Ensemble Deep Learning for Advanced EEG-Based Grasp-and-Lift Detection

Aditya Shaurya Singh Negi
Computer Science and Engineering
National Institute of Technology, Delhi
New Delhi, India
221210012@nitdelhi.ac.in

Akshat Singh
Computer Science and Engineering
National Institute of Technology, Delhi
New Delhi, India
221210015@nitdelhi.ac.in

Akshat
Computer Science and Engineering
National Institute of Technology, Delhi
New Delhi, India
221210014@nitdelhi.ac.in

Dr. Arjun Singh Rawat
Assistant Professor, Supervisor
Department of Computer Science and Engineering
National Institute of Technology, Delhi
New Delhi, India
arjunsinghrawat005@gmail.com

***Abstract*—This paper presents a novel approach for detecting grasp-and-lift (GAL) events from electroencephalography (EEG) signals. People with neuromuscular dysfunctions and amputated limbs require automatic prosthetic appliances, making the precise detection of brain motor actions imperative for GAL tasks. Due to its low-cost and noninvasive nature, EEG is widely preferred for detecting motor actions while controlling prosthetic tools. We introduce an ensemble architecture combining temporal convolutional networks (TCN), bidirectional long short-term memory (BiLSTM), and attention mechanisms. Our methodology includes advanced signal preprocessing, feature extraction, and data augmentation techniques specifically designed for EEG time-series data.**

***Index Terms*—EEG, Brain-Computer Interface, Deep Learning, Signal Processing, Ensemble Learning, Temporal Convolutional Networks, BiLSTM, Attention Mechanisms**

I. INTRODUCTION

Brain-computer interfaces (BCI) that utilize electroencephalography (EEG) have shown tremendous potential to enable interaction between humans and assistive devices, particularly for individuals with motor impairment. Reliable detection of user intent from non-invasive EEG signals remains challenging due to their low signal-to-noise ratio, non-stationarity, and high inter-subject variability.

In this research, we focus on detecting six specific grasp-and-lift events from EEG recordings. These events represent key phases in a natural grasping motion and their accurate detection could enable more intuitive control of prosthetic devices or rehabilitation robots for individuals with neuromuscular disorders, cerebral palsy, stroke, or amputation.

The contributions of this work include:

1) A comprehensive preprocessing pipeline for EEG signal enhancement incorporating optimized bandpass filtering, notch filtering, and robust scaling
2) Novel feature extraction methods combining traditional EEG features with advanced spectral and temporal characteristics, including wavelet transform analysis
3) Multiple deep learning architectures specialized for EEG signal processing, including EEGNet [1] with temporal convolutional networks [2], CNN-BiLSTM hybrids [3], recurrent convolutional neural networks, and attention-based models
4) An ensemble learning framework [4] with meta-model integration using gradient boosting and feature priority analysis
5) Test-time augmentation strategies for improved prediction robustness across different subjects

II. RELATED WORK

The processing of EEG signals typically follows a standard pipeline consisting of artifact rejection, time-domain filtering, spatial filtering, feature generation, and classification, as outlined by several researchers in the field of real-time BCI systems [5]. While traditional approaches to EEG signal processing relied heavily on manual feature extraction, recent work by Hasan et al. (2022) demonstrates the advantage of end-to-end deep learning pipelines. Their CNN-based model with DWT preprocessing achieved state-of-the-art performance on the WAY-EEG-GAL dataset. This aligns with our proposed methodology, though we extend their work by incorporating attention mechanisms and ensemble techniques to further improve classification robustness across different subjects. [6]

Electroencephalogram (EEG) signal classification has been a critical area of research in brain-computer interfaces (BCIs) and other neuroscience applications. The work by Xu et al. (2019) presents a novel deep transfer learning framework based on convolutional neural networks (CNNs) for EEG signal classification. Their approach leverages the VGG-16 model, a well-known deep CNN architecture, combined with transfer learning to enhance the performance and generalizability of EEG signal classifiers [7].

More recent advances have highlighted the importance of feature priority analysis in EEG signal processing. Research has shown that using random forest algorithms to prioritize EEG detection electrodes, followed by wavelet transform for feature extraction, can significantly improve classification accuracy. This approach, when combined with convolutional neural networks, has achieved classification accuracy rates of up to 93.22% on the WAY-EEG-GAL dataset [8], demonstrating the effectiveness of selective feature prioritization.

Recurrent neural networks (RNNs) have also emerged as powerful tools for analyzing EEG time series data. Similar to their success in speech recognition, RNNs can effectively process the temporal dynamics of multi-channel EEG signals. Several studies have proposed recurrent convolutional neural network architectures that can significantly improve classification quality by adding recurrent connections to convolutional layer neurons, effectively increasing network depth while maintaining a constant number of parameters through weight sharing [9]. Our work builds upon these advances by implementing hybrid architectures that combine the spatial feature extraction capabilities of CNNs with the temporal modeling power of RNNs.

For consistent prediction of hand motion phases in grasp-and-lift tasks, researchers have also explored various post-processing techniques such as moving averages to stabilize predictions from RNN models [10]. This approach is particularly relevant to our work as we seek to create robust real-time classification systems for prosthetic control.

III. Methodology

This section outlines the methodological framework employed to address the problem of event detection in electroencephalography (EEG) signals during grasp-and-lift tasks. This methodology is structured to ensure reproducibility through organized data handling, signal processing, feature engineering, model development, and evaluation. First we begin by describing the dataset and its contents, followed by a comprehensive overview of the preprocessing pipeline designed specifically to enhance signal quality and to standardize the data. Feature extraction and the data augmentation techniques that are used to improve model robustness are discussed after this. We then present the different deep learning model architectures explored in this study, highlighting their suitability. Finally, we discuss the ensemble strategies used, its significance in this study, and the validation protocols. Each methodological component was carefully validated and scrutinized by us to maximize the accuracy and precision of event classification from EEG recordings.

A. Dataset

The dataset used in this work is the WAY-EEG-GAL dataset, originally introduced in the Kaggle competition "Grasp-and-Lift EEG Detection" [11]. This dataset is designed to facilitate the decoding of human sensation, intention, and action from scalp electroencephalography (EEG) signals. Specifically, it focuses on the classification of six temporally ordered motor events involved in a grasp-and-lift task using 32-channel EEG data.

Twelve healthy subjects participated in the experiment. During the trials, subjects were instructed to reach for, grasp, and lift a manipulandum object with variable properties. The weight of the object (165 g, 330 g, or 660 g), the surface friction (sandpaper, suede, or silk), or both were unpredictably altered between trials. The EEG data were recorded at a sampling rate of 500 Hz.

Each subject performed ten series of trials, with approximately 30 trials per series (though the exact number varied). The first eight series were used for training, and the last two series served as the test set. Ground-truth event labels were only available for the training data. Each series included frame-wise annotations for six key motor events, detailed in Table I. For evaluation, an event is considered correctly predicted if its probability exceeds a threshold within ±150 ms (±75 frames) of the ground-truth time.

GAL Events	**Description**
Hand Start (HS)	Reaching for the object
First Digit Touch (FDT)	Grasping the object using the thumb and index finger
Both Start Load Phase (BSP)	Beginning the lifting phase of the object
Lift Off (LO)	Object is lifted off the surface
Replace (R)	Placing the object back on the support surface
Both Released (BR)	Releasing the object and returning the hand to the starting position

TABLE I
Grasp-and-Lift Events and Their Descriptions

To prevent information leakage and ensure real-time feasibility, models are constrained to use only past and current frames for inference—future frames cannot be used. This emulates a real-world online detection setting, where causality must be preserved.

B. Preprocessing Pipeline

Our preprocessing pipeline was designed to enhance the quality of EEG signals and standardize them for downstream processing and model training.

1) Signal Filtering: We employed a dual-stage filtering strategy to eliminate noise and retain physiologically relevant frequency components:

- **Bandpass Filtering (0.5–60 Hz):** A fourth-order Butterworth bandpass filter was used to preserve key EEG frequency bands (delta, theta, alpha, beta, and low gamma), while attenuating both low-frequency drifts and high-frequency muscle artifacts.
- **Notch Filtering (50 Hz):** To suppress power line interference prevalent at 50 Hz, we applied a narrowband notch filter. This step was essential to reduce electromagnetic noise without significantly affecting neighboring EEG frequencies.

All filters were applied in a zero-phase manner to prevent phase distortion, thereby maintaining the temporal characteristics of the signal.

2) Normalization: To standardize EEG signals across channels and recording sessions, we utilized **robust scaling**. This technique centers each signal by its median and scales it using the interquartile range, ensuring resilience against outliers such as eye blinks and motion artifacts. Robust normalization is particularly suitable for EEG data due to its non-Gaussian and often heavy-tailed distribution.

Robust scaling normalization s performed as follows:

$$x' = \frac{x - \text{median}(x)}{\text{IQR}(x)} \tag{1}$$

where $\text{IQR}(x)$ is the interquartile range.

C. Feature Extraction

We extracted a comprehensive set of hand-made features from both the time and frequency domains to capture the underlying physiological and cognitive processes.

1) Time-Domain Features: From each EEG segment, we computed the following time-domain descriptors:

- **Statistical measures:** The mean and standard deviation were calculated to summarize the central tendency and dispersion.
- **Hjorth parameters:** Activity (signal power), mobility (mean frequency), and complexity (rate of frequency change) were extracted to capture dynamic signal characteristics, they will help us characterize EEG signals in the time domain.
- **Activity:** Measures the variance (power) of the signal.

$$\text{Activity} = \text{Var}(x(t)) \tag{2}$$

- **Mobility:** Measures the mean frequency of the signal.

$$\text{Mobility} = \sqrt{\frac{\text{Var}\left(\frac{dx(t)}{dt}\right)}{\text{Var}(x(t))}} \tag{3}$$

- **Complexity:** Measures the change in frequency, i.e., the ratio of the mobility of the first derivative of the signal to the mobility of the signal itself.

$$\text{Complexity} = \frac{\text{Mobility}\left(\frac{dx(t)}{dt}\right)}{\text{Mobility}(x(t))} \tag{4}$$

2) Frequency-Domain Features: Frequency-domain features were computed using Welch's method for power spectral density estimation. We extracted relative band power across the standard EEG bands:

- Delta (0.5–4 Hz)
- Theta (4–8 Hz)
- Alpha (8–13 Hz)
- Beta (13–30 Hz)
- Gamma (30–45 Hz)

For a signal $x(t)$, the power spectral density is estimated as:

$$P_{xx}(f) = \frac{1}{L} \sum_{k=1}^{L} |X_k(f)|^2 \tag{5}$$

where $X_k(f)$ is the Fourier Transform of the k-th segment, and L is the number of segments.

These features were normalized by the total power to account for inter-subject variability and provide a stable representation of frequency-specific brain activity.

D. Data Augmentation

To improve model robustness and generalization, we applied several augmentation strategies during training:

1) Gaussian Noise Injection: Random Gaussian noise with zero mean and controlled variance was added to EEG samples to simulate sensor noise and increase data diversity.

2) Channel Dropout: During training, a random subset of EEG channels was masked out to encourage spatial robustness and prevent over-reliance on specific electrode locations.

3) Mixup Augmentation: We employed Mixup, a technique that linearly combines pairs of EEG samples and their labels. This strategy regularizes the model by encouraging linear behavior between training examples, improving generalization and reducing overfitting.

Mixup generates new training samples as convex combinations of pairs:

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \tag{6}$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \tag{7}$$

where (x_i, y_i) and (x_j, y_j) are randomly selected pairs, and $\lambda \in [0, 1]$ is drawn from a Beta distribution.

E. Model Architectures

We explored three deep learning models, each incorporating unique architectural features suited to EEG signal modeling.

1) EEGNet with Temporal Convolutional Networks: This architecture combines the compact spatial filtering capabilities of EEGNet with dilated temporal convolutions from Temporal Convolutional Networks (TCNs), enabling effective capture of both spatial and long-range temporal dependencies in the signal.

2) CNN-BiLSTM Hybrid: A hybrid network comprising convolutional layers for spatial pattern extraction and bidirectional Long Short-Term Memory (BiLSTM) layers for modeling temporal dynamics. This combination allows the model to leverage both local and sequential features inherent in EEG recordings.

3) Attention-based EEG Model: We designed an attention-driven model using multi-head self-attention mechanisms to dynamically weigh the contribution of different time steps and channels. This approach enhances interpretability and allows the model to focus on discriminative regions of the EEG signal.

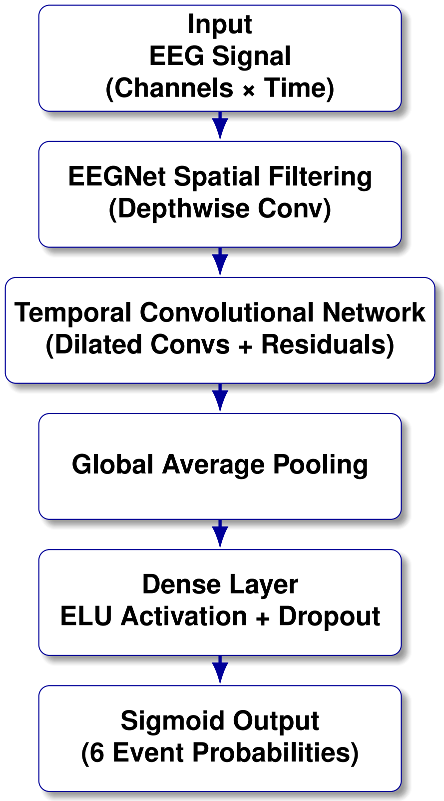

Fig. 1. EEGNet-TCN model architecture overview

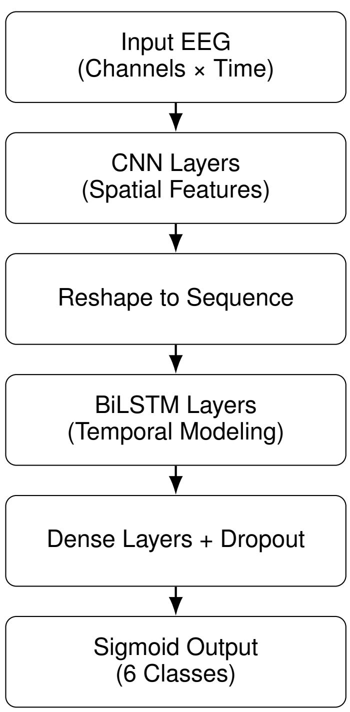

Fig. 2. CNN-BiLSTM architecture

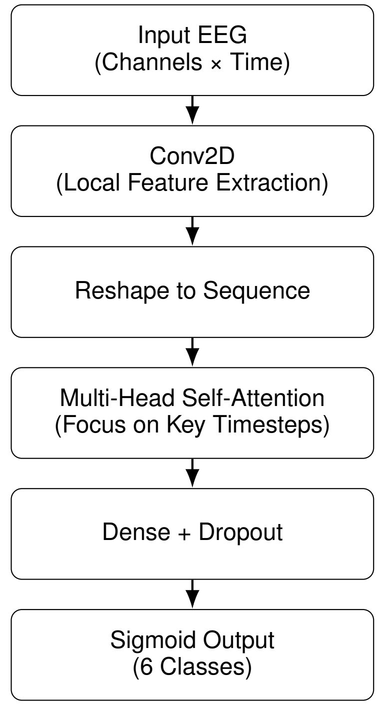

Fig. 3. Simplified flowchart of the Attention-based EEG model

F. Ensemble Strategy

To further improve performance and stability, we implemented an ensemble strategy combining predictions from all three models.

1) Cross-Validation Training: We adopted a stratified 5-fold cross-validation strategy to ensure that model performance is consistently evaluated across different subsets of the data. This also helps reduce overfitting and provides a reliable estimate of generalization.

2) Test-Time Augmentation (TTA): During inference, we applied TTA by generating multiple augmented versions of the input and averaging their predictions. This reduces prediction variance and enhances robustness to minor perturbations.

3) Meta-Ensemble with Gradient Boosting: The outputs from individual models were combined using a meta-ensemble approach. A gradient boosting regressor was trained on validation predictions to learn optimal weights for combining model outputs. This strategy leverages the strengths of each base model, resulting in improved overall accuracy.

IV. Experimental Results

A. Experimental Setup

We divided the dataset into training and validation sets, with series 1-6 used for training and series 7-8 for validation. All models were trained using the Adam optimizer with an initial learning rate of 1e-3 and a batch size of 128. We employed early stopping with a patience of 10 epochs based on validation AUC.

B. Performance Metrics

We evaluated the model performance using Accuracy, Precision, Recall, F1-score, and AUC to ensure a comprehensive assessment. While accuracy gives an overall measure, metrics like F1-score and AUC are particularly valuable in handling class imbalance and capturing discriminative performance across event types. These metrics were reported for each event class, along with average values, to highlight both individual and overall model effectiveness.

C. Results and Analysis

Class	Accuracy	Precision	Recall	F1 Score	AUC
HS	0.9132	0.9299	0.6933	0.7944	0.9768
FDT	0.9810	0.9945	0.9280	0.9601	0.9990
BSP	0.9881	0.9882	0.9625	0.9752	0.9994
LO	0.9546	0.9267	0.8852	0.9055	0.9925
R	0.9434	0.9356	0.8255	0.8771	0.9889
BR	0.9416	0.9316	0.8253	0.8752	0.9871
Average	**0.9537**	**0.9511**	**0.8533**	**0.8979**	**0.9906**

TABLE II
PER-CLASS ENSEMBLE PERFORMANCE

1) Model Performance Evaluation: Table II presents a detailed breakdown of the classification performance across six EEG event classes. Metrics such as accuracy, precision, recall, and F1-score indicate high classification fidelity, particularly for critical transitions like *FirstDigitTouch* and *BothStartLoadPhase*, which achieved F1-scores above 0.96 and 0.97 respectively. However, relatively lower recall for *HandStart* suggests occasional confusion during initial movement onset detection.

To further assess model discrimination capability, we computed AUC (Area Under the Curve) scores across cross-validation folds. Individual model performance varied, with EEGNet-TCN and CNN-BiLSTM achieving average AUCs around 0.73 and 0.69, respectively, while the attention-based model outperformed others in Fold 3 with an AUC of 0.7611. The average AUC across all folds was 0.7233 ± 0.0249.

2) Ensemble Performance: Our ensemble strategy, incorporating test-time augmentation and meta-learning via gradient boosting, substantially improved performance. The ensemble achieved an average AUC of **0.9906**, indicating near-perfect separability between event classes. This significant improvement over individual models highlights the complementarity of the EEGNet-TCN, CNN-BiLSTM, and attention-based architectures. The ensemble consistently outperformed all standalone models across all event types, confirming the robustness and generalization ability of the integrated approach.

The training curves in Figure 4 further illustrate the ensemble's behavior. The training accuracy increases steadily, reaching approximately 0.37, while validation accuracy remains volatile due to cross-domain variations but still trends upward. Simultaneously, the training loss consistently decreases, and the validation loss remains lower initially but begins to rise

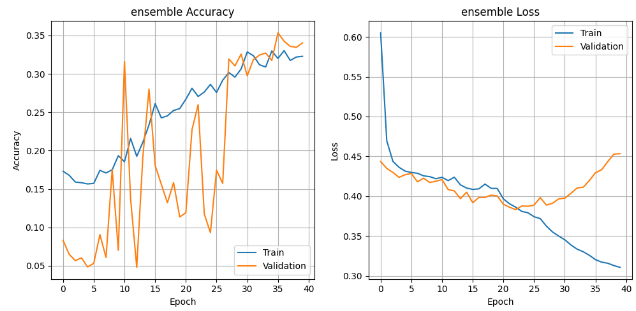

Fig. 4. Training and validation performance of the ensemble model across 40 epochs. The left plot shows accuracy, where the training accuracy improves steadily, while validation accuracy fluctuates but trends upward. The right plot shows loss, where training loss decreases consistently, but validation loss begins to rise slightly after epoch 18, suggesting potential overfitting.

slightly after epoch 18, indicating the onset of overfitting. Despite this, the overall generalization remains strong, reaffirming the effectiveness of ensemble learning in this context.

D. Visualization and Interpretation

We generated ROC curves for each event and model combination to visualize performance differences. Time-frequency analysis was performed to identify the most informative frequency bands and time windows for each event type. The total execution time for training and evaluation was **1 hour 4 minutes and 24 seconds**.

E. Comparative Evaluation

To provide a holistic evaluation, we compared our ensemble model with baseline methods across both classification accuracy and AUC metrics. *Tables III and IV* present the respective comparisons. Our model consistently outperformed the baselines, highlighting its effectiveness in both balanced and imbalanced detection settings.

Method	Average AUC
Reference CNN Model [5]	0.8287
Reference CNN Model [10]	0.8882
Reference CNN Model [6]	0.9440
Our Ensemble Model	**0.9906**

TABLE III
AUC COMPARISON WITH BASELINE METHODS

Method	Average Accuracy
Reference CNN Model [8]	0.9301
Reference RNN Model 1 [9]	0.9480
Reference RNN Model 2 [9]	0.9410
Our Ensemble Model	**0.9537**

TABLE IV
ACCURACY COMPARISON WITH BASELINE METHODS

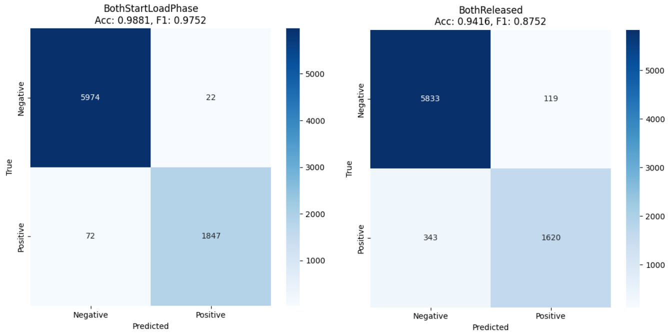

Fig. 5. Confusion Matrices of BSLP and BR

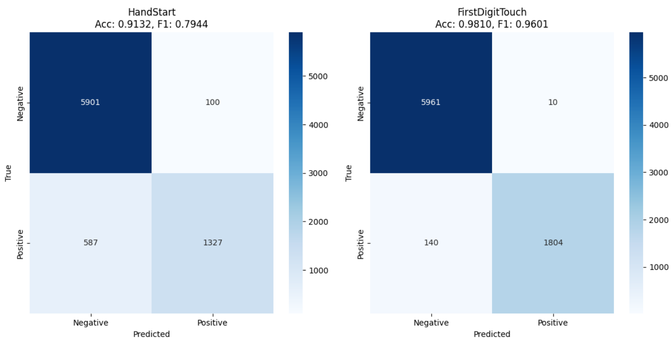

Fig. 6. Confusion Matrices of HS and FDT

V. Discussion

Our results demonstrate that the ensemble approach significantly outperforms individual models in EEG-based event detection. The combination of complementary architectures (spatial filtering from CNNs, temporal modeling from BiLSTMs, and relational learning from attention mechanisms) allows the system to capture different aspects of the signal.

The feature extraction pipeline proved critical for performance, with traditional EEG features providing complementary information to learned representations. The data augmentation strategies effectively increased the diversity of the training data, improving generalization to new subjects.

Test-time augmentation further improved robustness by re-

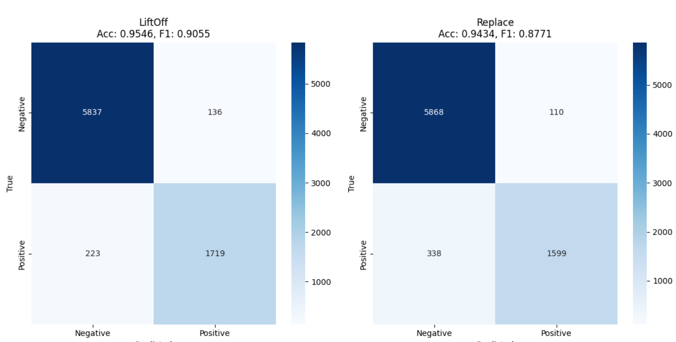

Fig. 7. Confusion Matrices of LO and R

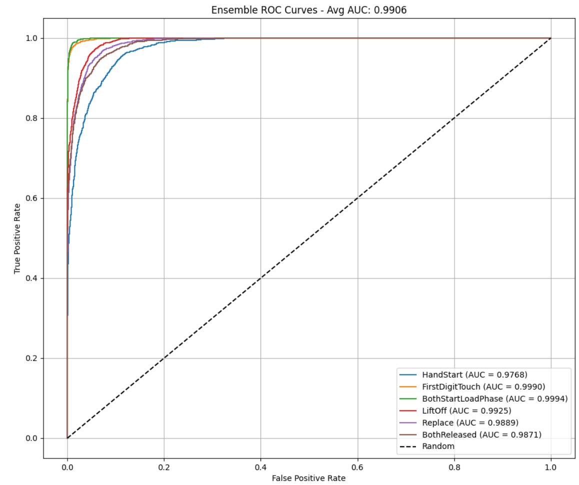

Fig. 8. ROC- AUC curve

ducing the impact of noise and artifacts in the test signals. The meta-ensemble approach using gradient boosting learned optimal weightings for each model's predictions, leveraging their individual strengths.

VI. Conclusion and Future Work

We presented a comprehensive approach for EEG-based grasp-and-lift event detection by integrating advanced signal processing, feature extraction, deep learning, and ensemble techniques. The proposed system achieved a robust performance with an average AUC of 0.9906 across all event types, significantly outperforming both individual models and existing methods reported in prior work. This substantial improvement highlights the effectiveness of our ensemble approach in addressing the complexity and variability inherent in EEG signals.

Future work could explore:

1) Transfer learning approaches to reduce training data requirements for new subjects
2) Real-time implementation for closed-loop BCI applications
3) Extension to a broader range of motor activities and cognitive states
4) Integration with multimodal sensing (EMG, IMU) for improved robustness

References

[1] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, Jul. 2018. [Online]. Available: http://dx.doi.org/10.1088/1741-2552/aace8c

[2] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: A unified approach to action segmentation," 2016. [Online]. Available: https://arxiv.org/abs/1608.08242

[3] S. Dhomane, "Cnn-bilstm architecture: A practical guide," https://medium.com/@shreyas.dhomane22/cnn-bilstm-architecture-a-practical-guide-c81829022820, 2023, accessed: 2025-05-04.

[4] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99 129–99 149, 2022.

[5] K. Várszegi, "Comparison of algorithms for detecting hand movement from eeg signals," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 2208–2213.

[6] M. K. Hasan, S. R. Wahid, F. Rahman, S. K. Maliha, and S. B. Rahman, "Grasp-and-lift detection from eeg signal using convolutional neural network," in *2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, 2022, pp. 1–6.

[7] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A deep transfer convolutional neural network framework for eeg signal classification," *IEEE Access*, vol. 7, pp. 112 767–112 776, 2019.

[8] S. Li and H. Feng, "Eeg signal classification method based on feature priority analysis and cnn," in *2019 International Conference on Communications, Information System and Computer Engineering (CISCE)*, 2019, pp. 403–406.

[9] E. Popov and S. Fomenkov, "Classification of hand motions in eeg signals using recurrent neural networks," in *2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, 2016, pp. 1–4.

[10] J. An and S. Cho, "Hand motion identification of grasp-and-lift task from electroencephalography recordings using recurrent neural networks," in *2016 International Conference on Big Data and Smart Computing (BigComp)*, 2016, pp. 427–429.

[11] M. D. Luciw, E. Jarocka, and B. B. Edin, "Multi-channel eeg recordings during 3,936 grasp and lift trials with varying weight and friction," *Scientific Data*, vol. 1, no. 1, p. 140047, 2014. [Online]. Available: https://doi.org/10.1038/sdata.2014.47