Developing an IoT Driven BCI Framework for Real-Time Neural Signal Decoding to Speech Conversion

Prateek Malagund
Dept. of CSE
SCEM
Mangalore

Gagan V
Dept. of CSE
SCEM
Mangalore

Misbah Zohar
Dept. of CSE
SCEM
Mangalore

Neha P Achar
Dept. of CSE
SCEM
Mangalore

Dr. Mustafa Basthikodi
Dept. of CSE
SCEM
Mangalore

Abstract-This study introduces an innovative method for converting EEG brain signals into phonemes, facilitating communication for those with speech disorders. The research utilizes a machine learning system that analyzes multi-channel EEG data from 14 electrodes to predict phonemes associated with intended speech. Unlike traditional approaches that require expensive EEG equipment, this project incorporates a budget-friendly simulation framework, combining random signal generation for dynamic authenticity and dataset-driven signal replication for accurate predictions. The methodology involves EEG data preprocessing, feature extraction, and training a fusion model to achieve effective phoneme classification. The findings show considerable accuracy in phoneme prediction, underscoring the potential of EEG-based systems in augmentative and alternative communication (AAC) technologies. Additionally, a simulated hardware prototype and an interactive graphical user interface are created to offer a realistic system demonstration, addressing the limitations of restricted access to EEG hardware. This research addresses a crucial need for accessible speech synthesis systems and paves the way for affordable, scalable solutions in brain-computer interface technology.

***Index Terms*—Electroencephalography (EEG), Brain-Computer Interface (BCI), Phoneme Recognition, Signal Processing, Neural Signal Analysis, Model Performance Metrics**

I. Introduction

In Brain-Computer Interface (BCI) development, accurate and efficient models are crucial for transforming raw neural signals into actionable insights. BCIs bridge human neural activity and external devices, enabling applications in medical diagnostics, augmentative communication, and beyond. However, neural data is inherently complex, characterized by high dimensionality, noise, and variability across subjects and sessions, which poses significant challenges for traditional computational approaches.

This project aims to advance BCI systems' accuracy, reliability, and interpretability by leveraging state-of-the-art machine learning and deep learning techniques. A suite of models—including Random Forest, Gradient Boosting, Recurrent Neural Networks, Long Short-Term Memory Networks, and Transformers—is evaluated to determine the most effective approaches for neural signal processing. Each model is rigorously assessed using accuracy, precision, recall, F1 score, and visualizations like confusion matrices and performance curves, providing a comprehensive evaluation framework.

By identifying models that excel under specific conditions and uncovering critical features that drive effective neural signal interpretation, this work contributes to the development of robust, generalizable BCI systems. The insights gained from this study lay a foundation for next-generation applications in neurotechnology, fostering more reliable human-machine interactions and expanding possibilities for assistive communication and beyond.

A. Scope

This project focuses on EEG-based word recognition, utilizing simulated brainwave data for feature extraction, model training, and real-time processing. It includes the development of a signal-cleaning pipeline, application of machine learning techniques for accurate word classification, and optimization of system configurations to replicate real-world brain activity patterns. These approaches enable scalable, efficient data collection and testing, offering a cost-effective alternative to live EEG experiments. Key benefits include improved model validation, reduced operational complexity, and accelerated development cycles, paving the way for future integration with live EEG systems for assistive applications.

The project also encompasses the evaluation and comparison of diverse machine learning and deep learning models, including Random Forest, Gradient Boosting (XGBoost), Recurrent Neural Networks, Long Short-Term Memory networks, 1D CNNs, and Transformers (BERT). Performance is assessed using metrics like accuracy, precision, recall, and F1 score, alongside visualizations such as ROC Curves, Precision-Recall Curves, Learning and Loss Curves, and Feature Importance plots. This evaluation framework helps identify optimal models for Brain-Computer Interface (BCI) applications, guiding

feature selection, model tuning, and performance enhancement.

B. Objectives

- Collect or create datasets using sensors to measure the brain's analog signals, then amplify and filter these signals to remove noise and convert them into a digital format for further processing
- Implement advanced algorithms for feature extraction and train multiple machine learning (ML) and deep learning (DL) models to classify EEG signals into corresponding phonemes.
- Develop a robust system using a weighted average fusion approach to combine predictions from multiple models, enhancing classification accuracy and reliability.
- Propose a pathway for future integration of real EEG hardware, such as the BioAmp EXG Pill, to adapt signal acquisition and classification to individual users' neural patterns.

II. Literature Review

The development of brain-computer interface (BCI) systems for speech decoding and synthesis has been a topic of significant research interest. Luo *et al.* [1] explored the role of BCIs in augmenting communication by decoding neural activity into speech signals using advanced machine learning models. Similarly, Peksa and Mamchur [2] provided a comprehensive review of the state-of-the-art BCI technologies, emphasizing their applications in healthcare, assistive communication, and the challenges associated with signal processing.

Angrick *et al.* [3] demonstrated the feasibility of online speech synthesis using a chronically implanted BCI in individuals with ALS, highlighting the system's ability to produce real-time speech. Brumberg *et al.* [4] focused on the design of BCI systems for speech communication, presenting approaches that map neural signals to phoneme sequences using robust processing pipelines.

Allison *et al.* [5] reviewed advancements in BCI systems, emphasizing improvements in signal acquisition, filtering techniques, and the growing accuracy of machine learning algorithms. Warshi [6] proposed a thought-to-speech framework using BCIs, highlighting the need for high-quality signal clarity to achieve reliable speech decoding.

Zhang *et al.* [7] introduced Cascade and Parallel Convolutional Recurrent Neural Networks for EEG-based intention recognition, demonstrating how spatio-temporal feature extraction improves BCI performance. Similarly, Hong *et al.* [8] discussed artificial speech generation using invasive brain stimulation, identifying the challenges in precision and ethical considerations of brain intervention technologies.

Zhang *et al.* [9] applied a multiple generator Wasserstein GAN to augment EEG data, addressing the challenge of limited datasets and improving model generalization for emotion recognition. Kübler *et al.* [10] developed a P300-based spelling system for locked-in patients, showcasing the practical application of auditory event-related potentials for communication.

Vansteensel *et al.* [11] successfully implemented a fully implanted BCI for locked-in ALS patients, demonstrating its effectiveness for long-term use. Bartels *et al.* [12] detailed the design and implantation of neurotrophic electrodes into the human motor cortex, offering a reliable approach for speech-related signal acquisition.

Mridha *et al.* [13] highlighted advancements and ongoing challenges in BCI systems, such as the need for improved signal processing and real-time execution. Värbu *et al.* [14] discussed the evolution of EEG-based BCIs, emphasizing their applications in healthcare, communication, and assistive systems, while addressing signal variability and noise issues.

Finally, Bauer *et al.* [15] provided a foundational classification of the locked-in syndrome, which serves as a basis for BCI research aimed at restoring communication in severely paralyzed patients.

The Fourteen-Channel EEG with Imagined Speech (FEIS) dataset, introduced by Wellington and Clayton [16], provides EEG recordings for imagined speech recognition tasks. This dataset includes EEG signals captured using 14 electrodes placed according to the 10-20 system, ensuring optimal spatial coverage of brain regions associated with speech processing.

Participants were instructed to imagine specific phonemes or words during the signal acquisition process. The dataset offers clean, high-quality EEG data with minimal noise, making it ideal for experiments in imagined speech decoding, machine learning, and brain-computer interface (BCI) applications.

Vlek et al. [17] explore the ethical challenges in Brain-Computer Interface (BCI) research, development, and dissemination. Key concerns include ensuring informed consent, protecting user autonomy, and addressing privacy issues regarding brain data. The study also highlights the importance of equitable access to BCI technologies and the need to anticipate long-term societal and psychological implications of their use. Ethical frameworks and interdisciplinary collaboration are crucial to mitigate these challenges as BCI systems advance.

The P300 wave, as explored by Picton [18] [30], is a prominent component of the event-related potential (ERP) associated with cognitive processes like attention and decision-making. It is widely utilized in Brain-Computer Interfaces (BCIs) for detecting user responses, particularly in speller systems and stimulus-based paradigms. The P300's robustness makes it a reliable signal for human-computer interaction, aiding in assistive communication systems.

The standardized 10-20 electrode system proposed by Klem et al. [19] [29] provides a systematic method for EEG electrode placement on the scalp. It ensures consistent spatial coverage of brain regions, enabling reproducibility and comparability of EEG recordings across studies. The system divides the scalp into proportional distances, optimizing signal acquisition for clinical and research applications.

The Emotiv EPOC+ [20] [33] is a portable, 14-channel EEG headset designed for real-time brain signal acquisition and analysis. It features wireless connectivity, making it suit-

able for applications such as Brain-Computer Interface (BCI) research, cognitive studies, and emotional state detection. Its user-friendly interface and signal-processing capabilities provide a robust solution for neuroscience experimentation and development.

Giudice et al. [21] explored the interpretability of deep convolutional neural networks (CNNs) for detecting eye blinks in EEG signals. Using visual explanation techniques, the study identified significant EEG features contributing to blink detection, improving the transparency and trustworthiness of CNN models in EEG-based BCI systems.

Li et al. [22] provide a systematic review of EEG-based mobile robots, focusing on the integration of EEG signals for real-time robot control. The study discusses key advancements in system architecture, signal decoding methods, and real-time performance optimization, highlighting challenges like signal quality and processing efficiency for practical applications.

Rakhmatulin et al. [23] [36] explored various Convolutional Neural Network (CNN) architectures to extract spatial and temporal features from EEG signals. The study highlights that optimized CNNs enhance EEG feature representation, leading to improved classification accuracy in EEG-based systems.

Sharma and Meena [24] [32] provide a systematic review of recent advancements in EEG signal processing. The study explores noise removal techniques, feature extraction methods, and the integration of machine learning and deep learning models. The review underscores the importance of real-time processing, improved classification accuracy, and the growing applications of EEG in healthcare and brain-computer interfaces.

Sun and Mou [25] [28] provide a detailed survey on advancements in EEG-based signal processing. They discuss key research areas, including feature extraction, classification techniques, and real-time processing. The study underscores the role of deep learning and hardware improvements in addressing challenges like noise, variability, and computational efficiency, paving the way for robust EEG applications.

III. Methodology

A. *Data Preprocessing*

The FEIS (Fourteen-channel EEG for Imagined Speech) dataset, comprises EEG recordings of 21 English-speaking participants recorded with a lightweight, 14-channel mobile headset with dry electrodes (the Emotiv EPOC+). Recordings are time-aligned with phone stimuli, consisting of three stimulus types: heard, spoken internally (imagined), and spoken overtly.

1) Participants: 21 participants were recruited at the University of Edinburgh. Participants are either native or near-native speakers of English, with no known neurological disorders. Three participants are left-handed, one ambidextrous, and the remaining 17 are right-handed. (FEIS metadata available at [16]).

2) Stimuli: Sixteen English phonemes were chosen to represent a balanced categorical spread of binary phonological features ([±nasal], [±back], [±voice], etc.). These are shown in Table 1

TABLE I
PHONEME TYPES IN THE FEIS DATASET

A. Consonants

	Labial	**Alveolar**	**Postalveolar/ Velar**
Plosive (-voice)	/p/	/t/	/k/
Fricative (-voice)	/f/	/s/	/ʃ/
Fricative (+voice)	/v/	/z/	/ʒ/
Nasal (+voice)	/m/	/n/	/ŋ/

B. Vowels

	Front	**Back**
High	/i/	/u/
Low	/æ/	/ɔ/

3) Recording Procedure: High-quality audio of the phonemes listed in Table 1 was recorded in the participants' own voices. A single instance of each of the 16 phones was recorded at 44.1 kHz with a cardioid microphone. We used audio processing software to convert these single-phone prompts into stimuli consisting of five repetitions of each phone. For plosives (e.g. /p/), participants were instructed to form a neutral release

Participants carried out the experiment alone, sitting in a comfortable chair in front of a laptop screen, inside a hemianechoic chamber. Our intention behind these choices in methodology is to mitigate contamination from brainwave components resulting from unexpected audio or visual stimuli. (such as the P300 event-related potentials (ERPs) [18]
The Emotiv EPOC+ is a mobile headset with semi-flexible sensor "arms" which allow for universal fitting, within a fixed configuration. While this allows ease of use, it means that electrode positions are inconsistent relative to the international 10-20 montage system [19], due to participants' different head sizes. For reasonable consistency, we ensure F3/F4 sensors are 20mm above each subject's eyebrows, and M1/M2 dummy electrodes placed on the mastoid process

Figure 1 shows the electrode positions used in our experiment, following the international 10-20 system. Selected electrodes were chosen based on their relevance to speech processing tasks, as summarized in Table II.
The electrode placements are crucial for capturing relevant brain activity during the experiment, specifically for heard, imagined, and overt speech.
The selected electrodes target brain regions critical for auditory processing, motor planning, speech imagination, and articulation, enabling comprehensive EEG data collection during the experimental epochs.

The EEG recordings consist of 160 trials, comprising 6 phonemes × 10 repetitions, randomized to maintain participant attention. Each trial has four 5-second "epochs," as illustrated in Figure 2. First, a "resting" epoch, in which participants are

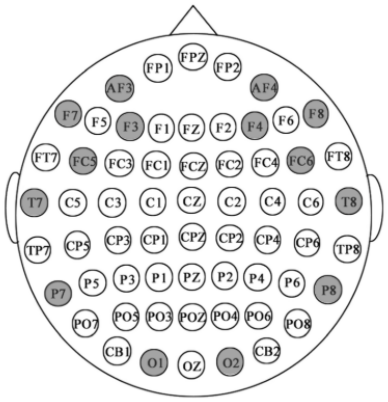

Fig. 1. Electrode positions in the 10-20 EEG montage. Key electrodes like F3, F4, T7, T8, and others are highlighted for their importance in speech processing.

TABLE II
SELECTED ELECTRODES AND THEIR USES

Electrode(s)	Use
F3, F4	Core electrodes for speech production and motor cortex activity.
T7, T8	Capture signals from the temporal lobes, essential for auditory processing and speech perception.
C3, C4	Detect motor planning and articulation activity, particularly during overt speech.
O1, O2	Measure mental imagery signals during visualization of phonemes.
FP1, FP2, AF3, AF4	Monitor cognitive effort, attention, and prefrontal activity.
P3, P4, P7, P8	Track sensory integration, spatial focus, and cognitive load.

shown the word "REST" on screen, and attempt to clear their mind (resting state measurement can be used for task-specific feature extraction, and also reduces cognitive load). Next, a "stimuli" epoch, in which participants are played their own vocalisation of a single phone looped five times, and shown a corresponding IPA representation (which participants were familiar with). Next, a "thinking" epoch, in which participants are presented with a blank screen, and imagine repeating the phone, but without any articulator movement. Finally, a "speaking" epoch, in which participants are prompted with an image of a mouth to then vocalize the phone. In each of the two latter epochs, subjects imagine/speak the phone five times in a steady rhythm, imitating the recording played in the stimuli epoch.

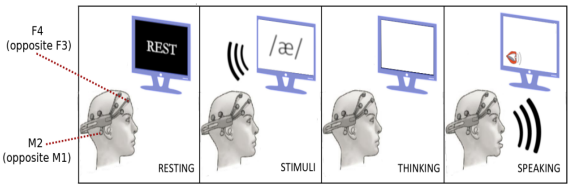

Fig. 2. Illustration of the recording procedure. Participants listen to five repetitions of a phone (recorded in their own voice), then imagine speaking the phone five times (with the same rhythm), then overly speak the phone five times.

4) Noise Removal: The built-in software of the Emotiv EPOC+ performs notch filtering at 50 Hz and 60 Hz to remove power line noise. [20] No signal preprocessing was carried out to remove physiological artifacts (such as blinks or saccades). Often, an independent component analysis (ICA) pipeline is used to remove such artifacts from the data, however, since the Emotiv EPOC+ lacks the ocular channels typically used to isolate noise components through correlation, this was not carried out. Future work could perform ICA on FEIS by using ICA solutions from datasets collected using other devices.

B. *BioAmp EXG Pill*

The BioAmp EXG Pill is a compact, versatile biosignal amplifier designed for capturing electrical signals such as EEG, EMG, ECG, and EOG. Developed by Upside Down Labs, it provides an efficient and cost-effective solution for physiological signal acquisition, making it ideal for applications like brain-computer interfaces (BCI).

1) Features and Specifications: The BioAmp EXG Pill is equipped with the following features and specifications, making it a versatile tool for biosignal acquisition:

- Input Voltage: 4.5 – 40 V
- Input Impedance: 10^{12} ohm
- Compatible Hardware: Any ADC input
- Biopotentials: Configurable for ECG, EMG, EOG, or EEG | Default configuration: EEG and EOG
- Number of Channels: 1
- Electrodes: Configurable for 2 or 3 electrodes | Default configuration: 3 electrodes
- Dimensions: 25.4 x 10.0 mm
- Open Source: Both hardware and software are open-source.

2) BioAmp Circuit Design: The BioAmp EXG Pill features a compact design with key components that enable signal amplification, noise reduction, and wireless transmission. Its architecture includes:

- Instrumentation Amplifier (INA): Amplifies the biopotential signals with high accuracy and noise reduction.
- Electrode Reference Configuration: Configurable for 2-electrode or 3-electrode modes using solder jumpers.
- Bandpass Filter: Configurable for wide or narrow frequency bands to suit EMG, EOG, ECG, or EEG signals.
- Power Supply Filtering: Ensures clean and stable power for signal acquisition.
- Amp Ref + Driven Right Leg (DRL): Provides stability and minimizes common-mode noise in EEG signals.
- Header Pins and Connectors: Designed for seamless connections to external devices such as microcontrollers or ADC inputs.

The circuit design is depicted in Figure 3, which showcases its components and their roles in signal acquisition.

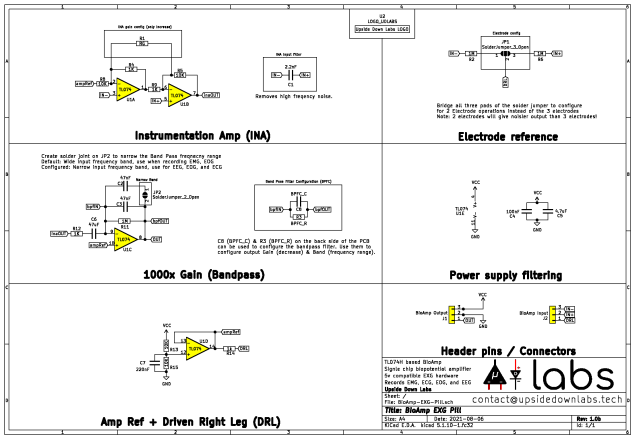

Fig. 3. BioAmp EXG Pill - Circuit Design and Functional Blocks.

3) Circuit Design and Flow: The circuit diagram of the BioAmp EXG Pill is presented in Figure 3. The design can be broken down into the following key functional blocks:

- Instrumentation Amp (INA): Amplifies low-level biopotential signals while rejecting noise.
- 1000x Gain (Bandpass): A gain stage coupled with bandpass filtering to remove high-frequency noise and optimize the signal for EEG acquisition.
- Electrode Reference Configuration: Allows for 2-electrode or 3-electrode setups based on requirements.
- Power Supply Filtering: Provides smooth and noise-free operation of the amplifier circuitry.
- Amp Ref + Driven Right Leg (DRL): Reduces noise and stabilizes the input signals.

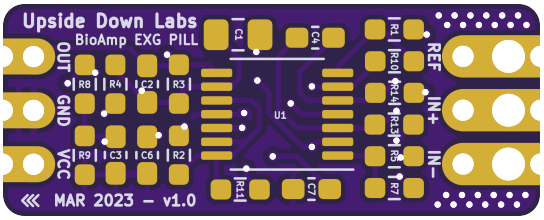

Fig. 4. BioAmp EXG Pill - Front View of the PCB

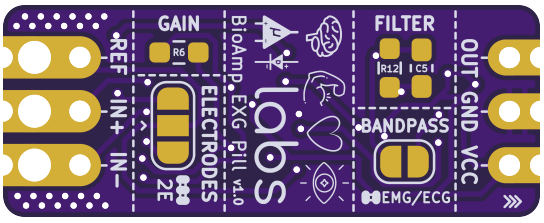

Fig. 5. BioAmp EXG Pill - Rear View of the PCB

4) Role of BioAmp EXG Pill in the Project: In this project, the BioAmp EXG Pill plays a crucial role in real EEG signal acquisition and preprocessing. By capturing high-quality brainwave data from 14 strategically placed electrodes, the BioAmp Pill enables precise measurement of brain activity associated with phoneme processing. Key benefits include:

- Cost-Effective Solution: The BioAmp Pill offers an affordable alternative to expensive EEG systems, enabling accessibility for research and prototyping.
- High Signal Quality: Its amplification and filtering capabilities ensure that noise-free signals are fed into the machine learning models for accurate classification.
- Real-Time Data Transmission: The integration with ESP modules ensures seamless wireless data flow, supporting real-time BCI applications.

C. *Proposed System Design*

The proposed system is designed to classify EEG signals into phonemes using a robust architecture that integrates signal acquisition, preprocessing, feature extraction, machine learning (ML) and deep learning (DL) models, a weighted fusion model, and output generation. The detailed workflow of the system is illustrated in the flow diagram (Figure 6).

Fig. 6. Flow diagram illustrating signal acquisition, preprocessing, feature extraction, model training, fusion, and output generation.

The system consists of the following components:

1) Signal Acquisition Layer: EEG signals are captured using the BioAmp EXG Pill hardware integrated with 14 electrodes strategically placed on the scalp. These electrodes detect the brain's analog signals and convert them into digital signals, which are transmitted wirelessly using ESP modules.

2) Signal Preprocessing: The raw EEG signals undergo preprocessing to improve their quality:

- Noise Removal: Filters out artifacts, interference, and background noise.
- Bandpass Filtering: Focuses on the relevant EEG frequency range (e.g., 0.5–40 Hz) for effective analysis.

3) Feature Extraction: Temporal and spatial features are extracted from the clean EEG data to serve as inputs to machine learning and deep learning models. These features provide meaningful representations of brain activity corresponding to phoneme classes.

4) Model Training and Predictions: The extracted features are passed to multiple ML and DL models for training and prediction:

- Random Forest: Baseline ensemble learning model.
- XGBoost: Gradient boosting algorithm for efficient classification.
- RNN: Captures sequential dependencies in EEG signals
- 1D-CNN: Captures spatial and temporal dependencies.
- LSTM: Models long-term dependencies in sequential EEG data.
- Transformer: Leverages self-attention mechanisms to capture global signal relationships.

5) Fusion Model: Predictions from the Random Forest, XGBoost, 1D-CNN, LSTM, and Transformer models are combined using a **weighted average fusion model**. The final prediction probabilities (p_{final}) are calculated as:

$$p_{\text{final}} = w_1 \cdot p_{\text{rf}} + w_2 \cdot p_{\text{xgb}} \tag{1}$$

where w_1 and w_2 are optimized weights for the models (e.g., Random Forest = 0.7, XGBoost = 0.3).

6) Final Prediction and Outputs: The final predictions are utilized for two key outputs:

- Phoneme Prediction: The classified phonemes are synthesized into audible speech through speech synthesis techniques.
- Performance Metrics: The model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. Visualizations, including *Confusion Matrices*, *ROC Curves*, and *Precision-Recall Curves*, are generated for comparative analysis.

7) Speech Synthesis: Predicted phonemes are processed to generate speech output, enabling real-time communication.

D. Experimentation and Discussions

To develop a robust system for decoding neural signals into phonetic representations, six machine-learning models were implemented and evaluated. Each model was designed to handle the high-dimensional nature of EEG data recorded from 14 electrodes and map these signals to the corresponding phonetic labels.

*1) **Machine Learning Models**:*

a) Random Forest (RF): The Random Forest model was implemented to establish baseline performance. Random Forest works as an ensemble method that combines outputs from multiple decision trees to make predictions, offering interpretability and robustness. The key parameters for this model included the number of trees (*n_estimators* = 100), maximum depth set to *None* to allow full tree growth, and a random state of 42 to ensure reproducibility.

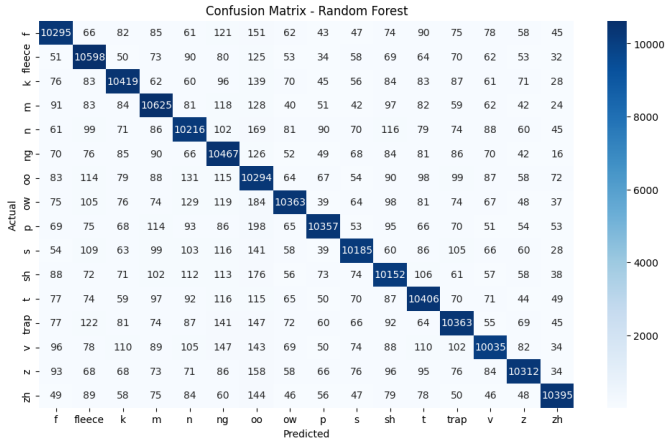

Fig. 7. Confusion Matrix for Random Forest

The confusion matrix in Figure 7 provides a detailed evaluation of the Random Forest model's predictions. The diagonal elements represent correctly classified instances for each of the 16 phoneme classes, while the off-diagonal values indicate the number of misclassifications. High diagonal values demonstrate the model's effectiveness in accurate classification.

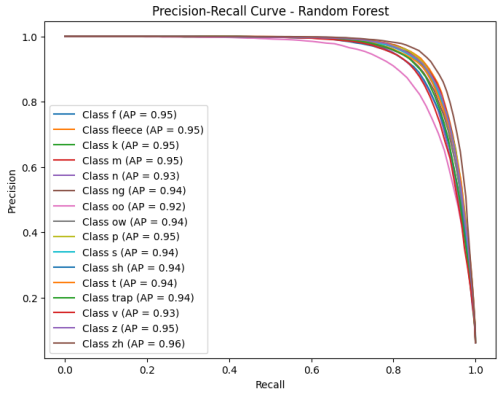

Fig. 8. Precision-Recall Curve for Random Forest

Figure 8 depicts the Precision-Recall curve, showcasing the trade-off between precision and recall for all phoneme classes. The Average Precision (AP) scores range between 0.92 and 0.96, reflecting the model's strong predictive capabilities and consistent performance across all classes.

The ROC curve shown in Figure 9 illustrates the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) for all phoneme classes. The Area Under the Curve

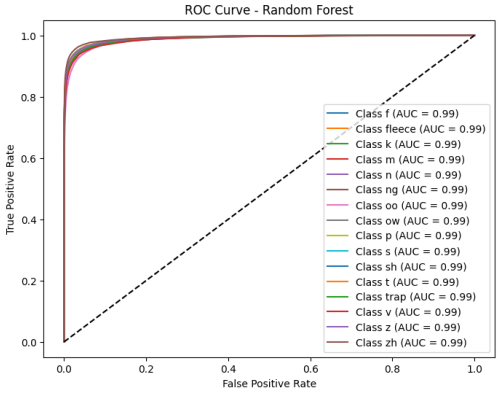

Fig. 9. ROC Curve for Random Forest

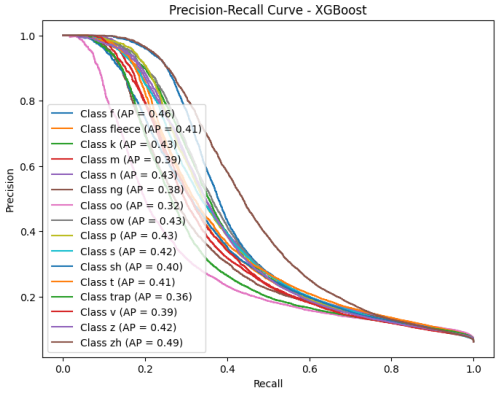

Fig. 11. Precision-Recall Curve for XGBoost

(AUC) for each class is approximately 0.99, indicating that the Random Forest model demonstrates excellent classification performance.

b) Gradient Boosting (XGBoost): XGBoost was chosen due to its efficiency and accuracy, particularly in handling tabular EEG data. XGBoost employs an iterative boosting approach to improve model performance and minimize errors. The key parameters for this model were a learning rate of 0.1, 100 boosting rounds, a maximum depth of 6 to control model complexity, and a random state of 42 for reproducibility.

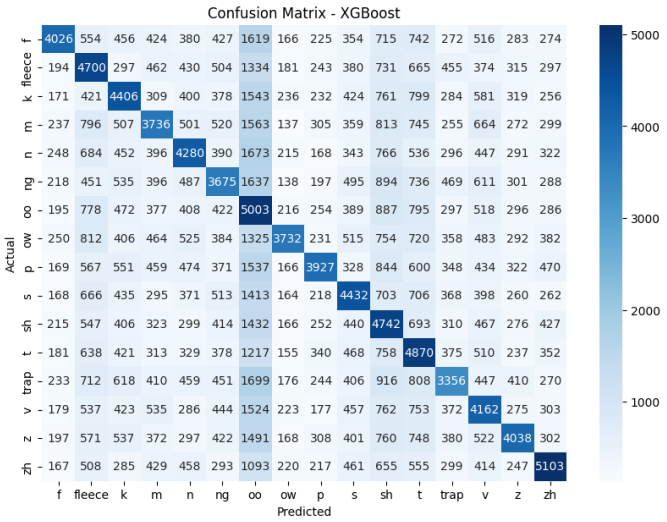

Fig. 10. Confusion Matrix for XGBoost

The confusion matrix in Figure 10 provides a detailed evaluation of the XGBoost model's predictions. The diagonal elements represent correctly classified instances for each phoneme class, while off-diagonal elements indicate misclassifications. The results show that certain classes, such as "zh" and "sh", were predicted with higher accuracy compared to others, reflecting the model's strengths and weaknesses.

Figure 11 illustrates the Precision-Recall curve for the XGBoost model, which depicts the trade-off between precision and recall for each phoneme class. The Average Precision (AP) values range between 0.32 and 0.49, indicating that while the model performs reasonably well, certain classes such as "zh" achieve higher precision compared to others like "oo" and "trap".

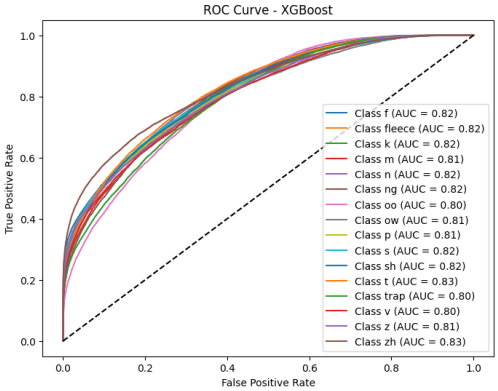

Fig. 12. ROC Curve for XGBoost

The ROC curve shown in Figure 12 presents the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) for all phoneme classes classified using the XGBoost model. The Area Under the Curve (AUC) values range between 0.80 and 0.83, indicating moderate classification performance compared to other models.

*2) **Deep Learning Models**:*

a) 1D Convolutional Neural Network (1D-CNN): The 1D-CNN model was implemented to extract spatial and temporal features from EEG signals. The architecture consists of the following components: EEG signals are first reshaped into a 1D format at the input layer. Subsequently, convolutional layers with 64 filters and a kernel size of 1 apply localized feature extraction, using the ReLU activation function. A max-pooling layer with a pool size of 1 is used to reduce feature dimensionality. The network then includes a dense layer with 32 neurons and ReLU activation, followed by a dropout layer with a rate of 0.2 for regularization. Finally, the output layer

consists of a softmax activation function with 16 neurons, corresponding to the phoneme classes.

b) Recurrent Neural Network (RNN): The RNN model was designed to capture sequential dependencies in EEG signals. The architecture begins with an input layer, where EEG data is reshaped into a 3D format to ensure compatibility with recurrent layers. A bidirectional Simple RNN layer with 64 units and ReLU activation processes the input sequences in both forward and backward directions. To prevent overfitting, a dropout layer with a rate of 0.3 is incorporated. The network concludes with a dense layer containing a softmax activation function for multi-class classification. For optimization, the Adam optimizer with a learning rate of 0.001 was used, and the sparse categorical cross-entropy loss function was employed to train the model effectively.

c) Long Short-Term Memory (LSTM): The LSTM model was implemented to capture long-term dependencies in EEG signals. The architecture consists of an input layer, where EEG sequences are reshaped for compatibility with LSTM layers. A bidirectional LSTM layer with 64 units and tanh activation models the sequential dependencies in both directions. To prevent overfitting, a dropout layer with a rate of 0.3 is applied. The final dense layer includes a softmax activation function to produce class probabilities for the 16 phoneme classes.

d) Transformer-Based Model: The Transformer-based model leverages the self-attention mechanism to capture global dependencies between EEG time steps. The architecture begins with an input embedding layer, where EEG signals are encoded into a 1D sequence. The encoded sequence is then passed through 2 transformer encoder blocks, each containing 4 attention heads and feedforward layers with a size of 64. A global average pooling layer reduces the dimensionality of the features extracted by the transformer encoders. The network concludes with a dense layer containing 32 neurons with ReLU activation, followed by a softmax output layer with 16 neurons for classification of the phoneme classes.

*3) **Fusion Model**:* To improve the classification performance, the predictions from the Random Forest and XGBoost models were combined using a weighted average ensemble method. Figure 13 illustrates the fusion approach, where the prediction probabilities from each model (p_{rf} and p_{gb}) are weighted and combined. The ensemble method works as follows:

- The prediction probabilities (p_{rf} and p_{gb}) are generated by the Random Forest and XGBoost models, respectively.
- Each model's prediction is assigned a weight, with $w_1 = 0.7$ for Random Forest and $w_2 = 0.3$ for XGBoost. These weights are optimized based on validation accuracy.
- The final prediction probability p_{final} is calculated using the weighted average formula:

$$p_{\text{final}} = w_1 \cdot p_{\text{rf}} + w_2 \cdot p_{\text{gb}}$$

- The combined prediction probabilities are used to determine the final class label, which is saved as the output of the ensemble model.

a) Significance of the Weighted Average Fusion:: The weighted average fusion method allows the strengths of both models to be utilized effectively. By assigning a higher weight to the Random Forest model ($w_1 = 0.7$) due to its superior individual performance, and a smaller weight to XGBoost ($w_2 = 0.3$), the ensemble model improves overall robustness and accuracy. This approach mitigates the limitations of individual models while combining their predictive capabilities.

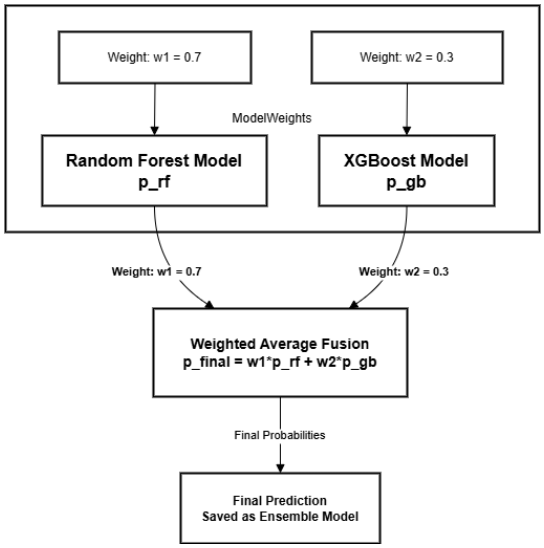

Fig. 13. Weighted Average Fusion Approach for Ensemble Model

b) Building an Ensemble Model:

- Step 1: Predictions (p_{rf} and p_{gb}) are obtained from the Random Forest and XGBoost models, respectively.
- Step 2: These probabilities are combined using the weighted average formula $p_{\text{final}} = 0.7p_{\text{rf}} + 0.3p_{\text{gb}}$, where the weights w_1 and w_2 are based on the validation performance of each model.
- Step 3: The final prediction probabilities (p_{final}) are used to determine the class labels.

The weighted average approach ensures that the Fusion Model combines the strengths of both Random Forest and XGBoost models, resulting in improved classification accuracy and robustness.

The confusion matrix in Figure 14 evaluates the performance of the Fusion Model across the 16 phoneme classes. The diagonal elements represent correctly classified instances, while the off-diagonal elements denote misclassifications. The Fusion Model demonstrates high accuracy, as indicated by the prominent diagonal, with minimal misclassifications across most classes.

The Precision-Recall curve shown in Figure 15 illustrates the trade-off between precision and recall for each phoneme class. The Fusion Model achieved consistently high Average Precision (AP) values across all classes, demonstrating its reliability in maintaining high precision and recall, even for challenging classes.

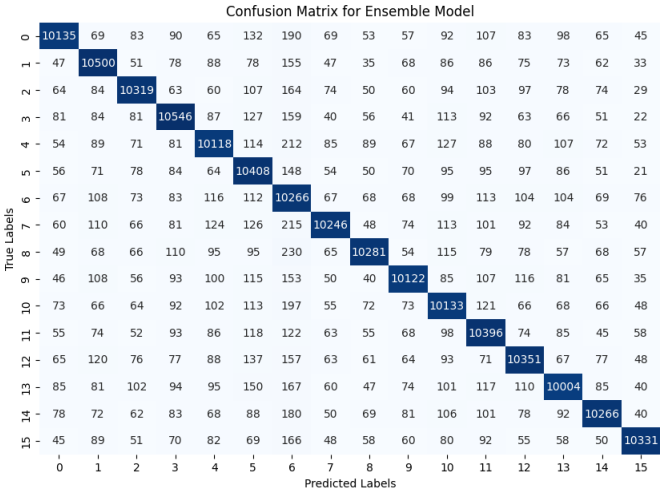

Fig. 14. Confusion Matrix for Fusion Model

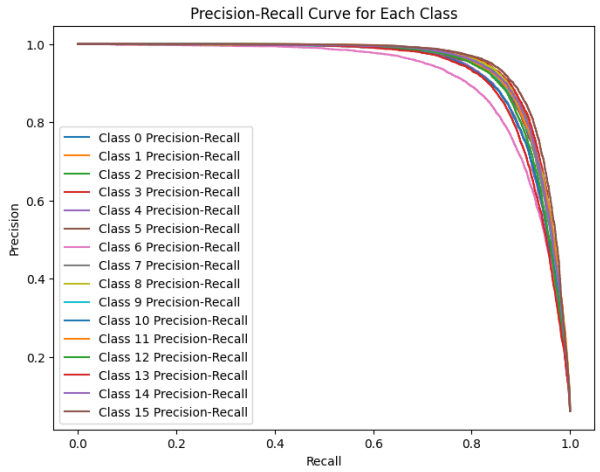

Fig. 15. Precision-Recall Curve for Fusion Model

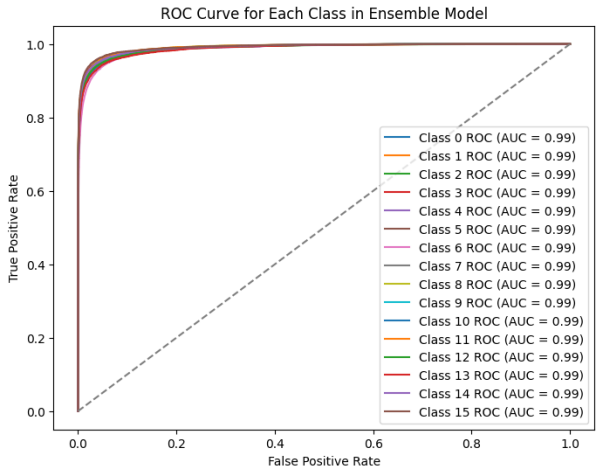

Fig. 16. ROC Curve for Fusion Model

The ROC curve presented in Figure 16 shows the True Positive Rate (TPR) against the False Positive Rate (FPR) for each phoneme class. The Fusion Model achieved an Area Under the Curve (AUC) of 0.99 across all classes, reflecting its exceptional ability to distinguish between the 16 phoneme classes with near-perfect performance.

IV. Results and Discussion

The performance of the implemented models is compared across four key evaluation metrics: Accuracy, Precision, Recall, and F1 Score. The results are summarized in Table III, and visualized in Figures 17, 18, 19, and 20.

TABLE III
Performance Metrics of Implemented Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score
Random Forest	84.77	84.55	87.43	84.79
XGBoost	36.99	40.73	36.99	37.73
RNN	7.48	38.74	7.48	4.2
LSTM	7.38	42.66	7.38	2.86
1D CNN	6.19	3.8	6.19	7.2
Transformer	6.19	3.8	6.89	8.9
Fusion Model	89.42	89.54	89.42	89.92

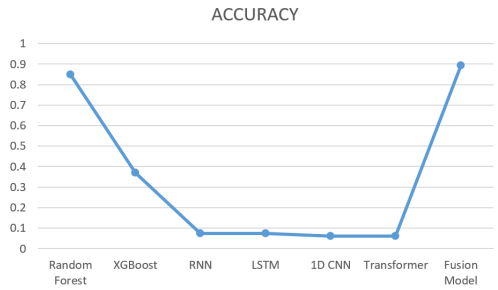

Fig. 17. Accuracy Comparison of Implemented Models

A. Accuracy Comparison

As shown in Figure 17, the Fusion Model achieved the highest accuracy of 89.42%, outperforming all other models. The Random Forest model follows with an accuracy of 84.77%, demonstrating its robustness. On the other hand, XGBoost achieved moderate accuracy (36.99%), while deep learning models (RNN, LSTM, 1D CNN, and Transformer) exhibited significantly lower accuracies, with all scoring below 10%. This highlights the limitations of deep learning models when handling smaller datasets or high-dimensional EEG data.

B. Precision Comparison

The Precision values, depicted in Figure 18, show a similar trend. The Fusion Model achieved the highest precision of 89.54%, indicating its strong ability to minimize false positives. Random Forest also performed well, achieving 84.55% precision. In contrast, the deep learning models, particularly 1D CNN and Transformer, recorded near-zero precision values (3.8%). This underperformance indicates their inability to generalize effectively with the given data.

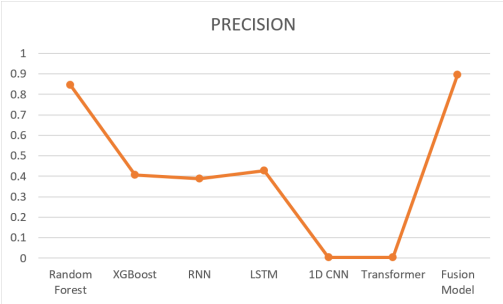

Fig. 18. Precision Comparison of Implemented Models

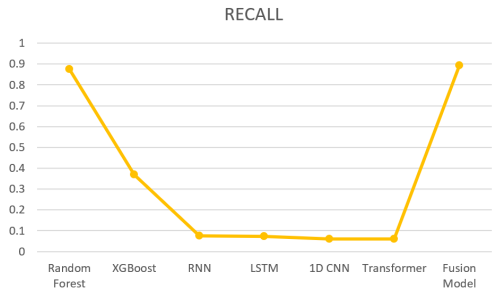

Fig. 19. Recall Comparison of Implemented Models

C. Recall Comparison

Figure 19 illustrates the Recall performance, where the Fusion Model and Random Forest excelled with values of 89.42% and 87.43%, respectively. XGBoost recorded a moderate recall of 36.99%. However, the deep learning models again underperformed, with recall values below 10%, except for Transformer (6.89%). This suggests that the deep learning models struggled to capture the temporal dependencies and discriminative features in EEG signals.

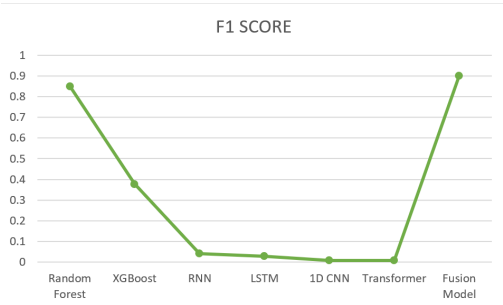

Fig. 20. F1 Score Comparison of Implemented Models

D. F1 Score Comparison

The F1 Score, which balances Precision and Recall, is presented in Figure 20. The Fusion Model achieved the best F1 Score of 89.92, followed by Random Forest with 84.79. The XGBoost model achieved a moderate F1 score of 37.73%, while deep learning models performed poorly, with 1D CNN and Transformer recording near-zero F1 scores (7.2% and 8.9%, respectively).

E. Comparative Analysis

- Superiority of the Fusion Model: The Fusion Model consistently outperformed all other models across all metrics. By leveraging the strengths of Random Forest and XGBoost, the Fusion Model achieved the highest accuracy, precision, recall, and F1 score, demonstrating its robustness and reliability for EEG-based phoneme classification.
- Performance of Random Forest: Random Forest emerged as the best standalone model, achieving competitive results across all metrics. Its ensemble learning approach allowed it to capture the variability in EEG signals effectively.
- Limitations of Deep Learning Models: The RNN, LSTM, 1D CNN, and Transformer models underperformed due to challenges in handling the high-dimensional EEG data with limited training samples. This highlights the need for larger datasets or improved optimization techniques for deep learning models.
- XGBoost as a Moderate Performer: While XGBoost did not match the performance of Random Forest or the Fusion Model, it achieved moderate results, particularly in precision and recall, making it a valuable standalone method.

V. Conclusion and Future Work

The comparative analysis demonstrates that ensemble models, particularly the Fusion Model, offer significant advantages for EEG-based phoneme classification tasks. The superior performance of the Fusion Model highlights the effectiveness of combining predictions from multiple models using optimized weights.

1) Conclusion: This study demonstrated the classification of EEG signals into phonemes using traditional machine learning models, deep learning architectures, and ensemble techniques. The key findings of this research are as follows:

- Traditional models, such as Random Forest and XGBoost, provided strong baseline performance for EEG classification, with Random Forest achieving an accuracy of 84.77%.
- Deep learning models, including 1D-CNN, RNN, LSTM, and Transformer, faced challenges in learning meaningful patterns from the EEG dataset due to its limited size and high variance.
- The Fusion Model, combining predictions from Random Forest and XGBoost through a weighted averaging approach, achieved the highest accuracy of 89.42%. This result highlights the effectiveness of ensemble techniques in leveraging the strengths of multiple models to improve robustness and accuracy.

The findings of this study underscore the potential of machine learning and deep learning techniques for Brain-Computer Interface (BCI) applications. The results provide a strong foundation for further advancements in EEG signal classification, particularly for speech and communication technologies.

A. Future Work

This project successfully demonstrated the feasibility of EEG-based phoneme classification and achieved promising results. With additional resources and advanced hardware, the following enhancements can further improve the system's robustness and practical applicability:

1) Development of a Custom EEG Machine: While the current study utilized simulated EEG data, future advancements can focus on designing a custom EEG system using 14 BioAmp EXG Pill electrodes. These electrodes, known for their precision and efficiency, can be leveraged to collect real EEG signals. The proposed enhancements include:

- Signal Acquisition: Strategically placing 14 BioAmp electrodes over the scalp to capture high-quality brainwave signals from key regions associated with speech processing.
- Data Transmission: Integrating ESP modules to enable wireless transmission of EEG signals to a central processing unit, ensuring seamless data flow.
- Signal Preprocessing Unit: Implementing advanced filtering, noise removal, and amplification techniques to ensure signal clarity before classification.

The development of a custom EEG machine would enable real-time data acquisition and strengthen the system's practical usability in real-world applications.

2) Real-Time Processing and Speech Synthesis: Building on the success of the current classification framework, future efforts can focus on implementing real-time processing pipelines to facilitate end-to-end communication. Enhancements include:

- Latency Optimization: Refining machine learning models to reduce processing time for real-time phoneme classification.
- Speech Synthesis: Incorporating speech synthesis techniques to convert the predicted phonemes into audible words, enabling seamless communication.

With these advancements, the system can evolve into a fully functional brain-computer interface for practical applications, such as assistive communication for individuals with speech impairments.

REFERENCES

[1] Luo, S., Rabbani, Q. and Crone, N.E., 2023. Brain-computer interface: applications to speech decoding and synthesis to augment communication. *Neurotherapeutics*, 19(1), pp.263-273.

[2] Peksa, Janis, and Dmytro Mamchur. 2023. "State-of-the-Art on Brain-Computer Interface Technology" Sensors 23, no. 13: 6001. https://doi.org/10.3390/s23136001.

[3] Angrick, M., Luo, S., Rabbani, Q. *et al.* Online speech synthesis using a chronically implanted brain-computer interface in an individual with ALS. *Scientific Reports* 14, 9617 (2024). https://doi.org/10.1038/s41598-024-60277-2.

[4] Brumberg, Jonathan, Nieto-Castanon, Alfonso, Kennedy, Philip, Guenther, Frank. (2010). Brain-Computer Interfaces for Speech Communication. *Speech Communication*, 52, 367-379. https://doi.org/10.1016/j.specom.2010.01.001.

[5] Allison, Brendan Z., Elizabeth Winter Wolpaw, and Jonathan R. Wolpaw. "Brain-computer interface systems: progress and prospects." *Expert Review of Medical Devices* 4, no. 4 (2007): 463-474.

[6] Warshi, A. (2023). Brain-Computer Interface for Converting Thoughts to Speech.

[7] Zhang, Dalin, Yao, Lina, Zhang, Xiang, Wang, Sen, Chen, Weitong, Boots, Robert. (2017). EEG-based Intention Recognition from Spatio-Temporal Representations via Cascade and Parallel Convolutional Recurrent Neural Networks.

[8] Hong, Y., Ryun, S., Chung, C. K. (2024). Evoking artificial speech perception through invasive brain stimulation for brain-computer interfaces: Current challenges and future perspectives. *Frontiers in Neuroscience*, 18, 1428256. https://doi.org/10.3389/fnins.2024.1428256.

[9] Zhang, A., Su, L., Zhang, Y., Fu, Y., Wu, L., and Liang, S., 2021. EEG data augmentation for emotion recognition with a multiple generator conditional Wasserstein GAN. *Complex Intelligent Systems*, pp.1-13.

[10] Kübler, A., Furdea, A., Halder, S., Hammer, E.M., Nijboer, F. and Kotchoubey, B., 2009. A brain–computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients. *Annals of the New York Academy of Sciences*, 1157(1), pp.90-100.

[11] Vansteensel, M.J., Pels, E.G., Bleichner, M.G., Branco, M.P., Denison, T., Freudenburg, Z.V., *et al.*, 2016. Fully implanted brain–computer interface in a locked-in patient with ALS. *New England Journal of Medicine*, 375(21), pp.2060-2066.

[12] Bartels, Jess, Andreasen, Dinal, Ehirim, Princewill, Mao, Hui, Seibert, Steven, Wright, E. Joe, and Kennedy, Philip. Neurotrophic electrode: Method of assembly and implantation into human motor speech cortex. *Journal of Neuroscience Methods*, Volume 174, Issue 2, 2008, Pages 168-176. https://doi.org/10.1016/j.jneumeth.2008.06.030.

[13] Mridha, M.F., Das, S.C., Kabir, M.M., Lima, A.A., Islam, M.R., and Watanobe, Y., 2021. Brain-computer interface: Advancement and challenges. *Sensors*, 21(17), p.5746.

[14] Värbu, K., Muhammad, N., and Muhammad, Y., 2022. Past, present, and future of EEG-based BCI applications. *Sensors*, 22(9), p.3331.

[15] Bauer, G., Gerstenbrand, F., and Rumpl, E., 1979. Varieties of the locked-in syndrome. *Journal of Neurology*, 221, pp.77-91.

[16] S. Wellington and J. Clayton, "Fourteen-channel EEG with imagined speech (FEIS) dataset," Aug 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3369179

[17] Vlek, Rutger J. MSc; Steines, David MSc; Szibbo, Dyana MSc; Kübler, Andrea Prof.; Schneider, Mary-Jane PhD; Haselager, Pim PhD; Nijboer, Femke PhD. Ethical Issues in Brain–Computer Interface Research, Development, and Dissemination. Journal of Neurologic Physical Therapy 36(2):p 94-99, June 2012. | DOI: 10.1097/NPT.0b013e31825064cc

[18] T. W. Picton, "The P300 wave of the human event-related potential," Journal of clinical neurophysiology, vol. 9, no. 4, pp. 456– 479, 1992.

[19] G. H. Klem, H. O. Luders, H. Jasper, C. Elger ¨ et al., "The tentwenty electrode system of the international federation," Electroencephalogr Clin Neurophysiol, vol. 52, no. 3, pp. 3–6, 1999.

[20] Emotiv EPOC+. [Online]. Available: https://www.emotiv.com/ epoc/

[21] M. L. Giudice et al., "Visual Explanations of Deep Convolutional Neural Network for eye blinks detection in EEG-based BCI applications," 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 2022, pp. 01-08, doi: 10.1109/IJCNN55064.2022.9892567.

[22] H. Li, X. Li and J. d. R. Millán, "Noninvasive EEG-Based Intelligent Mobile Robots: A Systematic Review," in IEEE Transactions on Automation Science and Engineering, doi: 10.1109/TASE.2024.3441055

[23] Rakhmatulin, I.; Dao, M.-S.; Nassibi, A.; Mandic, D. Exploring Convolutional Neural Network Architectures for EEG Feature Extraction. Sensors 2024, 24, 877. https://doi.org/10.3390/s24030877

[24] Sharma, R., Meena, H.K. Emerging Trends in EEG Signal Processing: A Systematic Review. SN COMPUT. SCI. 5, 415 (2024). https://doi.org/10.1007/s42979-024-02773-w

[25] Sun C and Mou C (2023) Survey on the research direction of EEG-based signal processing. Front. Neurosci. 17:1203059. doi: 10.3389/fnins.2023.1203059

[26] Subha, G. & Priya.R, Hema & Warshi, Alam & yappan, M.I. (2023). BRAIN-COMPUTER INTERFACE FOR CONVERTING THOUGHTS TO SPEECH. International Scientific Journal of Engineering and Management. 02. 10.55041/ISJEM00110.

[27] 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 Feb 14;9(1):52. doi: 10.1038/s41597-022-01147-2. PMID: 35165308; PMCID: PMC8844234.

[28] Foteini Simistira Liwicki, Vibha Gupta, Rajkumar Saini, Kanjar De, Nosheen Abid, Sumit Rakesh, Scott Wellington, Holly Wilson,

Marcus Liwicki, Johan Eriksson bioRxiv 2022.05.24.492109; doi: https://doi.org/10.1101/2022.05.24.492109

[29] Saha, P., & Fels, S. (2019). Hierarchical Deep Feature Learning for Decoding Imagined Speech from EEG. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 10019-10020. https://doi.org/10.1609/aaai.v33i01.330110019

[30] F. Huang, Y. He, X. Deng and W. Jiang, "A Novel Discount-Weighted Average Fusion Method Based on Reinforcement Learning For Conflicting Data," in IEEE Systems Journal, vol. 17, no. 3, pp. 4748-4751, Sept. 2023, doi: 10.1109/JSYST.2022.3228015.

[31] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," in IEEE Access, vol. 10, pp. 99129-99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[32] Yang, Y., Duan, Y., Zhang, Q., Jo, H., Zhou, J., Lee, W. H., Xu, R., & Xiong, H. (2024). NeuSpeech: Decode Neural Signal as Speech. *arXiv preprint arXiv:2403.01748*. Retrieved from https://arxiv.org/abs/2403.01748.

[33] Faruk, M. J. H., Valero, M., & Shahriar, H. (2022). An Investigation on Non-Invasive Brain-Computer Interfaces: Emotiv Epoc+ Neuroheadset and Its Effectiveness. *arXiv preprint arXiv:2207.06914*. Retrieved from https://arxiv.org/abs/2207.06914.

[34] Zhang, Wenchang & Tan, Chuanqi & Sun, Fuchun & Wu, Hang & Zhang, Bo. (2018). A Review of EEG-Based Brain-Computer Interface Systems Design. Brain Science Advances. 4. 156-167. 10.26599/BSA.2018.9050010.

[35] Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. Nature. 2019 Apr;568(7753):493-498. doi: 10.1038/s41586-019-1119-1. Epub 2019 Apr 24. PMID: 31019317; PMCID: PMC9714519.

[36] Makin JG, Moses DA, Chang EF. Machine translation of cortical activity to text with an encoder-decoder framework. Nat Neurosci. 2020 Apr;23(4):575-582. doi: 10.1038/s41593-020-0608-8. Epub 2020 Mar 30. PMID: 32231340; PMCID: PMC10560395.

[37] Moses DA, Leonard MK, Makin JG, Chang EF. Real-time decoding of question-and-answer speech dialogue using human cortical activity. Nat Commun. 2019 Jul 30;10(1):3096. doi: 10.1038/s41467-019-10994-4. PMID: 31363096; PMCID: PMC6667454.

[38] Angrick M, Herff C, Mugler E, Tate MC, Slutzky MW, Krusienski DJ, Schultz T. Speech synthesis from ECoG using densely connected 3D convolutional neural networks. J Neural Eng. 2019 Jun;16(3):036019. doi: 10.1088/1741-2552/ab0c59. Epub 2019 Mar 4. PMID: 30831567; PMCID: PMC6822609.

[39] Dash, D., Wisler, A., Ferrari, P., Adjeroh, D. A., & Obeid, I. (2020). A Review on Real-Time EEG-Based Brain-Computer Interface Systems for Speech Communication. *Brain Sciences*, 10(11), 930. https://doi.org/10.3390/brainsci10110930.

[40] Abiri R, Borhani S, Sellers EW, Jiang Y, Zhao X. A comprehensive review of EEG-based brain-computer interface paradigms. J Neural Eng. 2019 Feb;16(1):011001. doi: 10.1088/1741-2552/aaf12e. Epub 2018 Nov 15. PMID: 30523919.

[41] Gu X, Cao Z, Jolfaei A, Xu P, Wu D, Jung TP, Lin CT. EEG-Based Brain-Computer Interfaces (BCIs): A Survey of Recent Studies on Signal Sensing Technologies and Computational Intelligence Approaches and Their Applications. IEEE/ACM Trans Comput Biol Bioinform. 2021 Sep-Oct;18(5):1645-1666. doi: 10.1109/TCBB.2021.3052811. Epub 2021 Oct 7. PMID: 33465029.

[42] Fodor, M. A., Csapó, T. G., & Arthur, F. V. (2024). Towards Decoding Brain Activity During Passive Listening of Speech. ArXiv. https://doi.org/10.15775/Besztud.2024.1.158-184