

Research Article

Imagined Speech Classification Using EEG and CBAM-CNN Model

Meenakshi Bisla^{1,*} , Radhey Shyam Anand²

¹School of Computer Science, UPES, Dehradun, India

²Electrical Engineering Department, Indian Institute of Technology, Roorkee, India

Abstract

This work introduces a novel Convolutional Block Attention Module (CBAM) along with Convolutional Neural Network (CNN) architecture for the recognition of imagined speech electroencephalography (EEG) data. The real-time data is recorded in Biomedical Signal Processing Laboratory, IIT Roorkee. The Pearson correlation (P) method is utilized to gain understanding of neural activity during speech imagination. The proposed CBAM model leverages both spatial and channel attention, allowing the model to selectively focus on the most distinguishing regions. The incorporation of the CBAM mechanism enhances feature representation by adaptively emphasizing the most informative spatial and channel-wise EEG components, thereby improving both model performance and interpretability. This model also utilizes binary and multiclass classification of imagined speech from correlation-based EEG feature images. The subject independent and subject-dependent classification accuracies obtained from P feature images range from $52.72 \pm 7.1\%$ to $68.20 \pm 5.3\%$ and $67.47 \pm 5.8\%$ to $88.09 \pm 4.2\%$ respectively. The results suggest that correlation-based feature representation effectively captures the underlying neural dynamics associated with imagined speech. Comparative analysis with existing state-of-the-art methods indicates that the proposed model achieves improved classification accuracy and generalization, validating its effectiveness for EEG-based imagined speech decoding. These findings indicate the potential of the proposed approach for reliable and scalable brain-computer interface (BCI) applications in real-world scenarios.

Keywords

EEG, Deep Learning, Signal Processing, Imagined Speech, Attention

1. Introduction

Imagine speech being a natural and daily life activity of an individual, this arena is blessed with unlimited number words, prompts and sentences and each one of them refers to a class, leading to an excellent degree of freedom. These characteristics of the imagination modality enhance its usefulness over existing brain computer interface (BCI) based

on modalities like motor imagery or event related potentials with few drawbacks such as limited degree of freedom, more training required, and user discomfort [1, 2]. The choice of imagination classes like particular words or sentences can potentially enhance the decoding performance and user com-

*Correspondence: Meenakshi Bisla (meenakshi.bisla@ddn.upes.ac.in)

Received: 25 February 2026; Accepted: 25 March 2026; Published: 15 April 2026



Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

munication. The previous studies validate the potential reliability of an EEG based imagined speech real time BCI designs but fails to attain actual communication along with the restrictions on a number of imagination categories limited to few words, vowels or syllables [3-5].

The era of EEG-based decoding of imagined speech started Long ago in 2009 when C. S. DaSalla et al. invented an EEG-based imagined speech decoding system based on binary classification of vowel categories like /a/, /u/ and rest using CSP (common spatial filter) based spatial filter as features and nonlinear SVM as classifier and obtained an accuracy of 68-79% [6]. After DaSalla, many researchers have designed enhanced EEG-based systems for decoding vowel speech imagery [7-9]. In 2010, Brigham et al. derived an algorithm for classifying two imagined syllables, /ba/ and /ku/, using autoregressive coefficients and k-nearest neighbour (kNN) classifier and reported an accuracy between 46% to 88% [10]. In [11] deep CNN (62.37%) and shallow CNN (60.88%) are applied to EEG data for 15 pairs of binary classifications of imagined Spanish words. Ashwin Kamble et al. [12] achieved maximum classification accuracy achieved is $89.6 \pm 4.6\%$ and $61.1 \pm 5.1\%$ in binary and multiclass (seven classes) signals, respectively using Machine-learning-enabled adaptive signal decomposition for a brain-computer interface using EEG. In later works of Ashwin Kamble et al., techniques like optimized rational dilation wavelet transform [13], Spectral analysis [14] and Smoothed Pseudo Wigner-Ville Distribution [15] were explored with appreciable classification accuracies.

This study presents the classification of imagined EEG signals for imagination of three language categories named vowels, words and sentences for potential applications into more feasible BCI system design. In recent years, numerous researches have explored the use of correlation analysis [16-18], to gain insights into various aspects of human neuronal activity and cognitive processes, validating the significance of correlation-based connectivity analysis in disentanglement of complex associations between imagined speech EEG signals.

Correlation analysis of the EEG signals is performed to explore the association between various cortical regions during imagination of speech. The Channel * Channel connectivity matrix images will be generated for all trials of EEG signals and provided to an attention-based optimized CNN algorithm for classification of imagined work. We hypothesize that the unification of the deep learning (DL) algorithm with the correlation feature EEG images can provide enhanced analysis and classification of imagined speech characteristics than the conventional machine learning (ML) based methods existing in the literature [12, 19]. The remainder of this article is organized as follows. Section II describes the dataset, Feature extraction, and CBAM-CNN based proposed model. Experimental results and discussion are presented in Section III. Finally, Section IV presents the conclusion of the proposed method.

2. Methods and Material

This section presents the feature extraction and classification methodology adopted to accomplish high accuracy decoding of sentences, words and vowels from imagined speech dataset. The comprehensive architecture of the proposed methodology is presented in Figure 1. It shows the proposed prototype is alienated into five steps: 1) Acquisition of the EEG dataset; 2) Preprocessing and segmentation of the EEG signals; 3) Obtaining correlation-based EEG feature images using Pearson correlation, 4) Designing and training of attention-based convolution neural network classifier architecture for classification, and 5) Binary and multiclass classification is performed with from the images of obtained correlation matrices.

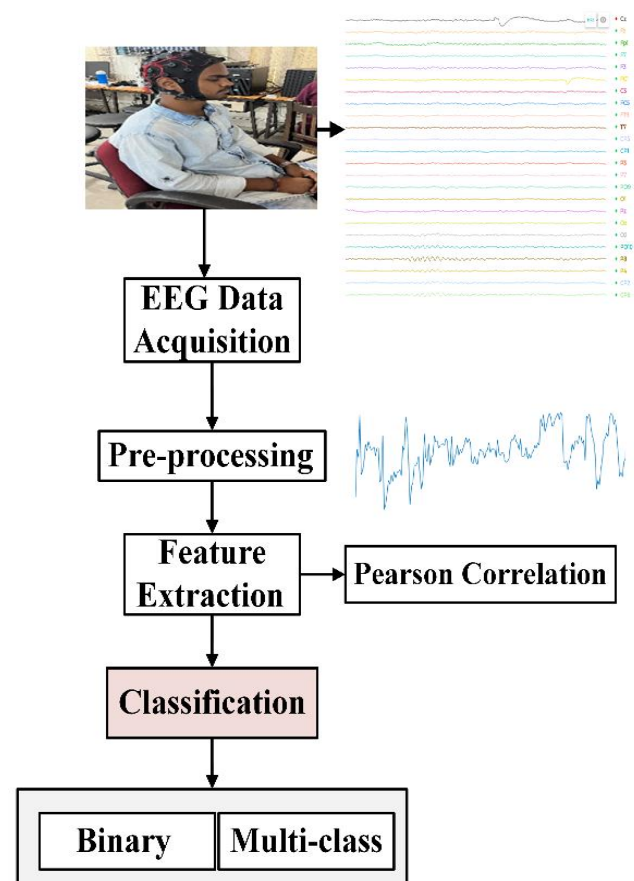


Figure 1. The comprehensive architecture of the proposed methodology.

A. EEG Data Acquisition

The original EEG dataset of ten healthy volunteers (S1–S10), all male, with a mean age of 29 is recorded at Biomedical lab located in Electrical Engineering department of IIT, Roorkee. The dataset utilized in this work was generated by the authors and has been made publicly accessible via IEEE DataPort to ensure transparency and reproducibility [20]. The experimental instruction about the imagination of desired

word, vowel or sentence were presented to the participants from a computer monitor. The EEG experimental procedure adopted for each subject is shown in **Figure 2**.

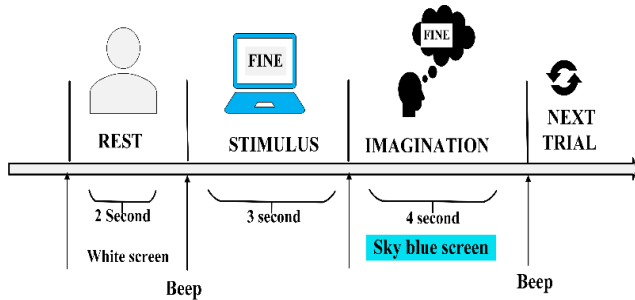


Figure 2. Experimental protocol of Dataset 1.

The wireless 32 channel EMOTIVE EPOC Saline system is used for EEG data acquisition. EEG signals were recorded at a sampling rate of 128Hz. The experiment consists of three imagination categories: 1) Vowels ('/a/' and '/i/'), 2) Words ("WATER," "FINE," "PROBLEM," "SLEEP,") and 3) Sentences ("I AM FINE," "I HAVE PROBLEM,").

Each subject performed around 60 repeated imagination trials for each of the above-mentioned classes. The recording session started with a relaxation stage of 2s, followed by a stimulus stage of 3 s, and an imagination stage of 4 s. Every imagination stage was divided by a beep sound. The participants were instructed to keep their eyes closed during the imagination stage until they heard the beep sound.

B. Pre-processing and Segmentation

EEGLAB toolbox from MATLAB is utilized for the purpose of basic preprocessing of data. EEG data was filtered in the range 0.5 Hz to 60 Hz using the fifth order Butterworth bandpass filter to suppress noisy components. Power line noise is eliminated with the help of 50 Hz notch filter. At the end, EEG data was rereferenced using common average referencing technique. After pre-processing, data recordings were further segmented into trials, and each trial was additionally divided into rest, stimulus, imagined state data. Each trial is associated with a data matrix with 32 rows referring to 32 channel electrodes and 512 columns for time samples (sampling rate 128Hz). The 4-second EEG epochs of speech imaginary state of dataset have been taken out for analysis, counting to a total of 480 (60 trials * 8 imagination classes) epochs per subject.

C. Correlation based Feature Extraction

Correlation is a measure of how two or more variables are related. It provides insights into the strength and direction of the relationship. In imagined speech decoding, identifying and analyzing correlation between EEG time series data is critical for accurately interpreting and understanding the brain's activity related to internal speech processes.

Pearson Correlation is a straightforward method to assess the linear relationship between two signals. Extracted features

are converted to images and are fed to classification process. This study utilizes Pearson's correlation as a fundamental statistical tool to trace the linearity between two signals. It measures the extent to which two signals vary in relation to each other [16]. It is exhaustively used in research and analysis, to qualify the stretch and directionality of association between two continuous signals. However, it is necessary to consider its underlying assumptions and potential limitations such as linearity, bidirectionality, sensitivity to outliers etc. Pearson correlation is used in the study with the purpose of capturing underlying connectivity trends between various cortical regions to enhance imagined speech classification. The Pearson correlation (γ) coefficient is calculated using expression (1).

$$\gamma = \frac{\sum_i^n (N_i - \bar{N})(K_i - \bar{K})}{\sqrt{\sum_i^n (N_i - \bar{N})^2 \sum_i^n (K_i - \bar{K})^2}} \quad (1)$$

Where N_i and K_i are individual sample points of signal N and K, \bar{N} and \bar{K} represents the mean values of signal N and K and n refers to number of sample points. The value of ' γ ' ranges from [0 to 1], where $|\gamma| = 1$ represents perfect linear relationship, $\gamma = 0$ represents no relationship and $|\gamma| \geq 0.5$ signifies strong linear correlation. We have applied Pearson correlation to imagined speech EEG signals to find overall correlation between EEG channel during entire duration of a trial. **Figure 3** contains channel * channel Pearson feature image corresponding to imagination of one trial of word 'Fine'. The process is repeated for all the 60 trials of an imagination class leading to 60 images per class per subject.

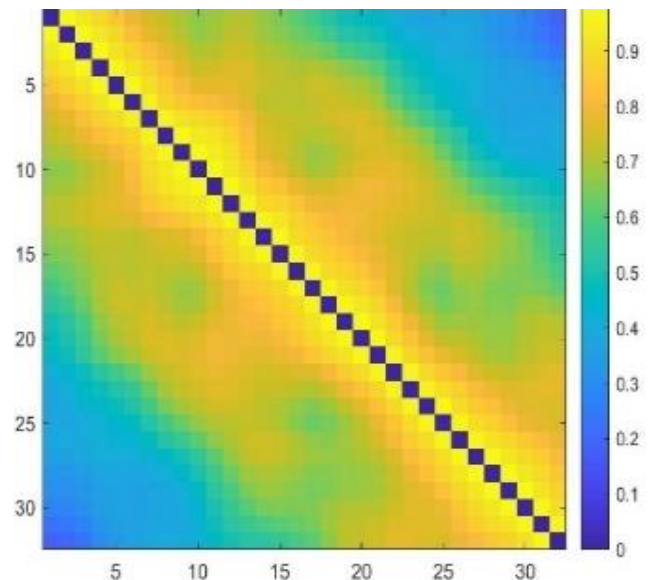


Figure 3. Pearson correlation EEG feature Image. The x-axis and y-axis correspond to 32 EEG channels.

D. CBAM CNN Architecture for Classification

CBAM based convolutional neural networks architecture is designed for purpose of improving classification performance.

It leverages both spatial and channel attention, personalized precisely for EEG signal feature images allowing the model to selectively focus on the most informative regions and channels.

The pre-processed EEG data was utilized for feature extraction to extract EEG feature images using Pearson correlation, leading to 60 images per class per subject. These feature images were used as input for a CBAM CNN classifier.

CNNs are the deep neural network that captures the spatial hierarchies of EEG data after converting EEG features into images. An attention-based CBAM CNN can further boost feature illustration by concentrating on the most relevant parts of

an image EEG feature image, enhancing the classification performance in complex cases like imagined speech recognition where specific details are crucial.

CBAM comprises two sub-sections named channel attention module (CAM) and spatial attention module (SAM) [21]. CAM highlights channel significances and SAM highlights spatially significant regions. The functional architecture of CBAM attention module is shown in Figure 4.

CAM forms channel descriptors by concentrating on the feature map along the spatial dimension (height, width). SAM detects the significant spatial regions inside the feature maps to focus on assured spatial associations or patterns that might correspond to imagined speech task.

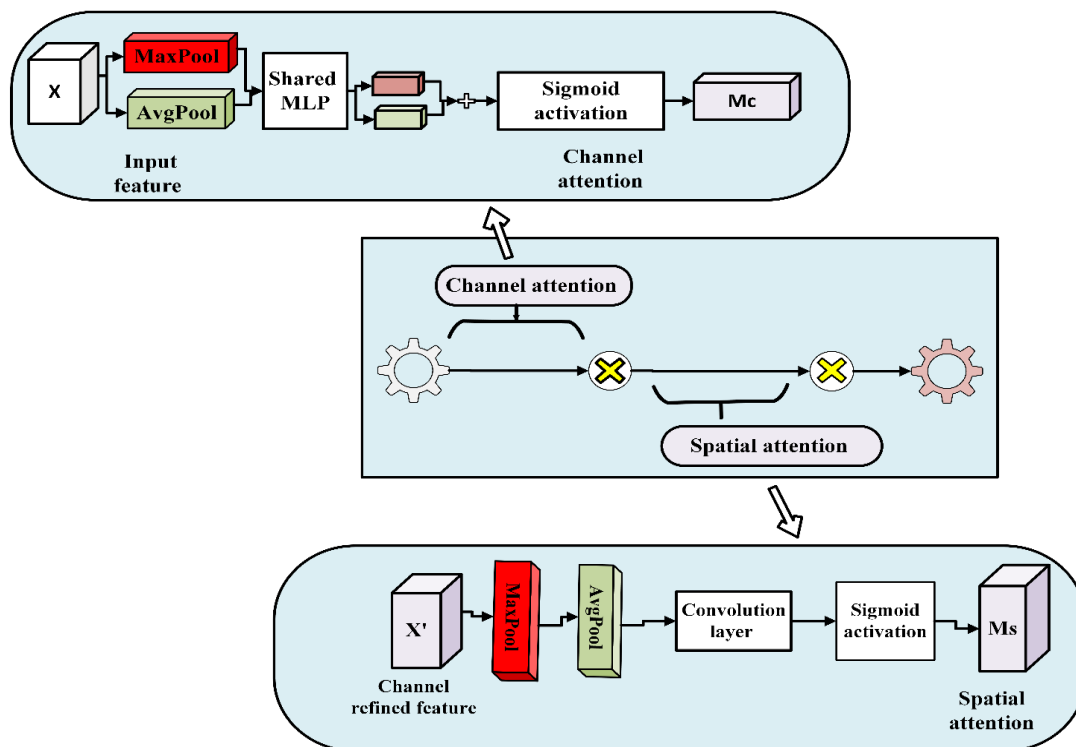


Figure 4. The functional architecture of CBAM attention module.

E. Proposed Model

Raw and original EEG was acquired corresponding to diverse categories of speech. The recorded dataset is pre-processed and segmented into trials, and Each trial is associated with a data matrix with 32 rows referring to 32 channel electrodes and 512 columns for time samples (sampling rate 128Hz). Correlated based EEG feature images are extracted each trial. leading to 60 images per class. The same feature extraction process is repeated for all 10 subjects. The EEG feature images of size $100 \times 100 \times 3$ are inputted to proposed CBAM based CNN model for classification. The architecture of the CBAM CNN model is presented in Figure 5.

The 32×512 EEG signals were transformed into a 32×32 correlation matrix by computing pairwise channel correlations. This matrix was visualized using the 'imagesc' function from

MATLAB and resized to a 100×100 image. To ensure compatibility with CNN models, the image was represented in a 3-channel format ($100 \times 100 \times 3$), where the channels correspond to colormap-based representations of the correlation values.

The input dataset consists of EEG-derived feature images used for training the classification models. To analyze the impact of input resolution, multiple image sizes were evaluated during experimentation. It was observed that an input size of $100 \times 100 \times 3$ provides an optimal balance between computational efficiency and classification performance. Larger image sizes did not result in any significant improvement in accuracy, while smaller sizes led to a noticeable degradation in performance. Therefore, a fixed input size of $100 \times 100 \times 3$ was adopted for all experiments.

The hyperparameters of the model were optimized using Bayesian optimization via the Keras Tuner library, enabling efficient exploration of the search space and selection of the best-performing configuration based on validation performance. The final configuration employed the Adam optimizer with an initial learning rate of 0.0005, trained for 40 epochs

with a batch size of 16 using a categorical cross-entropy loss function. Tables 1-7 presents the specific hyperparameters used for classification of imagined vowels, words and sentences. The performance of the proposed imagined speech classification system is assessed using numerous performance measures, including accuracy, precision, F1 score, recall.

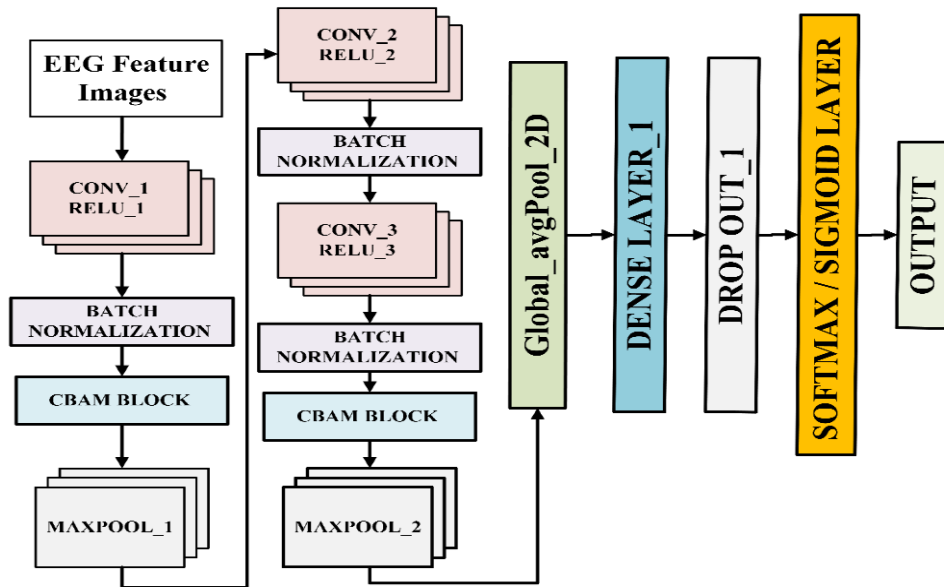


Figure 5. The architecture of proposed CBAM CNN model.

To further ensure the robustness and reliability of the reported results, 10-fold cross-validation was employed for both binary and multi-class classification tasks. In this process, the dataset is divided into 10 equal folds, with 9 folds used for training and 1-fold reserved for testing in each iteration. In each fold, 90% of the data was used for training and 10% for testing, ensuring that every sample was used for testing exactly once. Furthermore, 10% of the training data was set aside as a validation set for hyperparameter tuning and model selection. This procedure is repeated 10 times, ensuring that every fold is used once as a test set. The average accuracy across all

folds is then reported as the final performance score, minimizing the risk of overfitting and reducing bias in performance estimation.

All experiments were conducted in a GPU-enabled computing environment to ensure efficient model training. Feature extraction was performed in MATLAB, where raw EEG signals were processed to generate feature-based images. These EEG feature images were subsequently used as input to deep learning models implemented in Python using the TensorFlow framework. This experimental setup minimizes overfitting and provides a reliable estimate of the model's generalization capability.

Table 1. Convolutional Layers Used in the Proposed CNN Architecture.

Layer	Layer Type	No. of Filters	Kernel Size	Stride	Activation
Conv_1	Conv2D	32	3x3	1	ReLU
Conv_2	Conv2D	64	3x3	1	ReLU
Conv_3	Conv2D	128	5x5	1	ReLU

Table 2. Pooling Layers Used in the Proposed CNN Architecture.

Layer	Layer Type	Pool Size	Stride
MaxPooling_1	MaxPooling2D	2x2	2
MaxPooling_2	MaxPooling2D	2x2	2

Table 3. Regularization Layer Configuration.

Layer	Layer Type	Drop Rate
Dropout_1	Dropout	0.25

Table 4. Fully Connected Layer Configuration.

Layer	Layer Type	Neurons	Activation
Dense_1	Dense	128	Tanh

Table 5. Output Layer Configuration for Different Classification Tasks.

Task	Layer Type	Neurons	Activation
Binary	Dense	2	Sigmoid
4-Class	Dense	4	Softmax
8-Class	Dense	8	Softmax

Table 6. Attention Module Configuration.

Module	Type	Kernel Size
CBAM	Attention	5x5

Table 7. Training Hyperparameters Used for Model Optimization.

Parameter	Value
Optimizer	Adam
Learning Rate	0.0005
Batch Size	16
Epochs	40
Loss (Binary)	Binary Cross-Entropy
Loss (Multi-class)	Categorical Cross-Entropy
Validation	10-fold CV

3. Results and Discussion

In this study, classification was carried out at multiple levels to comprehensively evaluate the discriminative capability of the proposed framework. First, Subject independent binary classification was performed across all possible pairwise combinations of prompts as shown in the binary classification heatmap presented in Figure 6, resulting in a total of 28 binary classification tasks. The subject-independent binary classification consists of 28 tasks, obtained by considering all possible pairwise combinations of the eight classes: ‘water’, ‘help’, ‘fine’, ‘problem’, ‘/a/’, ‘/i/’, ‘i am fine’, and ‘i have problem’. This results in a total of 28 unique binary classification problems, where each task involves distinguishing between two specific classes. The subject independent classification accuracies obtained from P feature images ranges from $52.72 \pm 7.1\%$ to $68.20 \pm 5.3\%$. This approach enabled the analysis of separability between every possible pair of imagined speech prompts, thereby highlighting the relative difficulty of distinguishing specific word pairs.

In addition to binary tasks, a multi-class classification was conducted in two stages. The first stage focused on 4-class classification (Water vs Fine vs Problem vs Sleep) involving only the word prompts, which provided insights into the framework’s ability to distinguish among imagined words. The second stage extended this evaluation to an 8-class classification, incorporating all prompts namely Water vs Fine vs Problem vs Sleep vs I am fine vs I have problem vs /a/ vs /i/, thus representing the most challenging scenario with the maximum number of categories. Table 8 contains the accuracy for four and eight class classification for all subjects obtained from P feature Images with CBAM-CNN classifier. For subject dependent multi-class classification, The overall 4 class classification accuracy ranges from $70.12 \pm 1.36\%$ to $75.48 \pm 2.7\%$ and overall ‘8’ class classification accuracy ranges from $49.62 \pm 3.41\%$ to $53.38 \pm 2.67\%$.

The performance variation observed across different subjects in Table 8 can be attributed to inherent inter-subject variability in EEG signals. EEG patterns associated with imagined speech are highly individual-specific, leading to differences in feature representation and classification performance. Additionally, factors such as subject compliance, attention levels during data acquisition, and cognitive differences may influence the quality and consistency of the recorded signals. Variations in signal-to-noise ratio and the presence of artifacts further contribute to performance differences across subjects. Moreover, the model may exhibit slight overfitting or underfitting for certain subjects due to limited subject-specific data. Despite these variations, the proposed CBAM-CNN model demonstrates relatively stable performance across subjects, indicating its robustness in handling subject-dependent variability.

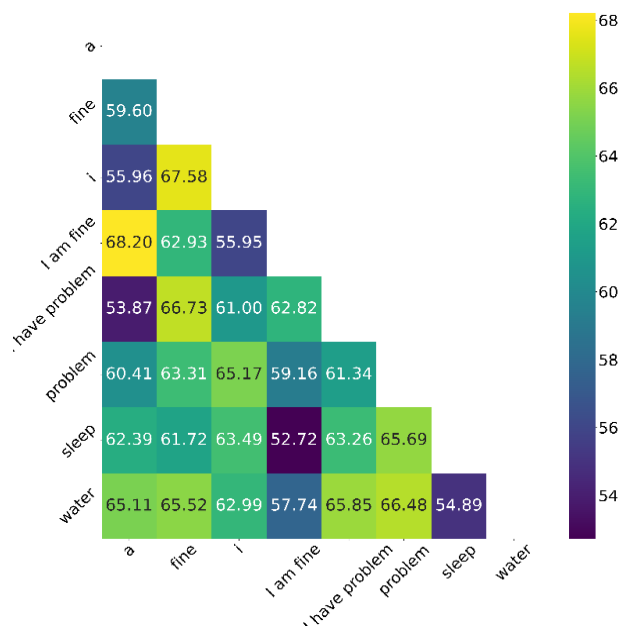


Figure 6. Heatmap of subject-independent binary classification of imagined words.

Table 8. Percentage accuracy ± standard deviation for four and eight class classification for all subjects obtained from P feature Images with CBAM-CNN classifier.

Subject	4 Class	8 Class
S1	70.12±1.36	50.42±1.54
S2	71.83±3.2	52.53±3.12
S3	74.15±1.9	51.64±2.54
S4	72.17±3.1	50.31±1.54
S5	70.69±4.1	52.31±4.5
S6	75.48±2.7	53.38±2.67
S7	73.37±1.78	49.62±3.41
S8	71.12±1.39	51.17±5.12
S9	72.67±2.4	52.63±2.43
S10	75.32±1.67	49.93±1.58

Performance of the proposed model is compared among different categories such as vowel, sentences, words, vowel vs sentences, words vs vowels and words vs sentences in terms of the performance evaluation measures, such as Precision, Recall, F1-score, Kappa and accuracy as shown in Figure 7. The comparative analysis is performed using subject independent approach with 10-fold cross validation. Words vs sentences, words vs vowel and words classification show significantly better performance in comparison to other categories. The outcomes consistently show higher values for these performance metrics, signifying the appreciable completeness in apprehending the desired performance.

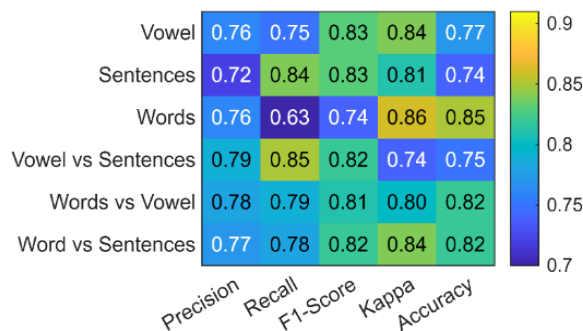


Figure 7. Comparison of performance of proposed model among different language categories in terms of the performance evaluation measures, such as Precision, Recall, F1-score, Kappa and Accuracy.

To further validate the effectiveness of the proposed CBAM-CNN model, a comparative analysis was performed with baseline models, including CNN [22], CNN with self-attention [22], ResNet-50 [22, 23], and DenseNet-121 [4]. The results, illustrated in Figure 8, demonstrate that the proposed model consistently outperforms the other architectures across all classification tasks. The improved performance highlights the effectiveness of the CBAM module in enhancing feature representation by focusing on relevant spatial and channel-wise information significantly and complement each other for improved imagined speech classification.

To evaluate the contribution of each component, an ablation study was performed by comparing four configurations: baseline CNN, CNN with Pearson correlation (P) features, CNN with CBAM, and the full proposed model. All experiments were conducted under identical 10-fold cross-validation settings. As shown in Table 9, the baseline CNN achieved the lowest accuracy, while the inclusion of Pearson correlation improved performance by capturing inter-channel relationships in EEG signals. The addition of CBAM further enhanced accuracy by enabling attention-driven feature refinement. The proposed combined model achieved the highest accuracy, demonstrating that both correlation-based feature representation and attention mechanisms contribute.

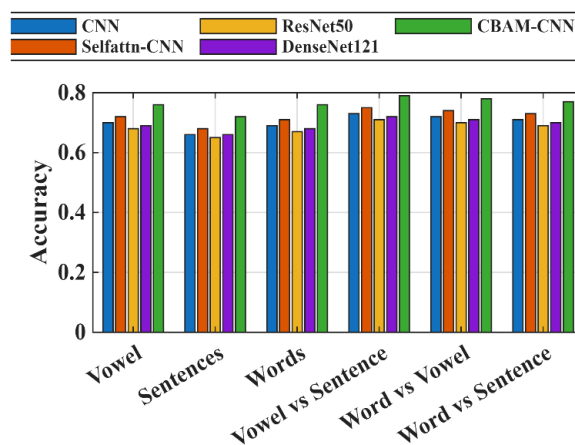


Figure 8. The comparative performance of different models for vowel, word, sentence, and cross-category classification tasks.

Table 9. Ablation study showing the impact of Pearson correlation (P) and CBAM on classification accuracy of the proposed EEG-based imagined speech recognition model.

Model Variant	Accuracy (%)
CNN (Baseline)	62.3
CNN + Pearson Correlation	69.2
CNN + CBAM	73.4
Proposed	76.7

Despite the promising results, this study has certain limitations. The performance in the subject-independent setting remains comparatively lower, highlighting challenges in generalizing across subjects due to inherent inter-subject variability in EEG signals. Additionally, the dataset size is relatively limited, which may affect the robustness of the deep learning model. The current work is also restricted to Pearson correlation-based feature representation and a single attention mechanism. In future work, larger and more diverse datasets will be considered along with advanced data augmentation techniques to enhance model generalization. Furthermore, incorporating rigorous statistical analysis will help validate the significance of the obtained results, while exploring additional feature extraction methods and hybrid attention mechanisms may further improve classification performance. To improve generalization, future work will include a larger and more diverse population encompassing different genders, age groups, and cognitive conditions.

4. Conclusion

This paper offers numerous key contributions to imagined speech classification. EEG feature images were created using correlation method to carefully capture brain connectivity patterns. The research utilized in-house EEG data recorded at the Biomedical Signal Processing Laboratory, IIT Roorkee, focusing on the classification of imagined words, sentences, and vowels. A novel attention-based CBAM-CNN architecture was proposed to enhance spatial and channel-wise feature learning. This CBAM-CNN model was applied to both binary and multiclass classification tasks, showing robust performance across various imagined speech categories. The inclusion of attention mechanisms significantly boosted the model's capability to identify informative EEG regions and channels, outperforming traditional CNN-based methods in recognition accuracy. The proposed attention-based CNN model has greatly improved classification accuracy and reduced the computational cost. The model has performed well with simple correlation-based features, hence reducing the need for exhaustive frequency-based, phase-based exploration and time frequency decomposition.

Abbreviations

CBAM	Convolutional Block Attention Module
CNN	Convolutional Neural Network
EEG	Electroencephalography
P	Pearson Correlation
BCI	Brain Computer Interface
DL	Deep Learning
ML	Machine Learning

Author Contributions

Meenakshi Bisla: Conceptualization, Data curation, Methodology, Visualization, Conceptualization, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft

Radhey Shyam Anand: Investigation, Supervision, Validation, Writing – review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] J. Fumanal-Idocin et al., "Supervised penalty-based aggregation applied to motor-imagery based brain-computer-interface," *Pattern Recognit.*, vol. 145, p. 109924, Jan. 2024, <https://doi.org/10.1016/j.patcog.2023.109924>
- [2] A. Mobaeni, R. Boostani, and S. Sanei, "Improving the performance of P300-based BCIs by mitigating the effects of stimulus-related evoked potentials through regularized spatial filtering," *J. Neural Eng.*, vol. 21, no. 1, p. 016023, Feb. 2024, <https://doi.org/10.1088/1741-2552/ad2495>
- [3] A. Singh and A. Gumaste, "Decoding Imagined Speech and Computer Control using Brain Waves," *J. Neurosci. Methods*, vol. 358, p. 109196, Jul. 2021, <https://doi.org/10.1016/j.jneumeth.2021.109196>
- [4] M. Bisla and R. Shyam Anand, "Transfer Learning Enabled Imagined Speech Interpretation Using Phase-Based Brain Functional Connectivity and Power Analysis," *IEEE Access*, vol. 12, pp. 108399–108413, 2024, <https://doi.org/10.1109/ACCESS.2024.3438873>
- [5] A. A. Torres-García, C. A. Reyes-García, L. Villaseñor-Pineda, and G. García-Aguilar, "Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification," *Expert Syst. Appl.*, vol. 59, pp. 1–12, Oct. 2016, <https://doi.org/10.1016/j.eswa.2016.04.011>
- [6] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural Networks*, vol. 22, no. 9, pp. 1334–1339, Nov. 2009, <https://doi.org/10.1016/J.NEUNET.2009.05.008>

- [7] B. M. Idrees and O. Farooq, "Vowel classification using wavelet decomposition during speech imagery," in 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, Feb. 2016, pp. 636–640. <https://doi.org/10.1109/SPIN.2016.7566774>
- [8] B. Min, J. Kim, H. Park, and B. Lee, "Vowel Imagery Decoding toward Silent Speech BCI Using Extreme Learning Machine with Electroencephalogram," *Biomed Res. Int.*, vol. 2016, pp. 1–11, 2016, <https://doi.org/10.1155/2016/2618265>
- [9] C. Cooney, R. Folli, and D. Coyle, "Optimizing Layers Improves CNN Generalization and Transfer Learning for Imagined Speech Decoding from EEG," in 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), IEEE, Oct. 2019, pp. 1311–1316. <https://doi.org/10.1109/SMC.2019.8914246>
- [10] K. Brigham and B. V. K. V. Kumar, "Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy," in 2010 4th International Conference on Bioinformatics and Biomedical Engineering, IEEE, Jun. 2010, pp. 1–4. <https://doi.org/10.1109/ICBBE.2010.5515807>
- [11] M. N. I. Qureshi, B. Min, H.-J. Park, D. Cho, W. Choi, and B. Lee, "Multiclass Classification of Word Imagination Speech With Hybrid Connectivity Features," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 10, pp. 2168–2177, Oct. 2018, <https://doi.org/10.1109/TBME.2017.2786251>
- [12] A. Kamble, P. Ghare, and V. Kumar, "Machine-learning-enabled adaptive signal decomposition for a brain-computer interface using EEG," *Biomed. Signal Process. Control*, vol. 74, p. 103526, Apr. 2022, <https://doi.org/10.1016/J.BSPC.2022.103526>
- [13] A. Kamble, P. H. Ghare, and V. Kumar, "Optimized Rational Dilation Wavelet Transform for Automatic Imagined Speech Recognition," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023, <https://doi.org/10.1109/TIM.2023.3241973>
- [14] A. Kamble, P. H. Ghare, V. Kumar, A. Kothari, and A. G. Keskar, "Spectral Analysis of EEG Signals for Automatic Imagined Speech Recognition," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–9, 2023, <https://doi.org/10.1109/TIM.2023.3300473>
- [15] A. Kamble, P. H. Ghare, and V. Kumar, "Deep-Learning-Based BCI for Automatic Imagined Speech Recognition Using SPWVD," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023, <https://doi.org/10.1109/TIM.2022.3216673>
- [16] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson Correlation Coefficient," 2009, pp. 1–4. https://doi.org/10.1007/978-3-642-00296-0_5
- [17] A. Zamm, S. Debener, A. R. Bauer, M. G. Bleichner, A. P. Demos, and C. Palmer, "Amplitude envelope correlations measure synchronous cortical oscillations in performing musicians," *Ann. N. Y. Acad. Sci.*, vol. 1423, no. 1, pp. 251–263, Jul. 2018, <https://doi.org/10.1111/nyas.13738>
- [18] W. Hesse, E. Möller, M. Arnold, and B. Schack, "The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies," *J. Neurosci. Methods*, vol. 124, no. 1, pp. 27–44, Mar. 2003, [https://doi.org/10.1016/S0165-0270\(02\)00366-7](https://doi.org/10.1016/S0165-0270(02)00366-7)
- [19] meenakshi bisla and R. S. Anand, "Machine learning based classification of imagined speech electroencephalogram data from the amplitude and phase spectrum of frequency domain EEG signal," *Biomed. Phys. Eng. Express*, Sep. 2025, <https://doi.org/10.1088/2057-1976/ae04ee>
- [20] R.S. Anand, Meenakshi Bisla, Anand Mohan, and Dilnawaz, "A Multi-Class Electroencephalography Dataset for Imagined Speech Decoding.," Jan. 25, 2026, *IEEE Dataport*.
- [21] Z. Ji et al., "CBAM-DeepConvNet: Convolutional Block Attention Module-Deep Convolutional Neural Network for asymmetric visual evoked potentials recognition," *Brain-Apparatus Communication: A Journal of Bacomics*, vol. 4, no. 1, Dec. 2025, <https://doi.org/10.1080/27706710.2025.2489396>
- [22] M. Bisla and R. S. Anand, "Speech imagery decoding from electroencephalography signals using an amalgamation of convolutional and recurrent neural networks," *Engineering Research Express*, vol. 7, no. 2, p. 025232, Jun. 2025, <https://doi.org/10.1088/2631-8695/adcff1>
- [23] M. A. Morid, A. Borjali, and G. Del Fiol, "A scoping review of transfer learning research on medical image analysis using ImageNet," *Comput. Biol. Med.*, vol. 128, p. 104115, Jan. 2021, <https://doi.org/10.1016/j.combiomed.2020.104115>