Proceedings of ICBioMed 2025 Symposium: Computational Modelling and Simulation for Biology and Medicine
DOI: 10.54254/2753-8818/2025.LD26088

A Review of Language BCIs: Comparing Input Methods and Machine Learning Models

Yixuan Wu

Sir Winston Churchill Secondary School, Vancouver, Canada
shinewuyx@gmail.com

Abstract. Language BCIs are promising technologies that offer alternative ways of communication for people with disabilities. The most common language BCIs can be classified by their input methods as invasive or non-invasive and by their underlying machine learning methods as supervised learning or deep learning. The supervised methods examined include LDA and SVM. The deep learning methods, including CNNs and RNNs, are explained and discussed. Many other studies have reviewed those methods broadly or focused on comparing particular mechanisms; however, this review details and compares research into language BCIs specifically. This review aims to explain the mechanisms and nuances behind frequently studied language BCI methods classified based on those two criteria, discuss each of their strengths and limitations, and anticipate future trends. This review identifies the current gap in BCI research as the tradeoff between fidelity and invasiveness when it comes to input methods and the lack of data-efficient, transparent machine-learning algorithms.

Keywords: Brain-Computer Interface, Language BCIs, Machine Learning, Deep Learning, Supervised Learning

1. Introduction

Brain-computer interfaces (BCIs) are technologies that enable communication between the brain and automated devices. They bypass traditional human outputs such as movement, language, and even emotions. In doing so, BCIs have emerged as an emerging alternative method of control and communication for patients with severe impairments. In recent years, alongside the development of motor BCIs, such as those that allow patients to control artificial arms to complete daily tasks, research into language BCIs has opened the possibility of translating imagined words into speech. The path taken by the language BCIs represents the general trajectory of the BCI field, which is moving beyond basic motor intention to more complex cognitive functions such as language. These systems primarily assist individuals with locked-in syndrome, cerebral palsy, amyotrophic lateral sclerosis (ALS), severe traumatic brain injuries, or brainstem strokes to reclaim their ability to communicate through external means.

Language BCIs can be broadly classified based on differences at each stage of the procedure: 1) input method, 2) decoding technique. In terms of input, they are divided into invasive and non-invasive BCIs. Invasive approaches, such as electrocorticography (ECoG), offer high spatial and

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.LD26088

temporal resolution but require surgery. Non-invasive methods, electroencephalography (EEG) in particular, are safer and more widely accessible, though they offer lower levels of resolution. Machine learning models used to decode these inputs vary from classical methods of linear discriminant analysis (LDA) and support vector machines (SVM) to deep learning systems such as convolutional and recurrent neural networks (CNNs & RNNs).

This review will explore research on language BCIs in detail, addressing each of the categories with an emphasis on the machine learning strategies in Section 2. It will be followed by an evaluative discussion of current limitations and future trends.

2. Input methods: EEG and ECoG

The first step of BCI requires the collection and input of neurological data through a device. Those devices could either be invasive or non-invasive. Electrocortiography (ECoG) and electroencephalography (EEG) represent the two most common methods of input acquisition from either category. To determine which input device to use for a particular study, these factors are often considered: temporal resolution, spatial resolution, and noise. This section will provide a comparative overview of both EEG and ECoG against those measurements.

2.1. Electroencephalography (EEG)

EEG is a non-invasive technique that records the postsynaptic electrical potentials of pyramidal neurons through electrodes placed outside the skull and scalp. These pyramidal neurons, especially those in the superficial layers (II or III) of the neural cortex, are vertically aligned with one another and oriented perpendicular to the surface of the brain. They have long dendrites that allow neural currents to summate to form a detectable field on the scalp that registers on an EEG device. EEG captures oscillations between 1 to 150 Hz, corresponding to brainwaves ranging from delta and gamma.

While EEG offers excellent temporal resolution—typically between 1 to 5 milliseconds, it has relatively low spatial resolution, typically between 2 to 3 centimeters. Temporal resolution refers to how precisely a recording device can detect changes over time, or the smallest distinguishable time interval for the device. Spatial resolution is the ability of a method of recording to distinguish between two points in space, which determines how precisely a researcher using the data can localize the origin of the detected signal. EEG's low spatial resolution stems from the fact that the electrical signals the EEG electrodes detect must pass many layers of tissue, including the skull and the cerebrospinal fluid in the ventricles, before reaching the EEG device. These layers are resistive and make the origins of the signal less clear [1]. Another reason for the low spatial resolution of EEGs comes from the fact that action potentials are too short and small to be detected by EEG; instead, it summarizes the slower activity of coordinated dendritic currents in larger areas on the cortex, hence the low spatial resolution.

Due to the low spatial resolution, EEG data is laced with noises that come from blinking, heartbeat, and environmental interference. The low signal-to-noise ratio (SNR) of EEG data results in the low accuracy of the data obtained from one EEG trial. Hence, more than one trials are usually combined to correctly identify the data. The result of this is that speed is often compromised [2]. Optimizations of EEG data are closely integrated with research into EEG-based language BCIs. In particular, the optimization of the traditional EEG-based P300 speller has been studied. In parallel with the optimization of outputs of EEG-based devices, improvements in the processing of EEG signals through machine learning have also been mitigating noise in the data through algorithms

DOI: 10.54254/2753-8818/2025.LD26088

such as stepwise linear discriminant analysis (SWLDA), support vector machines (SVMs), and deep learning (Section 2).

2.2. Electrocochleography (ECoG)

Alternatively, ECoG involves putting grids of electrodes on the surface of the cortex inside the dura mater. This is typically done on patients who are undergoing monitoring and require surgical implantation. Signals acquired by the ECoG device can include high-frequency neural activity between 70 to 500 Hz and low-frequency ones, such as theta and alpha waves, which EEG is not able to adequately detect. This property of ECoGs can prove to be essential. For example, high-frequency activity in the superior temporal gyrus is strongly associated with articulatory planning and inner speech, which are inextricable from BCIs that aim to translate intentions of speech into words.

ECoGs achieve spatial resolution in the millimeter scale and temporal resolution in the millisecond scale [3]. Its reduced noise and increased precision mainly stem from their proximity to the neuron sources and the absence of layers between the electrodes and the cortex. In particular, the high-gamma activity (70 to 500 Hz) that is linked to certain cognitive tasks in Broca's and Wernicke's areas during speech can be adequately analyzed using ECoG graphs [4], but only vaguely using EEGs. Leuthardt et al. showed that BCIs based on ECoGs displayed cortical activity in the high gamma range (70 to 170 Hz) with high fidelity [5].

Additionally, since data obtained from EEGs are often noisier than ECoGs, fewer trials are needed for ECoGs to achieve a certain outcome than EEGs. This often makes real-time implementation more favorable. Research into ECoG-based BCIs has also tried increasing the precision of the devices even more by decreasing the size of the electrodes. An example is the micro-ECoG (μECoG), which refines the traditional ECoG to be high-density and micro-scale. These small electrodes are around 200 μm in diameter. In a landmark study published on Nature in 2023, Duraivel et al. implemented μECoG with a 57× higher electrode density compared to standard ECoG [6]. This resolution allowed the researchers to analyze micro-level activity in the sensorimotor part of the speech cortex.

Overall, comparing the two methods, ECoG technology, as Ritaccio et al. suggest, is a promising method that allows for high spatial and temporal resolution with less noise and more precision [7]. However, many studies still use EEGs due to their non-invasiveness. Invasive procedures raise ethical issues of consent, particularly as most participants suffer from difficulty in regular communication. Hence, a limitation of ECoG studies is the small sample size Studies such as Schalk et al. only featured five participants and might be difficult for the results to be generalized [8]. The challenge for future research lies in non-invasive methods that can match or surpass the precision of ECoGs.

3. Machine learning models: supervised and deep learning

The choice of a machine learning model often directly impacts the accuracy or efficiency of a BCI's performance. Research into the machine learning aspect of language BCIs is essential to the improvement of the "software" aspect of BCIs beyond research into primary methods of neural data collection. In language BCIs, especially semantics, grammar, spelling, and meaning interact in ways that are more complex and context-dependent than motor BCIs, for example. As such, language BCIs require a more robust set of machine learning strategies to attack these multi-dimensional problems. Though different machine learning methods approach the specifics in numerous and

DOI: 10.54254/2753-8818/2025.LD26088

nuanced ways, the overarching problem that they try to tackle is classification: how can a BCI system correspond a certain pattern in brain activity shown on the EEG or ECoG graphs to a word, concept, intention, and so on.

In motor BCIs, binary classifications such as between the user's intention to move the left hand versus the right [9] and regression models that predict a continuous variable of how much to the left or right the person intends to move [10] are often studied. However, language BCIs increasing focus on finding and optimizing methods of decoding brain activity to the 26-group classification task of the English alphabet for spelling or the multi-group classification task of phoneme or word-based methods. Various machine learning methods are used to decode that data. Some of them include supervised learning such as LDA and SVM and deep learning such as CNNs and RNNs. Many specific methods and optimizations exist within each category, and only the most frequently implemented ones are compared in this review.

3.1. Supervised learning

Supervised machine learning is fundamental in computer science. The supervised algorithm is trained based on labeled data and is then used to predict the labels of a different set of data. Its accuracy is measured by comparing the model's predictions to the pre-assigned labels. Hence, in computer science, supervised models require large, high-quality, and annotated datasets that are pre-collected and standardized. With these datasets, supervised models often perform effectively; however, the development of such a model involves millions of samples.

In BCI systems, especially language BCIs, supervised learning operates in a slightly different way. There is no million upon millions of available datapoints to train the model on; instead, the labeled EEG or ECoG data is calibrated to be specific to the user and must be collected in real-time as the BCI operates. As discussed in the previous Section, brain activity is highly individual and subject to artifacts and noise, resulting in relatively limited labeled data. Accordingly, classical supervised methods like Linear Discriminant Analysis (LDA) and Support Vector Machines (SVMs) are commonly used in BCIs due to their data-efficiency. These models require fewer parameters and retain higher accuracy in small sample experiments. In these experiments, labeled data is obtained by asking the user to perform a specific task, such as focusing on a word or concept in imagination and extracting the corresponding neural signals.

3.1.1. Linear Discriminant Analysis (LDA)

LDA is commonly used by the P300 spellers because of its simplicity and efficiency even with nigh-noise data such as EEG data. The goal of LDA is to project the data into two or more classifications. In P300s, LDA reduces the high-dimensional data input to lower lower-dimensional space. For example, to perform the binary classification between neural activities relating to the desired letter displayed on the screen and those that correspond to non-desired letters, LDA often reduces the data to one dimension. Figure 1 shows a P300 speller screen. that lights up a row or column for 100 ms every 175 ms. If the target letter were "D," the algorithm should be able to classify column 4 and row 1 with the label "target" when they are lit up.

DOI: 10.54254/2753-8818/2025.LD26088

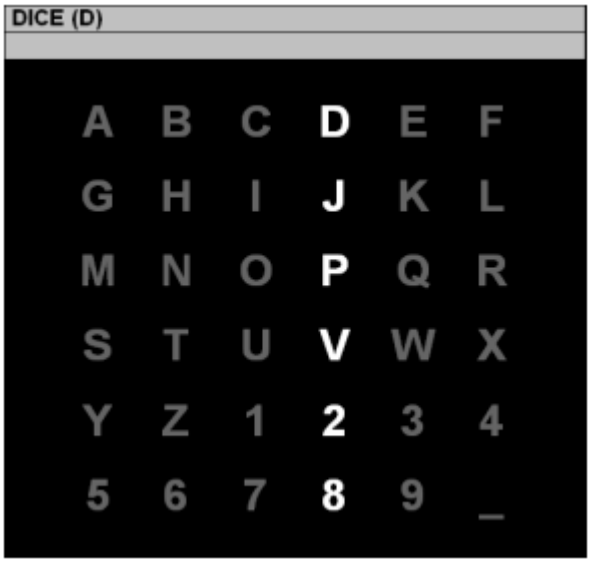


Figure 1. A P300 speller screen [11]

The set-up of the P300 is well-suited to the LDA classification method, as each trial requires the model to perform a binary classification of whether the neural signal corresponded to the desired letter being in the row/column illuminated or not. Naturally, LDA is used for classification beyond the early research into P300 spellers. In combination with canonical correlation analysis (CCA), LDA was shown by Zhang et al. to optimize the SSVEP spellers [12]. In language BCIs based on the user's intended movements of their limbs to control a cursor to spell, LDA is often coupled with the use of Common Spatial Patterns (CSP) that extract filtered EEG data. LDA, though most utilized as a binary classifier, can also be modified to be implemented in a one-to-one or one-to-many manner to classify data into more than two classes.

The limitations of LDA arise from dimensionality. The basic principle behind LDA is to identify a hyperplane in high-dimensional space that can separate the data into 2 or more clusters with the maximum distance between clusters while minimizing the distance between points within clusters. Mathematically, it tries to find a weight vector θ with which each data point in the matrix x is multiplied to give the labeled matrix y.

$$y = sign(\theta x + b) \tag{1}$$

Each point in the matrix x is multiplied by θ and added to the threshold value b to produce a data point in y. The hyperplane in LDA that separates the two classes is found using the calculated values of the means of each class and their covariance matrices. It is worth noting that the covariance matrix has the dimensionality of the feature space squared. This means that in a standard 32-electrode EEG or ECoG data feature, the covariance matrix of the LDA is 32x32. A large covariance matrix such as this means that in the training phase of this supervised model, much more data needs to be collected to make reliable predictions or estimations.

3.1.2. Support Vector Machines (SVMs)

While LDA assumes the data as normally distributed and that all classes have the same covariance matrix, SVM, on the other hand, does not make those assumptions. Although both methods aim to identify a boundary hyperplane that best distinguishes the classes, SVM achieves this through

DOI: 10.54254/2753-8818/2025.LD26088

calculating the support vectors, which are the points of data that lie closest to the hyperplane boundary. Then, SVM aims to optimize to maximize the margin distance between the hyperplane boundary and the support vectors. The second goal of SVM is to optimize for minimizing the norm squared of the weight vector θ, which is often expressed as $\frac{1}{2}||\theta||^2$. From this equation, it is also deduced that minimizing |(|θ|) | implies maximizing the margin $\frac{2}{||\theta||}$, as they are inversely proportional. Hence, the two goals of SVM are one.

SVMs can also be enhanced to non-linear boundaries using kernel functions, a capability not available in LDA methods. In BCI practices, SVM often works well with sparse data. Because SVM only calculates the support vectors instead of all data points like LDA, it is less affected by noise or outliers. In EEG-based BCIs, Li et al. found that semi-supervised SVM can reduce the training time in P300 spellers, increasing their accessibility [13]. Gaussian-kernel versions of SVM have performed among the top classifiers [14]. In the same study, Ghane and Hossain applied PCA dimensionality reduction and then used a decision-tree SVM to successfully classify imagined vowel sounds. These studies showed that SVMs, when combined with task-specific features, become efficient decoding tools for language BCIs.

3.2. Deep learning

Unlike supervised learning, deep learning systems can learn features from raw data, without the need for labels. Because language BCIs depend on subject-specific data do not benefit from large public datasets in the same way as image or text classification tasks do, deep learning seems like a sensible alternative.

3.2.1. Convolutional Neural Networks (CNNs)

CNNs are a type of deep learning model that is adept at processing grid-like data such as images, time signals, and in BCIs, the spatiotemporal patterns from EEG or ECoG data. CNNs, which are developed originally for image recognition purposes, are inspired by the structure of the visual system in animals. Neurons are organized in a hierarchy, and each responds only to a local space in the input pool. The “convolution” in CNN comes from a filtering kernel that is applied to the data to extract features for identification. These filters and features are not inputted or labeled but instead emerge from backpropagation automatically. CNNs can detect multiple features at once. Upon receiving an input like EEG signal , the CNN passes it forward through its layers to produce an output, such as a label prediction. Afterwards, the model checks how different this is from the truth using a loss function, computes the derivatives of the loss function, and works back from the output to the input here. Through the repetition of forward pass and backpropagation over many trials, the model makes predictions that are more accurate over time.

CNNs have been used to analyze motor 1D EEG data [15], in 2D EEG data to reflect both spatial and temporal aspects of motion [16], and 3D ECoG data to decode finger movement in 3D space [17]. In language BCIs specifically, a recent study presented a multi-channel CNN that integrated the outputs of multiple CNNs that processed EEG activity and achieved 97.67% accuracy [18]. CNNeeg1-1 can classify Spanish vowels with EEG activity as input and, in a dataset of 50, its performance was superior to baseline models [19].With CNN operating on ECoG data, researchers produced high-quality audio reconstructions of imagined speech [20]. Methods like EEGNet are proposed to specifically tackle the problem of the difference between neural data’s small sample

DOI: 10.54254/2753-8818/2025.LD26088

size, noise, and variability and traditional image-classifying CNN training data. EEGNet extrapolates spatial and frequency patterns in EEG using more compact layers.

3.2.2. Recurrent Neural Networks (RNNs)

RNNs are specialized for model sequential data like time-series, text, or speech, which is instrumental in language BCIs. Unlike CNNs that feed forward, RNNs retain a memory that changes over time according to $h_t = f_W(h_{t-1}, x_t)$, where h_t represents the current state, h_{t-1} the preceding state, x_t an input vector at some timestamp, and f_W a non-linear function that is defined by its parameter W. It represents how the RNN updates its memory over time. Traditional RNNs do suffer from issues with gradients that limit memory length. Optimizations such as Long Short-Term Memory (LSTM) networks have been developed to regulate the inputting, processing, and forgetting of memories while keeping gradients intact. These methods are selective in what memories the RNN system retains and forgets. Mahapatra et al. performed an imagined number classification with a bidirectional LSTM model that is able to generate a 96.18% accuracy in predictions [21]. Similarly, Agarwal & Kumar compared LSTM networks, GRUs, and CNNs in a speech identification task across 13 patients. The deep LSTM network achieved the highest accuracy [22].

Despite high accuracy, deep learning models have great limitations. The limitation of CNN used in language BCIs is that, as Wei et al. put it, “CNNs cannot directly use multiple subjects' EEG to enhance model performance directly” [23]. Additionally, for deep learning specifically, the hidden layers in between the input and the output are difficult to understand and interpret. And, for BCI research, which is still in its infancy, it is instrumental to understand the how behind the results obtained, for example, whether the results correspond to previous neuroscience research or physiological understanding.

4. Discussion

In terms of input data, the challenge remains the tradeoff between invasive but higher-resolution ECoGs or the non-invasive but low-resolution EEGs. The invasive techniques naturally carry risks to participants and limit applicability and generalizability. EEGs are also limited by their high noise, which either requires complex annotations or is otherwise unable to analyze complex or fine tasks. The ethical implications of long-term implantation and adoption or commercialization, even of non-invasive BCI methods, remain a topic of debate.

From a machine learning perspective, the scarcity of data, partially caused by the rarity of patients but also due to invasive measures, is another challenge. Most machine learning models are only trained on a few subjects, and variability, especially in EEG data, exists. Consequently, retraining and calibration are required for each user, which is not scalable or efficient. The data needed for supervised learning to train and be tested on, moreover, requires labeling. Hours of EEG or ECoG data need to be timestamped, which is not efficient. Those labeled data are usually produced in a highly controlled environment to limit noise or artifact, but this also contributes to the limited ecological validity.

Whereas supervised models lack reliable and scalable data, deep learning methods lack interpretability, which complicates the application. Often, the more iterations and optimizations a deep learning neural network undergoes, the more difficult it is for their developers to understand them. Especially as the BCI system operates on a human, when a detrimental error occurs, legal and moral accountability can be difficult to place. Furthermore, neural networks are computationally

DOI: 10.54254/2753-8818/2025.LD26088

costly and often need the participant to be connected to a device with a GPU hardware, which can be infeasible in daily actions outside of laboratories.

To address some of these limitations, new research can be and has been heading in the following directions.

For example, Li, Zhang & Ming combined EEG and fNIRS signals to improve motor imagery model's performance [24]. A similar approach could be applied to language BCIs with the integration of multiple signal input methods. Transfer learning models pretrained with one modality of data, for example, EEG data, could be fine-tuned with another modality to reduce pretraining. Unsupervised methods such as BENDR, which is a BERT algorithm developed by Kostas et al. at the University of Toronto, can use unsupervised methods to check portions of BCI performance, reducing supervised learning and labeling datasets [25]. As for deep learning models, although the accuracy in the classifications may be high, the researchers do not know what features of the input data resulted in that classification, and it therefore makes mistakes difficult to correct. Efforts have been made towards the deep learning black-box comprehensible, though, through "a modification of the activation maximization approach for regression tasks" [26] and through visualization [27].

Overall, research into input methods often either admits the tradeoff between fidelity and invasiveness or attempts to integrate other methods with EEG. Meanwhile, research into machine learning methods tends to incorporate bidirectional reinforcement or unsupervised learning to streamline the inefficient process of labeling training data.

5. Conclusion

The goal of language BCIs is to restore natural communication to people who have lost that vital ability. EEGs, while non-invasive, are noise-ridden; ECoGs, though invasive, decrease artifacts. Simple supervised methods offer data-efficiency, but are hardly scalable, whereas deep learning achieves highly accurate, but remains difficult to interpret. As new researchers experiment with diverse and integrative solutions to these challenges, and through more research into input methods, the "hardware" of language BCIs, and machine learning models, the "software", we can reasonably foresee a future where non-invasive, high-quality, and efficient systems can help those who need them most.

Acknowledgments

The author expresses gratitude to Professor Joni Wallis from the University of California, Berkeley, for all her valuable and insightful guidance related to the project.

References

[1] P. L. Nunez and R. Srinivasan. A theoretical basis for standing and traveling brain waves measured with human EEG with implications for an integrated consciousness. Clinical Neurophysiology, vol. 117, no. 11, pp. 2424–2435, Nov. 2006, doi: https: //doi.org/10.1016/j.clinph.2006.06.754.

[2] J. E. Huggins, P. A. Wren, and K. L. Gruis. What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis, vol. 12, no. 5, pp. 318–324, May 2011, doi: https: //doi.org/10.3109/17482968.2011.572978.

[3] S. Martin, I. Iturrate, J. del R. Millán, R. T. Knight, and B. N. Pasley. Decoding inner speech using Electrocorticography: progress and challenges toward a speech prosthesis. Frontiers in Neuroscience, vol. 12, Jun. 2018, doi: https: //doi.org/10.3389/fnins.2018.00422.

[4] V. L. Towle et al. ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain, vol. 131, no. 8, pp. 2013–2027, Aug. 2008, doi: https: //doi.org/10.1093/brain/awn147.

DOI: 10.54254/2753-8818/2025.LD26088

[5] E. C. Leuthardt et al. Using the electrocorticographic speech network to control a brain–computer interface in humans. Journal of Neural Engineering, vol. 8, no. 3, p. 63–71, Apr. 2011, doi: https: //doi.org/10.1088/1741-2560/8/3/036004.
[6] S. Duraivel et al. High-resolution neural recordings improve the accuracy of speech decoding. Nature Communications, vol. 14, no. 1, p. 6938, Nov. 2023, doi: https: //doi.org/10.1038/s41467-023-42555-1.
[7] A. L. Ritaccio et al. Proceedings of the eighth international workshop on advances in Electrocorticography. Epilepsy & Behavior, vol. 64, pp. 248–252, Nov. 2016, doi: https: //doi.org/10.1016/j.yebeh.2016.08.020.
[8] G. Schalk et al. Two-dimensional movement control using electrocorticographic signals in humans. Journal of Neural Engineering, vol. 5, no. 1, pp. 75–84, Feb. 2008, doi: https: //doi.org/10.1088/1741-2560/5/1/008.
[9] M. Tangermann et al. Review of the BCI Competition IV. Frontiers in Neuroscience, vol. 6, 2012, doi: https: //doi.org/10.3389/fnins.2012.00055.
[10] C. Herff, G. D. Johnson, L. Diener, J. J. Shih, D. J. Krusienski, and T. Schultz. Towards direct speech synthesis from ECoG: a pilot study. Proc. Interspeech, San Francisco, CA, USA, Aug. 2016, doi: https: //doi.org/10.1109/embc.2016.7591004.
[11] D. J. Krusienski et al. A comparison of classification techniques for the P300 Speller. Journal of Neural Engineering, vol. 3, no. 4, pp. 299–305, Oct. 2006, doi: https: //doi.org/10.1088/1741-2560/3/4/007.
[12] Y. Zhang, P. Xu, T. Liu, J. Hu, R. Zhang, and D. Yao. Multiple frequencies sequential coding for SSVEP-based Brain-Computer Interface. vol. 7, no. 3, pp. e29519–e29519, Mar. 2012, doi: https: //doi.org/10.1371/journal.pone.0029519.
[13] Y. Li, C. Guan, H. Li, and Z. Chin. A self-training semi-supervised SVM algorithm and its application in an EEG-based brain computer interface speller system. Pattern Recognition Letters, vol. 29, no. 9, pp. 1285–1294, Jul. 2008, doi: https: //doi.org/10.1016/j.patrec.2008.01.030.
[14] P. Ghane and G. Hossain. Learning patterns in imaginary vowels for an intelligent brain computer interface (BCI) design. arXiv.org, 2020. https: //arxiv.org/abs/2010.12066.
[15] F. Mattioli, C. Porcaro, and G. Baldassarre. A 1D CNN for high accuracy classification and transfer learning in motor imagery EEG-based brain-computer interface. Journal of Neural Engineering, vol. 18, no. 6, p. 066053, Dec. 2021, doi: https: //doi.org/10.1088/1741-2552/ac4430.
[16] J. Yang, S. Gao, and T. Shen. A two-branch CNN fusing temporal and frequency features for motor imagery EEG decoding. Entropy, vol. 24, no. 3, p. 376, Mar. 2022, doi: https: //doi.org/10.3390/e24030376.
[17] M. Śliwowski, M. Martin, A. Souloumiac, P. Blanchart, and T. Aksenova. Decoding ECoG signal into 3D hand translation using deep learning. Journal of Neural Engineering, vol. 19, no. 2, p. 026023, Mar. 2022, doi: https: //doi.org/10.1088/1741-2552/ac5d69.
[18] S. Tiwari, S. Goel, and A. Bhardwaj. Classification of imagined speech of vowels from EEG signals using multi-headed CNNs feature fusion network. Digital Signal Processing, vol. 148, pp. 104447–104447, May 2024, doi: https: //doi.org/10.1016/j.dsp.2024.104447.
[19] L. C. Sarmiento, S. Villamizar, O. López, A. C. Collazos, J. Sarmiento, and J. B. Rodríguez. Recognition of EEG signals from imagined vowels using deep learning methods. Sensors, vol. 21, no. 19, p. 6503, Sep. 2021, doi: https: //doi.org/10.3390/s21196503.
[20] M. Angrick et al. Speech synthesis from ECoG using densely connected 3D convolutional neural networks. Journal of Neural Engineering, vol. 16, no. 3, pp. 036019, Mar. 2019, doi: https: //doi.org/10.1088/1741-2552/ab0c59.
[21] N. C. Mahapatra and P. Bhuyan. EEG-based classification of imagined digits using recurrent neural network. Journal of Neural Engineering, vol. 19, no. 5, p. 056040, Mar. 2023, doi: https: //doi.org/10.1088/1741-2552/acc976.
[22] P. Agarwal and S. Kumar. Electroencephalography-based imagined speech recognition using deep long short-term memory network. ETRI Journal, vol. 44, no. 2, pp. 306–317, Jun. 2022, doi: https: //doi.org/10.4218/etrij.2021-0118.
[23] X. Wei, P. Ortega, and F. A. Aldo. Inter-subject deep transfer learning for motor imagery EEG decoding. arXiv.org, 2021. https: //arxiv.org/abs/2103.05351.
[24] Y. Li, X. Zhang, and D. Ming. Early-stage fusion of EEG and fNIRS improves classification of motor imagery. Frontiers in Neuroscience, vol. 16, p. 551, Jan. 2023, doi: https: //doi.org/10.3389/fnins.2022.1062889.
[25] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz. BENDR: using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. arXiv: 2101.12037 [cs, q-bio], Jan. 2021, Available: https: //arxiv.org/abs/2101.12037.
[26] M. Angrick, C. Herff, G. D. Johnson, and T. Schultz. Interpretation of Convolutional Neural Networks for speech regression from Electrocorticography. ResearchGate, Apr. 25, 2018. https:

DOI: 10.54254/2753-8818/2025.LD26088

//www.researchgate.net/publication/325781566_Interpretation_of_Convolutional_Neural_Networks_for_Speech_Regression_from_Electrocorticography.

[27] R. T. Schirrmeister et al. Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping, vol. 38, no. 11, pp. 5391–5420, Aug. 2017, doi: https: //doi.org/10.1002/hbm.23730.