

Comparison of Lightweight EEG Pipelines for Silent Speech Interfaces

E. Nikolopoulos^{1,2}, G. Kontaxakis^{3,4}, E. J. Gómez^{1,2,5}, I. Oropesa^{1,2,5}

¹ Biomedical Engineering and Telemedicine Centre, ETSI Telecomunicación, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain

² Instituto de Investigación Hospital 12 de Octubre (imas12), Hospital Universitario 12 De Octubre, Madrid, Spain

³ ETSI Telecomunicación, Universidad Politécnica de Madrid (UPM), Madrid, España

⁴ Information Processing and Telecommunications Center, ETSI Telecomunicación, UPM, Madrid, Spain

⁵ Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, Madrid, Spain

E-mail: manosnikol@gmail.com

ABSTRACT: This study investigates various machine learning (ML) and deep learning (DL) methodologies for decoding inner speech using non-invasive electroencephalography (EEG). Utilizing a publicly available high-density 128-channel dataset from ten participants, nine distinct computational scenarios were tested, ranging from traditional spatial filtering and statistical ensembles to Riemannian geometry and convolutional neural networks. The research aims to compare and identify the most effective signal processing strategies to overcome the inherently low signal-to-noise ratio of EEG signals generated during inner speech. Results demonstrate that Riemannian manifold-based approaches and compact deep learning architectures significantly outperform traditional Euclidean-based methods, achieving average accuracies across the ten participants of 32.41% and 30.48%, with peak accuracies reaching 44.50% and 34.50% respectively, compared with the 25% chance level of a four-class inner speech task.

INTRODUCTION

A brain-computer interface (BCI) serves as a computational bridge between neural activity and external hardware. These systems function by acquiring brain signals, decoding the underlying intent, and translating that data into digital commands to operate devices such as computers or prosthetic limbs [1,2]. Beyond assistive control, BCIs can also be used to monitor cognitive states such as fatigue and mental workload and support a wide range of applications. Silent Speech Interfaces (SSIs) represent a class of communication systems that enable speech processing in the absence of an intelligible acoustic signal [3]. Unlike traditional speech recognition that relies on microphone input, an SSI captures physical or physiological data

produced during the speech production process, such as brain signals, movement of the articulators or surface muscle activation. By decoding these internal precursors to sound, the system can generate text or synthesized speech, effectively bypassing the requirement for vocalization [4].

The development of SSIs relies on a diverse array of sensing modalities designed to capture articulatory intent without the need for acoustic output. These technologies are broadly categorized into methods that measure physiological activity directly and those that track physical articulatory movements [5]. A critical frontier in this domain is the intersection of SSIs with BCIs, which bypass peripheral articulators entirely to decode the cognitive intent to speak. While invasive neural approaches like electrocorticography capture high-resolution signals with superior signal-to-noise ratio (SNR) directly from the speech-motor cortex [6–9], non-invasive electroencephalography (EEG) remains the preferred practical choice for monitoring scalp-level brain activity due to its safety, affordability, and high temporal resolution [10–13]. EEG-based BCIs capture the cortical electrical activity generated during the cognitive formulation of words, which is a process widely referred to in the literature as 'covert' or 'imagined' speech [14]. By measuring voltage fluctuations at the scalp, researchers can record the high-temporal-resolution neural dynamics associated with language production, completely bypassing the need for physical articulation. This non-invasive approach offers a safe, portable, and highly accessible alternative to surgical neural implants.

The state-of-the-art (SoA) processing pipeline for these EEG-based covert speech decoders heavily relies on robust artifact removal, such as Independent Component Analysis (ICA) [15], followed by advanced feature extraction [14]. Recently, classification strategies have experienced a paradigm shift; contemporary research

increasingly leverages deep learning (DL) architectures, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, which autonomously extract complex, non-linear spatiotemporal patterns from raw or minimally preprocessed multichannel EEG data to significantly improve decoding accuracy [14].

Despite these algorithmic advancements, the practical deployment of EEG-based BCI systems for silent speech remains constrained by fundamental physiological and hardware challenges. The inherently low SNR of EEG signals and high inter-subject variability restrict current systems primarily to offline, synchronous environments that utilize highly limited vocabularies, such as isolated vowels, phonemes, or simple directional commands [13]. Furthermore, multi-class classification accuracy significantly deteriorates as the vocabulary size expands, and many experimental protocols still rely on the unnatural mental repetition of prompts to achieve viable decoding rates [14]. Crucially, the implementation of advanced deep learning architectures demands immense computational processing power, which is exceedingly difficult to achieve in real-world, standalone neuroprosthetic devices. This stringent hardware requirement severely restricts the portability of these systems and limits their practical, everyday applications outside of controlled laboratory settings.

The goal of this study is to investigate different methodologies used in machine learning (ML) and DL in the field of EEG-driven SSIs, with a focus on minimizing computational requirements for real-world deployment. Using an open access dataset, we compare the classification performance of several algorithms across different methodological scenarios. More specifically, we perform intrasubject classification, and compare the average accuracy across participants and the maximum accuracy achieved among nine different scenarios, while maintaining low computational requirements to ensure feasibility for real-world applications.

MATERIALS AND METHODS

Dataset description: In this work, we analyzed the publicly available dataset by Nieto et al. [16]. The study provides a comprehensive, open-access EEG resource specifically designed to advance the study of inner speech and BCIs. The dataset contains over nine hours of continuous EEG recordings. It significantly contributes in the neuroengineering research by offering a publicly available, high-density EEG database of inner, pronounced and visualized speech commands that can be used to better understand speech-related brain mechanisms and train robust classification models.

The dataset features ten healthy, right-handed, native Spanish-speaking adults (mean age 34; four females, six males) with no history of neurological or speech disorders and no prior BCI experience. The data acquisition was performed in an electrically shielded room utilizing a high-resolution, 136-channel BioSemi Active Two system. The hardware setup prioritized

dense cortical coverage with 128 active EEG channels, supplemented by 8 external channels for electrooculography (EOG) and electromyography (EMG) to rigorously monitor and control for eye and articulatory movement artifacts. The EEG signals were initially recorded at a high sampling rate of 1024 Hz with 24-bit resolution. To ensure optimal signal quality, researchers fitted each subject with appropriately sized head caps, applied conductive gel, and strictly maintained electrode impedances below 40 ohms prior to all recording sessions.

The experimental protocol, as described on Fig. 1, yielded a total of 5,640 EEG trials across all participants and across the three distinct paradigms: inner speech, pronounced speech, and a visual spatial condition. During the tasks, participants responded to visual cues corresponding to four directional commands ("arriba," "abajo," "derecha," "izquierda").

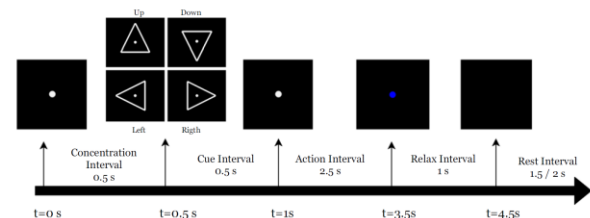


Figure 1: The experimental procedure described in [16]. Each trial spanned 4.5 seconds, starting with a 0.5-second concentration interval where participants focused on a white circle. This was followed by a 0.5-second cue interval displaying a directional triangle (up, down, left, or right). At $t = 1.0$ s, the cue disappeared, initiating a 2.5-second action interval where the subject performed the designated mental or vocal task. At $t = 3.5$ s, the circle turned blue to signal a 1.0-second relax interval, concluding the trial before a variable rest period began. Modified from Nieto et al. [16] under the Creative Commons CC-BY 4.0 license.

In this work, we focused exclusively on the data obtained during the inner speech condition. This subset comprised a total of 2,236 trials, yielding an average of approximately 224 total inner speech trials per participant (averaging 56 trials per class). Inner speech can be defined as an internalized cognitive process in which individuals think in pure meanings, typically accompanied by the auditory imagery of their own inner “voice”. Unlike overt speech, inner speech occurs without physical articulation or acoustic output. Specifically, during these trials, participants were instructed to mentally repeat the cued word using their own voice as a direct command to the computer until the on-screen circle turned blue. To ensure purely covert generation, they were not provided with any rhythmic cues, and they were explicitly directed to remain completely still, avoiding any tongue or mouth movements or focus on articulatory gestures.

Preprocessing of the data: A major advantage of this dataset is that it provides both the continuous raw data and precisely preprocessed EEG epochs. The data processing pipeline, as provided by [16], implemented primarily using the MNE Python library, began with re-referencing the signals using a virtual channel created by averaging the two earlobe electrodes (EXG1 and EXG2)

to eliminate common-mode voltage. The signals were then band-pass filtered between 0.5 and 100 Hz using a zero-phase finite impulse response filter, along with a 50 Hz notch filter to remove power line noise interference. Next, the continuous data was decimated to a 256 Hz sampling rate and epoched into 4.5-second segments corresponding to individual trials. Finally, ICA using the infomax algorithm was applied strictly to the 128 EEG channels to isolate and remove artifacts related to blinks, gaze, and mouth movements, resulting in ready-to-use data tensors tailored for BCI classification and time-frequency analysis [17].

Word Classification: To provide a comprehensive analysis of inner speech classification, the following methodologies represent a diverse range of EEG signal processing strategies. These scenarios span from traditional spatial filtering to advanced adaptive denoising, Riemannian geometry, and deep learning.

Scenario 1: Multi-Band Common Spatial Patterns (CSP). This approach utilizes CSP to find spatial filters that maximize the variance for one class while simultaneously minimizing it for another [18]. The methodology tests three distinct frequency bands, motor-related (8–30 Hz), broadband (1–45 Hz), and gamma (30–80 Hz), to determine which oscillatory rhythms provide the most discriminative information for inner speech. The spatially filtered features are then fed into a Linear Discriminant Analysis (LDA) classifier to perform the final classification of the inner speech targets [19].

Scenario 2: Region of Interest (ROI) Spatial Filtering. This methodology employs a physiologically driven channel reduction strategy, focusing exclusively on 39 specific channels that correspond to the left hemisphere's speech-dominant regions and the midline, such as Broca's area, Wernicke's area, the primary motor cortex, and the supplementary motor area. Consequently, 89 of the 128 total channels are explicitly removed to eliminate irrelevant spatial data. This removal includes the entire right hemisphere (all 32 'B' channels), the majority of the 'A' block (channels A3–A32), as well as specific peripheral 'C' (C3–C11, C28–C32) and 'D' channels (D1–D5, D25–D32). By filtering out this irrelevant spatial data before applying CSP, the pipeline concentrates the classifier's attention on the cortical areas most likely to generate language-related signals during inner speech. After applying CSP to these targeted spatial regions, the extracted features are classified using a Support Vector Machine (SVM) with an RBF kernel to decode the specific inner speech commands [20].

Scenario 3: Statistical Feature Ensemble. This strategy treats EEG signals as a source for high-dimensional statistical modeling by extracting ten specific features per channel, including time-domain measures (mean, skewness, kurtosis, root mean square) and frequency-domain band powers. It uses mutual information to select the top 500 features and employs a soft-voting ensemble of five logistic regression models to improve classification stability and mitigate the risk of overfitting.

Scenario 4: Multi-Wavelet and Autoregressive Modeling. This methodology captures the complexity of neural signals through a combination of temporal and frequency-scale analysis. It extracts autoregressive (AR) coefficients to model temporal dynamics and uses wavelet packet shannon entropy and wavelet variance to quantify the complexity and variability of the signal across different frequency scales. A third-degree polynomial SVM is then used to handle the non-linear boundaries of these features.

Scenario 5: Optimized ROI and Spectral Feature Engineering. This methodology refines the ROI approach by combining dynamic channel subsetting with targeted spectral extraction during the active "thought" window (1.5–3.5s). It extracts statistical features alongside power in the alpha (8–13 Hz) and beta (13–30 Hz) bands from the left hemisphere. A mutual information selector then retains the top 100 features, which are then fed to an SVM.

Scenario 6: Variational Mode Decomposition (VMD). This advanced signal processing methodology uses VMD to adaptively decompose EEG signals into Intrinsic Mode Functions (IMFs). By reconstructing the signal using only relevant middle modes that capture alpha and beta activity, it acts as a non-linear filter to remove low-frequency drift and high-frequency noise. The cleaned signal is then processed for statistical and spectral features before classification via a logistic regression ensemble.

Scenario 7: Riemannian Geometry and Filter-Banks. This methodology approaches EEG classification by treating signal covariance matrices as points on a curved Riemannian manifold. It utilizes a filter-bank approach across three frequency bands, applying principal component analysis (PCA) for dimensionality reduction and Xdawn covariances to enhance the signal-to-noise ratio. The matrices are then mapped into a tangent space, allowing for robust linear classification [21].

Scenario 8: Linear Predictive Coding (LPC) Formant Analysis. Borrowing from speech processing, this methodology treats EEG channels as if they were audio sources capturing the "inner voice". It applies a pre-emphasis filter to flatten the spectrum and then calculates LPC coefficients to model the spectral envelope of the signal. These coefficients, representing "formants" of inner speech, are classified using a gradient-boosted XGBoost model [22].

Scenario 9: Deep Learning with EEGNet. This methodology employs EEGNet [23], a compact CNN designed specifically for EEG data. It uses depthwise and separable convolutions to learn both temporal and spatial filters directly from raw, broadband EEG data (1–40 Hz). This allows the model to automatically discover discriminative patterns across the array without manual feature engineering [24].

To evaluate the robustness of the proposed models, the classification performance was assessed using a 5-fold stratified cross-validation strategy following an intra-subject approach. This approach ensures that each training and testing fold maintains an equal

representation of the four inner speech classes, providing a reliable estimate of the models' generalization capabilities across different data subsets while preventing class imbalance bias [25]. While the stratified cross-validation utilized shuffled trials to maximize training set diversity, the temporal integrity of the data was preserved through prior ICA artifact removal and the use of strictly unsupervised spatial filtering within each fold to prevent data leakage from future signals. Given the finite sample size, the theoretical chance level of 25% is insufficient; therefore, the adjusted statistical significance threshold for non-random classification was calculated to be 29.8% based on the binomial distribution ($\alpha = 0.05$) for an average of 224 trials per subject, following the framework described by [26].

RESULTS

Tab. 1 presents a comparative performance analysis of the abovementioned nine distinct classification scenarios for inner speech, highlighting the average testing and maximum accuracy achieved by each methodology across a 10-subject dataset. The results indicate that Scenario 7 (Riemannian Geometry) is the most effective methodology for this dataset, achieving the highest average accuracy of 32.41% and a peak max accuracy of 44.5%. By treating EEG covariance matrices as points on a curved manifold and utilizing Xdawn covariances, this approach successfully managed the low signal-to-noise ratio inherent in inner speech. It significantly outperformed traditional spatial filtering methods like Scenario 1 (CSP), which struggled to exceed the adjusted 29.8% random chance threshold for this 4-class problem.

Table 1: Performance metrics for inner speech classification across nine experimental scenarios, detailing the grand average accuracy (%), and peak subject performance (%).

Object	Average Accuracy	Max Accuracy
Scenario 1	24.43 (%)	33.83 (%)
Scenario 2	23.3 (%)	28.9 (%)
Scenario 3	26.2 (%)	31.1 (%)
Scenario 4	24.9 (%)	27.5 (%)
Scenario 5	25.3 (%)	28.9 (%)
Scenario 6	24.4 (%)	28.5 (%)
Scenario 7	32.41 (%)	44.5 (%)
Scenario 8	24.5 (%)	30.4 (%)
Scenario 9	30.48 (%)	34.5 (%)

Scenario 9 (EEGNet) emerged as the second-strongest performer with an average accuracy of 30.48%. The success of this DL approach suggests that the native spatial-temporal filters learned by the CNN are better at capturing the complex, non-linear patterns of broadband EEG data than most manual feature engineering methods. While Scenario 3 (Statistical Ensemble) and Scenario 5 (Optimized ROI/Stats) showed slight improvements over the baseline by reaching roughly 25-26% average accuracy, they remained notably behind the Riemannian and Deep Learning pipelines in terms of

consistency and peak performance.

Interestingly, specialized signal processing techniques such as Scenario 6 and Scenario 8 did not yield the expected performance breakthroughs, hovering around 24.4% to 24.5% average accuracy.

DISCUSSION

The experimental results highlight a significant performance gap between manifold-based or automated feature extraction methods and traditional manual feature engineering. Scenario 7, which utilized Riemannian geometry and Xdawn covariances, proved the most robust, suggesting that treating EEG signal covariance as points on a curved manifold better captures the subtle neural dynamics of inner speech than standard spatial filters like CSP. CSP-based methods (Scenarios 1) struggled to maintain performance above the adjusted random chance threshold, likely due to the high spatial attenuation and non-stationarity of the EEG signal across the 128-channel array.

The insufficient accuracies of the specialized signal processing techniques of Scenarios 6 and 8 suggest that while VMD is effective at cleaning signals and LPC is theoretically sound for speech-related brain activity, the high dimensionality of the 128-channel data may require the more robust manifold-based or automated feature extraction found in Scenarios 7 and 9. Ultimately, the transition from Euclidean-based statistics to Riemannian geometry provided the most substantial leap in classification reliability for this inner speech task.

Furthermore, the results of Scenario 9 (EEGNet) indicate that DL models can more effectively discover latent spatial-temporal patterns in broadband EEG data without the need for manual physiological priors. While specialized techniques such as VMD (Scenario 6) and LPC (Scenario 8) are theoretically grounded in speech and signal processing, they did not yield substantial performance gains in this context. This suggests that the high dimensionality of 128-channel EEG requires either the robust geometric frameworks of Riemannian manifold mapping or the hierarchical feature learning of neural networks to reliably differentiate between complex linguistic cognitive intents.

Our best models, Riemannian Geometry (32.41%) and EEGNet (30.48%), successfully replicate the standard ~30% baseline for the highly noisy Nieto et al. dataset [16,27,28], but fall short of the 45–81% accuracies achieved by the current SoA [29,30]. This performance gap highlights a strict computational trade-off. SoA DL methods, such as 3D-CNNs and spatial-temporal cross-attention networks, achieve high accuracy but require massive parameter counts, billions of FLOPs, and intensive multi-GPU training environments [29,31]. Meanwhile, top-performing ML approaches, like ensemble logistic regression or multi-wavelet SVMs, avoid deep learning's training costs but shift the heavy computational burden onto exhaustive, resource-intensive feature extraction and selection processes [30,32]. Ultimately, bridging the gap between our ~32%

peak accuracy and the SoA will require sacrificing computational efficiency to implement either complex spatial neural modeling or heavy feature engineering. Despite the promise of Riemannian mapping and convolutional architectures, key limitations regarding model capacity and generalization remain. The small cohort of ten participants yields a restricted number of trials, which inherently increases the risk of high-capacity models like EEGNet overfitting to this specific public dataset. As demonstrated by recent cross-session BCI competitions [33], complex neural network architectures frequently suffer drastic performance regression on unseen data compared to traditional methods. Therefore, before committing to high-capacity hypotheses for real-world neuroprosthetic deployment, these architectures must be rigorously evaluated on independent, out-of-distribution data to verify true generalization. Additionally, the 32.41% peak average accuracy underscores the need for broader algorithmic exploration. Future research must investigate advanced architectures to achieve the robust accuracy required for functional BCI communication. Most notably, the emerging field of Geometric Deep Learning [34–36] offers a highly promising trajectory by explicitly combining the manifold robustness of Riemannian geometry with the automated feature extraction capabilities of neural networks. Crucially, future explorations must balance high accuracy with the computational efficiency required for real-world deployment. High-density 128-channel setups and complex mathematical operations, such as Riemannian covariance calculations, impose severe burdens on portable neuroprosthetic hardware. A primary limitation of current research is sustaining classification accuracy while drastically reducing this computational overhead. To realize accessible and lightweight EEG-based silent speech interfaces, future work must prioritize intelligent channel reduction, algorithmic optimization, and edge-computing frameworks.

CONCLUSION

This research concludes that Riemannian geometry provides the most substantial leap in classification reliability for EEG-driven silent speech interfaces, achieving a peak accuracy of 44.5%. Deep learning via EEGNet also offers a highly competitive and automated alternative for decoding imagined speech. The findings suggest that future development in non-invasive SSIs should move away from traditional Euclidean-based statistics in favor of geometric manifold learning and spatial-temporal neural architectures to handle the complex computational hurdles of neural language production.

REFERENCES

[1] Cruz MV, Jamal S, Sethuraman SC. A Comprehensive Survey of Brain-Computer Interface Technology in Health care: Research

- Perspectives. *J Med Signals Sens* 2025;15.
- [2] Zhang H, Jiao L, Yang S, Li H, Jiang X, Feng J, et al. Brain-computer interfaces: the innovative key to unlocking neurological conditions. *Int J Surg* 2024;110:5745–62.
- [3] Ferreira D, Silva S, Curado F, Teixeira A. Exploring Silent Speech Interfaces Based on Frequency-Modulated Continuous-Wave Radar. *Sensors* 2022;22.
- [4] Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. *Nature* 2019;568:493–8.
- [5] Denby B, Schultz T, Honda K, Hueber T, Gilbert JM, Brumberg JS. Silent speech interfaces. *Speech Commun* 2010;52:270–87.
- [6] Understanding and Decoding Imagined Speech using Electrographic Recordings in Humans Stéphanie MARTIN. n.d.
- [7] SpringerBriefs in Speech technology. n.d.
- [8] Chakrabarti S, Sandberg HM, Brumberg JS, Krusienski DJ. Progress in speech decoding from the electrocorticogram. *Biomed Eng Lett* 2015;5:10–21.
- [9] Luo S, Rabbani Q, Crone NE. Brain-Computer Interface: Applications to Speech Decoding and Synthesis to Augment Communication. *Neurotherapeutics* 2022;19:263–73.
- [10] Vorontsova D, Menshikov I, Zubov A, Orlov K, Rikunov P, Zvereva E, et al. Silent eeg-speech recognition using convolutional and recurrent neural network with 85% accuracy of 9 words classification. *Sensors* 2021;21.
- [11] H Chowdhury ME, Sarmun R, Bushnaq D, Chabbouh M, Aljindi R, Pedersen S, et al. IJABC: International Journal of Activity and Behavior Computing Summary of the Silent Speech Decoding Challenge Using EEG Data for Arabic Inner Speech Recognition Summary of the Silent Speech Decoding Challenge Using EEG Data for Arabic Inner Speech Recognition IJABC: International Journal of Activity and Behavior Computing. n.d.
- [12] Lopez-Bernal D, Balderas D, Ponce P, Molina A. A State-of-the-Art Review of EEG-Based Imagined Speech Decoding. *Front Hum Neurosci* 2022;16.
- [13] Rahman N, Khan DM, Masroor K, Arshad M, Rafiq A, Fahim SM. Advances in brain-computer interface for decoding speech imagery from EEG signals: a systematic review. *Cogn Neurodyn* 2024;18:3565–83.
- [14] Panachakel JT, Ramakrishnan AG. Decoding Covert Speech From EEG-A Comprehensive Review. *Front Neurosci* 2021;15.
- [15] Chuang CH, Ko LW, Lin YP, Jung TP, Lin CT. Independent component ensemble of EEG for brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 2014;22:230–8.
- [16] Nieto N, Peterson V, Rufiner HL, Kamienkowski

- JE, Spies R. Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition. *Sci Data* 2022;9.
- [17] Bell AJ, Sejnowski TJ. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Comput* 1995;7:1129–59.
- [18] Moufassih M, Tarahi O, Hamou S, Agounad S, Idrissi Azami H. Boosting motor imagery brain-computer interface classification using multiband and hybrid feature extraction. *Multimed Tools Appl* 2024;83:49441–72.
- [19] Almufareh MF, Kausar S, Humayun M, Tehsin S, Farooq A. Inner Speech Decoding: A Comprehensive Review. *Wiley Interdiscip Rev Cogn Sci* 2025;16.
- [20] Gonzalez-Lopez JA, Gomez-Alanis A, Martín Doñas JM, Pérez-Córdoba JL, Gomez AM. Silent speech interfaces for speech restoration: A review. *IEEE Access* 2020;8:177995–8021.
- [21] Nguyen CH, Karavas GK, Artemiadis P. Inferring imagined speech using EEG signals: A new approach using Riemannian manifold features. *J Neural Eng* 2018;15.
- [22] Shah R, Riaz A, Quershi S, Zia A, Iqbal U. Enhancing Silent Speech Decoding Using EEG and Gradient Boosting Classifiers: A Comparative Analysis of XGBoost and LightGBM on a Native Arabic Silent Speech Dataset. 2025 International Conference on Activity and Behavior Computing, ABC 2025, Institute of Electrical and Electronics Engineers Inc.; 2025.
- [23] Milyani AH, Attar ET. Deep learning for inner speech recognition: a pilot comparative study of EEGNet and a spectro-temporal Transformer on bimodal EEG-fMRI data. *Front Hum Neurosci* 2025;19.
- [24] Kobayashi N, Nemoto T, Morooka T. High Accuracy Silent Speech BCI Using Compact Deep Learning Model for Edge Computing. International Winter Conference on Brain-Computer Interface, BCI, vol. 2023- February, Institute of Electrical and Electronics Engineers Inc.; 2023.
- [25] Alzahrani S, Banjar H, Mirza R. Systematic Review of EEG-Based Imagined Speech Classification Methods. *Sensors* 2024;24.
- [26] Müller-Putz GR, Scherer R, Pfurtscheller G, Brunner C, Leeb R. Better than Random? A closer look on BCI results. vol. 10. 2008.
- [27] Radwan YA, Ahmed Mohamed E, Metwalli D, Barakat M, Ahmed A, Kiroles AE, et al. Stochasticity as a solution for overfitting—A new model and comparative study on non-invasive EEG prospects. *Front Hum Neurosci* 2025;19.
- [28] Lopez-Bernal D, Balderas D, Ponce P, Molina A. Exploring inter-trial coherence for inner speech classification in EEG-based brain-computer interface. *J Neural Eng* 2024;21.
- [29] Wang Y, Cai H, Zhou L, Yang Z, Wang L, Zhao J, et al. Decoding inner speech via frequency-specific cortical EEG representations. *Inf Process Manag* 2026;63:104676.
- [30] Tasin SM, Chowdhury MEH, Pedersen S, Chabbouh M, Bushnaq D, Aljindi R, et al. Ensemble Machine Learning Model for Inner Speech Recognition: A Subject-Specific Investigation. *IEEE Access* 2026.
- [31] Dong L, Shao H, Zhang L, Li L. STCCA: Spatial–Temporal Coupled Cross-Attention Through Hierarchical Network for EEG-Based Speech Recognition. *Sensors* 2025;25.
- [32] Abdulghani MM, Walters WL, Abed KH. Enhancing the Classification Accuracy of EEG-Informed Inner Speech Decoder Using Multi-Wavelet Feature and Support Vector Machine. *IEEE Access* 2024;12:147929–41.
- [33] Roy RN, Hinss MF, Darnet L, Ladouce S, Jahanpour ES, Somon B, et al. Retrospective on the First Passive Brain-Computer Interface Competition on Cross-Session Workload Estimation. *Frontiers in Neuroergonomics* 2022;3.
- [34] Aristimunha B, Ju C, Collas A, Bouchard F, Mian A, Thirion B, et al. SPD Learn: A Geometric Deep Learning Python Library for Neural Decoding Through Trivialization 2026.
- [35] Paillard J, Hipp JF, Engemann DA. GREEN: A lightweight architecture using learnable wavelets and Riemannian geometry for biomarker exploration with EEG signals. *Patterns* 2025;6.
- [36] Kobler RJ, Hirayama J, Zhao Q, Kawanabe M. SPD domain-specific batch normalization to crack interpretable unsupervised domain adaptation in EEG 2022.