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

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.

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.

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] [24] 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] [23] is a portable, 14-channel EEG headset designed for real-time brain signal acquisition and analysis. It features wireless connectivity, making it suitable 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] [28] 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] [25] 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.

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.

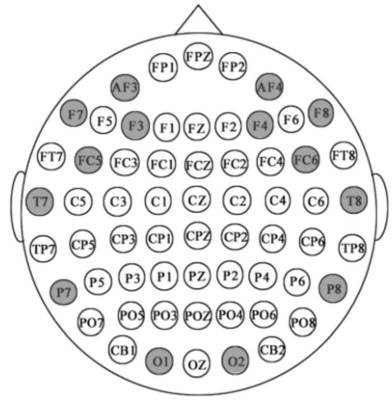


Fig. 1. Electrode positions in the 10-20 EEG montage. Key electrodes like F3, F4, T7, T8, and others are highlighted for their importance in speech processing.

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.

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.

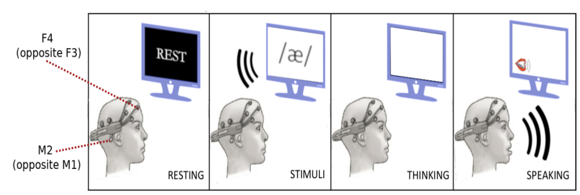


Fig. 2. Illustration of the recording procedure. Participants listen to five repetitions of a phone (recorded in their own voice), then imagine speaking the phone five times (with the same rhythm), then overly speak the phone five times.

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 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.

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, developed by Upside Down Labs, is a compact and cost-effective biosignal amplifier for EEG, EMG, ECG, and EOG signal acquisition. Its versatility and open-source nature make it ideal for BCI applications.

1) Features and Specifications: Key features and specifications include:

- Input Voltage: 4.5 – 40 V
- Input Impedance: 10^{12} ohm
- Biopotentials: Configurable for ECG, EMG, EOG, or EEG (default: EEG and EOG)
- Electrodes: Configurable for 2 or 3 electrodes (default: 3 electrodes)
- Compatible Hardware: Any ADC input
- Channels: 1
- Dimensions: 25.4 x 10.0 mm

2) BioAmp Circuit Design: The BioAmp EXG Pill's architecture includes:

- Instrumentation Amplifier (INA): Accurate signal amplification with noise reduction.
- Electrode Reference Configuration: Flexible for 2-electrode or 3-electrode setups.
- Bandpass Filter: Removes noise and optimizes signals for target applications.
- Power Supply Filtering: Ensures clean and stable signal acquisition.
- Amp Ref + DRL: Reduces noise and stabilizes input signals.
- Header Pins: Easy integration with microcontrollers and ADCs.

Figures 3, 4, and 5 depict the circuit design, front, and rear views of the PCB.

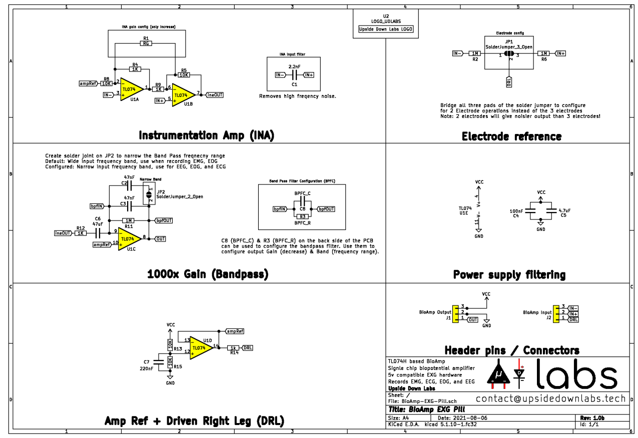


Fig. 3. BioAmp EXG Pill - Circuit Design and Functional Blocks.

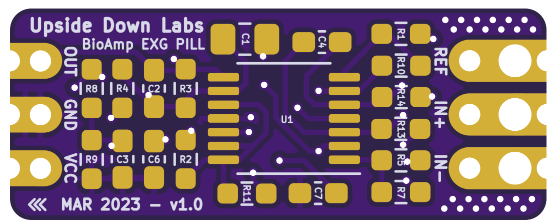


Fig. 4. BioAmp EXG Pill - Front View of the PCB

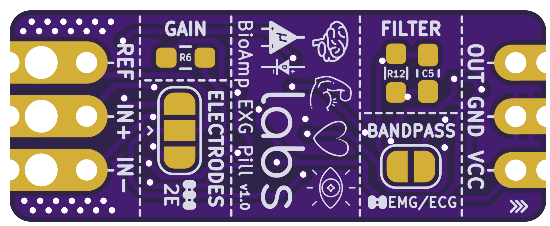


Fig. 5. BioAmp EXG Pill - Rear View of the PCB

3) Role in the Project: The BioAmp EXG Pill is integral for EEG data acquisition and preprocessing. Its key contributions include:

- Cost-Effective: Affordable alternative to high-end EEG systems(electrodes).
- High Signal Quality: Noise-free signals improve classification accuracy.
- Real-Time Data Transmission: Wireless ESP modules enable seamless data flow for real-time BCI applications.
- Open Source Hardware and Software: Backyard Brains or Chords from Upside Down Labs

Using 14 strategically placed BioAmp electrodes, this system enables precise measurement of brain activity for phoneme processing.

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**:*

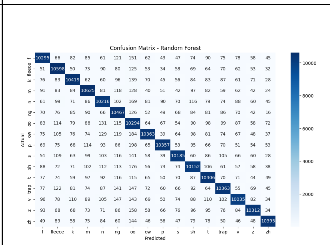
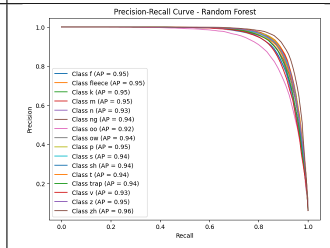
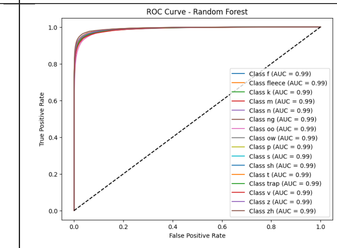
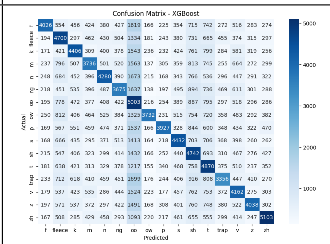
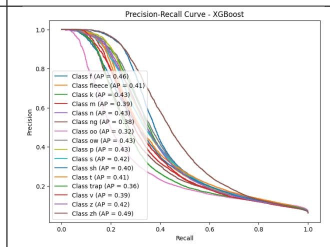
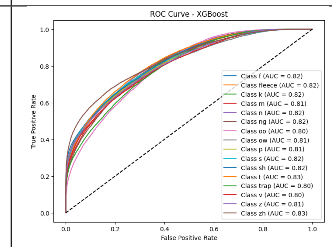
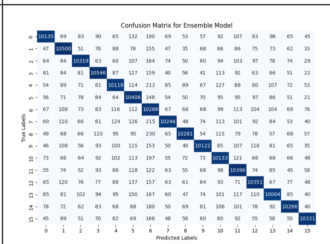
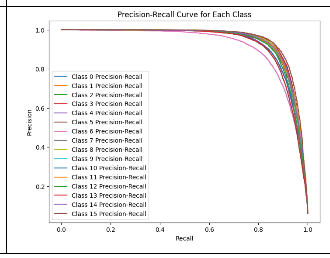
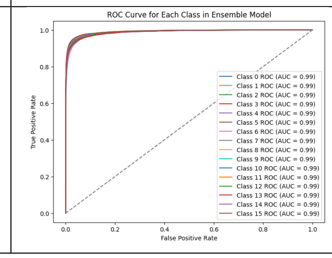
a) Random Forest (RF): The Random Forest model was implemented to establish baseline performance. Random Forest works as an ensemble method that combines outputs from multiple decision trees to make predictions, offering interpretability and robustness. The key parameters for this model included the 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.

The performance of the Random Forest model is summarized in Table III, which includes the confusion matrix, Precision-Recall curve, and ROC curve. The model achieved an Average Precision (AP) of 0.92–0.96 and an Area Under the Curve (AUC) of 0.99, reflecting its strong predictive performance and robustness.

b) Gradient Boosting (XGBoost): XGBoost, known for its efficiency in tabular EEG data, was configured with a learning rate of 0.1, 100 boosting rounds, a maximum depth of 6, and a random state of 42 for reproducibility.

Table III highlights the performance of the XGBoost model. While it achieved moderate performance with an AP of

TABLE III
SUMMARY OF MODELS, CONFUSION MATRICES, PRECISION-RECALL CURVES, AND ROC CURVES

Model	Confusion Matrix	Precision-Recall Curve	ROC Curve
Random Forest (RF)			
Gradient Boosting (XGBoost)			
Fusion Model			

0.32–0.49 and an AUC of 0.80–0.83, it showed strengths in specific classes like "zh" and "sh." However, its overall performance was inferior to Random Forest and the Fusion Model.

*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. While effective in reducing feature dimensionality, its performance was suboptimal, failing to generalize well for all phoneme classes.

b) Recurrent Neural Network (RNN): The RNN model aimed to capture sequential dependencies in EEG signals using bidirectional layers. However, it struggled with overfitting and delivered lower accuracy than Random Forest and XGBoost.

c) Long Short-Term Memory (LSTM): The LSTM model was designed to handle long-term dependencies in EEG data. Despite capturing temporal features effectively, it exhibited higher computational cost and lower classification accuracy.

d) Transformer-Based Model: The Transformer model used self-attention to capture global dependencies in EEG time steps. However, its performance was limited, particularly for imbalanced classes, and it required significant computational resources.

e) Comparative Analysis: As summarized in Table III, deep learning models (1D-CNN, RNN, LSTM, Transformer) underperformed compared to machine learning models like Random Forest and XGBoost. Their lower accuracy and high computational cost made them less viable for this project, ultimately leading to the development of the Fusion Model.

*3) **Fusion Model**:* The Fusion Model combines predictions from Random Forest and XGBoost using a weighted average ensemble method, with weights optimized to $w_1 = 0.7$ for Random Forest and $w_2 = 0.3$ for XGBoost. The combined prediction probabilities were computed as:

$$p_{\text{final}} = 0.7 \cdot p_{\text{rf}} + 0.3 \cdot p_{\text{gb}}$$

Table III highlights the performance of the Fusion Model, which achieved the best results across all metrics. With an AP of 0.92–0.96 and an AUC of 0.99, it demonstrated exceptional accuracy and robustness. The confusion matrix reflects minimal misclassifications, while the Precision-Recall and ROC curves illustrate its ability to maintain high precision and recall across all phoneme classes.

a) Significance of the Fusion Model: By combining the strengths of Random Forest and XGBoost, the Fusion Model outperformed all individual models in accuracy, precision, and recall. Its robust performance ensures reliable classification of EEG-based phoneme data, as shown in Table III.

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 IV and visualized in Figures 7 and 8.

TABLE IV
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

Table IV highlights the comparative analysis of all models across key metrics. The Fusion Model emerged as the best performer with the highest accuracy (89.42%), precision (89.54%), recall (89.42%), and F1 score (89.92). Random Forest followed closely, with robust performance across all metrics, making it the best standalone model. In contrast, the deep learning models, including RNN, LSTM, 1D CNN, and Transformer, showed significantly lower results due to challenges in handling high-dimensional EEG data with limited training samples.

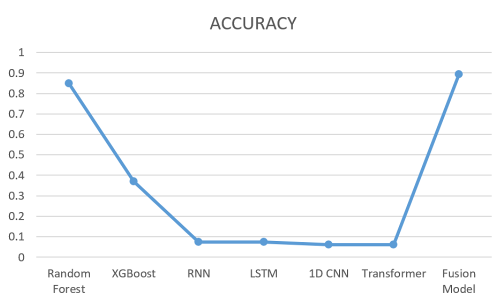


Fig. 7. Accuracy Comparison of Implemented Models

A. Accuracy and F1 Score Comparison

As shown in Figure 7, the Fusion Model achieved the highest accuracy of 89.42%, followed by Random Forest with 84.77%. XGBoost recorded moderate accuracy (36.99%), while the deep learning models all scored below 10%. This demonstrates the superiority of ensemble-based approaches for EEG-based phoneme classification.

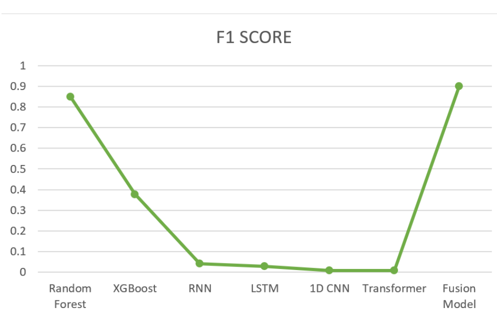


Fig. 8. F1 Score Comparison of Implemented Models

Figure 8 presents the F1 score, balancing precision and recall. The Fusion Model again outperformed all other models with an F1 score of 89.92, reflecting its consistent performance across all metrics. Random Forest performed well as a standalone model, while deep learning models failed to achieve competitive F1 scores due to their inability to generalize effectively.

B. Comparative Analysis

- Fusion Model: The Fusion Model achieved the best performance across all metrics, leveraging the strengths of Random Forest and XGBoost for robust and reliable phoneme classification.
- Random Forest: As the best standalone model, Random Forest demonstrated strong accuracy, precision, recall, and F1 scores, effectively capturing the variability in EEG signals.
- Deep Learning Models: RNN, LSTM, 1D CNN, and Transformer models struggled to handle high-dimensional EEG data with limited training samples, resulting in poor overall performance.
- XGBoost: While moderate in performance, XGBoost contributed significantly as part of the Fusion Model, providing valuable insights for ensemble-based classification.

V. CONCLUSION AND FUTURE WORK

The comparative analysis highlights the effectiveness of ensemble techniques, particularly the Fusion Model, for EEG-based phoneme classification. By leveraging the strengths of Random Forest and XGBoost, the Fusion Model achieved the highest accuracy of 89.42%, demonstrating its robustness and reliability.

1) Conclusion: This study explored EEG-based phoneme classification using machine learning, deep learning, and ensemble techniques. Key findings include:

- Random Forest provided a strong baseline with an accuracy of 84.77%.
- Deep learning models (1D-CNN, RNN, LSTM, Transformer) struggled due to the dataset's limited size and high variance.
- The Fusion Model outperformed all other approaches, showcasing the benefits of weighted ensemble techniques.

These results underline the potential of ensemble models in BCI applications, offering a foundation for advancing EEG signal classification for speech technologies.

A. Future Work

Future efforts can focus on enhancing the system's practicality and robustness through the following:

a) Custom EEG Machine Development: Developing a custom EEG system using 14 BioAmp EXG Pill electrodes for real-time data acquisition. Key improvements include:

- Enhanced signal acquisition from key speech-processing regions.
- Wireless data transmission using ESP modules.
- Advanced preprocessing for noise-free signal clarity.

b) Real-Time Processing and Speech Synthesis: Implementing real-time pipelines for end-to-end communication:

- Optimizing models to reduce latency for real-time classification.
- Integrating speech synthesis to convert classified phonemes into audible words.

These advancements would enable the system to function as a practical brain-computer interface, particularly for assistive communication.

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.