mutism and Speech Disorder Using Brain-Computer Interface

Mokhles M. Abdulghani, Wilbur L. Walters and Khalid H. Abed *

Jackson State University, MS, USA; mokhles.m.abdulghani@students.jsums.edu (M.M.A.); wilbur.l.walters@jsums.edu (W.L.W)

* Correspondence: khalid.h.abed@jsums.edu

Abstract: Brain-Computer Interface (BCI) systems can assist physically challenged people to interact with their surroundings and improve the quality of their lives. Decoding human thoughts is a powerful technique that can assist paralyzed people who have lost their speech production ability. Speaking is a combined process involving synchronizing the brain and the oral articulators. This paper proposed a high-accuracy brain wave pattern recognition based on inner speech using a novel feature extraction method. Only eight EEG electrodes were used in this study, and they were set on selected spots on the scalp. Support Vector Machine (SVM) was employed to decode the recorded EEG dataset into four internally spoken words which are: Up, Down, Left, and Right. The proposed approach achieved overall classification accuracy that ranges between 96.20% to 97.5%. In addition, more performance evaluation metrics were estimated to test the reliability of classifying the EEG-based inner speech data, and we obtained 97.61%, 97.50%, and 97.73% for F1-score, recall, and precision respectively. Furthermore, the Area Under Curve of the Receiver Operating Characteristic (AUC-ROC) proved the strength of the proposed approach for classifying the specified inner speech commands by achieving a macro-average amount of 99.32%. The inner speech classification method using electroencephalography proposed in this work can clinically help improve communication for patients with problems including speech disorder, mutism, cognitive development, executive function, and psychopathology.

Keywords: inner speech; brain-computer interface; imagined speech; support vector machine; SVM; autoregressive model; AR; wavelet variance; shannon entropy

1. Introduction

Brain signals were used in 1967 for secured text messages transmitting based on the transmission of letters of the alphabet using EEG and Morse code by Dewan and his research team [1]. Although studying inner speech is an enormous challenge, it is essential to understand the development of language capabilities and the advanced mental abilities to which language is linked [2]. Another skill that appears to be linked to inner speech is silent reading. Furthermore, inner speech assists with several brain disorders after a traumatic brain injury, brain stem infarcts, cerebral palsy, or amyotrophic lateral sclerosis, which affects verbal communication [3]. The concept of Brain-Computer Interface (BCI) offered great assistance for paralyzed people to interact directly with the environment surrounding them and improve the quality of their lives. More particularly, people with aphasia or speech disorder would be supported with a communication system that can recognize inner speech from their brain signals [4]. The brain signal used in BCI systems can be captured and monitored with different approaches, such as functional Near-Infrared Spectroscopy (fNIRS), Magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI), Electrocorticography (ECOG), and the Electroencephalography (EEG).

Among various brain signals monitoring technologies, EEG has been proven to be one of the most popular methods for monitoring brain activities due to its cost-effectiveness and noninvasiveness. Moreover, EEG offers the quantification and interpretation of cortical activity in several brain regions by measuring the summation of postsynaptic action potentials [5]. In [6,7], and [8] studies were conducted to develop EEG-based BCI systems that distinguish between different

inner speech activities and promising results were reported. The number of EEG sensors required to design any EEG-based BCI system is the main factor for determining the cost, setup, and maintenance complexity of the BCI system, which as a result will affect the possibility of its manufacturing such as a system designed to be used in real-time.

Several studies were able to achieve high overall classification accuracy of 97.66%, 98.60% and 99% for EEG-based BCI for different applications in [52–54], respectively. Most of the researchers have used high-cost EEG headsets to design BCI systems for inner speech processing. Recent research published in Scientific Data-Nature journals [9] used a costly 128 channels EEG headset to record inner speech-based brain activities from 10 subjects. The participants were trained to perform speech imagery for four commands: Up, Down, Right, and Left responding to a visual cue presented on a computer screen. In [10], the same 128-channel EEG dataset was used with a deep learning method to classify them according to their corresponding to the internally spoken word, and an average EEG-based inner speech classification accuracy of 29.67% was achieved. A subsequent study [11] achieved 51% EEG-based inner speech classification accuracy using the same 128-channel EEG datasets by considering only a specific number of channels (28-channel only) in the classification process depending on their location on the scalp that could be linked to the inner speech activity.

In [12], a 20-channel EEG headset was used for aircraft pilots' cognitive workload estimation, and an accuracy of 91.67% was achieved using a combination of multi-feature extraction and Support Vector Machine (SVM). In our very recent research [13], we were able to achieve 93% classification accuracy to design EEG-based Internet of Brain-Controlled Things (IoBCT) based on visual cues using only a 16-channel EEG headset. Moreover, in [14,15], we achieved excellent accuracy in classifying EEG signals with only an 8-channel EEG headset, where we classified EEG signals to control a drone and designed EEG-based IoBCT based on visual cues, respectively.

Our brain can be excited and stimulated by the external environment through the various senses we have such as hearing, touch, sight, smell, and taste. Visual and auditory cues play a great role in the excitement of central motor cortex [16] and [17], but still the functional effect of such as these are limited [18]. In [19], EEG was used to classify inner speech, and the EEG electrodes have been placed on different spots on the scalp. The results revealed that the most important EEG channels for classifying inner speech were the ones laid on Broca's and Wernicke's regions of the brain. Both regions continuously analyze and control the production of our speech, but Wernicke ensures that the speech makes sense, while the Broca controls the fluency of our speech.

The experimental procedure followed in an EEG-based inner speech classification research is essential for obtaining classifiable data. The most followed procedure is to ask the subjects to imagine speaking the commands only once. However, in [20] and [21], the participants were asked to imagine saying a specific command multiple times in the same recording session. In [22], four commands, namely, *up*, *down*, *right*, and *left*, were used to be internally spoken and discriminated based on the recorded EEG using Extreme Learning Machine (ELM) classifier. Overall classification accuracy of 49.77% and 85.57% were obtained, respectively.

Recent research revealed that EEG-based inner speech classification accuracy can be improved when auditory cues are used [23]. In [24], four participants imagined speaking without any subvocal or vocal activity while the audio cues were given to stimulate the brain. During the initial period of this experiment, the participants heard the audio cues through electrostatic earphones, either a spoken ("*ku*") or a spoken ("*ba*") followed by a train of clicks (arrows) indicating the rhythm to be reproduced. In [25], seven participants imagined speaking a cued syllable, and the cues were also submitted with auditory stimuli. Additionally, no motor activity analysis was conducted in those papers. Finally, in [26], the participants were trained to carefully listen to spoken words and try to comprehend them and then inner speaking them, or "Internal speaking," as called in the paper, immediately after seeing it. Practical research examined inner speech classification using EEG-based BCI systems and showed that inner speech could be recognized using words with high discriminatory pronunciation [27]. Feature extraction methods such as Autoregressive (AR) coefficient estimation, Shannon entropy, and wavelet variance estimation were used in several studies to model EEG to obtain a representation of the signal at each channel and improve the classification accuracy [28–31].

This paper is a follow up for our previous work in [44] where we used a low-cost 8- channels EEG headset, *g.tec Unicorn Hybrid Black*+ [32]. Audio cues were employed for the purpose of

stimulating motor imagery in this study. The question: *Where do you want to go* was used, and the participant replied with inertly spoken audio commands *Up, Down, Left, and Right*. Finally, we employed multi-feature extraction methods and SVM to discriminate between the four commands.

## 2. Materials and Methods

2.1. Subjects

Four native English speakers, right-handed and healthy individuals participated in this study, all of them had no neurological or movement disorders, no hearing loss, and no speech loss. Each participant signed his written informed consent. None of the participants had any previous BCI contribution or experience. The participants were two males and two females aged 20 to 56 and were named (sub-01) for the first subject (sub-02) for the second subject and so on. The study was accompanied in Jackson State University at the Department of Electrical and Computer Engineering and Computer Science. All subjects were voluntarily participated in the experiment, and the experimental procedure and data collection have been done according to the approved Institutional Review Board (IRB) procedures at Jackson State University [33].

2.1. Apparatus

An EEG *Unicorn Hybrid Black+* headset manufactured by *g.tec* [32] was used to record the EEG-based inner speech data. It is a low-cost EEG headset that records up to seventeen channels at up to 250 Hz sampling frequency, eight of them are EEG, a three-axis accelerometer, a three-axis gyroscope, a battery signal, a counter signal, and a validation signal. The recorded channels are on the positions: (FZ, C3, CZ, C4, PZ, PO7, OZ, and PO8). A cap with appropriate size was chosen to fit the participants head and all electrodes were placed in the required positions in the cap. A conductive gel was used to fill the gaps between the scalp and the electrodes.

2.3. Experimental setup and procedure

We considered the 10-20 electrode placement system recommended by the American clinical neurophysiology society [34]. The 10-20 system was first presented by Herbert Jasper at the 1957 Brussels IV International EEG Congress. The numbers (10) and (20) denote the distances between the electrodes position, which are either 10% or 20% of the total distance (front-back or right-left) of the skull. The head cap has been adjusted to ensure that there are electrodes placed, as much as possible, within the Wernicke and Broca regions, which are considered good spots for better quality inner speech-based EEG. Figure 1 presents the *g.tec Unicorn Hybrid Black+ headset*, Wernicke and Broca, and the electrode positions. Reference and ground electrodes are sticked on the mastoids using a one-use sticker.
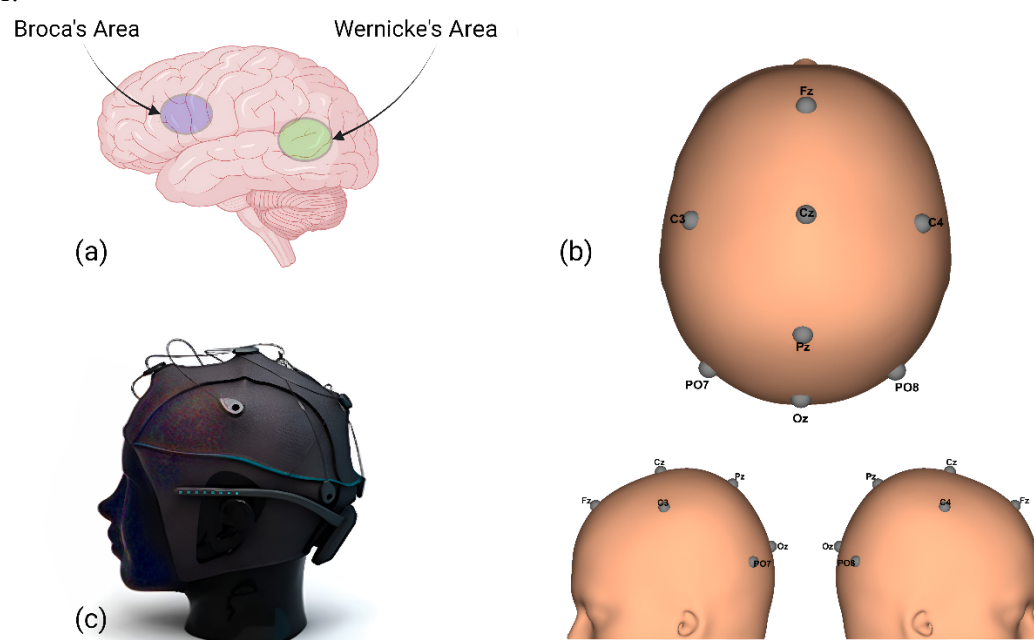
Figure 1. (a) Wernicke's and Broca's areas, (b) The electrode positioning layout, (c) The *g.tec Unicorn Hybrid Black+*.

The experiment has been designed to record EEG during performing inner speech. Inner speech, also called imagined speech, silent speech, covert speech, is thinking in the form of sound – "hearing" our own voice silently without the intentional movement of any extremities such as the lips, tongue, or hands. Two subjects participated in each recording session, where one subject was performing inner speech and the second was performing the audio cue. We think that getting the participants involved in announcing the audio cue will help prevent any bias in the results even if the spoken commands were in the same order each session, where the audio cue was announced by different subjects. Before the recording started, the two subjects were seated in a comfortable high-back chair in front of each other and to familiarize them with the recording procedures. All experiment steps were clarified during the setup of the EEG headset, and the external electrodes, which took about 15 minutes. The first subject was trained to imagine responding to the question: *Where do you want to go?* The question was said by the second subject who was sitting in front of the first subject. The response was an internally or an imaginary spoken command, which is either *Up* or *Down*, *Left* or *Right*. Each participant accomplished 25 recordings for each command. The recording procedure was implemented as follows. When the first 10 seconds (± 2 seconds) of the EEG recording passed, the audio cue was announced and by the end of the second 10 seconds (± 2 seconds), the participant started imagining the desired response as inner speech for 60 seconds. The participant was trained to keep repeating the internally spoken command until the end of the 60 seconds, and the recording was stopped after 10 seconds as illustrated in Figure 2.

A total of 400 recordings sessions were successfully completed, and the collected data has been merged without separating them according to their corresponding participants. We ended up having 100 recordings for the command *Up*, 100 recordings for *Down*, and so on. This way, we can examine the performance of the proposed classification method in distinguishing between the four commands using a dataset from four subjects altogether in one run. Additionally, by combining the data together and using them in the training and testing process, we can design a reliable algorithm that can be tested later in real-time on different subjects. For each command, the first 25 recordings were for subject 1, the second 25 recordings were for subject 2, the third 25 recordings were for subject 3, and the last 25 recordings were for subject 4. The recorded EEG dataset was spitted, labeled, and stored to be prepared for the preprocessing stage.
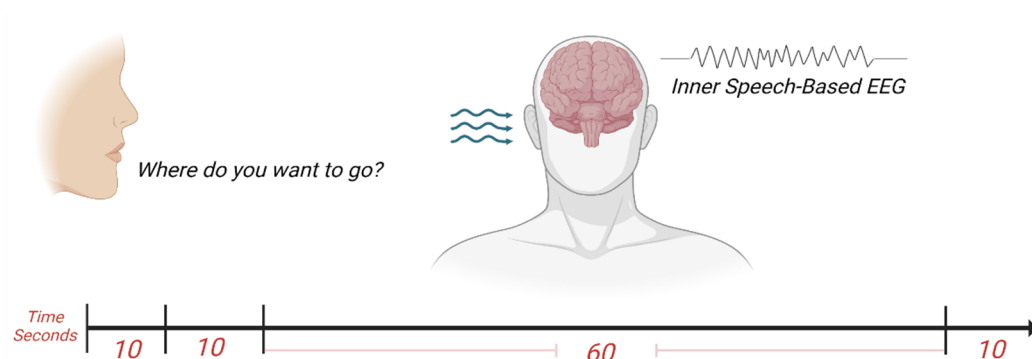
Figure 2. The recording procedure.

2.4. EEG Pre-Processing

Before the EEG- Preprocessing stage, 8 seconds from the start and the end of each recording was trimmed to make sure that we have EEG data that reflects the participants brain activities while they were purely performing imagined speech. In the EEG pre-processing stage, the recorded data forwarded to a combination of several noise attenuation and calibration approaches to prepare the EEG signals for further analysis. Pre-processing is a significant stage for EEG analysis to remove any expected noises. This noise can be categorized as environmental or instrumentation noise such as the noise from the power line and biological nose such as ECG and EMG signals arising from muscle movement. To attenuate biological artifacts such as EMG and ECG noise, researchers have developed

different methods. Traditional filtration technique worked well so far to eliminate electrical line noise and other biological artifacts with high frequency such as EMG. But removing ECG can result in damaging the EEG characteristics and losing its fundamental features because the ECG artifacts have a noteworthy spectral overlap with the original EEG signals [52–54].

The recorded EEG signals were analyzed using *gHIsys* MATLAB toolbox (https://www.gtec.at/product/ghisys). To ensure that we have only the performed speech imagery data, we considered removing the first and the last 8 seconds of the 60 seconds in each recording. A bandpass filter between 10 and 100 Hz was used to attenuate the baseline drift and the noisy signals from EEG signals. This filtering bandwidth delivers only the typical frequency bands corresponding to EEG in the human brain [35]. A bandstop (notch) filter at 60 Hz was used to reject the power frequency used in Mississippi, USA. The normalization (vectorization) technique was used to simplify the dataset and reduce the computing demand required to classify the four commands. The dataset was split into 360 recordings for training and 40 recordings for testing (90% for training and 10% for testing). The training and testing dataset was normalized by determining the mean and standard deviation for each of the eight EEG signals and apply the following formula to combine them in a single-vector signal:

$$EEG_{Norm} = \frac{x - \alpha}{\sigma} \tag{1}$$

where (x) is the filtered EEG signal, (α) is the mean, and (σ) is the standard deviation. The EEG dataset was then prepared for the feature extraction step. The result of pre-processing steps for one subject EEG is shown in Figure 3.
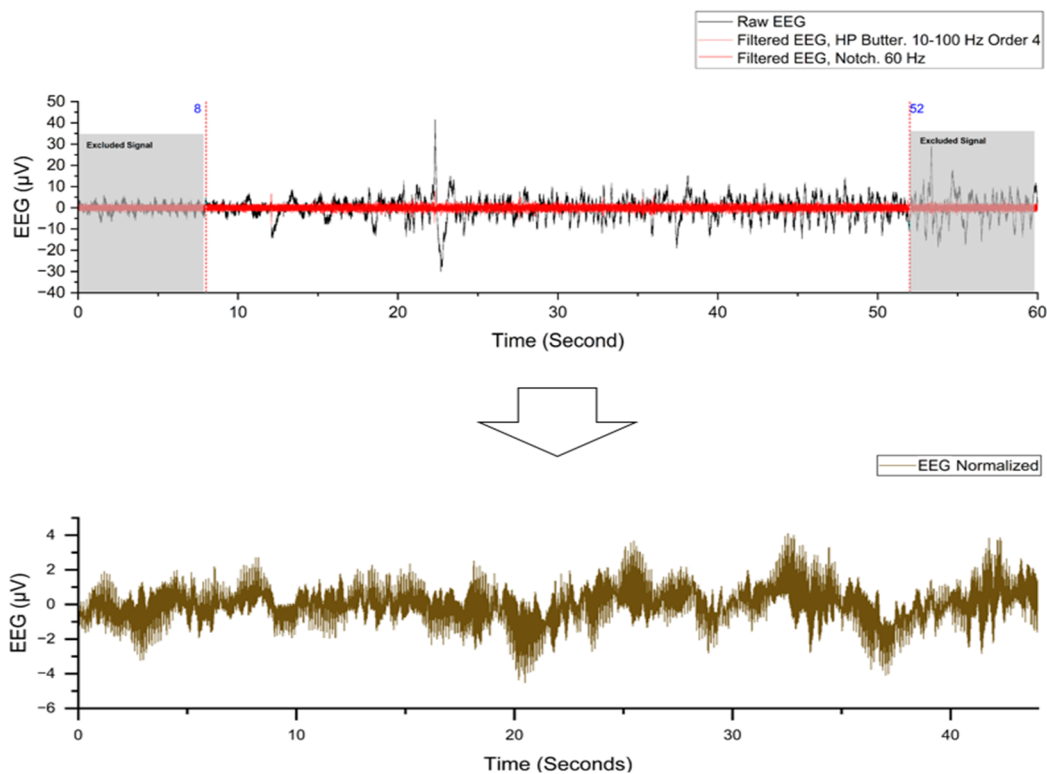
Figure 3. Eight-channel preprocessed EEG dataset at 250 HZ (250 sample per second).

2.5. Feature Extraction

Multi-feature extraction methods were applied on eight blocks for each recording with a time window of about 4 seconds (1024 samples). Autoregressive model (AR) coefficients, Shannon Entropy (SE), and multiscale wavelet variance estimates were used to extract features of the recorded data.

2.5.1. AR Coefficients

In an AR method of order p, the signal $X\{n\}$ at time n could be represented as a linear sequence of p prior estimates of the same signal. Specifically, the AR method is modeled as:

$$X\{n\} = \sum_{j=1}^{p} a\{i\}x(n - i) + e\{n\} \tag{2}$$

where $a\{i\}$ is i coefficients of the AR representation, $e\{n\}$ is added noise with zero mean value, and p is the order number of the AR model. Countless methods could be used to calculate the coefficients of an AR representation. The method we used to estimate the AR order in this work is the ARfit [36]. The 1st-order has been selected for the recorded EEG signals.

2.5.2. Shannon Entropy

Shannon entropy is one of the most attractive cost functions, which is a measure of signal complexity to wavelet coefficients generated by wavelet packet transform where larger entropy values represent higher process uncertainty and, therefore, higher complexity [37]. The representation of the Shannon entropy for the undecimated wavelet packet transform is formulated as follows:

$$SE_j = \sum_{k=1}^{n} P_{jk} \, log \, P_{jk} \tag{3}$$

where n is the subsequent coefficients in a j number of nodes and P_{jk} are the normalized squares of the wavelet packet coefficients in each node.

2.5.3. Multiscale Wavelet Variance Estimates

Wavelet variance measures the variability in EEG signal by scale or equivalently in EEG signal over octave-band frequency intervals. We adjusted the vectorized data to make the number of samples in each recording in the form of (2A). The biggest number of (A) we could get with the number of samples we have in each recording is 12. For the signal length of 8192 samples (2 ^ 12) and using the 'db2' wavelet with level 5 [38], 10 multiscale wavelet variance features were extracted from each recording using the following formula:

$$Number \; of \; Wavelet \; Variance = \; 2^{A-db} \tag{4}$$

A total of 170 features were extracted from the EEG data: 4 per time window (1024 sample) AR coefficients, 16 per time window SE values, and 10 wavelet variance estimations. After the multi-feature extracting stage, the EEG data was reconstructed to be a 360-by-170 features matrix for training and a 40-by-170 features matrix for testing. By employing Autoregressive coefficients, Shannon Entropy, and multiscale wavelet variance estimates, the data was reduced from 8192 to 170 element vectors. Representation of the difference in the wavelet variance for the 170 extracted features from the recorded and preprocessed EEG data will be presented in the final results.

2.6. Classification

In the classification stage, the data was processed with supervised learning, where the specified algorithm was employed to learn from the prepared data. In this study, the classification stage was defined as the determination of four different internally spoken commands (*Up*, *Down*, *Left*, and *Right*), which are considered a multiclass classification process. SVM is one of the most well-known supervised learning algorithms specialized in classification problems. Classification using SVM is powered through generating a best line or decision boundary that segregates an n-dimension space to multiclass to easily enable data sorting to the category to which they belong [39,40]. SVM works on picking the margin points that construct vectors which are called support vectors to assist with generating the best decision boundary.

The SVM architecture utilizes a set of mathematical functions that are known as the kernel functions. The kernel function performs a kind of similarity measure between input objects and transforms it into the required output [41]. We employed SVM, which is a machine learning algorithm for differentiation between the four chosen commands. Furthermore, k-fold cross-

validation (k = 10) was used to achieve a perfect estimate of the proposed model performance on the recorded inner speech data and to avoid overfitting in the classification process. The K-fold validation is an alternative to a fixed validation set. It does not affect the need for a separate held-out test set. So indeed, the data will be split into training, testing and cross-validation data and is performed on folds of training sets. With k-fold cross-validation of value 10, the model performance will be evaluated after dividing the data into 10 subsets (10 folds) while using the k-1 subsets for training the data. In this way, it can ensure that testing data will be entirely unknown to the classifier that is testing and training data are not coming from the same given group. Figure 4 illustrates the concept of using the K-fold for cross-validation during the data training process.
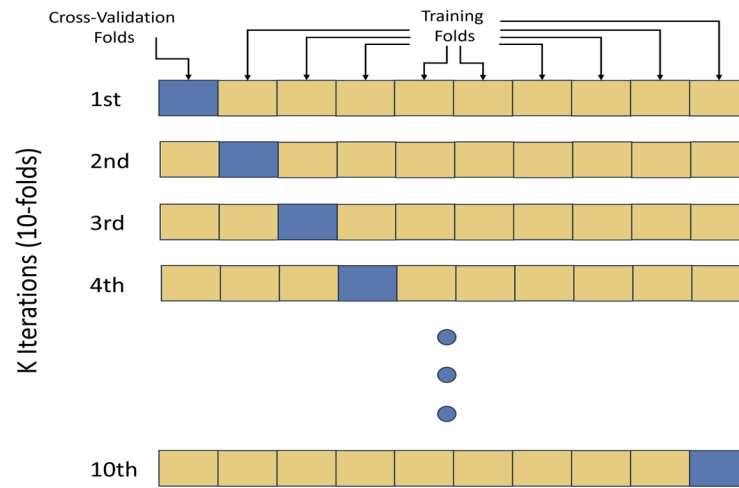

Figure 4. Cross-Validation Using 10-Fold.

2.7. Performance Evaluation

Evaluation metrics adopted within a variety of machine learning techniques are critical in examining the reliability of the designed classifiers. To evaluate the trained model performance, metrics following [42] have been considered. The classified EEG data using the proposed machine learning method was grouped into true positive T_P, false positive F_P, true negative T_N, and false negative F_N. The value of F_P and F_N are the samples that were misclassified, and the value of T_P and T_N are the samples that were correctly classified [42]. The most state-of-art metrics for classification are accuracy, precision, recall, and F1-score. Accuracy estimates the percentage of correct predicted outputs to the overall number of samples in the processed dataset. Recall (sometimes called Sensitivity) estimates the percentage of T_P to the summation of T_P and F_N. Precision estimates the percentage of T_P to the summation of T_P and F_P. Hence the F1-Score estimates the average between recall and precision.

Moreover, the Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC) (AUC-ROC) was plotted. AUC-ROC is a common ranking type of metric that is utilized to show comparisons between learning algorithms and create an optimal learning model by exposing the entire classifier ranking performance [43]. Furthermore, while we have a multiclass classification task, areas under the curve were calculated and presented by macro-averaging, in which each corresponding metric for each individual class was estimated. The following formula is used to estimate the AUC-ROC value for multiclass problems:

$$ACU = S_p - n_p(n_n - 1)/Nn_pn_n \qquad (5)$$

where S_p, n_p, n_n, and N represent the sum of all positive samples, positive and negative samples, and the number of classes, respectively.
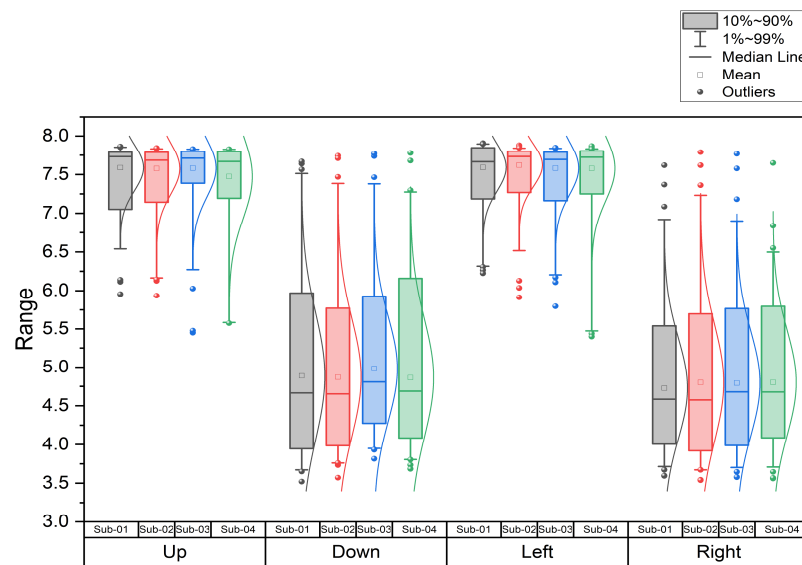
3. Results

In this section, we report the results of the proposed method for EEG signals classification, including the results of extracting multi-features from the preprocessed EEG signal and the results of the SVM machine learning model for the classification of the extracted features.

3.1. Feature Extraction Results

In the feature selection stage, the extracted features using Autoregressive coefficients, Shannon Entropy, and multiscale wavelet variance were compared using boxplot to examine the variance level between each individual command. The obtained results using the suggested features extraction methods showed a noticeable variation between the four commands which will assist with discriminating between them and improve the classification accuracy. Every feature vector of the same class should be closer in its representation point and in different classes they should be far from each other.

For precisely monitoring the variance in data distribution of all features in the four classes we have, a boxplot was used. Representation of the difference in the wavelet variance for the 170 extracted features from the sample of the recorded and preprocessed EEG data using a boxplot is shown in Figure 5.

Figure 5. A boxplot of the wavelet variance for the extracted features from sample of the recorded EEG data.

3.2. Classification Results

After the multi-features extraction stage, classification between the four internally spoken commands was carried out using machine learning to evaluate the model performance. The SVM with a polynomial kernel function, C = 2 and gamma = 0.1 was selected as the best estimator with the best margin size (M) after several trials based on trial-and-error as shown in Table 1 and Figure 6. Gamma and C are regularization parameters where gamma determines the width of the kernel function, and C controls the trade-off between achieving a simple decision boundary and an excellent fit to the data during the training process. The cross-validation splitting strategy was chosen as five-fold cross-validation. Moreover, class names were *Up*, *Down*, *Left*, and *Right*.

We trained and tested our model using the extracted features by SVM five-fold cross-validation. The highest performance of the model was achieved by feeding the features selected by Autoregressive coefficients, Shannon Entropy, and multiscale wavelet variance with an accuracy of 97.5%, precision of 97.73%, recall of 97.50%, and F1-score of 97.61%. In addition, the macro-average AUC-ROC of the model was 99.32%. The model showed excellent performance using the extracted features by the proposed feature extraction methods. The confusion matrix and AUC-ROC plot illustrate the performance of the proposed classifier in Figure 7 and Figure 8, respectively.

Table 1. Information of the participated subjects.

Iteration	Cost (C)	gamma	Accuracy %
1	2.75	0.18	96.2
2	2.7	0.17	96.2

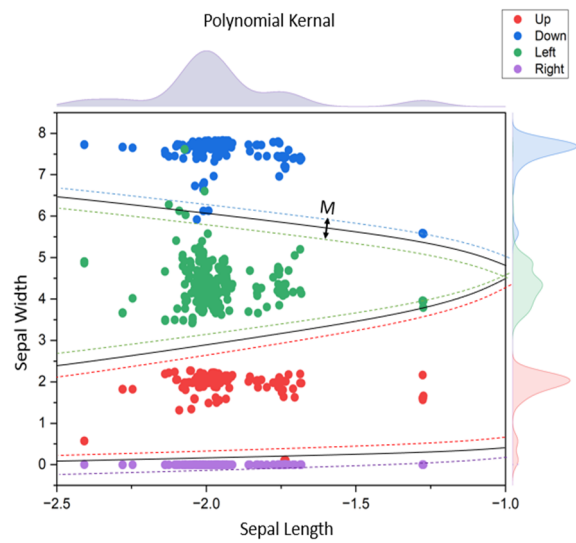| 3 | 2.2 | 0.15 | 96.6 |
| 4 | 2 | 0.1 | 97.5 |

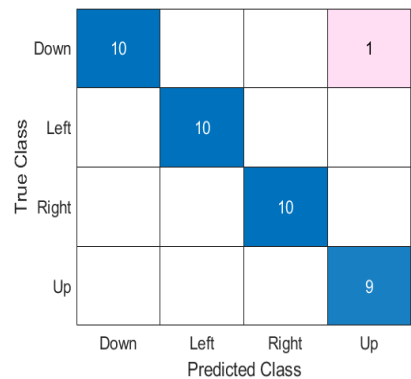**Figure 6.** Margine size in SVM classifier.

Figure 7. Confusion matrix for the classification performance of the SVM model.
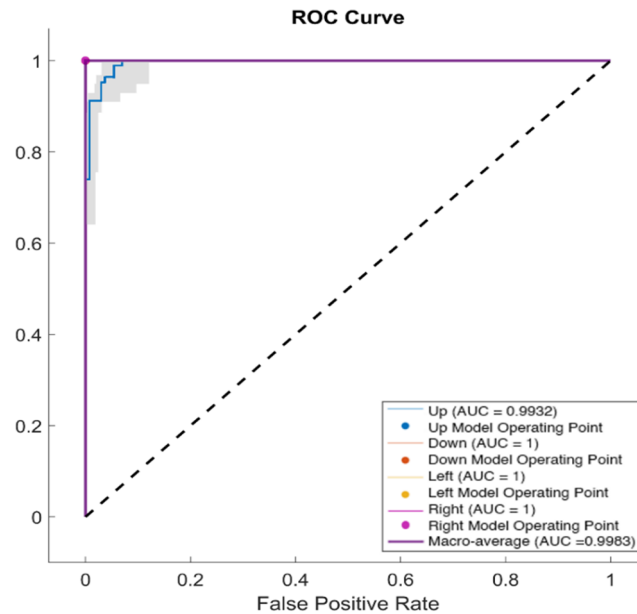

Figure 8. The AUC-ROC plot of the SVM model using the extracted features.

4. Discussion

Complex Applications of EEG in a realistic environment, such as decoding inner speech, generate dynamic and complicated responses in the EEG signals. As reported in [20], the classification accuracy of inner speech applications based on EEG can be affected by the type of cues used to stimuli the brain in the recording procedure. In [20], the participants were trained to keep repeating the internally spoken words for up to 14 seconds and they only responded when they heard a beep. In [21], visual cues were used in each recording, and the participants were told to keep repeating the internally spoken words for 30 seconds. Audio cues have been used to stimulate the brain by asking a question to one of the four subjects and let them imagine speaking one of the four specified commands. Unlike [20] and [21], we did not include cues during all the 60 seconds of response and the subjects were trained to keep repeating the specified command until the time ended.

Classifying inner speech using EEG requires robust and efficient classification approaches. Some researchers recommended autoregressive modeling, Shannon entropy, and wavelet variance estimation as powerful feature-extracting methods to classify EEG [28–31,43]. In [27], massive efforts were made to record EEG data for inner speech applications using an expensive 128-channel EEG headset. Nevertheless, this high number of channels did not allow subsequent researchers [10,11] to get good classification performance when they used the same 128-channel EEG data. In our research, with data recorded using a low-cost 8-channel EEG headset, carefully selected electrodes position on the skull, and the proposed multi-feature extraction method, the results showed that robust and accurate EEG classification could be implemented. Autoregressive modeling, Shannon entropy, and wavelet variance estimation were applied for detecting and classifying inner speech in EEG time series data. The data was reduced from 8192 to 170 element vectors by employing the suggested multi-feature extracting method. A total of 170 features were extracted from each recording and the EEG data was reconstructed as a 360-by-170 features matrix for training and a 40-by-170 features matrix for testing.

It is noteworthy that we combined the recorded data from all the subjects and applied the suggested multi-feature extraction method on them. Aanalyzing the results by splitting the data according to which participant it belongs and then averaging the results as in [45–48] by summing and dividing the results by the number of the participated subject may not be the most practical method for calculating the accuracy and efficiency of the designed classifier. For example, a 100% classification accuracy for the data from subject A, and 80% for data from subject B can be averaged to 90% by doing math calculation (100%+80%)/2. Since we know that the minimum accuracy is 80%, stating that the overall classification accuracy for this classifier is 90% is not really accurate, and this method will not provide any close number to what was averaged when testing this classifier online on different subjects. In our work, we were able to design a BCI system that can distinguish between four inner speech commands for four subjects at the same time rather than designing four systems, where each system is tailored for each participant if a designer considered the methods used by averaging the results. This makes our system more general and more practical because it allows multiple physically-challenged people to use it.

Besides the reduction in the size and complexity of data, the reported results showed a significant variance between the specified classes. Even though this is a significant reduction in data size and complexity, the main objective of using the proposed multi-features extraction method is not just a reduction in data. We aimed to re-represent the data with much smaller features that allow capturing the differences between the required classes so that a classifier can ideally separate the EEG signals. As explained in the result section, the extracted features resulted in high accuracy, precision, recall, F-score, and macro-average AUC. The resulting classifier can be converted to a *C++* or *Python* code using MATLAB code generation and uploaded to a microcontroller to be tested in real-time.

5. Limitation

Data for this study included EEG data from four participants, each around 1.6 hours in total. More participants would allow greater generalizability to indicate the reliability of the proposed classification method. Furthermore, for accurately performing the experiment procedure, the data collection requires the research team to train each participant to familiarize them with the procedure by conducting at least one recording session before the considered one in this paper.

6. Conclusion

This research aims to pave the way for a better understanding of processing and classifying inner speech using EEG and machine learning. Numerous numbers of people around the world need such an idea to improve the quality of their life. Signal processing was implemented to extract wavelet multi-feature from EEG signals and employ those features to classify four inner speech classes. Not only did the multi-feature extraction result in a substantial amount of data reduction, but it enabled capturing the differences between the *Up*, *Down*, *Left*, and *Right* classes as confirmed by the results of cross-validation and the performance of the support vector machine classifier on the test dataset as well. The experiment further demonstrated that applying autoregressive modeling, Shannon entropy, and wavelet variance estimation to the raw EEG data resulted in an excellent performance. Five-fold cross-validation was used to improve the classification performance and generalization. The achieved results range between 96.20% to 97.5% for overall classification accuracy. Other performance evaluation metrics were estimated, and we obtained 97.73% for precision, 97.50% for recall, and 97.61% for F1-score. Moreover, the macro-average AUC-ROC of 99.32% proved the efficacy and validity of the proposed approach, for classifying different inner speech commands using EEG.

References

1. E. M. Dewan, "Occipital alpha rhythm eye position and lens accommodation," Nature, vol. 214, pp. 975–977, 1967.
2. N. Eisenberg, A. Sadovsky, and T. Spinrad, "Associations of emotion-related regulation with language skills, emotion knowledge, and academic outcomes," New directions for child and adolescent development, vol. 2005, no. 109, pp. 109–118, 2005.
3. M. P. Jani, G. B. Gore, "Occurrence of communication and swallowing problems in neurological disorders: analysis of forty patients," Neuro Rehabilitation, vol. 35, no. 4, pp. 719–727, 2014.
4. S. Martin et al., "Decoding inner speech using electrocorticography: Progress and challenges toward a speech prosthesis" Frontiers in neuroscience, vol. 12, p. 422, 2018.
5. Biasiucci, A., Franceschiello, B. & Murray, M. M. Electroencephalography. Curr. Biol. 29, R80–R85 (2019).
6. P. Gaur, R.B. Pachori, H. Wang, G. Prasad, An automatic subject specific intrinsic mode function selection for enhancing two-class EEG based motor imagery-brain computer interface, IEEE Sens. J. 19 (2019) 6938–6947.
7. M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, R. Srinivasan, Toward EEG sensing of imagined speech, in: J.A. Jacko (Ed.), Human–Computer Interaction. New Trends: 13th International Conference, HCI

International 2009, San Diego, CA, USA, July 19–24, 2009, Proceedings, Part I, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 40–48.

8. M. Matsumoto, J. Hori, Classification of silent speech using support vector machine and relevance vector machine, Appl. Soft Comput. 20 (2014) 95–102.

9. Nieto, N., Peterson, V., Rufiner, H.L.et al. Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition.Sci Data 9, 52 (2022). https://doi.org/10.1038/s41597-022-01147-2.

10. B. v. d. Berg, S. v. Donkelaar and M. Alimardani, "Inner Speech Classification using EEG Signals: A Deep Learning Approach," 2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS), Magdeburg, Germany, 2021, pp. 1-4, doi: 10.1109/ICHMS53169.2021.9582457.

11. Wei X, Surjana AI, Söffker D, "Inner speech classification based on electroencephalography (EEG) signals and support vector machine (SVM)", Preprint, November 2023, DOI: 10.21203/rs.3.rs-2287259/v1.

12. Taheri Gorji, H., Wilson, N., VanBree, J. et al. Using machine learning methods and EEG to discriminate aircraft pilot cognitive workload during flight. Sci Rep 13, 2507 (2023). https://doi.org/10.1038/s41598-023-29647-0.

13. M. M. Abdulghani, O. Franza, F. Fargo and H. Raad, "Brain Waves Pattern Recognition Using LSTM-RNN for Internet of Brain-Controlled Things (IoBCT) Applications," 2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 2022, pp. 1-5, doi: 10.1109/IEMTRONICS55184.2022.9795715.

14. Mokhles M. Abdulghani, Arthur A. Harden, and Khalid H. Abed, "A Drone Flight Control Using Brain-Computer Interface and Artificial Intelligence," The 2022 International Conference on Computational Science and Computational Intelligence – Artificial Intelligence (CSCI'22–AI), IEEE Conference Publishing Services (CPS), Las Vegas, Nevada, December 14-16, 2022.

15. Mokhles M. Abdulghani, Wilbur L. Walters, and Khalid H. Abed, "Low-Cost Brain Computer Interface Design Using Deep Learning for Internet of Brain Controlled Things Applications," The 2022 International Conference on Computational Science and Computational Intelligence – Artificial Intelligence (CSCI'22–AI), IEEE Conference Publishing Services (CPS), Las Vegas, Nevada, December 14-16, 2022.

16. Riccio A, Mattia D, Simione L, Olivetti M, Cincotti F (2012) Eye gaze independent brain computer interfaces for communication. Journal Neural Eng 9: 045001.

17. Hohne J, Schreuder M, Blankertz B, Tangermann M (2011) A novel 9-class auditory ERP paradigm driving a predictive text entry system. Front Neuroscience 5: 99.

18. Panachakel, T., Vinayak, N., Nunna, M., Ramakrishnan, G., and Sharma, "An improved EEG acquisition protocol facilitates localized neural activation," in Advances in Communication Systems and Networks (Springer), pp. 267–281, 2020.

19. Wang, H. E., Bénar, C. G., Quilichini, P. P., Friston, K. J., Jirsa, V. K., and Bernard, C. (2014). A systematic framework for functional connectivity measures. Front. Neurosci. 8:405. doi: 10.3389/fnins.2014. 00405.

20. Nguyen, C. H., Karavas, G. K., and Artemiadis, P. (2017). Inferring Inner speech using EEG signals: a new approach using Riemannian manifold features. J. Neural Eng. 15:016002. doi: 10.1088/1741-2552/aa8235.

21. Koizumi, K., Ueda, K., and Nakao, M. (2018). "Development of a cognitive brain-machine interface based on a visual imagery method," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (Honolulu: IEEE), 1062–1065.

22. Pawar, D., and Dhage, S. (2020). Multiclass covert speech classification using extreme learning machine. Biomed. Eng. Lett. 10, 217–226. doi: 10.1007/s13534-020-00152-x.

23. H. Li and F. Chen, "Classify imaginary mandarin tones with cortical EEG signals," INTERSPEECH, pp. 4896-4900, 2020.

24. D'Zmura, M., Deng, S., Lappas, T., Thorpe, S. & Srinivasan, R. Toward EEG sensing of imagined speech. In International Conference on Human-Computer Interaction, 40–48 (Springer, 2009).

25. Deng, S., Srinivasan, R., Lappas, T. & D'Zmura, M. EEG classification of imagined syllable rhythm using Hilbert spectrum methods. Journal of Neural Engineering 7, 046006 (2010).

26. Suppes, P., Lu, Z.-L. & Han, B. Brain wave recognition of words. Proceedings of the National Academy of Sciences 94, 14965–14969 (1997).

27. K. Brigham, B.V.K.V. Kumar, Subject identification from electroencephalogram (EEG) signals during imagined speech, 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS) (2010) 1–8.

28. Möller E, Schack B, Arnold M, Witte H. Instantaneous multivariate EEG coherence analysis by means of adaptive high-dimensional autoregressive models. J Neu-rosci Methods 2001;105(2):143–58.

29. Franaszczuk P, Bergey G, Kami΄ nski M. Analysis of mesial temporal seizure onset and propagation using the directed transfer function method. Electroencephalogr Clin Neurophysiol 1994;91(6):413–27.

30. Malihe Sabeti; Serajeddin Katebi; Reza Boostani (2009). Entropy and complexity measures for EEG signal classification of schizophrenic and control participants, 47(3), 263–274. doi: 10.1016/j.artmed.2009.03.003.

31. Follis JL, Lai D. Variability analysis of epileptic EEG using the maximal overlap discrete wavelet transform. Health Inf Sci Syst. 2020 Sep 15;8(1):26. doi: 10.1007/s13755-020-00118-4. PMID: 32999715; PMCID: PMC7492322.

32. g.tec Medical Engineering GmbH (2020). Unicorn Hybrid Black. Find it at: https://www.unicorn-bi.com/brain-interface-technology/(Accessed 12 19, 2022).

33. Jackson State University, Institutional Review Board (IRB), https://www.jsums.edu/researchcompliance/irb-protocol/. 2023.

34. Acharya, J.N.; Hani, A.; Cheek, J.; Thirumala, P.; Tsuchida, T.N. American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature. J. Clin. Neurophysiol. 2016, 33, 308–311.

35. Liu, Qing, Liangtao Yang, Zhilin Zhang, Hui Yang, Yi Zhang, and Jinglong Wu. "The Feature, Performance, and Prospect of Advanced Electrodes for Electroencephalogram." Biosensors 13, no. 1 (2023): 101.

36. A. Neumaier and T. Schneider, 2001: "Estimation of parameters and eigenmodes of multivariate autoregressive models". ACM Trans. Math. Softw., 27, 27C57.

37. Deng Wang, Duoqian Miao, Chen Xie, "Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection", Expert Systems with Applications, Volume 38, Issue 11, 2011, Pages 14314-14320.

38. I. Daubechies, Ten Lectures on Wavelets, SIAM, 1992, p. 194.

39. Tan, Y., Wang, J.: A support vector machine with a hybrid kernel and minimal vapnik-chervonenkis dimension. IEEE Transactions on knowledge and data engineering 16(4), 385–395 (2004).

40. Ghosh, S., Dasgupta, A., Swetapadma, A.: A study on support vector machine based linear and non-linear pattern classification. In: 2019 International Conference on Intelligent Sustainable Systems (ICISS), pp. 24–28 (2019).

41. Grauman, K., Darrell, T.: The pyramid match kernel: Discriminative classification with sets of image features. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 2, pp. 1458–1465 (2005).

42. Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamar´ıa, J., Fadhel, M.A., Al-Amidie, M., Farhan, L.: Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. Journal of Big Data 8(1), 1–74 (2021).

43. Vernon Lawhern; W. David Hairston; Kaleb McDowell; Marissa Westerfield; Kay Robbins (2012). Detection and classification of subject-generated artifacts in EEG signals using autoregressive models., 208(2), doi: 10.1016/j.jneumeth.2012.05.017.

44. Abdulghani, Mokhles M., Wilbur L. Walters, and Khalid H. Abed. 2023. "Imagined Speech Classification Using EEG and Deep Learning" *Bioengineering* 10, no. 6: 649. https://doi.org/10.3390/bioengineering10060649.

45. S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, QLD, Australia, 2015, pp. 992-996, doi: 10.1109/ICASSP.2015.7178118.

46. Lee, D.-Y.; Lee, M.; Lee, S.-W. Classification of Imagined Speech Using Siamese Neural Network. 2020, arXiv:2008.12487.

47. Vorontsova, D.; Menshikov, I.; Zubov, A.; Orlov, K.; Rikunov, P.; Zvereva, E.; Flitman, L.; Lanikin, A.; Sokolova, A.; Markov, S.; et al. Silent EEG-Speech Recognition Using Convolutional and Recurrent Neural Network with 85% Accuracy of 9 Words Classification. *Sensors* **2021**, *21*, 6744. https://doi.org/10.3390/s21206744.

48. Pramit, S.; Muhammad, A.-M.; Sidney, F. SPEAK YOUR MIND! Towards Imagined Speech Recognition with Hierarchical Deep Learning. 2019, arXiv:1904.04358.

49. Suhaimi, N.S.; Mountstephens, J.; Teo, J. A Dataset for Emotion Recognition Using Virtual Reality and EEG (DER-VREEG): Emotional State Classification Using Low-Cost Wearable VR-EEG Headsets. *Big Data Cogn. Comput.* **2022**, *6*, 16. https://doi.org/10.3390/bdcc6010016.

50. Hashem, H.A.; Abdulazeem, Y.; Labib, L.M.; Elhosseini, M.A.; Shehata, M. An Integrated Machine Learning-Based Brain Computer Interface to Classify Diverse Limb Motor Tasks: Explainable Model. *Sensors* 2023, *23*, 3171. https://doi.org/10.3390/s23063171.

51. F. Ahmed, H. Iqbal, A. Nouman, H. F. Maqbool, S. Zafar and M. K. Saleem, "A non Invasive Brain-Computer-Interface for Service Robotics," *2023 3rd International Conference on Artificial Intelligence (ICAI)*, Islamabad, Pakistan, 2023, pp. 142-147, doi: 10.1109/ICAI58407.2023.10136672.

52. J. Sijbers, J. Van Audekerke, M. Verhoye, A. Van der Linden, and D. Van Dyck, 'Reduction of ECG and gradient related artifacts in simultaneously recorded human EEG/MRI data,'' Magn. Reson. Imag., vol. 18, no. 7, pp. 881–886, 2000.

53. S. Tong, A. Bezerianos, J. Paul, Y. Zhu, and N. Thakor,''Removal of ECG interference from the EEG recordings in small animals using independent component analysis,'' J. Neurosci. Methods, vol. 108, no. 1, pp. 11–17, 2001.

54. C. Dai, J. Wang, J. Xie, W. Li, Y. Gong and Y. Li, "Removal of ECG Artifacts from EEG Using an Effective Recursive Least Square Notch Filter," in IEEE Access, vol. 7, pp. 158872-158880, 2019, doi: 10.1109/ACCESS.2019.2949842.