ORIGINAL ARTICLE

Multiclass covert speech classification using extreme learning machine

Dipti Pawar[1] · **Sudhir Dhage**[1]

Abstract

The objective of the proposed research is to classify electroencephalography (EEG) data of covert speech words. Six subjects were asked to perform covert speech tasks i.e mental repetition of four different words i.e 'left', 'right', 'up' and 'down'. Fifty trials for each word recorded for every subject. Kernel-based Extreme Learning Machine (kernel ELM) was used for multiclass and binary classification of EEG signals of covert speech words. We achieved a maximum multiclass and binary classification accuracy of (49.77%) and (85.57%) respectively. The kernel ELM achieves significantly higher accuracy compared to some of the most commonly used classification algorithms in Brain–Computer Interfaces (BCIs). Our findings suggested that covert speech EEG signals could be successfully classified using kernel ELM. This research involving the classification of covert speech words potentially leading to real-time silent speech BCI research.

Keywords Multiclass classification · Covert speech · Brain–computer interface (BCI) · Electroencephalography (EEG) · Wavelet transform

1 Introduction

Verbal communication is the natural way by which humans interact. However, individuals having neuromuscular impairments like brain injury, brainstem infarcts, stroke and advanced amyotrophic lateral sclerosis (ALS), their conditions prevent normal communication. In some circumstances e.g due to security aspects, it would be desirable to communicate using brain signals. In this context, the brain–computer interface (BCI) is promising to use as an alternative communication technology that uses brain signals emerging from distinct mental tasks to control an assistive device or convey a message.

The neuroimaging technique electroencephalography (EEG) is the most common choice for researchers due to its convenience of recording and noninvasive nature. Several previous studies of EEG signal classification based on motor imagery [1–4]. The various actions required by motor imagery such as limb and foot movement, which is unintuitive for the locked-in syndrome individuals. In contrary

EEG based covert speech, BCI is the imminent modality for thought detection and is the utmost direct communication channel for locked-in individuals [5]. In addition, the speech BCI has also applications in cognitive biometrics, synthetic telepathy, and silent speech communication [6, 7].

Current research of covert speech classification using EEG signals was focused majorly on vowels and syllables. DaSalla et. al. [8] compared speech imagery BCI accuracy of English vowels /a/ and /u/. Brigham and Kumar [9] carried out the investigation on EEG signals of syllables /ba/ and /ku/ for speech imagery classification. Both of these studies are offline and reported accuracy is above chance level i.e (60%) across participants.

In terms of the covert articulation of words; several reports based on meaningful and complete non-English words. Torres-Garcia et al. [10] investigated the possibility of classifying imagined articulation of five distinct Spanish words: 'arriba' (up), 'abajo' (down), 'izquierda' (left), 'derecha' (right), and 'seleccionar' (select). In [11] Wang et. al. reported classification performance for Chinese characters corresponds to 'left' and 'one' in English. Both of these cases reported the average offline classification accuracy of (67%) across participants. In [12] Serekhseh et. al reported binary and ternary classification of EEG based covert rehearsal of the words 'Yes', 'No' and unconstrained

✉ Dipti Pawar
 dipti.rohan.pawar@gmail.com

1 Sardar Patel Institute of Technology, Andheri(W), Mumbai,
 India

rest. The binary classification surpassing the chance level i.e (57.8%) and ternary classification surpassing the chance level i.e (39.1%). With the emergence of electrocorticogram (ECoG) several studies of covert speech in terms of vowels, syllables, and words [13–15] are reported with accuracies significantly higher than EEG. The ECoG is not suitable for practical BCI due to its invasive nature.

The most prior EEG based covert speech studies have focused on very small units of language such as vowels and syllables. Moreover, the research is also lagging in the field of multiclass classification of words considering the covert speech. As the human speech production always originates in the brain first, the effort is made to develop EEG based covert speech BCI for multiclass classification of simple and meaningful English words 'left', 'right', 'up', and 'down'.

The covert speech recognition from EEG signals is an interesting challenge for researchers. The EEG neural signals of meaningful words could include additional information other than the acoustic features. The neuronal electrical activity in the brain causes time-varying potential differences on the scalp, which range from a few microvolts to several hundreds of microvolts. The EEG is a registration of these potential differences, measured with electrodes placed on standard positions on the head. In general EEG signals are mixed with other signals and may be distorted by artifacts like electrooculography (EOG) and electromyography (EMG). The artifact-free EEG signals are acquired using the independent component analysis (ICA) which is a powerful approach of EEG analysis which is used for data decomposition and artifact removal from neuronal activity.

There is a need to address multiclass classification in covert speech recognition of words. The use of an optimized classifier that categorizes EEG signals into different classes is essential. These classifiers are divided into various categories like the neural network, linear classifiers, nonlinear classifiers, nearest neighbor classifier and the combination of classifiers [16]. For instance, different classification algorithms have been used in speech classification such as support vector machine (SVM)[8, 11], the k-nearest neighbor classifier (KNN) [9, 17], and linear discriminant analysis (LDA) [18, 19]. However, these classifiers are expected hurdles to achieve local optimum or overfitting solution since their underlying binary classification mechanism. Henceforth, the selection of appropriate classifier which can provide generalized and optimal solutions for multiclass covert speech recognition of words is the first challenging issue.

Furthermore, these classifiers have high computational cost during both training (calibration) and testing process. For the real-time covert speech recognition the accuracy, as well as computational speed, are crucial to be concerned. Thus how to obtain a balance between enhancing computational efficiency and maintaining high accuracy is the second challenging issue in covert speech recognition.

In general, the word production commences with conceptual preparation (semantic), lemma retrieval (lexical), phonological code retrieval and syllabification (phonetic) linguistic processes followed by phonetic encoding i.e the movements of language muscles for articulation [20, 21]. The most notable language brain regions for word production are the prefrontal cortex (stimulus-driven executive control), Wernicke's area i.e the left superior temporal gyrus (phonological code retrieval), right, and left inferior frontal gyrus i.e Broca's area (syllabification) and primary motor cortex (articulation) [22]. Indeed the same language areas are activated in the covert and overt speech articulation [23, 24]. However, in covert speech, the activity of the primary motor cortex is highly attenuated [25] and thus might be difficult to detect by EEG. Thus our third challenge is to determine whether the data acquired from these language regions (including and excluding primary motor cortex) is sufficient compared to whole brain areas without compromising the classification accuracy.

Extreme Learning Machine (ELM) [26] is based on a single hidden layer feedforward neural networks architecture. The first most advantage of using ELM is that input weights (connection between the input and hidden layer) do not require any tuning due to random assignment of weights, whereas the output weights (connection between hidden layer and output layer) are trained without layer by layer backpropagation tuning and as a result having significant reduction in training time. The second advantage is ELM has good generalization as the cost function includes the norm of output weights. With the consideration of these advantages, ELM can be one of the best choices to obtain the generalized and optimal solution for multiclass recognition and is a better solution for our priorly mentioned first challenging issue.

Moreover, randomly assigned input weights can reduce the computational cost of training. ELM consists of one hidden layer also decreases the computational speed of the classification process. In addition, the kernel function of ELM can enhance the stability and generalization performance of the algorithm [27]. The wavelet transform among the time-frequency analysis method sticks out with efficiency and algorithmic elegance due to its multiresolution analysis, where a source signal is processed at multiple levels by decomposing it into diverse resolutions. Hence, as for the priorly mentioned second challenge, the combination of kernel ELM algorithm and dwt features can obtain a good balance between computational efficiency and maintaining robustness with high accuracy.

In this paper, we propose an efficient method for covert speech classification of words from the EEG signal. It first extracts wavelet features of covert speech EEG signals and then employs kernel extreme learning machine for multiclass classification. We present an investigation of how the choice of kernel ELM classifier against the previously

applied techniques influences the multiclass classification accuracy in the problem of covert speech words recognition from EEG data. To adhere to our third challenging issue, the comparative classification experiments were performed not only on the EEG data obtained from the electrodes with respect to most significant language processing brain regions also on the whole brain area EEG data.

The paper is organized as follows: Sect. 2 explains about materials and methods used for the research covering the EEG dataset, preprocessing, feature extraction, and classification techniques. Section 3 carries results followed by a discussion in Sect. 4. Finally, we conclude in Sect. 5 by discussing the future scope.

2 Materials and methods

2.1 Participants

In this study 6 healthy fluently English speaking human subjects were recruited (2 women and 4 men; mean age: 27.6 ± 3.2; 5 right-handed). All participants reported normal vision and auditory function. None of the participants had any history of neurological disorders or serious health problems. Written informed consent was acquired from all the participants and all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research ethics committee.

2.2 Choice of covert speech words

The choice of words was carefully selected to maximize acoustic features variability, no. of syllables and semantic categories. Four English words 'Left', 'Right', 'Up', 'Down' used in this study for the covert speech task. The /L/,/R/,/U/ and /D/ in 'left', 'right', 'up' and 'down' have diverse manners and places of articulation.

2.3 Experimental protocol

The subjects were comfortably seated in an armchair and covert speech trials for words 'left', 'right', 'up', 'down' are acquired in pseudorandom order. The subjects were coached beforehand and rehearsed. Hence conceptual preparation and lemma selection are fulfilled before stimulus onset. Figure 1 depicts the experimental paradigm of the covert speech trial. At first, the non-emotive question appeared on screen i.e what is the direction of the arrow with one of the direction arrows (left, right, up, down). Subsequently, two beeps were given to maintain a consistent time cue to covertly speak the words. Approximately 2 s after the second beep, subject starts the mental repetition of answer which is among the words 'left', 'right', 'up', 'down' for 10 s. The duration of a single trial was 15 s.
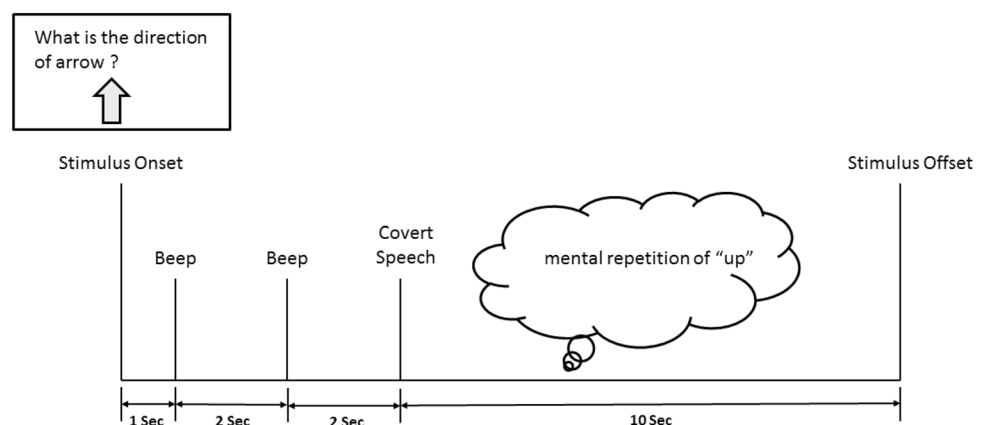
Trials are separated with a random rest within a time duration of 8 and 10 s, which prevents the subject to anticipate stimulus onset time. This ensures the inclusion of the remaining linguistic activities such as phonological code retrieval, syllabification, and covert articulation in a trial and perfect synchronization of system.

A single experimental session includes 100 trials divided equally into 'left', 'right', 'up' and 'down' trials. For each word total, 50 EEG responses were recorded for every participant during two separate sessions. E-Prime 2.0 software was used to design the experimental paradigm. The EEG data were recorded using 64 channel Neuroscan synamps 2 amplifier in a continuous mode at a sampling rate at 1000 Hz.

2.4 Signal processing

Prior to feature extraction, EEG signals are digital band filtered with cut-offs 0.5 Hz to 128 Hz. After that, the Independent component analysis (ICA) was applied for artifact removal from EEG signal based on the inference that EEG data obtained from multiple channels are linearly combined

Fig. 1 Schematic sequence of the experimental paradigm. The diagram shows capturing of a single covert trial of a particular word

What is the direction of arrow ?

Stimulus Onset

Stimulus Offset

Beep   Beep   Covert Speech   mental repetition of "up"

1 Sec   2 Sec   2 Sec   10 Sec

by temporally independent components. The adjust algorithm is used which is effective by means of removal of artifact components corresponds to EOG, blinks and eye movements [28].

2.5 Feature extraction

The effective analysis of nonstationary EEG signals is very difficult and araises the need of efficient features. Among the time-frequency analysis methods, wavelet transform stands out in algorithmic elegance and efficiency. Discrete wavelet transform features are using multiresolution analysis where a source signal is processed at multiple levels by decomposing it into diverse resolutions [29]. In general, a signal is decomposed into subbands, where each subband is distinct based on its characteristics. The Daubechies wavelet is used for multiresolution representation; since these wavelets are precisely supported with the maximal number of vanishing moments and the utmost phase for an inclined support width. Daubechies-4 (db4) was selected after various experiments with different levels of Daubechies wavelet such as db2, db4, and db6.

The entropy and energy computed at each decomposition level were used as features.

1. Energy: The energy for each frequency bands is calculated by squaring the wavelet coefficients as shown in Eq. 1.

$$ENG_i = \sum_{p=1}^{N} D_i(p)^2 \qquad (1)$$

2. Entropy: The entropy of the signal act as an eminent feature and is computed as in Eq. 2.

$$ENT_i = -\sum_{p=1}^{N} (D_i(p)^2) log(D_i(p)^2) \qquad (2)$$

where in $D_i(p)$, D stands for decomposition, i is the wavelet decomposition level and p is the number of wavelet coefficients varies from 1 to N.

Hence, for the whole brain area considering 62 electrodes (without consideration of ground and reference electrodes) from each 1200 trials 496 features were generated (62 electrodes × 4 decomposition levels × 2 features, i.e., Entropy and Energy).

2.6 Channel selection

The process of identification of language processing brain areas initiated in 1861, when french neurologist Paul Broca discovered that left inferior frontal gyrus (Broca's area) contains most expressive language centers [30]. A decade later,

Carl Wernicke investigated that left superior temporal gyrus (Wernicke's area) is responsible for language comprehension [31]. Other reserachers also carried out the studies to describe the neual areas of speech processing and their functional significance [20, 21, 32]. These studies have identified that the most prominent brain areas involved in perception and production of speech are prefrontal cortex, right inferior frontal gyrus, broca's L area, wernicke's area, and primary motor cortex.

This paper focuses on the classification of covert speech words. We investigated whether only the EEG channels from the brain regions that are responsible for language processing could be sufficient for the discrimination or not. The same language motors are activated in covert and overt speech. However, in the covert speech, the activity of the motor cortex is attenuated, we also investigated the classification performance by considering the language processing brain areas with inclusion and exclusion of primary motor cortex. We selected the channels with respect to three sets of brain areas (BA): BA1 consists of prefrontal cortex, Wernicke's area, right inferior frontal gyrus, Broca's area. BA2 consists of prefrontal cortex, wernicke's area, right inferior frontal gyrus, Broca's area and the primary motor cortex. The whole brain area is considered in BA3.

2.7 Classification

2.7.1 Extreme learning machine (ELM)

Extreme Learning Machine (ELM) [26] is a machine learning algorithm based on Single Layer Feedforward Network (SLFN). Figure 2 describes the structure of ELM with input layer nodes (P), hidden layer nodes (L), output layer nodes (V) and the $g(x)$ as the hidden layer activation function.

The feature vector x with dimension p is given as input to the input layer. The $g(x;w_i, b_i) = g(x.w_i + b_i)$ is the hidden node (index i) output where g represents the activation function, the w_i is the input weight vector between all input nodes and the ith hidden node, the bias of ith hidden node is denoted as b_i where $i = 1, 2, \ldots, L$.

The output vector of hidden layer is denoted as $h(x)$ which maps the input feature space (p-dimensional) to ELM feature space(L-dimensional).
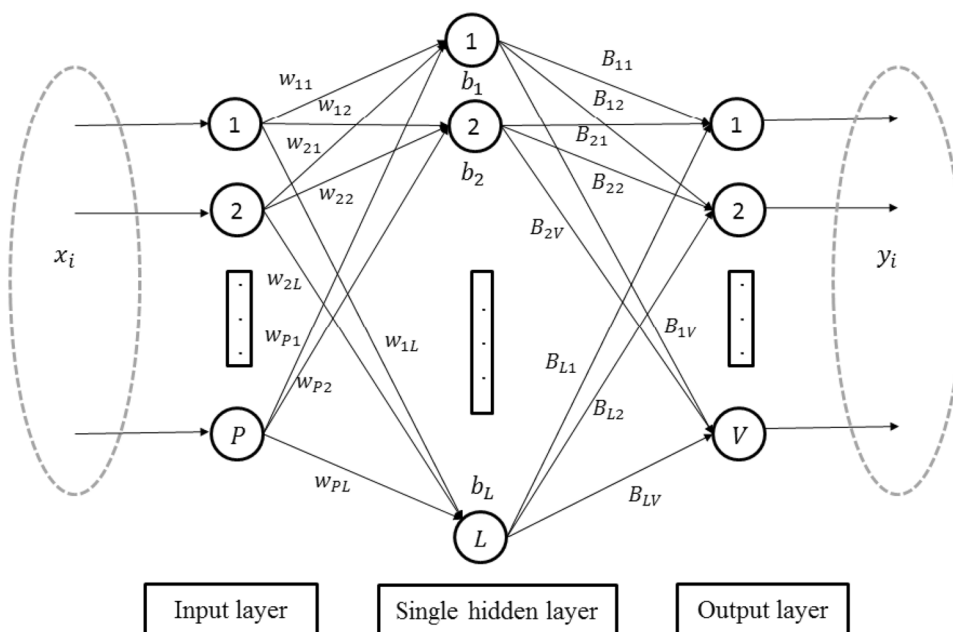
$$h(x) = [g(x;w_1, b_1), \ldots, g(x;w_L, b_L)] \qquad (3)$$

A generally effective feature mapping is the sigmoid function.

$$g(x;w_i, b_i) = \frac{1}{1 + exp[-(x.w_i + b_i)]} \qquad (4)$$

The number of output nodes V represents the covert speech words. $B_{i,j}$ is the output weight vector between the hidden node i and the output node j where $j = 1, 2, \ldots, V$.

Fig. 2 Structure of ELM

The value of output layer node with index j is computed as in Eq. 5

$$f_j(x) = \sum_{i=1}^{L} \beta_{i,j} \times g(x; w_i, b_i) \qquad (5)$$

Thus, for the input sample x, its output vector at the hidden layer can be written as:

$$f(x) = [f_1(x), \ldots, f_M(x)] = h(x)\beta \qquad (6)$$

where,

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_L \end{bmatrix} = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,V} \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,V} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{L,1} & \beta_{L,2} & \cdots & \beta_{L,V} \end{bmatrix} \qquad (7)$$

During the recognition process, the class label of test sample x can be determined as

$$label(x) = arg_{j=1\ldots V} max f_j(x) \qquad (8)$$

2.7.2 Training of ELM

ELM is a supervised machine learning algorithm requires N training sample pairs consists of feature vector x_k and its ground truth i.e binary class label vectors $t_k = [t_{k,1}, t_{k,2}, \ldots, t_{k,V}]$, where $k = 1, \ldots, N$. Each entry in the label vector indicates belonging of sample x_k to the

corresponding class. A matrix $T = [t_1, \cdot, t_N]^T$ is formed by all labels.

The ELM training parameters incorporate two parts such the input weights,biases $\{w_i, b_i\}_{i=1,2,\ldots,L}$ and the β as the output weight matrix as shown in Eq. 7.

The actual output vector is denoted by y_k for the input x_k. The linear representation can be formed by considering all training samples $\{x_k\}$ into 5

$$H\beta = Y \qquad (9)$$

where

$$H = \begin{bmatrix} h(x_1) \\ h(x_2) \\ \vdots \\ h(x_N) \end{bmatrix}$$

$$= \begin{bmatrix} g(x_1;w_1,b_1) & g(x_1;w_2,b_2) & \cdots & g(x_1;w_L,b_L) \\ g(x_2;w_1,b_1) & g(x_2;w_2,b_2) & \cdots & g(x_2;w_L,b_L) \\ \vdots & \vdots & \ddots & \vdots \\ g(x_N;w_1,b_1) & g(x_N;w_2,b_2) & \cdots & g(x_N;w_L,b_L) \end{bmatrix} \qquad (10)$$

and

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,V} \\ y_{2,1} & y_{2,1} & \cdots & y_{2,V} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N,1} & y_{N,1} & \cdots & y_{N,V} \end{bmatrix} \qquad (11)$$

For reduction in training error and the improvement in the generalization performance of neural net, the training error $||H\beta - T||$ and the output $||\beta||$ should get minimized at the same time i.e

$$Minimize : ||H\beta - T||, ||\beta|| \tag{12}$$

The least square solution of (12) based upon KKT conditions as in Eq. 13

$$\beta = H^T \left(\frac{I}{C} + HH^T \right)^{-1} T \tag{13}$$

If the number of training samples is less than the number of hidden neurons, Eq. 13 can be used; otherwise, Eq. 14 can be used.

$$\beta = \left(\frac{I}{C} + H^T H \right)^{-1} H^T T \tag{14}$$

where H denotes the hidden layer output matrix, T denotes the intended matrix of samples, I is unity matrix and C is used as the regulation coefficient to compromise between the smoothness of decision function and closeness of training data, hence improving the generalization performance.

2.7.3 Kernel-based ELM

The output function of ELM learning algorithm is:

$$f(x) = h(x)\beta \tag{15}$$

$$f(x) = h(x)H^T \left(\frac{1}{C} + HH^T \right)^{-1} T \tag{16}$$

The Mercer's conditions based ELM mapping for the unknown feature mapping $h(x)$ is as in Eq. 17, where kernel function of hidden neurons is denoted as $k(x_i, x_j)$.

$$h(x_i)h(x_j) = k(x_i, x_j) \tag{17}$$

Equation 18 represents the kernel based ELM output function $f(x)$.

$$f(x) = \begin{bmatrix} k(x, x_1) \\ k(x, x_2) \\ \vdots \\ k(x, x_{N_K}) \end{bmatrix} \left(\frac{1}{C} + k \right)^{-1} T \tag{18}$$

where

$$k = HH^T = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_{N_K}) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_{N_K}) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_{N_K}, x_1) & k(x_{N_K}, x_2) & \cdots & k(x_{N_K}, x_{N_K}) \end{bmatrix} \tag{19}$$

where N_k represents the number of training samples which are randomly selected for kernel ELM.

In this paper we used the Gaussian function as kernel k.

$$k(x_i, x_j) = exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2} \right) \tag{20}$$

where σ represents standard deviation (i.e spread) of the gaussian function. Figure 3 depicts the overall building of classification model including analytical steps.

3 Results

This research aims to classify the multiclass classification of covert speech words. Moreover, we were also interested in the brain area electrodes that play a significant role in Language and Speech processing. Therefore we consider the three different sets of brain areas BA(1–3) for multiclass and binary classification of covert speech words.

3.1 Multiclass classification results

Table 1 shows the multiclass classification of covert speech words with the mean accuracy, average standard deviation

Fig. 3 Analytical steps involved in building classification model
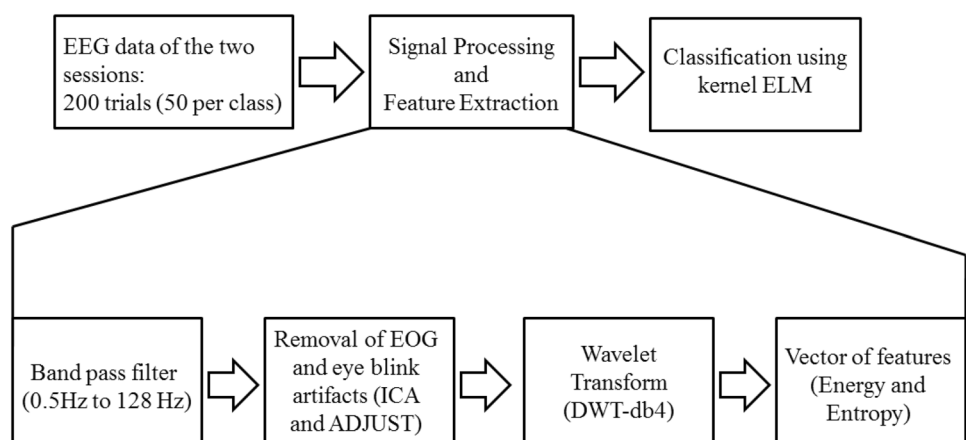
(SD), and kappa score for all the 6 subjects using kernel ELM where gaussian function is used as kernel function. Bold value depicts the highest scores. In the one versus all multiclass settings we achieved the highest classification mean accuracy (49.77%) for the subject S4. Among all the channel selections concerning brain areas (BA), the highest accuracy was obtained with BA3 and statistically significant difference was not found between the average accuracy for all subjects between the channel selection as per brain areas (BA1, BA2, BA3) by considering Tukey's posthoc test where $p < 0.05$ [33].

3.2 Comparison of different classification techniques

We compared the classification accuracies of kernel-based ELM (Gaussian-ELM) with the same wavelet features with ELM, linear support vector machine (Linear-SVM), SVM with the polynomial kernel (Polynomial-SVM) and regularized LDA, k-nearest neighbor (KNN), and navie bays (NB). The tuning parameters were chosen using same computation platform and tenfold cross-validation such as ELM (no. of hidden nodes L=78), kernel ELM (gaussian kernel spread sigma = 75), Linear-SVM (slack variable = 0.05), Polynomial-SVM (slack variable = 0.2, degree = 2), regularized LDA (gamma = 0.70, delta = 0.13), KNN (K=5).

We selected the best average cross-validated accuracy for each algorithm. Table 2 elaborates the multiclass classification accuracies for every subject using all these algorithms. Evidently in all channel selection concerning brain areas, Gaussian-ELM provides significant improvement in classification accuracy compared to all applied classifiers (Tukey's posthoc test where $p < 0.05$). Probable reasons for the variation in the performance of these algorithms are elaborated in the discussion section.

3.3 Binary classification results

We calculated the pairwise classification accuracies between each word for every subject. However, the article is focused

on multiclass classification, the highest accuracies among every class pair with a range between (80.17%) and (85.57%) are reported in Table 3.

The whole brain area channel selection (BA3) reported the highest mean classification accuracy, however, there are statistically significant differences were not found between averaged accuracy for all subjects between channel selection as per brain areas (BA1, BA2, BA3) by considering Tukey's posthoc test where $p < 0.05$. We obtained the utmost classification accuracy between 'left' and 'up' classes i.e (85.57%).

3.4 Computation time

The computational time was also compared among the Gaussian-ELM, ELM, Polynomial-SVM, linear SVM, and LDA under the MATLAB R2014a environment and hardware configuration as Intel(R) Core(TM) i3-3240 CPU (3.40 GHz) and 32 GB RAM. It should be noted that KNN and naive bays are not considered since they led to the lowest accuracy among applied classifiers. Table 4 lists the multiclass classification accuracy with computation time (calibration time) for each classifier. It is evident from the results that kernel ELM requires the least computation time and the highest classification accuracy compared to other classifiers.

4 Discussion

This section discusses the diverse factors which cause variation in observations and the potential reasons for differentiation in the performance of classification techniques.

4.1 Evaluation of channel selection and frequency band

As discussed earlier one of the aims of this paper is to determine whether the data acquired from the language regions of the brain is sufficient to compare to whole-brain areas without compromising the classification accuracy. Moreover, we

Table 1 Classification accuracies and kappa score in (%) for all channel selection with different brain areas using kernel ELM (Gaussian-ELM)

Subjects	BA1		BA2		BA3	
	Accuracy	kappa	Accuracy	kappa	Accuracy	kappa
S1	43.67	29.48	44.17	30.23	47.19	33.13
S2	41.50	27.20	42.20	28.42	45.89	31.81
S3	45.96	31.93	47.86	33.81	48.04	34.11
S4	46.23	32.21	48.18	34.38	**49.77**	**35.71**
S5	43.87	29.92	45.18	31.10	47.88	33.84
S6	44.89	30.87	42.82	28.91	48.77	34.81
AVG	44.35	30.27	45.07	31.14	47.92	33.90
SD	5.38	4.28	7.56	5.50	6.89	4.78

Table 2 Classification accuracies in (%) for all channel selection with different brain areas using different classification techniques

Subjects	BA1						BA2						BA3					
	ELM	Polynomial-SVM	Linear-SVM	LDA	KNN	NB	ELM	Polynomial-SVM	Linear-SVM	LDA	KNN	NB	ELM	Polynomial-SVM	Linear-SVM	LDA	KNN	NB
S1	33.17	38.50	29.37	21.15	18.20	15.21	34.27	39.12	30.32	22.43	18.76	17.60	37.28	40.32	32.23	24.23	20.34	18.34
S2	36.45	39.20	26.47	22.26	17.98	16.92	37.52	40.72	27.51	23.13	18.20	16.22	38.16	42.54	30.19	26.90	19.29	18.78
S3	37.80	36.50	30.13	23.00	16.45	13.30	38.16	38.50	31.23	22.67	16.40	16.30	40.27	40.23	33.13	24.80	17.15	16.78
S4	34.78	33.20	26.78	24.90	19.12	16.91	37.19	34.20	27.78	22.12	20.00	18.17	38.88	36.45	29.18	25.18	19.89	19.21
S5	35.93	35.89	30.97	21.60	18.56	16.29	36.91	36.81	29.17	24.19	19.45	17.78	38.23	37.88	32.50	27.13	19.14	19.10
S6	35.15	37.17	28.56	23.90	17.83	16.84	34.8	37.67	28.51	24.86	18.81	19.4	37.02	39.10	30.45	23.54	19.10	18.82
AVG	35.55	36.74	28.71	22.80	18.02	15.91	36.48	37.84	29.09	23.23	18.60	17.58	38.31	39.42	31.28	25.30	19.15	18.51
SD	9.45	7.36	8.91	9.29	10.34	8.23	5.38	6.36	9.67	10.39	6.40	11.32	9.67	5.76	10.12	8.17	7.39	9.55

Table 3 Binary classification accuracies in (%) for all channel selection with different brain areas using kernel ELM (Gaussian-ELM)

Subjects	BA1	BA2	BA3
S1	83.15	84.19	**85.57**
S2	80.18	82.18	83.88
S3	81.87	80.17	83.19
S4	82.67	83.24	85.39
S5	83.23	84.88	85.38
S6	81.88	83.56	83.89
AVG	82.16	83.04	84.55
SD	7.56	3.42	10.23

are also interested to determine the efficacy of attenuation of the primary motor cortex during channel selection on the classification accuracy.

As per the results in Tables 1 and 3, there is no statistically significant difference were found between the average accuracy for all subjects between the channel selection of diverse brain areas (BA1, BA2, BA3). Our findings suggest that prefrontal cortex, the left superior temporal gyrus (Wernicke's area), the right, and left inferior frontal gyrus (Broca's area) and primary motor cortex are the most prominent brain regions for covert speech recognition. Moreover, the primary motor cortex can be attenuated, since it is having a marginal penalty in classification accuracy.

Furthermore, we obtained the optimized results in the gamma frequency band (30–128 Hz) in comparison to other frequency bands such as alpha (8–13 Hz) and beta (13–30 Hz). However, several ECoG based studies were reported with optimized results for imagined speech classification in gamma frequency band [34, 35], this study might be the first report which reveals the significance of gamma frequency bands in EEG based covert speech classification of words.

4.2 Comparison of different classification techniques

We compared the performance of kernel ELM (Gaussian-ELM) among the previously reported classification algorithms using covert speech EEG data including SVM [8, 11], LDA [18, 19], KNN [9] and navie bays [36]. Among all the algorithms kernel ELM yields the highest classification accuracy and least computation time (Table 4). The various EEG studies using ELM reported better or similar classification accuracy with less computation time in comparison with other algorithms [37–40]. We achieved consistent results with the other research using ELM and better or similar classification accuracies for previous research of covert speech EEG data using different algorithms.

The reason for the superior performance of kernel ELM as it guarantees not only generalization but also provides fast learning speed. The hidden layer outputs calculation

Table 4 The Computation time in seconds and average classification accuracy in (%) among different Classifiers

Classifier	BA1		BA2		BA3	
	Computation time	Average accuracy	Computation time	Average accuracy	Computation time	Average accuracy
Gaussian-ELM	0.03	44.35	0.06	45.07	0.18	47.92
ELM	3.98	35.55	4.25	36.48	5.45	38.31
Polynomial-SVM	11.42	36.74	11.93	37.84	12.31	39.42
Linear-SVM	8.76	28.71	9.12	29.09	10.22	31.28
LDA	1.23	22.80	1.81	23.23	2.13	25.30

is avoided in kernel ELM since it is inherently encoded in the kernel matrix. The SVM with the polynomial kernel, ELM and linear SVM are the second, third and fourth best classifiers. SVM is not sensitive to the overfitting and dimensionality curse due to the margin maximization and regularisation term. The drawback of SVM is its sensitivity to noise and outliers and less suitability to model arbitrary classification boundaries. However, the polynomial kernel SVM has superior performance due to its ability to support nonlinearity. The ELM is an effective approach to minimize training error as well as the norm of output weights.

LDA achieved lower accuracy compared to kernel ELM. The major shortcoming of LDA is its linearity which restricts its performance with complex EEG data. Among the applied classification algorithms KNN and Naive bays led to the lowest accuracy. One possible reason is the sensitive nature of KNN to curse of dimensionality. The assumption of naive bays that features are independent which is often violates in EEG features.

5 Conclusion

In this paper, we investigated the multiclass classification of EEG-based covert speech signals of complete words. Experimental results have shown that with the Daubechies-dwt based features kernel ELM outperforms the most common classifiers in BCI in terms of classification accuracy and computational efficiency. It is evident from the results that the classification of EEG-based covert speech signals is possible using kernel ELM for the selected BCI users. Our findings suggested that data acquired from language processing regions are sufficient to discriminate covert speech signals. The future study includes the development of intelligent algorithms to classify a large number of words in real-time silent speech BCI research.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical statement All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Human participants informed consent The informed consent has been taken from all the involved human participants.

References

1. Hamedi M, Salleh SH, Noor AM. Electroencephalographic motor imagery brain connectivity analysis for BCI: a review. Neural Comput. 2016;28(6):999–1041.
2. Neuper C, Scherer R, Wriessnegger S, Pfurtscheller G. Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a brain–computer interface. Clin Neurophysiol. 2009;120(2):239–47.
3. Hwang HJ, Kwon K, Im CH. Neurofeedback-based motor imagery training for brain–computer interface (BCI). J Neurosci Methods. 2009;179(1):150–6.
4. Ahn M, Jun SC. Performance variation in motor imagery brain–computer interface: a brief review. J Neurosci Methods. 2015;243:103–10.
5. Chaudhary U, Xia B, Silvoni S, Cohen LG, Birbaumer N. Brain–computer interface-based communication in the completely locked-in state. PLoS Biol. 2017;15(1):e1002593.
6. Mohanchandra K, Saha S, Lingaraju GM. EEG based brain computer interface for speech communication: principles and applications. In: Brain–computer interfaces. Springer, Cham; 2015. p. 273–93.
7. Chaudhary U, Birbaumer N, Ramos-Murguialday A. Brain–computer interfaces for communication and rehabilitation. Nate Rev Neurol. 2016;12(9):513.
8. DaSalla CS, Kambara H, Sato M, Koike Y. Single-trial classification of vowel speech imagery using common spatial patterns. Neural Netw. 2009;22(9):1334–9.
9. Brigham K, Kumar BV. Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy. In: 4th IEEE international conference on bioinformatics and biomedical engineering 2010; p. 1–4.

10. Torres-Garcia AA, Reyes-Garcia CA, Villasenor-Pineda L, Garcia-Aguilar G. Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification. Expert Syst Appl. 2016;59:1–2.

11. Wang L, Zhang X, Zhong X, Zhang Y. Analysis and classification of speech imagery EEG for BCI. Biomed Signal Process Control. 2013;8(6):901–8.

12. Sereshkeh AR, Trott R, Bricout A, Chau T. Eeg classification of covert speech using regularized neural networks. IEEE/ACM Trans Audio Speech Lang Process. 2017;25(12):2292–300.

13. Brumberg JS, Krusienski DJ, Chakrabarti S, Gunduz A, Brunner P, Ritaccio AL, Schalk G. Spatio-temporal progression of cortical activity related to continuous overt and covert speech production in a reading task. PLoS ONE. 2016;11(11):e0166872.

14. Mugler EM, Patton JL, Flint RD, Wright ZA, Schuele SU, Rosenow J, Shih JJ, Krusienski DJ, Slutzky MW. Direct classification of all American English phonemes using signals from functional speech motor cortex. J Neural Eng. 2014;11(3):035015.

15. Martin S, Brunner P, Holdgraf C, Heinze HJ, Crone NE, Rieger J, Schalk G, Knight RT, Pasley BN. Decoding spectrotemporal features of overt and covert speech from the human cortex. Front Neuroeng. 2014;7:14.

16. Lotte F, Congedo M, Lecuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain–computer interfaces. J Neural Eng. 2007;4(2):R1.

17. Pawar D, Dhage SN. Recognition of unvoiced human utterances using brain-computer interface. In: Fourth IEEE international conference on image information processing (ICIIP). 2017; p. 1–4.

18. Kim J, Lee SK, Lee B. EEG classification in a single-trial basis for vowel speech perception using multivariate empirical mode decomposition. J Neural Eng. 2014;11(3):036010.

19. Deng S, Srinivasan R, Lappas T, D'Zmura M. EEG classification of imagined syllable rhythm using Hilbert spectrum methods. J Neural Eng. 2010;7(4):046006.

20. Indefrey P, Levelt WJ. The spatial and temporal signatures of word production components. Cognition. 2004;92(1–2):101–44.

21. Leuthardt E, Pei XM, Breshears J, Gaona C, Sharma M, Freudenburg Z, Barbour D, Schalk G. Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task. Front Hum Neurosci. 2012;6:99.

22. Jahangiri A, Sepulveda F. The relative contribution of high-gamma linguistic processing stages of word production, and motor imagery of articulation in class separability of covert speech tasks in EEG data. J Med Syst. 2019;43(2):20.

23. Numminen J, Curio G. Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. Neurosci Lett. 1999;272(1):29–32.

24. Chakrabarti S, Sandberg HM, Brumberg JS, Krusienski DJ. Progress in speech decoding from the electrocorticogram. Biomed Eng Lett. 2015;5(1):10–21.

25. Pei X, Leuthardt EC, Gaona CM, Brunner P, Wolpaw JR, Schalk G. Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. Neuroimage. 2011;54(4):2960–72.

26. Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing. 2006;70(1–3):489–501.

27. Huang GB, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification. IEEE Trans Syste Man Cybern Part B (Cybernetics). 2011;42(2):513–29.

28. Mognon A, Jovicich J, Bruzzone L, Buiatti M. ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology. 2011;48(2):229–40.

29. Patidar S, Pachori RB, Upadhyay A, Acharya UR. An integrated alcoholic index using tunable-Q wavelet transform based features extracted from EEG signals for diagnosis of alcoholism. Appl Soft Comput. 2017;50:71–8.

30. Broca P. Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauche du cerveau. Bull Soc Anthropol. 1861;2:235–8.

31. Wernicke C. Der aphasische symptomenkomplex. Berlin: Springer; 1974. p. 1–70.

32. Hickok G. Computational neuroanatomy of speech production. Nat Rev Neurosci. 2012;13(2):135–45.

33. Alpaydin E. Introduction to machine learning. Cambridge: MIT Press; 2014.

34. Martin S, Brunner P, Iturrate I, Millan JD, Schalk G, Knight RT, Pasley BN. Word pair classification during imagined speech using direct brain recordings. Sci Rep. 2016;6:25803.

35. Pei X, Hill J, Schalk G. Silent communication: toward using brain signals. IEEE Pulse. 2012;3(1):43–6.

36. Chi X, Hagedorn JB, Schoonover D, D'Zmura M. EEG-based discrimination of imagined speech phonemes. Int J Bioelectromagn. 2011;13(4):201–6.

37. Peng Y, Lu BL. Discriminative manifold extreme learning machine and applications to image and EEG signal classification. Neurocomputing. 2016;174:265–77.

38. Shi LC, Lu BL. EEG-based vigilance estimation using extreme learning machines. Neurocomputing. 2013;102:135–43.

39. Liang NY, Saratchandran P, Huang GB, Sundararajan N. Classification of mental tasks from EEG signals using extreme learning machine. Int J Neural Syst. 2006;16(01):29–38.

40. Yuan Q, Zhou W, Li S, Cai D. Epileptic EEG classification based on extreme learning machine and nonlinear features. Epilepsy Res. 2011;96(1–2):29–38.