Received: 17 June 2021 | Revised: 31 August 2021 | Accepted: 9 September 2021

DOI: 10.1002/ima.22655

WILEY

RESEARCH ARTICLE

Electroencephalography based imagined alphabets classification using spatial and time-domain features

Prabhakar Agarwal | **Sandeep Kumar**

Department of Electronics & Communication Engineering, National Institute of Technology Delhi, New Delhi, India

Correspondence
Sandeep Kumar, National Institute of Technology Delhi, Department of Electronics & Communication Engineering, A-7, Institutional Area, Near Satyawadi Raja Harish Chandra Hospital, New Delhi, 110040, Narela, India.
Email: sandeep@nitdelhi.ac.in

Abstract

Imagined speech is a neuro-paradigm that can provide an alternative communication channel for patients in a locked-in syndrome state. We have performed an experiment in which a 32 channel industry-standard electroencephalography (EEG) device was used to record 26 imagined English alphabets from 13 subjects. We denoised the imagined signals by discrete wavelet transform and extracted the spatial filters by common spatial pattern method, and time-domain features. Spatial features when classified with linear support vector machine, and time-domain features classified by random forest gave the best results. Alpha, beta, and theta bands could classify imagined alphabets better than other bands and had average classification accuracies of 88.59%, 87.39%, and 88.97%, respectively by using spatial features and 81.88%, 76.72%, and 79.25%, respectively, by time-domain features. The grand average accuracies of all the 26 alphabets in six EEG frequency bands was found to be 77.97% in a subject independent binary classification framework.

KEYWORDS

brain–computer interface, classification, electroencephalogram, imagined alphabets

1 | INTRODUCTION

Brain–computer interface (BCI) has been made possible by recording brain signals using various methods such as electroencephalography (EEG), electrocorticography, functional magnetic resonance imaging, functional near-infrared spectroscopy, magnetoencephalography, and so on. EEG measures the electrical activity of the brain in the form of potentials directly from the scalp in a non-invasive manner. It is relatively cheaper and a portable technique than its other counterparts. BCI, using EEG, is performed by mental control signals such as visual evoked potentials, motor imagery (MI), slow cortical potentials, and their hybrid equivalents.[1] For generating visual evoked potentials, the subject needs repetitive external stimuli which can be tiring and uncomfortable for them. MI[2] is limited by its degree of freedom and slow cortical potentials need more time for training. Speech imagery (SI) is a relatively new and more natural paradigm for BCI to provide a means for speech prosthesis for patients having medical conditions such as advanced amyotrophic lateral sclerosis (ALS), traumatic brain injury, and pseudo-coma in which the patients cannot speak but are mentally active. BCI can also be useful for other non-medical applications like gaming, autonomous driving, cognitive biometrics, prediction of web advertisements,[3] data mining,[4] and so on. Silent/imagined/covert/inner speech or verbal thoughts are some of the terms used to describe the silent articulation of one's thought linguistically, though its behavioral, physiological, and cerebral correlates with overt speech are still debatable.[5] Classifying imagined speech using EEG is still a difficult task may be due to its low signal to noise ratio (SNR).

Most of the work done related to EEG based SI can be sub-divided into classifying vowels,[6–11] syllables,[12,13]

short words,[10,14–21] long words,[10] shapes,[22] and a combination of characters, digits, and objects.[23] Some classical and noteworthy EEG based imagined speech work is now described. DaSalla et al.[6] applied common spatial patterns (CSP) to distinguish between two tasks taken pairwise from imagined English vowels /a/, /u/, and rest interval. They designed spatial filters such that their variance carried the most distinctive information. The accuracies using variances as features and support vector machine (SVM) as a classifier varied between 68% and 78%. Matsumoto and Hori,[7] classified silent speech of five Japanese vowels, /a/,/i/,/e/,/u/, and /o/. They chose suitable electrodes and the time duration of vowel signals and then calculated CSP for each vowel combination. They achieved classification accuracies of 77% and 79% by using relevance and support vector machines with Gaussian kernel, that is, (RVM-G), and (SVM-G), respectively. Torres-Garcia et al.[18] classified five imagined words /down/, /up/, /left/, /right/, and /select/ by extracting features such as minimum, maximum, standard deviation (SD), average, and relative wavelet energy in different frequency bands decomposed by discrete wavelet transform (DWT). Then relevant channels were selected by using Mamdani fuzzy inference system. They classified the features using a random forest (RF) classifier and achieved average accuracies of 68.18% and 70.33% taking the selected and all channels, respectively. Qureshi et al.[15] classified five imagined words /back/, /go/, /right/, /left/, and /stop/. They used covariance-based, maximum linear cross correlation-based, and phase-only information feature extraction techniques on channels from Broca's and Wernicke's areas of the brain. They achieved maximum classification accuracies of 40.3% and 87.97% for five and two classes, respectively, using an extreme learning machine. Bakshali et al.[21] classified four imagined words /gnaw/, /knew/, /pot/, and /pat/ in a pairwise manner. They obtained the distance between the correntropy spectral density matrices from these words. By using the k nearest neighbor (kNN) algorithm they classified these distances and achieved a maximum accuracy of 90.25%. Agarwal and Kumar[14] classified three imagined words /sos/, /stop/, and /medicine/. They formed covariance connectivity matrices and the top three eigenvalues of these matrices contained more than 95% of the variance. These features were classified with many classifiers like decision tree (DT), naïve Bayes, SVM, kNN, and RF, and maximum accuracy of 76.4% was achieved.

1.1 | Research gaps

The existing work on the MI based classification of imagined alphabets and then converting to text as given in Zhang et al.[24] have five degrees of freedom. So, the subject has to repeat certain imagined actions to complete 26 alphabets. Due to these repetitive actions, the time consumed for the display of each alphabet is large.[25] BCI with SI can overcome the above challenges. Currently, alphabet imagery has been largely confined to vowels and few characters only. Moreover, EEG based BCI has been evaluated on individual subject performance metrics and subject independence issues have been poorly addressed till now. The proposed work overcomes the abovementioned issues and brings new insights into the basic understanding of imagined alphabets in time and spatial domain with the use of simple classifiers.

1.2 | Contributions

In the existing work on SI, each subject has repeated the same task several times due to which temporal artifacts may have been detected[26] by the classifiers, thus obtaining high classification accuracy. In this paper, we have done the experiment for classifying EEG based imagined alphabets from A to Z. We have accounted for subjective, linguistic, and articulatory variability in our imagined speech data by choosing subjects, the first tongue of whom were of different language, that is, either Hindi, Odiya, or Bengali language. However, English was known to all of them. In our work, each subject imagined an alphabet only once, which is the limiting case from the practical aspects of the BCI. The important contributions of the paper are mentioned below:

a. We have conducted an experiment in which a noninvasive modality (EEG) is used to collect 26 imagined English alphabet signals and classify them. This analysis is performed in individual EEG frequency bands.
b. We have extracted spatial and time-domain statistical features of the imagined alphabets and classified them against their corresponding no task intervals by using RF and linear SVM (SVM-L) in a pairwise manner. The features of all the subjects were concatenated to obtain a cross-subject BCI.
c. We have shown that spatial features in combination with a linear classifier, and statistical features together with an RF classifier give good classification results. Also, we have suggested a suitable frequency band and type of feature extraction technique which can provide the maximum classification accuracy for each alphabet. Alpha, beta, and theta bands could classify imagined alphabets with more accuracy.

To the best of the author's knowledge, the electrical activity of a complete set of imagined English alphabets has not been studied and analyzed before in a subject

independent, SI-BCI framework. The paper consists of four sections. Section 2 explains the experimental design and methodology used in this work. In Section 3 results and discussion is presented followed by a conclusion and future work in Section 4.

2 | MATERIALS AND METHODS

2.1 | Procedure of the experiment

Thirteen subjects (1 female, 12 males) aged between 26 and 37 years participated in the experiment. All the subjects were right-handed and did not possess any mental or physical illness. They were given training before the start of the experiment. The subjects gave their written consent to participate in the experiment and their identity was maintained confidential. The permission for conducting the experiment was approved by the competent authorities of the institute. At the onset of the experiment, the subjects sat on a chair comfortably and a computer monitor was placed at a comfortable distance. The participants were instructed to remain in a calm and relaxed state and avoid any eye or muscle movement. Initially, a fixation cross was shown to them on the computer screen for 5 s before each trial to focus. Then, an alphabet was displayed on the screen for 2 s after which the display was turned off. Then, another 2 s time was given in which the subject silently imagined that alphabet only once without any lip movement. Then, a 3 s rest interval (no task) was given and the experiment continued for other alphabets. The EEG of the subjects was recorded simultaneously. All the 26 English alphabets were presented to the subjects in random order. A trial consisted of imagining an English alphabet only once by each subject. So each alphabet had a total of 13 trials by 13 subjects. A total of 338 trials were recorded for all the alphabets considering all subjects. The recording session was completed in 2 days. The EEG was recorded by a 32-channel MOBITA wireless physiological data acquisition device developed by BIOPAC Systems, Inc., Goleta, California. This device uses water based electrodes and the data was recorded from the locations Fp1, Fp2, AFz, F3, F7, Fz, F4, F8, FC1, FC5, FC2, FC6, T7, C3, C4, Cz, T8, CP1, CP5, CP2, CP6, P3, P7, Pz, P4, P8, POz, O1, O2, and Oz with left mastoid A1 as reference. This placement of electrodes is according to the 10–20 International system.[27] The sampling frequency (F_s) was 250 Hz and the impedance of electrodes was below 20 kΩ. The recorded data was transmitted by Wi-Fi and was stored in another computer nearby. Figure 1 shows the workflow of the experiment.

2.2 | Preprocessing

All the EEG signals were visually inspected and the recording of three subjects was removed as they were excessively corrupted by noise. Then we performed common average referencing (CAR) on the raw EEG data to cancel out the effect of any common signal which may have occurred during recording from all the channels simultaneously. In CAR, from each sample of the data, we subtracted the mean value of all the samples of the data from all the channels according to Equation (1).

$$x_j^{\mathrm{CAR}} = x_j - \sum_{i=1}^{M} x_i / M \tag{1}$$

Here, x_j^{CAR} is the CAR signal, x_j is the signal of jth channel, and M is the total number of channels. In this manner, SNR will be improved by performing CAR.[28] Then, a Butterworth notch filter of second order was applied to remove the power line interference at 50 and 100 Hz frequencies, respectively. The EEG signals are non-stationary and have many artifacts apart from power line interference like electromyographic movement, white noise, and so on, in which the power spectral density of noise sources can fully overlap with that of EEG signal. These noises cannot be removed by conventional bandpass filtering. DWT is a powerful tool for denoising the non-stationary signals due to the shrinkage property of its coefficients. The process of denoising is implemented in four steps, namely, obtaining DWT coefficients (approximation coefficients $[A_j]$ and detailed coefficients $[D_j]$ in each level j), denoising, reconstruction, and calculation of performance measures. Six different methods given in MATLAB, namely empirical Bayes (EB),[29] block James–Stein (BJS),[30] false discovery rate (FDR),[31] Stein's unbiased risk estimate (sure), minimax estimation (ME), and universal threshold (UT)[32] were used to obtain threshold values for denoising. These threshold values modify the detailed coefficients D_j at each level. The smaller D_j are considered as noise and are removed. The larger D_j values are retained as they contribute to signal values.We calculated two performance measures, that is, mean square error (MSE) and SNR between the raw EEG signal $x(n)$ and the reconstructed signal $x_d(n)$ from Equations (2) and (3), respectively, as conformed by Hazarika and Dasgupta.[33]

$$\mathrm{MSE} = \frac{1}{N}\sum_{n=1}^{N} (x(n) - x_d(n))^2 \tag{2}$$

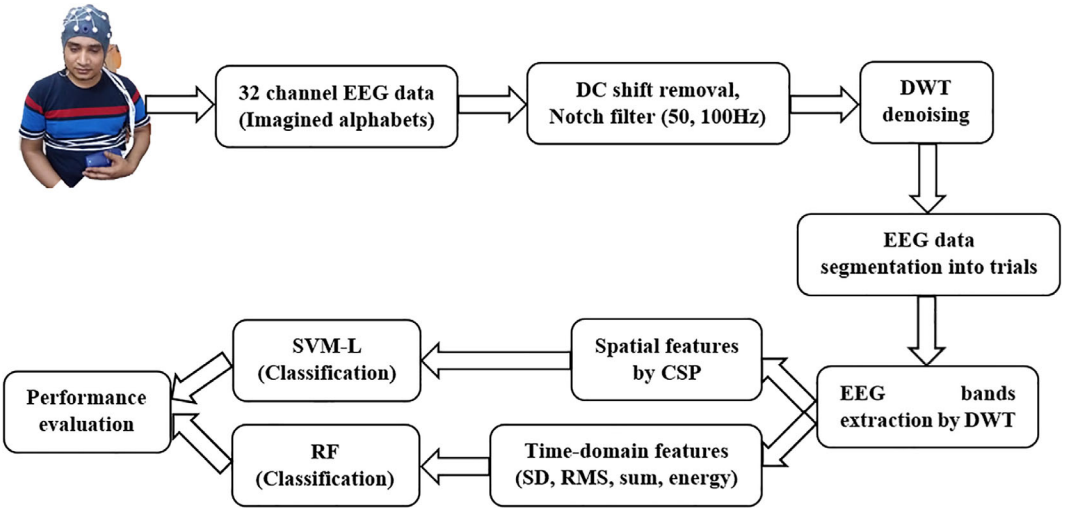

FIGURE 1 Imagined alphabets analysis framework

$$\mathrm{SNR} = 20 \times \log \frac{\sum_{n=1}^{N} x_d(n)^2}{\sum_{n=1}^{N} (x(n) - x_d(n))^2} \tag{3}$$

The shape of the mother wavelet affects the output of the wavelet transform. We investigated 25 different wavelet functions[34] namely Haar, Daubechies (db2, db4, db6, db8, db10), symlet (sym1, sym2, sym4, sym6, sym8, sym10), coiflets (coif1, coif2, coif4, coif5), biorthogonal (bior1.1, bior2.2, bior3.9, bior6.8), reverse biorthogonal (rbio1.1, rbio3.9, rbio5.5, rbio6.8) and discrete meyer (dmeyer). We chose eight decomposition levels and obtained a resolution of 0.5 Hz at eighth level for $F_s = 250$ Hz. We calculated MSE and SNR from each of the six different denoising methods, per subject, for each wavelet. Table 1 shows the average results for 10 subjects. Dmeyer wavelet gave minimum MSE when denoised with "EB" and maximum SNR with "sure" method. The minimum MSE and maximum SNR characterize the best wavelet function as it indicates better filtering action.[33,35] We chose the denoising method as "sure" based upon majority voting of individual subject's performance. Nine subjects out of ten outperformed with "sure" method individually, and hence this method along with "dmeyer" wavelet was chosen for denoising.

2.3 | Feature extraction

Feature extraction is an essential step in machine learning. It reduces the dimensionality of data and prevents overfitting. The denoised EEG signals were decomposed into five levels by dmeyer wavelet. The EEG bands corresponding to wavelet coefficients at different levels were highgamma (HG) (62.5–125 Hz, D_1), gamma (31.25–62.5 Hz, D_2), beta (15.62–31.25 Hz, D_3), alpha (7.81–15.62 Hz, D_4), theta (3.90–7.81 Hz, D_5), and delta (0–3.9 Hz, A_5). These frequency components are natural frequencies of the brain. From each of the frequency bands, we extracted spatial and time-domain features and used them for classification as discussed below.

2.3.1 | Spatial features by CSP method

CSP[36,37] method can extract spatial features from time series signals like EEG. Let imagined alphabet /a/ followed by /no task/ interval are two groups of EEG time series denoted by matrix $\left(X_g^i\right)_{N\times T}$, where, $g = 1$ and 2 are the two groups, respectively, i denotes a particular trial in each group, N is the number of EEG channels (30), and T denotes the number of samples (500) in each trial. One EEG trial of any group is a distribution of T points in N dimensional space. Using the CSP method, spatial filters are designed, which are common to both groups. When EEG data is projected onto these CSPs, one obtains another time series whose variance is maximally suited to distinguish between the two groups. We implemented the CSP algorithm, the summarized steps of which are described below:

a. Calculate normalized averaged covariance matrices for each group, that is, $\overline{C_1}$ and $\overline{C_2}$ as shown in

TABLE 1 MSE and SNR of the denoised signals using different wavelet thresholding techniques

Wavelets/ thresholding method	MSE						SNR (dB)					
	BJS	Sure	ME	UT	EB	FDR	BJS	Sure	ME	UT	EB	FDR
db1 or Haar	967.50	63.19	313.41	490.19	71.62	93.70	73.79	77.52	69.83	67.93	76.90	75.64
db2	90.50	48.97	197.64	301.76	52.91	66.65	78.14	82.01	73.32	71.27	81.25	80.25
db4	81.20	50.05	175.45	266.65	49.45	60.96	80.10	83.99	74.93	72.82	83.03	82.11
db6	75.10	49.41	166.39	254.18	45.57	56.45	81.58	85.11	75.85	73.67	84.19	83.38
db8	71.93	49.91	161.38	247.15	44.41	53.76	81.34	84.82	75.80	73.63	83.84	83.11
db10	69.84	49.71	157.75	241.80	43.36	52.83	81.33	84.38	75.74	73.59	83.67	82.95
sym1	130.08	63.19	313.41	490.19	71.62	93.70	73.79	77.52	69.83	67.93	76.90	75.64
sym2	90.50	48.97	197.64	301.76	52.91	66.65	78.14	82.01	73.31	71.27	81.25	80.25
sym4	81.35	50.70	175.35	266.17	49.78	61.70	80.32	83.95	75.04	72.93	83.09	82.17
sym6	76.61	50.95	166.68	253.44	47.75	58.28	81.12	83.89	75.62	73.32	83.74	82.92
sym8	72.88	50.62	160.92	245.29	45.32	54.84	81.57	84.18	75.91	73.75	84.11	83.34
sym10	69.91	50.05	156.61	239.24	43.83	53.02	81.87	85.10	76.14	73.95	84.36	83.60
coif1	91.22	52.14	197.12	299.81	54.70	68.58	78.20	82.17	73.36	71.33	81.25	80.16
coif2	80.74	51.27	174.39	264.59	49.73	61.13	80.77	84.28	75.30	73.17	83.48	82.59
coif4	72.27	50.35	159.42	242.34	45.64	55.42	81.81	85.22	76.07	73.90	84.33	83.55
coif5	69.87	50.44	155.42	236.96	44.49	53.88	82.01	85.24	76.25	74.06	84.48	83.73
bior1.1	130.09	63.19	313.41	490.19	71.68	93.70	73.79	77.52	69.83	67.93	76.90	75.64
bior2.2	74.71	50.68	160.93	241.35	49.35	58.46	80.68	83.30	75.68	73.64	82.74	82.27
bior3.9	56.36	43.64	120.01	176.31	41.84	50.51	84.02	85.73	79.32	77.27	85.01	84.77
bior6.8	71.93	51.22	158.92	241.41	46.14	55.44	82.02	85.07	76.35	74.18	84.34	83.70
dmeyer	62.21	47.80	141.86	217.87	**39.80**	48.20	82.74	**85.79**	76.78	74.57	84.99	84.24
rbio1.1	130.09	63.19	313.41	490.49	71.62	93.70	73.79	77.52	69.83	67.93	76.90	75.57
rbio3.9	173.19	106.92	450.83	709.92	126.34	167.14	75.45	80.66	70.39	68.40	78.88	76.29
rbio5.5	116.22	75.34	243.45	368.65	74.53	90.59	79.02	81.86	74.40	72.37	81.09	80.36
rbio6.8	132.74	89.96	303.32	470.81	81.31	100.84	79.20	82.20	73.70	71.57	81.99	80.99

Note: The bold values indicates best performance numerical values.

Equation (4), n denotes the total number of trials, that is, 10 for each group.

$$\overline{C_g} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_g^i \left(X_g^i\right)^\top}{\text{trace}\left(X_g^i \left(X_g^i\right)^\top\right)}, \top \text{ denotes the transpose operation.} \tag{4}$$

b. Factorize the composite normalized covariance matrix (C_c) into its eigenvectors (E_c) and eigenvalues (λ_c) as shown in Equation (5).

$$C_c = \overline{C_1} + \overline{C_2} = E_c \lambda_c E_c^\top \tag{5}$$

Calculate linear whitening transform (W) as shown in Equation (6).

$$W = {\lambda_c}^{-1/2} {E_c}^\top \tag{6}$$

Apply W to the two normalized averaged covariance matrices such that $S_1 = W\overline{C_1}W^\top$ and $S_2 = W\overline{C_2}W^\top$. S_1 and S_2 have the same eigenvectors (U).

c. Define a projection matrix $P^\top = \left(U^\top W\right)$, here columns of ${P^\top}^{-1}$ are the CSPs. These can be seen as time-invariant EEG source distribution vectors. Choose only first m and last m columns of $P^\top$ and form a new projection matrix ${P^\top}_{2m}$.

d. Decompose each EEG trial to a new time series as calculated from Equation (7).

$$Z_g^i = {P^\top}_{2m}\left(\frac{X_g^i}{\sqrt{\text{trace}\left(X_g^i \left(X_g^i\right)^\top\right)}}\right) \tag{7}$$

This new series is optimized for discriminating between the two groups. From each Z_g^i time series, we calculated the logarithm of the variance of the number of samples in each row. In this manner, we obtained a feature vector of the 2 m dimension for each trial. In this work, we have selected m as five as it gave the maximum accuracy.

A task consisted of pairwise classification of two EEG trials containing an imagined alphabet and its consecutive no task interval of the first 2 s duration, having a dimension of 30×500 each. In this manner, 26 tasks for each subject were formed consisting of 26 English imagined alphabets followed by 26 consecutive no task intervals. For each task, it's both trials were combined, and this was done for all 10 subjects. So, 10 tasks for each alphabet called as a group were formed, for example, 10 tasks for imagined alphabet /a/ versus /no task/. In the case of spatial features vector set, for each trial, the log variance feature vector was of dimension 10×1 when m was selected as 5. One feature set consists of 10 tasks for each similar alphabet, and all these values were concatenated to form a combined total feature vector set of dimension 200×1 to obtain a cross-subject BCI. In this manner, we had 26 feature sets for 26 alphabets each having a size of 200×1 and these were classified using SVM-L.

2.3.2 | Time-domain features

From the pre-processed EEG trials we also calculated time-domain statistical features like SD $\left(\sqrt{\frac{\sum_{i=1}^{T}\left(x_i(n)-\overline{x(n)}\right)^2}{n}}\right)$, sum $(\sum_{i=1}^{T} x_i(n))$, root mean square (RMS) $\left(\sqrt{\frac{\sum_{i=1}^{T} x_i(n)^2}{n}}\right)$, and energy $(\sum_{i=1}^{T} x_i(n)^2)$.[11,23] Here, $x(n)$, $x_i(n)$, $\overline{x(n)}$, and T denotes pre-processed EEG trial, ith sample of the trial, mean, and total number of samples in the trial, respectively. For time-domain features, one feature set consisted of 10 tasks for each similar alphabet. A single channel corresponds to four values for SD, RMS, sum, and energy, each of dimensions 1×1. For all 30 channels, the feature vector per subject after concatenation of these four values was of size 30×4. For each alphabet consisting of 10 tasks, the final feature vector set corresponded to a dimension of 600×4. In this manner, we had 26 feature sets for 26 alphabets having dimensions of 600×4 each and these were classified using RF.

2.4 | Classification of imagined alphabets using SVM and RF classifier

2.4.1 | Support vector machine

SVM is a type of supervised learning algorithm that can tune multiple parameters, minimize the generalized error,[38] and can support both types of linear and non-linear classification. For binary classification and linearly separable data, SVM-L finds an optimal decision boundary known as hyperplane to separate the feature set of the two classes. The data samples closest to the separating hyperplane are called support vectors. SVM maximizes the distance between separating hyperplane and support vectors. Non-linear SVM maps the non-linear data into a higher dimension in which the data is linearly separable using a kernel function. An example of kernel used for non-linear SVM is radial basis function $K(x_j, x_k) = e^{-\gamma \left\| x_j - x_k \right\|^2}$, where K is a kernel function of support vectors x_j and pattern vectors x_k,[6] and γ is a non-negative parameter that controls the spread of the kernel. We have applied SVM-L to classify spatial feature vectors.

2.4.2 | Random forest

RF[39] is a set of DT in which each tree makes its prediction for the most suitable class. Each tree in RF learns by drawing a random sample of data points with replacement, known as bootstrapping, and the final prediction is made by averaging the prediction of each DT (Bagging).[40] So, each tree may have a high variance w.r.t. its training data set, but the overall forest will have lower variance and bias. In a DT, only a part of all the features is used for splitting at each node. Gini index[40] was used as a feature selection metric at each node of the DT. We have used RF to classify time-domain statistical features.

3 | RESULTS AND DISCUSSION

For each alphabet, spatial feature vector is of dimension 200×1. These feature vector sets were calculated separately in HG, gamma, beta, alpha, theta, and delta bands. The classification results using spatial features are shown in Figure 2. For each alphabet, accuracies were calculated in six EEG bands. The average accuracies of all the alphabets in HG, gamma, beta, alpha, theta, and delta bands are 70.96%, 81.47%, 87.39%, 88.59%, 88.97%, and 50.45%, respectively. The individual average accuracies of alphabets /a/, /b/, /c/, /d/, /e/, /f/, /g/, /h/, /i/, /j/, /k/, /l/, /m/,

/n/, /o/, /p/, /q/, /r/, /s/, /t/, /u/, /v/, /w/, /x/, /y/, and /z/ in all the bands are 81.67%, 81.75%, 75.28%, 83.06%, 81.11%, 83.89%, 81.39%, 81.39%, 79.45%, 73.06%, 71.11%, 74.72%, 78.89%, 74.72%, 67.78%, 66.95%, 66.67%, 77.78%, 75.55%, 75.83%, 84.72%, 81.95%, 83.05%, 81.39%, 81.39%, and 82.78%, respectively. It can be seen from the results that top five maximum average accuracies in various bands are from the alphabets /u/, /f/, /d/, /w/, and /z/, respectively. Alphabet /u/ has been classified with the highest average accuracy of 84.72% among all other alphabets. When we consider only alpha, beta, and theta bands there is an increase of classification accuracy. The individual average accuracies of alphabets /a/, /b/, /c/, /d/, /e/, /f/, /g/, /h/, /i/, /j/, /k/, /l/, /m/, /n/, /o/, /p/, /q/, /r/, /s/, /t/, /u/, /v/, /w/, /x/, /y/, and /z/ in these three bands are 93.89%, 97.94%, 85%, 97.78%, 93.89%, 95%, 90.55%, 88.89%, 88.34%, 83.89%, 78.33%, 82.78%, 85%, 81.67%, 76.11%, 76.11%, 80.55%, 91.11%, 88.89%, 82.22%, 95%, 91.11%, 90%, 95.56%, 94.44%, and 92.22%, respectively. The average accuracy of all the alphabets in alpha, beta, and theta bands is 88.32%. Grand average accuracy for all the alphabets taking into consideration all the frequency bands is 77.97%. From these results, it is evident that alpha, beta, and theta waves can be used to classify imagined alphabets better than other frequencies. It has been shown[41] that alpha activity is related to idle arousal of the brain system, beta activity shifts the system to attention state, and gamma-band corresponds to synchronization. Other studies have also shown that alpha, beta, and theta bands have significant features related to EEG based BCI.[42,43] Borghini et al.[44] have summarized the findings related to changes in physiological signals (EEG, electrooculography, and heart rate) while performing heavy mental workload[45] such as driving an aircraft or a car. Researchers have shown that power in theta band increases and in alpha band, it decreases while doing heavy mental tasks. During mental fatigue, an increase in power is associated in theta, delta, and alpha bands. In another study by Babiloni et al.,[46] both EEG and functional magnetic resonance imaging signals were used to show that parietal, premotor, and prefrontal areas of the brain were involved in MI tasks. These results compliment the findings of the current study as the accuracy is higher in specific EEG bands during mental tasks. We tried different classifiers for classifying spatial features, but SVM-L gave the best results. So, it can be observed that by using CSP, we obtain a new time series whose variances are maximally suited to distinguish between the two groups and so a linear classifier gave good accuracy compared to other nonlinear classifiers.

The time feature vector for each alphabet is of dimension 600 × 4. These feature vector sets were calculated separately in HG, gamma, beta, alpha, theta, and delta bands. The classification results using time-domain features are shown in Figure 3. For each alphabet, accuracies were calculated in six EEG bands. The average accuracies in HG, gamma, beta, alpha, theta, and delta bands are 70.29%, 72.81%, 76.72%, 81.88%, 79.25%, and 61.82%, respectively. The individual average accuracies of alphabets /a/, /b/, /c/, /d/, /e/, /f/, /g/, /h/, /i/, /j/, /k/, /l/, /m/, /n/, /o/, /p/, /q/, /r/, /s/, /t/, /u/, /v/, /w/, /x/, /y/, and /z/ in all the bands are 74.31%, 72.84%, 72.15%, 72.72%, 71.82%, 68.6%, 75.25%, 71.32%, 72.66%, 75.98%, 73.57%, 74.10%, 73.9%, 75.87%, 75.94%, 77.37%, 77.87%, 72.73%, 73.63%, 74.42%, 73.57%, 71.55%, 71.17%, 73.37%,

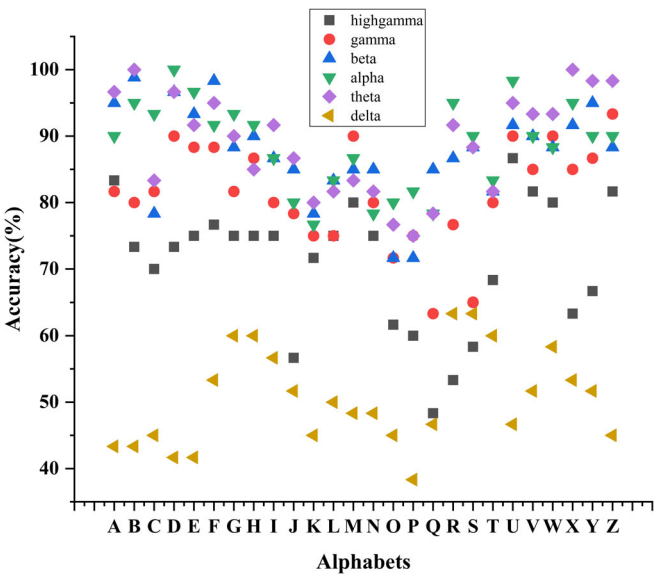

FIGURE 2 Accuracy (%) of alphabets using SVM-L classifier

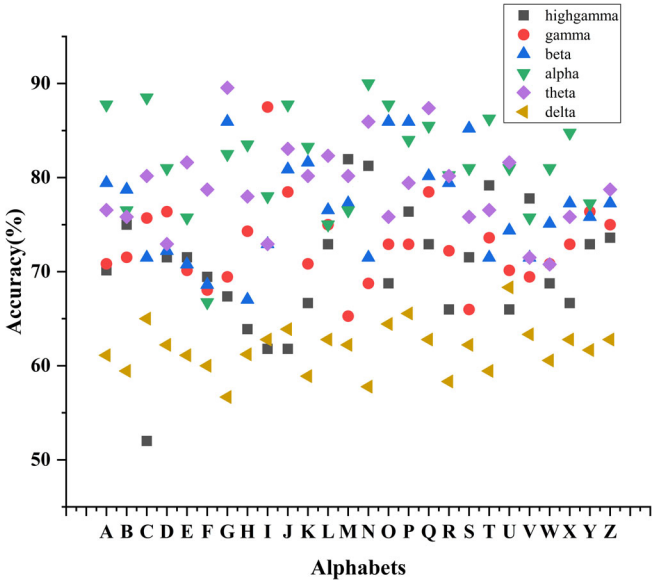

FIGURE 3 Accuracy (%) of alphabets using RF classifier

75.48%, and 76.48%, respectively. When we consider only alpha, beta, and theta bands there is an increase of classification accuracy. The individual average accuracies of alphabets /a/, /b/, /c/, /d/, /e/, /f/, /g/, /h/, /i/, /j/, /k/, /l/, /m/, /n/, /o/, /p/, /q/, /r/, /s/, /t/, /u/, /v/, /w/, /x/, /y/, and /z/ in these three bands are 81.25%, 77.02%, 80.06%, 75.39%, 76.04%, 71.36%, 86%, 76.18%, 74.63%, 83.9%, 81.68%, 77.96%, 77.98%, 82.48%, 83.17%, 83.13%, 84.35%, 79.95%, 80.69%, 78.1%, 79%, 72.92%, 75.63%, 79.28%, 80.64%, and 82.50%, respectively. The average accuracy of all the alphabets in alpha, beta, and theta bands is 79.28%. Grand average accuracy for all the alphabets taking into consideration all the frequency bands is 73.79%. These results also prove the effectiveness of alpha, beta, and theta waves to contain a good amount of EEG based mental work features.[41–44] The accuracy of time-domain features, in general was less than spatial features. Linear classifiers were also used for the time-domain feature set, but they gave poor results. We used 70% and 30% of the data for training and testing, respectively, with five-fold cross-validations. The analysis of our work was performed offline in MATLAB environment.

Table 2 shows the frequency band and the type of feature extraction technique suitable for classifying each alphabet with maximum accuracy. Out of 26 alphabets, 19 alphabets were classified with high accuracies when spatial features were used and only 7 alphabets were better classified with time-domain features. These results indicate superior performance of spatial features over time-domain features. Apart from its advantages, the CSP algorithm has some limitations. CSP is sensitive to noise due to which outliers and noise may be enhanced[47] and it may not give desired results. This may be the reason for the poor classification performance of CSP in some alphabets. The bootstrapping and bagging properties of the RF classifier can also lead to enhanced accuracy for time-domain features in some of the alphabets. As a matter of fact that the behavioral and physiological correlations of imagined speech signals are still not completely understood,[5] unlike overt speech. This work is a small step in bridging the gap between covert and overt speech.

We have also compared some recent work on the classification of imagined vowels and alphabets with the proposed work as shown in Table 3. This comparison is based on various signal processing techniques, classifiers, how data was recorded, and performance evaluation. DaSalla et al.[6] proposed a control scheme based on two imagined vowels /a/ and /u/. Three subjects were shown the visual cue of open mouth and rounded lips for imagining /a/ and /u/, respectively, and 50 trials were recorded for each task. The accuracy was calculated for each subject separately. In another work on vowel imagery by Matsumoto and Hori,[7] the subjects were provided with the audio cue of the vowels. Each vowel was imagined 13 times per subject. The classification accuracy was calculated by all five subjects individually. In Nguyen et al.[10] visual cue was shown for 7 s during which the subjects had to imagine the vowels and a beep sound was given regularly after 1 s. Hundred such trials were recorded individually by 15 subjects and accuracy was calculated in a subject dependent framework. Kumar et al.[23] classified imagined EEG signals of characters,

TABLE 2 Frequency bands and feature extraction technique for the maximum classified value of each alphabet

Alphabets	Frequency band	Feature extraction technique	Alphabets	Frequency band	Feature extraction technique
A	Theta	Spatial	N	Alpha	Time
B	Theta	Spatial	O	Alpha	Time
C	Alpha	Spatial	P	Beta	Time
D	Alpha	Spatial	Q	Theta	Time
E	Alpha	Spatial	R	Alpha	Spatial
F	Beta	Spatial	S	Alpha	Spatial
G	Alpha	Spatial	T	Alpha	Time
H	Alpha	Spatial	U	Alpha	Spatial
I	Theta	Spatial	V	Theta	Spatial
J	Alpha	Time	W	Theta	Spatial
K	Alpha	Time	X	Theta	Spatial
L	Alpha/beta	Spatial	Y	Theta	Spatial
M	Gamma	Spatial	Z	Theta	Spatial

TABLE 3 EEG based imagined vowels and alphabets recognition methods and comparison

Year [refs.]	Imagined data	Subjects, trials/ samples per task per subject	Feature extraction	Classifier	Average accuracy	Subject independent
2009[6]	Vowels (a, u) and rest	3, 50 trials	CSP	SVM-L	68%–78%	No
2014[7]	5 Japanese vowels (a, e, i, o, u)	5, 52 trials	Adaptive collection, CSP	RVM-G, SVM-G	77%–79%	No
2017[10]	Vowels (a, i, u)	15, 100 trials	CSP, tangent vectors between covariance matrices	RVM-G	49%	No
2018[23]	10 characters, 10 digits, and 10 objects	23, 10 s continuously	SD, RMS, sum, energy	3 parallel RF	67.03%	Yes
2018[24]	5 MI tasks for 26 alphabets	10, 28 000 samples	CNN and RNN	XG Boost	95.53% for 5 MI tasks	No
2020[11]	Vowels (a, u) and rest	3, 50 trials	CSP, DWT, time-domain	SVM-L, SVM-G,RF	52%–76%	No
2021[25]	5 MI tasks for 26 alphabets	1, 28 000 samples	CNN	XG Boost	92.87% for 5 MI tasks	No
2021[48]	26 alphabets	10, 1346 samples	Power spectrum by Morlet wavelet	CNN	95.2%	No
Proposed work	26 alphabets	10, 1 trial (500 samples)	CSP, time-domain	SVM-L, RF	88.32%	Yes

digits, and objects. The subjects imagined each item for 10 s continuously after the corresponding visual cue was shown to them. Zhang et al.[24] classified five MI actions (eyes closed, left hand, right hand, both hands, and both feet). Recurrent neural network (RNN) and convolutional neural networks (CNN) were used to exploit the advantages from temporal and spatial features of the two networks, respectively, and the features were classified using extreme gradient boost (XG Boost). The authors reported an average accuracy of 95.53%. The classification decision was used by a graphical user interface (GUI) to display the imagined alphabets. The subject had to perform five MI tasks for the display of each alphabet which may be uncomfortable for ALS patients and also this method is time consuming. In a similar work on the same dataset, Kumar et al.[25] has evaluated the display time on GUI of all the alphabets employing MI using CNN and XG Boost classifier. The display time of the alphabets varied between 10.9 and 36.82 s which is a considerable amount of time. The MI tasks have limited degrees of freedom (maximum four or five) and hence MI modality cannot be a feasible solution for displaying all the alphabets. Agarwal et al.[11] classified silent imagined vowels /a/, /u/, and rest by using spatial, time, and joint time-frequency feature extraction techniques. They achieved the best classification accuracy by using CSP in different frequency bands by using an RF classifier. In recent work, Ullah and Halim[48] classified 26 imagined EEG signal alphabets using CNN. They calculated powers in various EEG frequency bands using Morlet wavelet transform and then used CNN for classification. The aforementioned work was subject dependent and the authors recorded 1346 samples for one alphabet with an F_s of 128 Hz, which means a time duration of approximately 10.51 s. The existing BCI models in the literature are based on individual subjects and hence the performance metrics have been calculated subject wise.[6,7,10,11,24,25,48] The proposed work is subject-independent and presents a more generalized BCI. The imagined trials of all the subjects were mixed and then the accuracy was calculated. In some of the work in the literature, the subjects performed the same action for a longer duration of time[23–25,48] or performed a large amount of trials[10] which can lead to high accuracy. We have performed the classification from very limited data as the subjects were made to imagine a word only once in a trial. The work in the literature for classifying all the imagined alphabets is very limited. Either the subject had to perform MI[24,25] which is limited in degrees of freedom and time consuming, or the BCI framework was subject dependent.[48] The existing work is also limited to the

classification of vowels or few characters.[6,7,10,11,23] So, considering the aforementioned factors the proposed work is better than the state-of-the-art methods for classification of imagined alphabets.

4 | CONCLUSION AND FUTURE WORK

We have performed an experiment to brain-map the imagined alphabets signals using EEG. This work will create a basic biophysical understanding of neural correlates for all the alphabets in English literature. The various activation states of the brain corresponding to different frequency sub-bands have the effect of varying accuracies in classification. Cross-subject BCI is still a challenging task with a lot of scope for improvement because the neural mechanism of imagined speech remains poorly defined due to the lack of its clear timing and subjective nature. In this work, the results showed that alpha, beta, and theta frequencies can classify imagined alphabets more efficiently than other bands. Our work may be an essential contribution in understanding the articulatory behavior of each English alphabet, which may provide a lot many degrees of freedom for patients who are in a complete locked-in syndrome state. The future work will be to induce imagined signals with the facial expressions, sentiments, gestures, etc. to improve the performance metrics and build a robust subject independent BCI. Also, the roadmap for recognition of complete imagined sentences will be established from which the medical community would be able to serve the speech impaired patients in a better way.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Prabhakar Agarwal https://orcid.org/0000-0003-4818-9351
Sandeep Kumar https://orcid.org/0000-0001-9922-2663

REFERENCES

1. Ramadan RA, Vasilakos AV. Brain computer interface: control signals review. *Neurocomputing*. 2017;223:26-44. https://doi.org/10.1016/j.neucom.2016.10.024
2. Mahamune R, Laskar SH. Classification of the four-class motor imagery signals using continuous wavelet transform filter bank-based two-dimensional images. *Int J Imaging Syst Technol*. 2021. https://doi.org/10.1002/ima.22593
3. Asad M, Halim Z, Waqas M, Tu S. An in-ad contents-based viewability prediction framework using artificial intelligence for web ads. *Artif Intell Rev*. 2021;54:5095-5125. https://doi.org/10.1007/s10462-021-10013-3
4. Halim Z, Ali O, Ghufran Khan M. On the efficient representation of datasets as graphs to mine maximal frequent itemsets. *IEEE Trans Knowl Data Eng*. 2021;33(4):1674-1691. https://doi.org/10.1109/TKDE.2019.2945573
5. Perrone-Bertolotti M, Rapin L, Lachaux J-P, Baciu M, Lœvenbruck H. What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behav Brain Res*. 2014;261:220-239. https://doi.org/10.1016/j.bbr.2013.12.034
6. DaSalla CS, Kambara H, Sato M, Koike Y. Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Netw*. 2009;22(9):1334-1339. https://doi.org/10.1016/j.neunet.2009.05.008
7. Matsumoto M, Hori J. Classification of silent speech using support vector machine and relevance vector machine. *Appl Soft Comput*. 2014;20:95-102. https://doi.org/10.1016/j.asoc.2013.10.023
8. Kim J, Lee S-K, 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. https://doi.org/10.1088/1741-2560/11/3/036010
9. Min B, Kim J, Park H, Lee B. Vowel imagery decoding toward silent speech BCI using extreme learning machine with electroencephalogram. *Biomed Res Int*. 2016;2016:2618265. https://doi.org/10.1155/2016/2618265
10. Nguyen CH, Karavas GK, Artemiadis P. Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features. *J Neural Eng*. 2018;15(1):016002. https://doi.org/10.1088/1741-2552/aa8235
11. Agarwal P, Kale RK, Kumar M, Kumar S Silent speech classification based upon various feature extraction methods. In: 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN); 2020:16–20.
12. D'Zmura M, Deng S, Lappas T, Thorpe S, Srinivasan R. Toward EEG sensing of imagined speech. In: Jacko JA, ed. *Human-Computer Interaction. New Trends*. Springer; 2009.
13. Brigham K, Kumar BVKV. Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy. In: 2010 4th International Conference on Bioinformatics and Biomedical Engineering; 2010:1–4.
14. Agarwal P, Kumar S. Transforming imagined thoughts into speech using a covariance-based subset selection method. *Indian J Pure Appl Phys*. 2021;59(03):180-183. http://nopr.niscair.res.in/handle/123456789/56517
15. Qureshi MNI, Min B, Park H-J, Cho D, Choi W, Lee B. Multiclass classification of word imagination speech with hybrid connectivity features. *IEEE Trans Biomed Eng*. 2018;65(10):2168-2177. https://doi.org/10.1109/TBME.2017.2786251
16. González-Castañeda EF, Torres-García AA, Reyes-García CA, Villaseñor-Pineda L. Sonification and textification: proposing methods for classifying unspoken words from EEG signals. *Biomed Signal Process Control*. 2017;37:82-91. https://doi.org/10.1016/j.bspc.2016.10.012

17. Sereshkeh AR, Trott R, Bricout A, Chau T. EEG classification of covert speech using regularized neural networks. *IEEE/ACM Trans Audio Speech Language Process*. 2017;25(12):2292-2300. https://doi.org/10.1109/TASLP.2017.2758164
18. Torres-García AA, Reyes-García CA, Villaseñor-Pineda L, García-Aguilar G. Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification. *Expert Syst Appl*. 2016;59:1-12. https://doi.org/10.1016/j.eswa.2016.04.011
19. Mohanchandra K, Saha S. A communication paradigm using subvocalized speech: translating brain signals into speech. *Augment Hum Res*. 2016;1(1):1-14. https://doi.org/10.1007/s41133-016-0001-z
20. Wang L, Zhang X, Zhong X, Zhang Y. Analysis and classification of speech imagery EEG for BCI. *Biomed Signal Process Control*. 2013;8(6):901-908. https://doi.org/10.1016/j.bspc.2013.07.011
21. Bakhshali MA, Khademi M, Ebrahimi-Moghadam A, Moghimi S. EEG signal classification of imagined speech based on Riemannian distance of correntropy spectral density. *Biomed Signal Process Control*. 2020;59:101899. https://doi.org/10.1016/j.bspc.2020.101899
22. Esfahani ET, Sundararajan V. Classification of primitive shapes using brain–computer interfaces. *Comput-Aided Des*. 2012;44(10):1011-1019. https://doi.org/10.1016/j.cad.2011.04.008
23. Kumar P, Saini R, Roy PP, Sahu PK, Dogra DP. Envisioned speech recognition using EEG sensors. *Pers Ubiquitous Comput*. 2018;22(1):185-199. https://doi.org/10.1007/s00779-017-1083-4
24. Zhang X, Yao L, Sheng QZ, Kanhere SS, Gu T, Zhang D. Converting your thoughts to texts: enabling brain typing via deep feature learning of EEG signals. In: 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom); 2018:1–10.
25. Kumar S, Verma PR, Bharti M, Agarwal P. A CNN based graphical user interface controlled by imagined movements. *Int J Syst Assur Eng Manag*. 2021. https://doi.org/10.1007/s13198-021-01096-w
26. Porbadnigk A, Wester M, Calliess J, Schultz T. EEG-based speech recognition- impact of temporal effects. In: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing - BIOSIGNALS, (BIOSTEC 2009); 2009:376–381.
27. Klem GH, Lüders HO, Jasper HH, Elger C. The ten-twenty electrode system of the The International Federation of Clinical Neurophysiology. *Electroencephalogr Clin Neurophysiol Suppl*. 1999;52:3-6. https://doi.org/10.1080/00029238.1961.11080571
28. Moctezuma LA, Torres-García AA, Villaseñor-Pineda L, Carrillo M. Subjects identification using EEG-recorded imagined speech. *Expert Syst Appl*. 2019;118:201-208. https://doi.org/10.1016/j.eswa.2018.10.004
29. Johnstone IM, Silverman BW. Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences. *Ann Stat*. 2004;32(4):1594-1649. https://doi.org/10.1214/009053604000000030
30. Cai TT. On block thresholding in wavelet regression: adaptivity, block size, and threshold level. *Stat Sin*. 2002;12(4):1241-1273.
31. Abramovich F, Benjamini Y, Donoho DL, Johnstone IM. Adapting to unknown sparsity by controlling the false discovery rate. *Ann Stat*. 2006;34(2):584-653. https://doi.org/10.1214/009053606000000074
32. Donoho DL. De-noising by soft-thresholding. *IEEE Trans Inf Theory*. 1995;41(3):613-627. https://doi.org/10.1109/18.382009
33. Hazarika J, Dasgupta R. Neural correlates of action video game experience in a visuospatial working memory task. *Neural Comput Appl*. 2020;32:3431-3440. https://doi.org/10.1007/s00521-018-3713-9
34. Daubechies I. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics; 1992.
35. Hazarika J, Kant P, Dasgupta R, Laskar SH. Neural modulation in action video game players during inhibitory control function: an EEG study using discrete wavelet transform. *Biomed Signal Process Control*. 2018;45:144-150. https://doi.org/10.1016/j.bspc.2018.05.023
36. Ramoser H, Muller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans Rehabil Eng*. 2000;8(4):441-446. https://doi.org/10.1109/86.895946
37. Müller-Gerking J, Pfurtscheller G, Flyvbjerg H. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin Neurophysiol*. 1999;110(5):787-798. https://doi.org/10.1016/S1388-2457(98)00038-8
38. Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. *Mach Learn*. 2002;46(1):131-159. https://doi.org/10.1023/A:1012450327387
39. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. https://doi.org/10.1023/A:1010933404324
40. Nandi A, Ahmed H. *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machine*. 1st ed. Wiley; 2019.
41. Wróbel A. Beta activity: a carrier for visual attention. *Acta Neurobiol Exp*. 2000;60(2):247-260.
42. Li Y, Long J, Yu T, et al. An EEG-based BCI system for 2-D cursor control by combining mu/beta rhythm and P300 potential. *IEEE Trans Biomed Eng*. 2010;57(10):2495-2505. https://doi.org/10.1109/TBME.2010.2055564
43. Babiloni C, Carducci F, Cincotti F, et al. Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study. *Neuroimage*. 1999;10(6):658-665. https://doi.org/10.1006/nimg.1999.0504
44. Borghini G, Astolfi L, Vecchiato G, Mattia D, Babiloni F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci Biobehav Rev*. 2014;44:58-75. https://doi.org/10.1016/j.neubiorev.2012.10.003
45. Halim Z, Rehan M. On identification of driving-induced stress using electroencephalogram signals: a framework based on wearable safety-critical scheme and machine learning. *Inf Fusion*. 2020;53:66-79. https://doi.org/10.1016/j.inffus.2019.06.006
46. Babiloni F, Cincotti F, Babiloni C, et al. Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. *Neuroimage*. 2005;24(1):118-131. https://doi.org/10.1016/j.neuroimage.2004.09.036

47. Ghanbar KD, Rezaii TY, Farzamnia A, Saad I. Correlation-based common spatial pattern (CCSP): a novel extension of CSP for classification of motor imagery signal. *PLoS One*. 2021; 16(3):e0248511. https://doi.org/10.1371/journal.pone.0248511
48. Ullah S, Halim Z. Imagined character recognition through EEG signals using deep convolutional neural network. *Med Biol Eng Comput*. 2021;59(5):1167-1183. https://doi.org/10.1007/s11517-021-02368-0

How to cite this article: Agarwal P, Kumar S. Electroencephalography based imagined alphabets classification using spatial and time-domain features. *Int J Imaging Syst Technol*. 2022;32(1): 111-122. doi:10.1002/ima.22655