EEG-based envisioned speech recognition system using spectral graph wavelet transform

Dhanhanjay Pachori
Department of ECE
IIIT, Nagpur
Nagpur, India
bt22ece024@iiitn.ac.in

Luciano Caroprese
Engineering and Geology Department
University G. d'Annunzio of Chieti-Pescara
Pescara, Italy
luciano.caroprese@unich.it

M. Sabarimalai Manikandan
Department of EE
IIT Palakkad
Palakkad, India
msm@iitpkd.ac.in

***Abstract*—This paper presents a new framework for the recognition of envisioned speech from electroencephalogram (EEG) signal. The proposed framework consists of preprocessing for dividing the signals into blocks, spectral graph wavelet transform (SGWT), feature smoothing using the moving average filter, and classifiers (extra trees (ET), random forest (RF), and k-nearest neighbors (KNN)) for identifying three tasks such as digits, characters, and objects. The performance of the proposed method is evaluated on a publicly available database which consists of three classification tasks namely, digits, characters, and objects tasks. The SGWT-based method with ET classifier had the highest accuracies of 92.4%, 92.7%, and 92.3% for digits, character, and objects tasks, respectively that outperforms other SGWT-based methods using the RF and KNN. Evaluation results show that the class-wise accuracies are better than the state-of-the-art methodologies. The proposed EEG-based framework for recognition of envisioned speech can enable seamless brain-computer interfaces (BCIs) for communication of people having speech impairments and can control the devices using envisioned speech in human-machine interaction (HCI) applications.**

***Index Terms*—Electroencephalogram (EEG) signals, envisioned speech recognition, spectral graph wavelet transform (SGWT), moving average SGWT (MASGWT), signal processing, machine learning.**

I. Introduction

In today's world, there has been an exponential growth of electronic devices that are present everywhere [1]. Because of this, an intuitive interaction between humans and technology is needed [2]. This need has stimulated developments in human-computer interfaces (HCIs), notably through speech and gesture recognition technologies, which leverage signal processing and machine learning (ML) methods for real-time interpretation [3]. Despite of these advancements in HCIs, there are various challenges [4]. People with speech impairments, conditions like locked-in syndrome, or situations demanding high levels of privacy often find these interaction methods insufficient [4]. This highlights the need to investigate alternative strategies which utilize signal processing and ML techniques to address these limitations effectively. Electroencephalogram (EEG) signals are very useful for recognition of envisioned speech [5].

Previously, various researchers have proposed methodologies for the envisioned speech recognition by using several signal processing techniques and ML algorithms on EEG signals. Tripathi [6] has recognized the envisioned speech using the EEG rhythms (δ, θ, α, β, and γ rhythms) derived from low-pass, high-pass, and band-pass filters. The authors in [7] have proposed multivariate dynamic mode decomposition for recognition of imagined speech by using multichannel EEG signals. The researchers in [8] have proposed multivariate swarm sparse decomposition-based joint time-frequency analysis for the recognition of imagined speech by using EEG signals. Naik et al. [9] proposed a methodology based on convolutional neural network (CNN), gated recurrent unit (GRU), and generative adversarial network (GAN) for perceiving the human imagination. Mishra and Bhavsar [10] proposed Siamese models (both online and offline) and CNN-based approach for the recognition of envisioned speech.

This paper introduces a novel framework for the recognition of imagined speech using spectral graph wavelet transform (SGWT). The Fig. 1 represents the block diagram of the proposed framework. Initially, the EEG signals are preprocessed and segmented. Further, SGWT is applied on the EEG signals to obtain SGWT features. The obtained SGWT features are smoothened using a moving average filter to obtain moving average SGWT (MASGWT) features. The obtained MASGWT features are given as input to various ML classifiers namely, extra trees (ET), random forest (RF), and k-nearest neighbours (KNN) in order to recognize the envisioned speech. The proposed methodology performs better than the pre-existing state-of-the-art methodologies.

The rest of this paper is organized as follows. Section II contains the description of the publicly available dataset used for our proposed methodology. Section III contains the information regarding the proposed methodology. Section IV describes the results obtained and their discussion. The Section V provides the conclusion of the studied framework for recognition of the envisioned speech.

II. Database description

A publicly available envisioned speech database that contains EEG signals from 23 participants ($15 - 40$ years old) has been considered for the proposed methodology [11]. The EEG signals were recorded with 14 channels namely, AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, and O2 [11]. The representation of the electrode placement has been

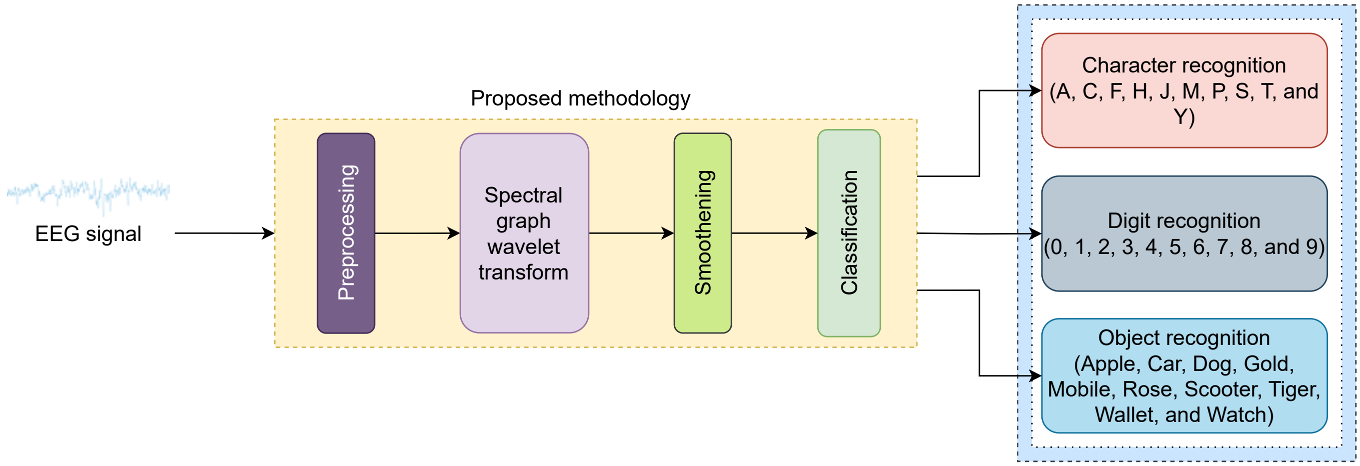

EEG signal
Proposed methodology
Preprocessing
Spectral graph wavelet transform
Smoothening
Classification
Character recognition (A, C, F, H, J, M, P, S, T, and Y)
Digit recognition (0, 1, 2, 3, 4, 5, 6, 7, 8, and 9)
Object recognition (Apple, Car, Dog, Gold, Mobile, Rose, Scooter, Tiger, Wallet, and Watch)

Fig. 1: Block diagram of the proposed framework for the classification of characters, digits, and objects task.

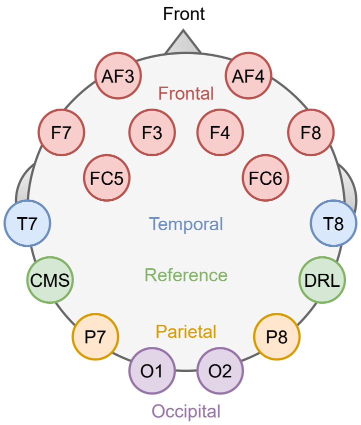


Fig. 2: Representation of the electrode placement.

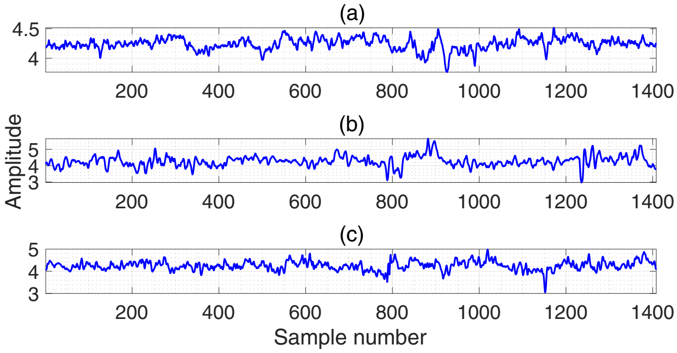


Fig. 3: Plots of EEG signals of (a) characters task, (b) digits task, and (c) objects task from the electrode AF3.

shown in the Fig. 2. The signals are downsampled from a sampling rate of 2048 Hz to 128 Hz by a factor of 16. The participants were shown characters (A, C, F, H, J, M, P, S, T, and Y), digits (0, 1, 2, 3, 4, 5, 6, 7, 8, and 9), and objects (Apple, Car, Dog, Gold, Mobile, Rose, Scooter, Tiger, Wallet, and Watch) on the screen. They were told to imagine the viewed stimuli for 10 seconds while keeping their eyes closed and in a resting state. Every two stimulus were separated by 20 seconds to allow the participant to regain his/her resting condition before considering the subsequent stimuli. During this process, the EEG signals of the subjected were recorded from Emotiv EPOC+ sensor [12]. The plots of the EEG signals corresponding to characters task, digits task, and objects task are represented in the Fig. 3.

III. Proposed Methodology

A. Preprocessing

The publicly available database used for the proposed framework contains EEG signals recorded for duration of 10 seconds for each class. Further, the EEG signals are segmented into 32 samples with a sliding increment of 8 samples [13]. Out of the 14 electrodes, 8 electrodes correspond to the frontal region, 2 electrodes correspond to the temporal region, 2 electrodes correspond to the parietal region, and 2 electrodes correspond to the occipital region. The inter-subject analysis has been performed for the proposed methodology.

B. SGWT

SGWT is an advanced signal processing method that is used to analyze the data that reside on the vertices of a graph [14]. This method is used to extend the concepts of wavelet transforms in the Euclidean domain to the domain of graphs [15]. It is formulated by using the Laplacian matrix $\boldsymbol{\Upsilon}$ which is derived from graph-based signals. The edges represent the relationships between any given nodes in the graph. These relationships are encoded in the adjacency matrix. The nodes are the samples of the EEG signals in our work. The edges connect node pairs with weights defined by the adjacency matrix, $W(i,j) = e^{(x(i)-x(j))^2}$, which is calculated separately for each electrode. The Laplacian matrix is computed by $\boldsymbol{\Upsilon} = D - W$, where D represents the diagonal matrix where each entry is the sum of the corresponding row in W. This is further used for SGWT feature extraction. A spectral graph wavelet at

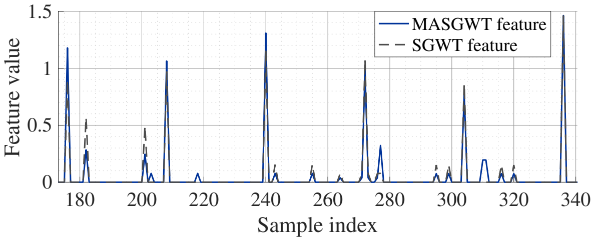


Fig. 4: Representation of SGWT feature and MASGWT feature of the EEG signal.

a specific scale ξ and localized at vertex ϑ can be expressed in vector form as, $\varphi_{\xi,\vartheta}(\chi) = \sum_{\kappa=0}^{N-1} \mu(\xi \cdot \Lambda_\kappa)\Phi_\kappa(\vartheta)\Phi_\kappa(\chi)$ [14]. Here, Λ_κ denotes the eigenvalues of $\mathbf{\Upsilon}$ and Φ_κ corresponds to the eigenvectors of $\mathbf{\Upsilon}$. The kernel μ is a band-pass filter which can be defined as follows [14], [15]:

$$\mu = \begin{cases} \eta, & \eta < 1, \\ \varpi(\eta), & 1 \leq \eta \leq 2, \\ \frac{2}{\eta}, & \eta > 2, \end{cases} \tag{1}$$

where $\varpi(\eta)$ is a cubic spline that adapts to the curve of μ. The SGWT coefficients of the signal $\mathbf{\Psi}$ are determined by calculating the inner product between the signal and the wavelet, i.e., $\langle \varphi_{\xi,\vartheta}, \mathbf{\Psi} \rangle$. An approximate SGWT is implemented by the help of the SGWT toolbox of MATLAB to reduce the computational complexity which has been described in the Appendix. From the EEG signals, the SGWT of each channel out of the 14 channels is combined to form a feature matrix. The Laplacian is computed from the feature matrix. The SGWT of the graph signal is calculated by using the Laplacian in the SGWT toolbox of MATLAB. The smoothening of the obtained features is essential as rapid fluctuations in the value of the features should be handled properly. The smoothing of the features using the moving average filter is found to perform better [16], [17]. The plots of the SGWT feature and MASGWT feature have been shown in the Fig. 4.

C. Classification

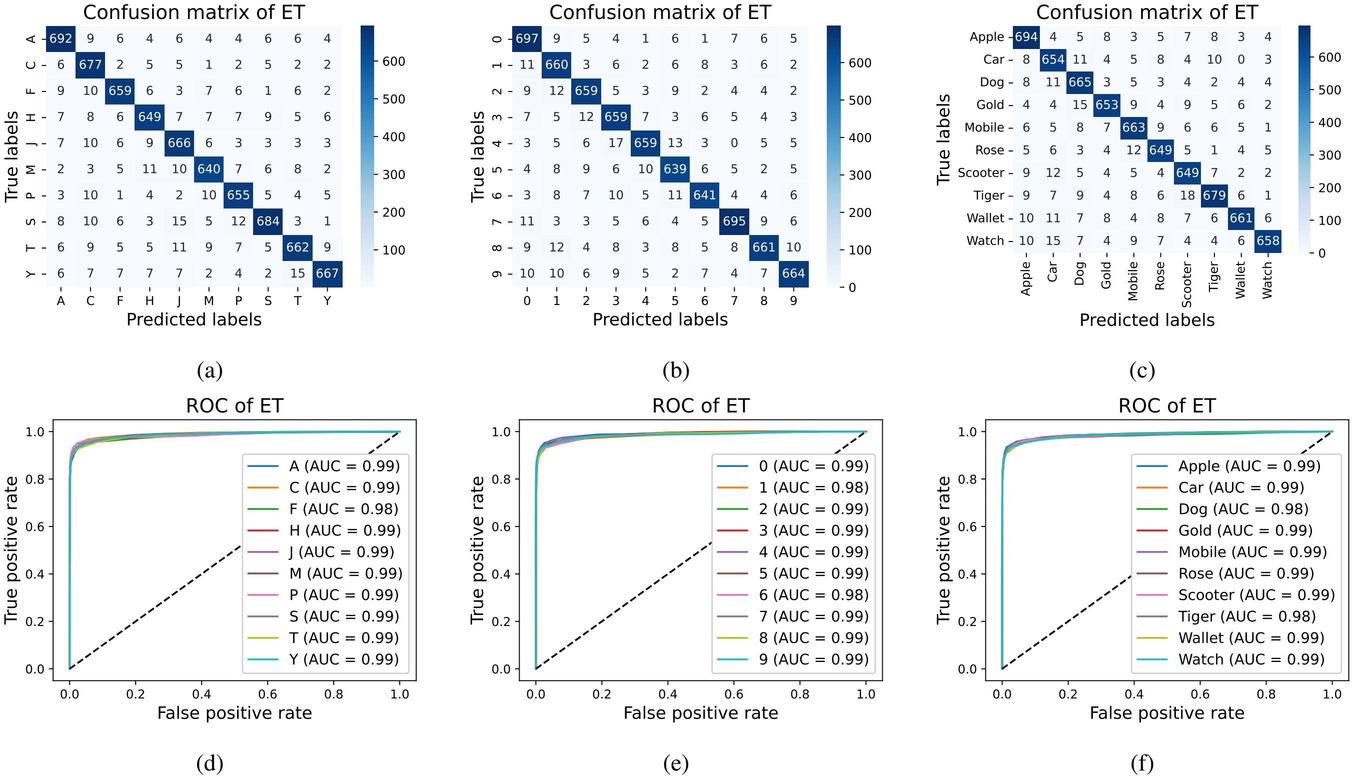
The obtained smoothened features are preprocessed using a median imputation to handle the missing values and also are scaled using standard normalization. By using label encoder, the labels are encoded into numerical form. The feature matrix is splitted into training (80%) and testing (20%) sets [18], [19]. For the proposed methodology, three well-studied ML classifiers namely, ET [20], RF [21], and KNN [22] have been considered. All the classifiers are trained using default hyperparameters. The random state and number of estimators considered for ET and RF classifiers are 42 and 100, respectively. The number of neighbors and metric used for KNN are 5 and Minkowski, respectively. The classifiers are evaluated on the testing set by using the performance metric accuracy which is mathematically represented as, Accuracy $= \frac{\phi_+ + \psi_-}{\phi_+ + \psi_- + \xi_+ + \zeta_-}$ [23]. Here, ϕ_+, ψ_-, ξ_+, and ζ_- represent the true positives, true negatives, false positives, and false negatives, respectively. The confusion matrix and the receiver operating characteristics (ROC) are useful tools for the performance assessment, where the area under the ROC curve (AUC) provides a measure of the model's ability to correctly distinguish between classes [24], [25]. The Fig. 5 represents the confusion matrices and ROC plots for the character, digit, and images tasks by using ET. The values (mean±standard daviation) of sensitivity (SN) & specificity (SP) for the characters task, digits task, and objects task are 0.929 ± 0.003 & 0.992 ± 0.000, 0.927 ± 0.002 & 0.992 ± 0.000, and 0.927 ± 0.003 & 0.992 ± 0.000, respectively on performing 30 iterations by using the ET classifier.

IV. Results and discussion

This section provides the results obtained at various steps of the proposed methodology along with their discussions. The proposed methodology presents a novel framework for the envisioned speech recognition using the SGWT and ML-based classifiers. The SGWT is computed on single-channel EEG signals of the database. Further, the SGWT-based features are smoothened and appended to form the final feature matrix. The size of the feature matrix for all the brain regions is (35880×448), where 35880 represents the total samples of the MASGWT-based features obtained by the overlapping EEG epochs and 448 is the multiplication of the EEG epoch (32) and number of channels (14). For different brain regions, different channels are considered. Out of the 14 electrodes, 8 electrodes correspond to the frontal region, 2 electrodes correspond to the temporal region, 2 electrodes correspond to the parietal region, and 2 electrodes correspond to the occipital region. This has also been shown in Fig. 2. The analysis of the proposed methodology has been shown in the Table I. The experiments were conducted on a MacBook M2 Air with an Apple M2 processor and 8 GB of RAM. MATLAB 2024b and Google Colab were utilized as the primary software platforms for this research work. The comparison of the performance metric accuracy of our proposed methodology with the pre-existing methodologies proposed by researchers has been shown in the Table II. It can be clearly seen from the Table II, a better accuracy is obtained by the proposed methodology compared to the previous studied for envisioned speech recognition. The results show that the frontal region and temporal region of the brain contribute significantly to the classification tasks. The occipital region and parietal region contribute comparatively less. Among the ML-based classifiers, ET consistently outperforms RF and KNN, which makes it the most suitable ML-based classifier for the proposed methodology. The runtime needed for the characters task, digits task, and objects task is 22.830 seconds, 22.243 seconds, and 22.899 seconds for the ET classifier.

V. Conclusion

This paper presented the recognition of envisioned speech using the EEG signal based on four stages, including the preprocessing, SGWT, feature smoothing, and classification. In this work, we study performance of three classifiers such as ET, RF, and KNN. Evaluation results showed that the proposed methodology with ET classifier outperforms of the



(a) (b) (c)

(d) (e) (f)

Fig. 5: Representation of (a)-(c) confusion matrices and (d)-(f) ROC curves for characters, digits, and images tasks, respectively by using ET.

TABLE I: Values of accuracy (in %) obtained for digits, characters, and objects tasks at different brain regions by using several ML-based classifiers

Task	ML-based classifier	Frontal region	Temporal region	Parietal region	Occipital region	All regions
Digits task	ET	87.2	72.1	63.0	41.5	92.4
	RF	79.1	60.4	56.3	38.6	82.9
	KNN	46.6	31.0	24.0	19.6	63.0
Characters task	ET	86.5	73.7	61.6	41.4	92.7
	RF	75.4	61.7	55.4	37.7	82.7
	KNN	48.3	30.9	24.9	19.9	57.9
Objects task	ET	86.6	73.5	61.6	39.0	92.3
	RF	76.7	63.2	53.7	36.2	82.6
	KNN	47.0	30.7	25.5	18.7	64.8

TABLE II: Comparison of the proposed methodology with the previous existing state-of-the-art methods

Author	Methodology	Characters task	Digits task	Objects task
Tirupattur et al. [26]	CNN	71.2%	72.9%	73.0%
Jolly et al. [27]	CNN and GRU	–	–	77.4%
Kumar et al. [11]	RF	66.9%	68.5%	65.7%
Kumar and Scheme [13]	CNN	71.0%	66.4%	72.0%
	Stacked LSTM	84.2%	75.7%	82.4%
	CNN and LSTM	87.1%	82.8%	86.6%
	CNN, LSTM, and MV	90.1%	85.1%	89.4%
Tripathi [6]	(δ rhythm + β rhythm + γ rhythm), CNN, and LSTM	87.3%	85.9%	87.5%
Mishra and Bhavsar [10]	Siamese model (online)	74.8%	76.2%	77.9%
	Siamese model (offline)	73.8%	75.2%	75.9%
	1D CNN-based approach	74.3%	75.6%	76.7%
Proposed methodology	**SGWT and ET**	**92.7**%	**92.4**%	**92.3**%

Abbreviations: CNN - convolutional neural network, GRU - gated recurrent unit, RF - random forest, LSTM - long short-term memory, MV - majority vote, SGWT - spectral graph wavelet transform, and ET - extra trees.

methods with RF and KNN classifiers and also the state-of-the-art methodologies by achieving highest accuracies of 92.7%, 92.4%, and 92.3% for recognition of characters task, digits task, and objects task, respectively. The methodology with SGWT and ET can be used for the applications of brain-computer interface (BCI) or HCI for communicating with the people having speech impairments. The proposed framework can be studied for subject-specific cases as a part of future work. In future, the proposed framework can be compared with the use of various other signal analysis methods, features, and classification techniques in order to get a comprehensive comparison. The use of statistical tests can be studied in the proposed framework as a part of future work.

Appendix

Approximate computation of SGWT

When SGWT is computed directly, its complexity is $O(N^3)$ with a memory requirement of $O(N^2)$ which makes it feasible only for the graphs that have lesser number of nodes. To rectify this computational issue, an approximate method based on truncated Chebyshev polynomials is used [28]. The kernel $\mu(t_j, \lambda)$ used in SGWT evaluation is approximated with low-dimensional Chebyshev polynomials which is given as, $\mu(t_j, \lambda) \approx \frac{1}{2}\left(c_{j,0} + \sum_{k=1}^{D_j} c_{j,k} T_k(\lambda - 1)\right)$. Here, D_j represents the degree of the approximation which is typically set to $D_j = 50$. The function $T_k(\lambda)$ denotes the k-th order shifted Chebyshev polynomial which satisfies the relation, $T_k(\lambda) = 2\lambda T_{k-1}(\lambda) - T_{k-2}(\lambda)$. The coefficients $c_{j,k}$ are the Chebyshev coefficients which are estimated using a spectrum upper bound λ_{max} [29]. In the approximated form, the transforms are expressed as [28], $\psi_{t_j}^T \approx \frac{1}{2}c_{j,0}x + \sum_{k=1}^{D_j} c_{j,k} T_k(L-1)x$ and $\varphi_x^{t_j} \approx \frac{1}{2}c_{j,0}x + \sum_{k=1}^{D_j} c_{j,k} T_k(L-1)x$, where $T_0(L) = I$ and $T_1(L) = L - I$. This approximation allows the Laplacian matrix L to be applied efficiently via matrix-vector multiplication which makes it fast for the sparse graphs.

References

[1] G. Riva, *Ambient intelligence: The evolution of technology, communication and cognition towards the future of human-computer interaction*, vol. 6. IOS Press, 2005.

[2] D. Fallman, "The new good: exploring the potential of philosophy of technology to contribute to human-computer interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1051–1060, 2011.

[3] C. Herff and T. Schultz, "Automatic speech recognition from neural signals: a focused review," *Frontiers in Neuroscience*, vol. 10, p. 429, 2016.

[4] L. Clark, P. Doyle, D. Garaialde, E. Gilmartin, S. Schlögl, J. Edlund, M. Aylett, J. Cabral, C. Munteanu, J. Edwards, *et al.*, "The state of speech in HCI: Trends, themes and challenges," *Interacting with Computers*, vol. 31, no. 4, pp. 349–371, 2019.

[5] M. Alsaleh, *Toward an imagined speech-based brain computer interface using EEG signals*. PhD thesis, University of Sheffield, 2019.

[6] A. Tripathi, "Analysis of EEG frequency bands for envisioned speech recognition," *arXiv preprint arXiv:2203.15250*, 2022.

[7] A. S. S. Reddy and R. B. Pachori, "Multivariate dynamic mode decomposition for automatic imagined speech recognition using multichannel EEG signals," *IEEE Sensors Letters*, vol. 8, no. 2, pp. 1–4, 2024.

[8] S. V. Bhalerao and R. B. Pachori, "Imagined speech–EEG detection using multivariate swarm sparse decomposition-based joint time–frequency analysis for intuitive BCI," *IEEE Transactions on Human-Machine Systems*, pp. 1–11, 2025.

[9] R. Naik, K. Chaudhari, K. Jadhav, and A. Joshi, "MindCeive: Perceiving human imagination using CNN-GRU and GANs," *Biomedical Signal Processing and Control*, vol. 100, p. 107110, 2025.

[10] R. Mishra and A. Bhavsar, "EEG classification for visual brain decoding via metric learning.," in *Bioimaging*, pp. 160–167, 2021.

[11] P. Kumar, R. Saini, P. P. Roy, P. K. Sahu, and D. P. Dogra, "Envisioned speech recognition using EEG sensors," *Personal and Ubiquitous Computing*, vol. 22, pp. 185–199, 2018.

[12] N. A. Badcock, P. Mousikou, Y. Mahajan, P. De Lissa, J. Thie, and G. McArthur, "Validation of the emotiv epoc® EEG gaming system for measuring research quality auditory ERPs," *PeerJ*, vol. 1, p. e38, 2013.

[13] P. Kumar and E. Scheme, "A deep spatio-temporal model for EEG-based imagined speech recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 995–999, IEEE, 2021.

[14] H. K. Meena, K. K. Sharma, and S. D. Joshi, "Facial expression recognition using the spectral graph wavelet," *IET Signal Processing*, vol. 13, no. 2, pp. 224–229, 2019.

[15] R. Krishna, K. Das, H. K. Meena, and R. B. Pachori, "Spectral graph wavelet transform-based feature representation for automated classification of emotions from EEG signal," *IEEE Sensors Journal*, vol. 23, no. 24, pp. 31229–31236, 2023.

[16] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 81–84, IEEE, 2013.

[17] G. K. P. Veeramallu, Y. Anupalli, S. kumar Jilumudi, and A. Bhattacharyya, "EEG based automatic emotion recognition using EMD and random forest classifier," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–6, IEEE, 2019.

[18] D. Pachori and T. K. Gandhi, "Automated emotion identification system utilizing EEG bands extracted via wavelet filter banks," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–6, IEEE, 2024.

[19] D. Pachori, R. K. Tripathy, and T. K. Jain, "Detection of atrial fibrillation from PPG sensor data using variational mode decomposition," *IEEE Sensors Letters*, vol. 8, no. 3, pp. 1–4, 2024.

[20] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, pp. 3–42, 2006.

[21] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[22] O. Kramer and O. Kramer, "K-nearest neighbors," *Dimensionality Reduction with Unsupervised Nearest Neighbors*, pp. 13–23, 2013.

[23] D. Pachori and T. K. Gandhi, "FBSE-based approach for discriminating seizure and normal eeg signals," *IEEE Sensors Letters*, vol. 8, no. 12, pp. 1–4, 2024.

[24] M. Heydarian, T. E. Doyle, and R. Samavi, "MLCM: Multi-label confusion matrix," *IEEE Access*, vol. 10, pp. 19083–19095, 2022.

[25] M. Majnik and Z. Bosnić, "ROC analysis of classifiers in machine learning: A survey," *Intelligent Data Analysis*, vol. 17, no. 3, pp. 531–558, 2013.

[26] P. Tirupattur, Y. S. Rawat, C. Spampinato, and M. Shah, "Thoughtviz: Visualizing human thoughts using generative adversarial network," in *Proceedings of the 26th ACM International Conference on Multimedia*, pp. 950–958, 2018.

[27] B. L. K. Jolly, P. Aggrawal, S. S. Nath, V. Gupta, M. S. Grover, and R. R. Shah, "Universal EEG encoder for learning diverse intelligent tasks," in *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pp. 213–218, IEEE, 2019.

[28] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[29] G. M. Phillips, *Interpolation and approximation by polynomials*, vol. 14. Springer Science & Business Media, 2003.