DCLL - A Deep Network for Possible Real-time Decoding of Imagined Words

Jerrin Thomas Panachakel[1⋆] and Ramakrishnan A G[2]

[1] MILE Lab, Indian Institute of Science, Bengaluru, India
jerrinp@iisc.ac.in
[2] MILE Lab, Indian Institute of Science, Bengaluru, India
agr@iisc.ac.in

Abstract. We present a novel architecture for classifying imagined words from electroencephalogram (EEG) captured during speech imagery. The proposed architecture employs a sliding window with overlap for data augmentation (DA) and common spatial pattern (CSP) in order to derive the features. The dimensionality of features is reduced using linear discriminant analysis (LDA). Long short-term memory (LSTM) along with majority voting is used as the classifier. We call the proposed architecture the DCLL (DA-CSP-LDA-LSTM) architecture. On a publicly available imagined word dataset, the DCLL architecture achieves an accuracy of 85.2% for classifying the imagined words "in" and "cooperate". Although this is around 7% less than the best result in the literature on this dataset, the DCLL architecture is roughly 300 times faster than the latter, making it a potential candidate for imagined word-based online BCI systems where the EEG signal needs to be classified in real-time.

Keywords: online BCI, speech imagery, LSTM, CSP, LDA, data augmentation, EEG, imagined word, imagined speech

1 Introduction

A brain–computer interface (BCI) converts the distinct neural activities into signals which can be used for controlling an external device. BCIs have several applications such as in developing devices using which a paralyzed person can control devices such as a wheelchair and a computer[1]. Different neuroimaging and electrophysiological modalities such as electrocorticogram (ECoG) [2, 3], electroencephalogram (EEG) [4, 5], intracortical electroencephalography (ICE) [6–8], magnetoencephalography (MEG) [9, 10], functional magnetic resonance imaging (fMRI) [11, 12] and functional near-infrared spectroscopy (fNIRS) [13–15] are used in BCI systems for capturing the neural activities. Currently available BCI systems using EEG use event-related potential (ERP) [16, 17, 5, 4], motor imagery [18, 19], or steady state visually evoked potentials (SSVEP) [20–22] for producing neural activity that needs to be recognized by the system.

⋆ Jerrin Thomas Panachakel is working in the Department of Electronics and Communication Engineering at College of Engineering, Trivandrum, Kerala, India

These BCI systems have limited degrees of freedom due to the finite number of distinguishable prompts available. Also, the prompts used in these systems are not intuitive leading to difficulties in training a patient to use them. An alternative approach is to use imagined words or speech imagery for causing distinct neural activities that can be recognized by a BCI system. BCI systems using imagined words have more intuitive prompts and have higher degrees of freedom. In addition to all the possible applications of a general BCI system based on motor imagery, a high-performance BCI system based on speech imagery, in conjunction with a text to speech (TTS) synthesis system, can be used by those with speech disabilities to communicate with others [23, 24].

Brain–computer interfaces are broadly classified into 1) online and 2) offline BCI systems [25]. An online system is one, where the EEG signal from the subject is analyzed in real-time whereas in an offline system, the processing does not happen and hence the outputs do not appear in real-time. Offline BCIs are used mainly in the laboratory environment and usually employ algorithms that have high computational cost. The fact that the outputs do not occur in real-time limits the practical use of these systems. Most of the works on speech imagery based BCI systems employ offline decoding strategies. A few exceptions are the works by Nguyen et al. [26] and Sereshkeh et al. [27] which employ EEG and the other works [15, 28, 29] in which fNIRS is used. The work by Nguyen et al. [26] uses both speech and motor imagery. We propose here a common spatial pattern - linear discriminant analysis - long short-term memory based architecture which we call the DCLL (DA-CSP-LDA-LSTM) architecture that has low computational cost. Hence it is a potential architecture for online BCI systems employing speech imagery. Lower computational cost is a critical design factor due to the following reasons:

1. An online BCI system based on EEG needs to process the EEG signal in real-time.
2. A practical BCI system might need to run on a mobile processing system which has limited processing power.
3. In case of a mobile BCI system running on batteries, lower computational cost of the algorithms translates to more running time before battery replacement or recharge.

2 Dataset used in the Study

We have tested the proposed DCLL system for deciphering the imagined words from the freely obtainable ASU speech imagery EEG data. This data was developed by HORC lab at Arizona State University [30] by recording EEG signal, sampling at 1000 Hz using a BrainProducts ActiCHamp amplifier and a 64-electrode EEG cap. The EEG signal had been downsampled later to 256 samples per second during the preprocessing stage. The low frequency trends and EMG artifacts were removed using a 5-th order Butterworth bandpass filter having the lower cut-off at 8 Hz and higher cut-off at 70 Hz. A notch filter was deployed to filter the line noise and adaptive filtering [31], to remove the ocular artifacts. This work focuses on distinguishing the short ("in") from the long ("cooperate") imagined word in the dataset.

3 Data Augmentation

Data augmentation is used in situations where the data available for training a model is not enough. In data augmentation, more data is generated from the existing data [32]. Techniques such as overlapping or sliding window [33–36] and generative adversarial networks (GAN) [37–39] are commonly used for augmenting an EEG data. Since the data studied comprises repeated imaginations in each trial, overlapping or sliding window is an ideal approach for data augmentation (DA). Without DA, one trial of 5 s duration contributes to one training or test frame. In this work, a window size of 256 samples (1 s) and a stride of 64 samples (0.25 s) are empirically selected, leading to an overlap of 192 samples (0.75 s). With these parameters, we get 17 analysis frames of length one second each, resulting in an augmentation factor of 17.

4 Extracting the Features and LSTM Classifier

Figure 1 gives the overall DCLL (DA-CSP-LDA-LSTM) architecture. This architecture is similar to the architecture employed in one of our previous works [40]. Common spatial pattern (CSP) [41, 42] is used to extract the features. CSP extremizes the objective function given below in order to derive a spatial filter $\mathbf{H}$:

$$J(\mathbf{H}) = \frac{trace(\mathbf{H}^T\mathbf{R}_i\mathbf{H})}{trace(\mathbf{H}^T\mathbf{R}_j\mathbf{H})} \tag{1}$$

where T refers to the transpose of the matrix and $\mathbf{R_k}$ is the normalized spatial covariance matrix of class k. $\mathbf{R_k}$ is computed as,

$$\mathbf{R_k} = \frac{\mathbf{Y}_k\mathbf{Y}_k^T}{trace(\mathbf{Y}_k\mathbf{Y}_k^T)} \tag{2}$$

where $\mathbf{Y}_k \in \mathbb{R}^{M\times N}$ is the class-k EEG signal matrix with M channels and N time samples per channel.

The filter $\mathbf{H}$, computed separately for each set of training data, is then used to obtain the spatially filtered data ($\mathbf{Z}_l$):

$$\mathbf{Z}_l = \mathbf{H}\mathbf{Y}_l \tag{3}$$

The number of spatial filter pairs is empirically chosen as 10. In our earlier work [43], we found that the accuracy does not improve for higher values of the filter pairs used. Thus, the dimension of $\mathbf{H}$ is 20×64, where 64 represents the EEG channels utilized and the dimension of Y_l is 64×64. In our previous

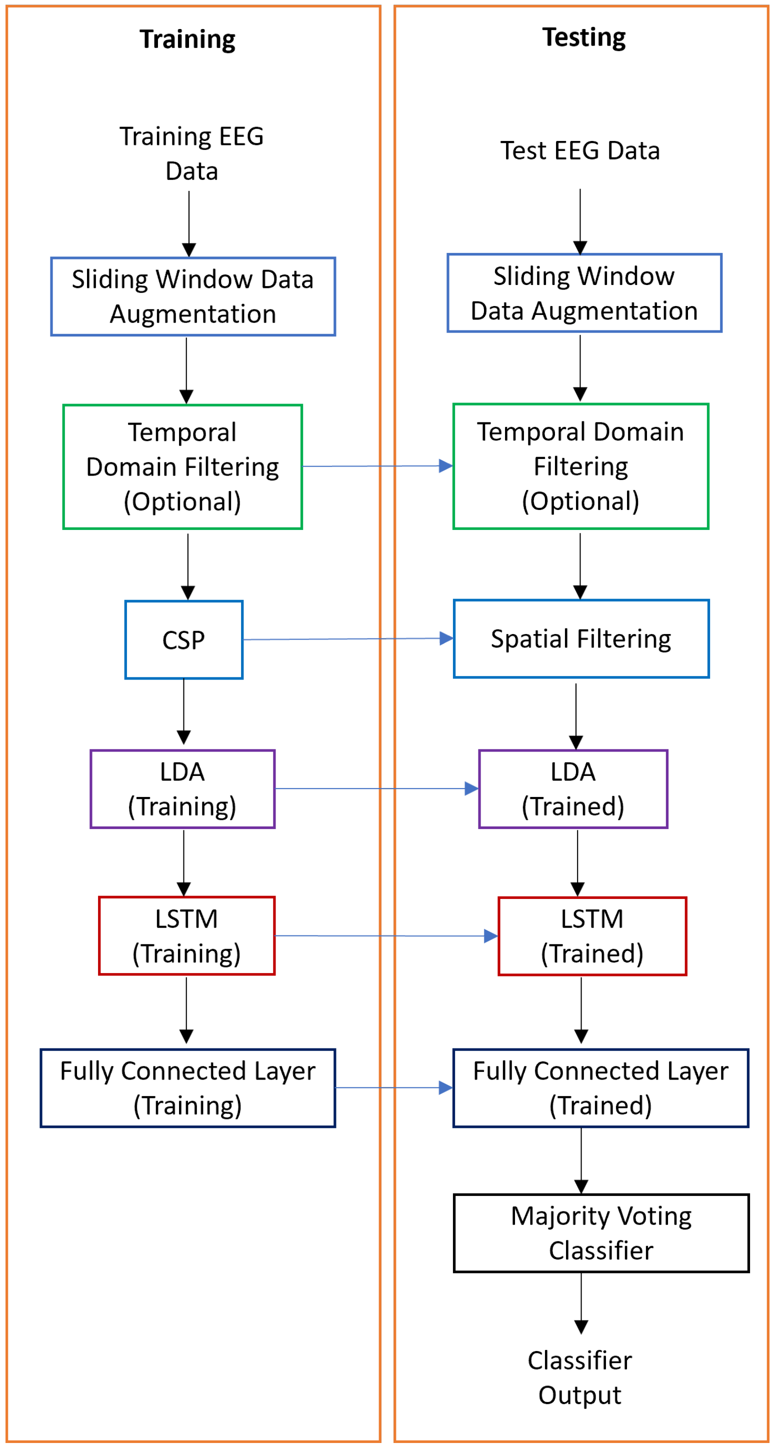

Fig. 1: Architecture of the DCLL method. CSP: common spatial pattern; LDA: linear discriminant analysis; LSTM: long short-term memory.

work [43], we had experimented with common spatial pattern with Tikhonov regularization (TR-CSP) but we did not observe any improvement is accuracy with TR regularization.

We employ the logarithm of the variance of each of the filtered vectors as a feature generating a feature vector of length 20. Linear discriminant analysis is made use of to reduce the dimension of these vectors to one. These values are provided as sequential input data to a long-short term memory network employed for classification. Two hundred hidden units are exploited in the classifier, and the number of epochs does not exceed 20 to avoid overfitting. Adam optimizer is utilized and the learning rate applied is 0.001.

During testing, the test EEG data is also augmented leading to 17 test frames corresponding to a single EEG data of 5s duration. Each of these augmented test frames is input to the LSTM classifier. Finally, majority voting with hard voting strategy is used to decide the class of the test EEG data based on the outputs of the fully connected layer for the 17 test frames.

5 Results

Two metrics are used in this work for evaluating the performance of the DCLL architecture: 1) classification accuracy and 2) execution time. The evaluation strategy followed for computing these metrics is 10-fold cross-validation. Although sometimes overlooked, execution time is an important performance metric since fast execution is required in an online BCI system.

5.1 Classification Accuracy

To identify which frequency band contributes the most to the classifier performance, we utilize three distinct EEG bands, namely alpha (8 - 13 Hz), beta (13 -30 Hz) and gamma (30 - 70 Hz); gamma band is restricted to 70 Hz since in the publicly available version of the ASU dataset, the signal has been bandpass filtered reducing the available bandwidth to 8 - 70 Hz. Entries from Sl. No. 1 to 7 in Table 1 compare the accuracies ($mean \pm S.D.$) of the proposed DCLL architecture for different EEG frequency bands and other methods in the literature for classifying the imagined words "in" and "cooperate". Sl. No. 8 in the same Table gives the accuracy of the system proposed by Sereshkeh et al. [27] for the task of online classification of the imagined prompts "yes" and "no". Among the bands, gamma band gives the highest accuracy of 78.2%. However, the system gives the best performance of 85.2% when the signal is not bandpasss filtered into different frequency bands. These results are inline with the earlier published results on decoding imagined words from EEG [43, 44]. The state-of-the-art performance reported in the literature is 92.8% by the transfer learning architecture [44], which uses magnitude squared coherence and mean phase coherence as features. However, this method is computationally complex as explained in the next subsection and cannot be used for real-time decoding of the imagined words.

Table 1: Sl. No. 1 to 7 compare the accuracies ($mean \pm S.D.$) of the proposed DCLL architecture for different EEG frequency bands and other methods in the literature for the classification of the imagined words "in" and "cooperate". Sl. No. 8 gives the accuracy of the system proposed by Sereshkeh et al. [27] for the task of online classification of the imagined prompts "yes" and "no". MSC: magnitude squared coherence, MPC: mean phase coherence.

Sl. No.	**Method**	**Accuracy (%)**
1	DCLL on alpha band	67.3 ± 6.7
2	DCLL on beta band	74.2 ± 4.3
3	DCLL on gamma band	78.2 ± 7.2
4	DCLL on the undivided EEG signal	85.2 ± 2.7
5	Tangent + relevance vector machine [30]	79.9 ± 8.2
6	Wavelet + deep neural network [45]	73.5 ± 8.2
7	MSC + MPC + trasnfer learning [44]	92.8 ± 1.9
8	Spectral and time-frequency features [27]	69.3 ± 14.1

Figure 2 compares the subject-wise performance of the proposed method with other methods for classifying the imagined prompts "in" and "cooperate". The evaluation strategy employed is 10-fold cross-validation. The performance of the proposed DCLL method is higher than the tangent+RVM [30] and Wavelet+DNN [45] techniques for the subjects S1, S5, S8, and S10 and higher than that of only [45] for S14. However, the technique MSC+MPC+TL [44] outperforms DCLL for all the subjects, even though it does not perform in real time.

5.2 Execution Time

The execution time of DCLL architecture (including data augmentation) and the architecture that gave the best results in classifying the imagined prompts "in" and "cooperate" are evaluated using MATLAB® 2021b running on a Windows 11 machine with Intel® Core™ i5-8250U processor and 16 GB DDR4 RAM. On an average, the DCLL architecture takes only 108.8 ± 33.8 ms for classifying a single trial of test EEG data of 5s duration. The state-of-the-art method [44] on the other hand takes 31.5 ± 3.4 s for the same task. Thus the DCLL architecture is around 300 times faster than the state-of-art method. Since the DCLL architecture takes only a fraction of a second for classifying one second of EEG data, it has the potential to be used for an online BCI system utilizing speech imagery. Although the DCLL architecture has lower classification accuracy than the state-of-the-art method, its performance is superior to all the other methods in the literature for distinguishing the short from the long words from the ASU dataset.

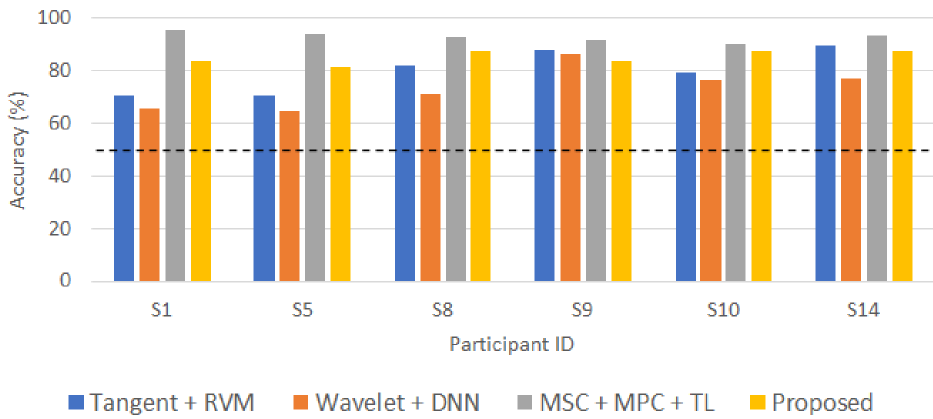

Fig. 2: Performance of the DCLL architecture compared with other work on the ASU dataset. The participant IDs are represented by the x-axis and the classification accuracies, by the y-axis. The dashed line indicates the chance level accuracy. RVM: relevance vector machine; DNN: deep neural network; MSC: magnitude squred coherence; MPC: mean phase coherence; TL: trasnfer learning.

6 Conclusion

This paper proposes a fresh algorithm for classifying imagined words from the EEG captured during speech imagery. The proposed DCLL architecture achieves an accuracy of more than 85% in distinguishing imagined word “in” from “cooperate”. Although this is 7% less than the performance of the best method on this data, the DCLL architecture has much lower computational cost. It is around 300 times faster than the state-of-the-art method and can classify 5s EEG data in about 100 ms. This addresses a key challenge in designing imagined word and EEG-based BCI systems that can classify EEG signals in real-time. In our future work, we intend to work on improving the design of the architecture to improve its accuracy to be comparable to the best in the literature, while retaining its speed. This will include improvements in the prepocessing steps to improve the signal-to-noise ratio of the EEG, dimensionality reduction using methods such as principal component analysis instead of LDA, replacement of majority voting classifier with other ensemble classifiers etc. We also intend to make the DCLL architecture a hierarchical architecture where the initial classification happens at the phonological level [46, 47]. The present work is based on the data collected from healthy subjects. The future work also includes testing the architecture on stroke-affected patients [48].

References

1. Abdulkader, S.N., Atia, A., Mostafa, M.S.M.: Brain computer interfacing: Applications and challenges. Egyptian Informatics Journal 16(2), 213–230 (2015)

2. Herff, C., Heger, D., De Pesters, A., Telaar, D., Brunner, P., Schalk, G., Schultz, T.: Brain-to-text: decoding spoken phrases from phone representations in the brain. Frontiers in neuroscience 9, 217 (2015)
3. Herff, C., Johnson, G., Diener, L., Shih, J., Krusienski, D., Schultz, T.: Towards direct speech synthesis from ECoG: A pilot study. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 1540–1543. IEEE (2016)
4. Mugler, E.M., Ruf, C.A., Halder, S., Bensch, M., Kubler, A.: Design and implementation of a p300-based brain-computer interface for controlling an internet browser. IEEE Transactions on Neural Systems and Rehabilitation Engineering 18(6), 599–609 (2010)
5. Xu, M., Xiao, X., Wang, Y., Qi, H., Jung, T.P., Ming, D.: A brain–computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli. IEEE Transactions on Biomedical Engineering 65(5), 1166–1175 (2018)
6. Brumberg, J.S., Wright, E.J., Andreasen, D.S., Guenther, F.H., Kennedy, P.R.: Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech motor cortex. Frontiers in neuroscience 5, 65 (2011)
7. Kennedy, P., Cervantes, A., Gambrell, C., Ehirim, P.: Advances in the development of a speech prosthesis. Direct and Indirect Benefits of Translingual Neurostimulation Technology for Neurorehabilitation of Chronic Stroke Symptoms p. 1 (2017)
8. Wilson, G.H., Stavisky, S.D., Willett, F.R., Avansino, D.T., Kelemen, J.N., Hochberg, L.R., Henderson, J.M., Druckmann, S., Shenoy, K.V.: Decoding spoken english from intracortical electrode arrays in dorsal precentral gyrus. Journal of Neural Engineering 17(6), 066007 (2020)
9. Dash, D., Wisler, A., Ferrari, P., Davenport, E.M., Maldjian, J., Wang, J.: MEG sensor selection for neural speech decoding. IEEE Access 8, 182320–182337 (2020)
10. Destoky, F., Philippe, M., Bertels, J., Verhasselt, M., Coquelet, N., Vander Ghinst, M., Wens, V., De Tiège, X., Bourguignon, M.: Comparing the potential of MEG and EEG to uncover brain tracking of speech temporal envelope. Neuroimage 184, 201–213 (2019)
11. Yoo, S.S., Fairneny, T., Chen, N.K., Choo, S.E., Panych, L.P., Park, H., Lee, S.Y., Jolesz, F.A.: Brain–computer interface using fMRI: spatial navigation by thoughts. Neuroreport 15(10), 1591–1595 (2004)
12. Abe, K., Takahashi, T., Takikawa, Y., Arai, H., Kitazawa, S.: Applying independent component analysis to detect silent speech in magnetic resonance imaging signals. European Journal of Neuroscience 34(8), 1189–1199 (2011)
13. Herff, C., Putze, F., Heger, D., Guan, C., Schultz, T.: Speaking mode recognition from functional near infrared spectroscopy. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 1715–1718. IEEE (2012)
14. Kamavuako, E.N., Sheikh, U.A., Gilani, S.O., Jamil, M., Niazi, I.K.: Classification of overt and covert speech for near-infrared spectroscopy-based brain computer interface. Sensors 18(9), 2989 (2018)
15. Sereshkeh, A.R., Yousefi, R., Wong, A.T., Chau, T.: Online classification of imagined speech using functional near-infrared spectroscopy signals. Journal of Neural Engineering 16(1), 016005 (2018)
16. Fouad, I.A., Labib, F.E.Z.M., Mabrouk, M.S., Sharawy, A.A., Sayed, A.Y.: Improving the performance of P300 BCI system using different methods. Network Modeling Analysis in Health Informatics and Bioinformatics 9(1), 1–13 (2020)

17. Sellers, E.W., Krusienski, D.J., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R.: A p300 event-related potential brain–computer interface (bci): the effects of matrix size and inter stimulus interval on performance. Biological psychology 73(3), 242–252 (2006)
18. Kevric, J., Subasi, A.: Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. Biomedical Signal Processing and Control 31, 398–406 (2017)
19. Onose, G., Grozea, C., Anghelescu, A., Daia, C., Sinescu, C., Ciurea, A., Spircu, T., Mirea, A., Andone, I., Spânu, A., et al.: On the feasibility of using motor imagery EEG-based brain–computer interface in chronic tetraplegics for assistive robotic arm control: a clinical test and long-term post-trial follow-up. Spinal cord 50(8), 599–608 (2012)
20. Ojha, M.K., Mukul, M.K.: Detection of target frequency from SSVEP signal using empirical mode decomposition for SSVEP based BCI inference system. Wireless Personal Communications pp. 1–13 (2020)
21. Müller-Putz, G.R., Scherer, R., Brauneis, C., Pfurtscheller, G.: Steady-state visual evoked potential (ssvep)-based communication: impact of harmonic frequency components. Journal of Neural Engineering 2(4), 123 (2005)
22. Han, C., Xu, G., Xie, J., Chen, C., Zhang, S.: Highly interactive brain–computer interface based on flicker-free steady-state motion visual evoked potential. Scientific Reports 8(1), 1–13 (2018)
23. Allison, B., Graimann, B., Gräser, A.: Why use a BCI if you are healthy. In: ACE Workshop-Brain-Computer Interfaces and Games. pp. 7–11 (2007)
24. Bogue, R.: Brain-computer interfaces: control by thought. Industrial Robot: An International Journal (2010)
25. Panachakel, J.T., Ramakrishnan, A.: Decoding covert speech from EEG-a comprehensive review. Frontiers in Neuroscience 15, 392 (2021)
26. Nguyen, C.H., Karavas, G.K., Artemiadis, P.: Adaptive multi-degree of freedom brain computer interface using online feedback: Towards novel methods and metrics of mutual adaptation between humans and machines for BCI. PloS One 14(3), e0212620 (2019)
27. Sereshkeh, A.R., Trott, R., Bricout, A., Chau, T.: Online EEG classification of covert speech for brain–computer interfacing. International Journal of Neural Systems 27(08), 1750033 (2017)
28. Naseer, N., Hong, M.J., Hong, K.S.: Online binary decision decoding using functional near-infrared spectroscopy for the development of brain–computer interface. Experimental Brain Research 232(2), 555–564 (2014)
29. Gallegos-Ayala, G., Furdea, A., Takano, K., Ruf, C.A., Flor, H., Birbaumer, N.: Brain communication in a completely locked-in patient using bedside near-infrared spectroscopy. Neurology 82(21), 1930–1932 (2014)
30. Nguyen, C.H., Karavas, G.K., Artemiadis, P.: Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features. Journal of Neural Engineering 15(1), 016002 (2017)
31. He, P., Wilson, G., Russell, C.: Removal of ocular artifacts from electro-encephalogram by adaptive filtering. Medical and Biological Engineering and Computing 42(3), 407–412 (2004)
32. Lashgari, E., Liang, D., Maoz, U.: Data augmentation for deep-learning-based electroencephalography. Journal of Neuroscience Methods p. 108885 (2020)
33. O'Shea, A., Lightbody, G., Boylan, G., Temko, A.: Neonatal seizure detection using convolutional neural networks. In: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP). pp. 1–6. IEEE (2017)

34. Kwak, N.S., Müller, K.R., Lee, S.W.: A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. PloS One 12(2), e0172578 (2017)
35. Ullah, I., Hussain, M., Aboalsamh, H., et al.: An automated system for epilepsy detection using EEG brain signals based on deep learning approach. Expert Systems with Applications 107, 61–71 (2018)
36. Majidov, I., Whangbo, T.: Efficient classification of motor imagery electroencephalography signals using deep learning methods. Sensors 19(7), 1736 (2019)
37. Luo, Y., Lu, B.L.: EEG data augmentation for emotion recognition using a conditional Wasserstein GAN. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 2535–2538. IEEE (2018)
38. Wei, Z., Zou, J., Zhang, J., Xu, J.: Automatic epileptic EEG detection using convolutional neural network with improvements in time-domain. Biomedical Signal Processing and Control 53, 101551 (2019)
39. Chang, S., Jun, H.: Hybrid deep-learning model to recognise emotional responses of users towards architectural design alternatives. Journal of Asian Architecture and Building Engineering 18(5), 381–391 (2019)
40. Panachakel, J.T., Kumar, G.P., Sharma, K., Ramakrishnan, A.: Automated classification of EEG into meditation and non-meditation epochs using common spatial pattern, linear discriminant analysis, and LSTM. In: TENCON 2021-2021 IEEE Region 10 Conference. pp. 1–6. IEEE (2021)
41. Goel, P., Joshi, R., Sur, M., Murthy, H.A.: A common spatial pattern approach for classification of mental counting and motor execution EEG. In: International Conference on Intelligent Human Computer Interaction. pp. 26–35. Springer (2018)
42. Panachakel, J.T., Vinayak, N.N., Nunna, M., Ramakrishnan, A.G., Sharma, K.: An improved EEG acquisition protocol facilitates localized neural activation. In: Advances in Communication Systems and Networks, pp. 267–281. Springer (2020)
43. Panachakel, J.T., Kumar, G.P., Sharma, K., Ramakrishnan, A.: Binary classification of meditative state from the resting state using EEG. In: 2021 IEEE 18th India Council International Conference (INDICON). pp. 1–6. IEEE (2021)
44. Panachakel, J.T., Ramakrishnan, A.: Decoding imagined speech from EEG using transfer learning. IEEE Access (2021)
45. Panachakel, J.T., Ramakrishnan, A., Ananthapadmanabha, T.: A novel deep learning architecture for decoding imagined speech from EEG. arXiv preprint arXiv:2003.09374 (2020)
46. Panachakel, J.T., Ramakrishnan, A., Anusha, A., Sharma, K.: Can we identify the category of imagined phoneme from EEG? In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE (2021)
47. Panachakel, J.T., Ramakrishnan, A.: Classification of phonological categories in imagined speech using phase synchronization measure. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE (2021)
48. Benzy, V., Vinod, A., Subasree, R., Alladi, S., Raghavendra, K.: Motor imagery hand movement direction decoding using brain computer interface to aid stroke recovery and rehabilitation. IEEE Transactions on Neural Systems and Rehabilitation Engineering 28(12), 3051–3062 (2020)