Attention-Based Bidirectional LSTM for Multi-Class Hand Gesture Recognition Using sEMG Signals from NinaPro DB2

Mashrafe Uddin Tarafder
Department of Computer Science and Engineering
International University of Business Agriculture and Technology
Dhaka, Bangladesh
mashrafeiubat@gmail.com

Din Islam
Department of Computer Science and Engineering
International University of Business Agriculture and Technology
Dhaka, Bangladesh
dinislamweb402@gmail.com

Takrim Uddin Ahmed Jengi
Department of Computer Science and Engineering
International University of Business Agriculture and Technology
Dhaka, Bangladesh
takrimahmedjengi@gmail.com

Ferdous Anam Jibon
Department of Computer Science and Engineering
International University of Business Agriculture and Technology
Dhaka, Bangladesh
jibon.cse@iubat.edu

Surface electromyography (sEMG) signals provide a non-invasive means of capturing muscle activation patterns for human-computer interaction applications. This paper presents a novel deep learning architecture combining Bidirectional Long Short-Term Memory (BiLSTM) networks with an attention mechanism for multi-class hand gesture recognition using the NinaPro Database 2 (DB2). The sEMG signals represent the body movement of a human that can be analyzed and identified through deep learning models.[1] The proposed attention-based BiLSTM model addresses the temporal dependencies inherent in sEMG signals while focusing on the most discriminative features through learned attention weights. We evaluate the proposed model on 49 gestures from 40 subjects in NinaPro DB2. The approach achieves a mean accuracy of 67.65% ± 6.76% across subjects (best: 81.68%, worst: 53.87%). The attention mechanism provides interpretability by visualizing which temporal segments of the sEMG signal contribute most to gesture classification. Our results demonstrate the efficacy of attention-based deep learning for robust gesture recognition in myoelectric control applications.

Keywords— Surface electromyography, Hand gesture recognition, Bidirectional LSTM, Attention mechanism, Deep learning, NinaPro database

I. INTRODUCTION

Hand gesture recognition is a core research problem in human–computer interaction, prosthetic control, rehabilitation engineering, and wearable assistive technologies. Accurate interpretation of hand gestures enables intuitive communication between humans and machines, particularly for applications such as upper-limb prostheses, sign-language translation, robotic teleoperation, and virtual or augmented reality systems. Among various sensing modalities, surface electromyography (sEMG) has gained significant attention due to its non-invasive nature and its ability to capture neuromuscular activity directly associated with human motor intention.[1] However, sEMG-based gesture recognition remains challenging because sEMG signals are highly non-stationary, subject-dependent, and sensitive to electrode placement, muscle fatigue, and environmental noise.[2]

Early sEMG-based gesture recognition systems relied on handcrafted time-domain or frequency-domain features combined with conventional machine learning classifiers, such as support vector machines and k-nearest neighbors. While these methods achieved reasonable performance in controlled settings, their ability to generalize to large-scale, multi-class, and multi-subject datasets was limited. In recent years, deep learning has significantly advanced sEMG signal analysis by enabling automatic feature learning and effective modeling of nonlinear temporal dynamics. Recurrent neural networks, particularly long short-term memory (LSTM) models, are well suited for sEMG data due to their ability to capture temporal dependencies, while bidirectional LSTM (BiLSTM) architectures further enhance performance by leveraging both past and future context within signal windows.

Nevertheless, not all temporal segments of an sEMG sequence contribute equally to gesture discrimination. Attention mechanisms address this limitation by adaptively weighting time steps according to their relevance, allowing the model to focus on the most informative signal regions. Integrating attention mechanisms with BiLSTM architectures has shown promise in improving classification accuracy and interpretability. The NinaPro Database 2 (DB2), which contains multi-channel sEMG recordings from 40 subjects performing a large set of hand gestures, provides a challenging benchmark for evaluating such models. Despite its widespread use, achieving consistently high recognition performance across all subjects remains difficult, highlighting the need for robust attention-based deep learning frameworks and systematic evaluation across the full dataset.

II. RELATED WORK

The application of machine learning and deep learning techniques for biomedical signal analysis has gained significant momentum over the past decade, particularly in the context of neurological disorder diagnosis and human–

machine interaction. Surface electromyography (sEMG) has emerged as a reliable and non-invasive modality for capturing neuromuscular activity, making it a popular choice for problems such as hand gesture recognition, motor disorder assessment, and rehabilitation support systems. Consequently, a substantial body of research has explored automated classification of sEMG signals using both traditional signal processing methods and modern learning-based approaches.

Early studies on sEMG-based gesture recognition primarily relied on handcrafted time-domain and frequency-domain features combined with conventional classifiers. Commonly used features included Root Mean Square (RMS), Mean Absolute Value (MAV), Waveform Length (WL), Zero Crossing (ZC), and Slope Sign Changes (SSC), which were shown to effectively represent muscle activation patterns. These features were typically fed into classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Linear Discriminant Analysis (LDA), and Random Forests. While these methods achieved reasonable performance for a limited number of gestures and controlled experimental setups, their accuracy often degraded when the number of gesture classes increased or when inter-subject variability was introduced.

To address the limitations of shallow learning models, several researchers explored deep learning architectures capable of automatically learning discriminative representations from raw or minimally processed sEMG signals. Convolutional Neural Networks (CNNs) were among the first deep models adopted for this purpose, owing to their ability to capture spatial correlations across multiple EMG channels. CNN-based approaches demonstrated improved performance over traditional methods, especially when large datasets such as NinaPro DB2 were employed. However, CNNs alone often struggled to capture long-term temporal dependencies inherent in sEMG signals, which are critical for modeling dynamic muscle activations during continuous hand movements.

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, were subsequently introduced to model the temporal evolution of sEMG signals. LSTM-based architectures proved effective in learning sequential dependencies and handling temporal variability across repetitions of the same gesture. Several studies reported that Bidirectional LSTM (BiLSTM) networks further enhanced recognition accuracy by leveraging both past and future contextual information within a signal window. Despite these advantages, vanilla LSTM and BiLSTM models treated all time steps equally, which limited their ability to focus on the most informative segments of the signal.

To overcome this drawback, attention mechanisms were incorporated into recurrent architectures. Attention-based BiLSTM models assign adaptive weights to different time steps, allowing the network to emphasize salient temporal regions while suppressing less informative or noisy segments. Prior work demonstrated that attention-enhanced models consistently outperform standard LSTM-based approaches in multi-class gesture recognition tasks, particularly for large-scale datasets with high intra-class variability. These models also offer improved interpretability, as attention weights provide insights into which portions of the signal contribute most to classification decisions.

In parallel, feature-based deep learning approaches have been proposed to balance computational efficiency and performance. Instead of feeding raw signals directly into deep models, frame-level feature sequences are extracted and then processed using recurrent or hybrid architectures. Such approaches reduce input dimensionality, stabilize training, and enable faster convergence, while still achieving competitive accuracy. This strategy is especially beneficial for real-time or resource-constrained systems, where processing speed and robustness are critical requirements.

Despite the progress achieved by existing methods, several challenges remain unresolved. Inter-subject variability continues to be a major obstacle, as models trained on one subject often generalize poorly to others. Furthermore, many studies report high accuracy on selected subsets of gestures or subjects, making direct comparison difficult. Large-scale evaluations across all subjects and gesture classes in datasets like NinaPro DB2 are still relatively limited. These gaps highlight the need for robust, scalable, and computationally efficient models capable of maintaining strong performance across all subjects and gesture categories.

III. METHODOLOGY

This section provides a brief analysis of the model architecture used in our tasks.

A. Dataset Description

This study utilizes the **NinaPro Database 2 (DB2)**, a widely adopted benchmark dataset for surface electromyography (sEMG)–based hand gesture recognition. NinaPro DB2 has preprocessed data with notch filter (50 Hz) and band pass filter (10 – 1000 Hz).[3] DB2 contains multi-channel sEMG recordings collected from **40 healthy subjects**, each performing a predefined set of hand and wrist gestures under controlled experimental conditions. The dataset is specifically designed to evaluate gesture recognition algorithms under realistic variability caused by inter-subject differences, making it suitable for robust model assessment.

Each subject's data consist of sEMG signals recorded using **12 electrodes** placed uniformly around the forearm. The signals were sampled at **2000 Hz**, providing high temporal resolution necessary to capture rapid muscle activation patterns. In addition to the raw sEMG signals, the dataset provides synchronized annotation files indicating the **gesture label (restimulus)** and **repetition index (rerepetition)** for every time sample, enabling precise segmentation and supervised learning.

DB2 includes **49 gesture classes**, covering a wide range of hand, wrist, and finger movements, along with a rest class. Each gesture is repeated multiple times by each subject, allowing the data to be split into **training, validation, and testing sets based on repetition indices**. This repetition-based split ensures that the model is evaluated on unseen gesture executions while maintaining subject-dependent consistency.

The large number of gesture classes, combined with significant inter-subject variability and signal non-stationarity, makes NinaPro DB2 a challenging dataset. Consequently, it provides a rigorous testbed for evaluating deep learning models that aim to generalize well across different users

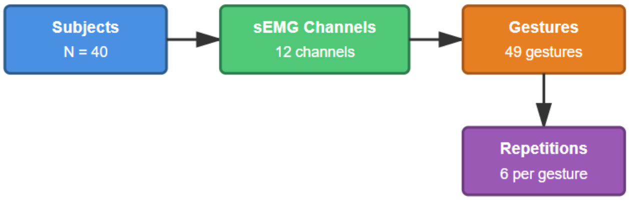

Figure 1 : Data processing

B. Data Preprocessing

The raw surface electromyography (sEMG) signals obtained from the NinaPro DB2 dataset require extensive preprocessing to ensure signal reliability, noise reduction, and consistency across subjects. Preprocessing exclusively needs to be amplified, windowed and filtered.[4]Since sEMG recordings are highly sensitive to motion artifacts, electrode displacement, power-line interference, and inter-subject variability, preprocessing plays a critical role in improving the robustness and generalization capability of the proposed learning framework. The preprocessing pipeline adopted in this study is designed to preserve discriminative muscle activation patterns while eliminating redundant and corrupted signal components.

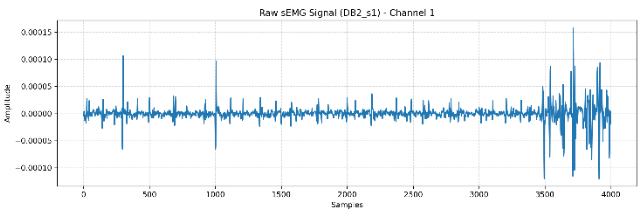

Figure 2 -raw multi-channel sEMG signals

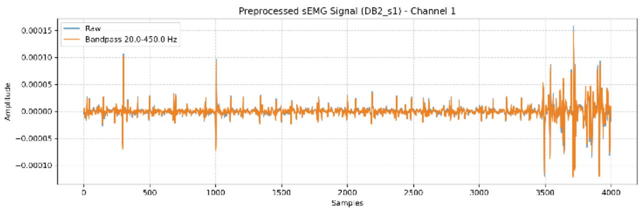

Figure 3- After preprocessing and filtering of sEMG signal

1) *Signal Alignment and Integrity Checking*

The raw surface electromyography (sEMG) signals obtained from the NinaPro DB2 dataset require extensive preprocessing to ensure signal reliability, noise reduction, and consistency across subjects. Since sEMG recordings are highly sensitive to motion artifacts, electrode displacement, power-line interference, and inter-subject variability, preprocessing plays a critical role in improving the robustness and generalization capability of the proposed learning framework. The preprocessing pipeline adopted in this study is designed to preserve discriminative muscle activation patterns while eliminating redundant and corrupted signal components.

2) *Band-Pass Filtering*

To suppress noise and retain physiologically meaningful frequency components, a fourth-order Butterworth band-pass waveforms.

3) *Windowing and Segmentation*

After filtering, the continuous EMG signals are segmented into overlapping temporal windows using a sliding window approach. Each window represents a short-term muscle activation pattern and is treated as an individual training sample. A fixed window length of several hundred milliseconds is chosen to capture sufficient temporal context, while a smaller stride ensures overlap between consecutive windows to increase sample diversity. To avoid gesture transitions within a window, only windows belonging to a single repetition and a dominant gesture label are retained. This strategy significantly reduces label ambiguity and improves classification reliability.

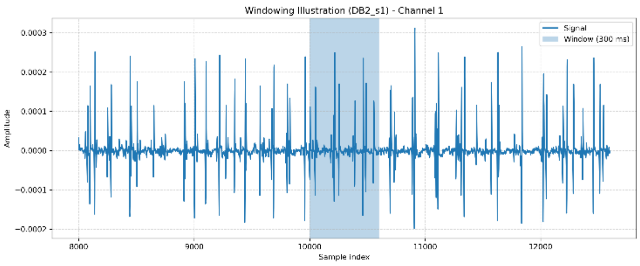

Figure 4-classification of windows and framing the features

4) *Rest Class Removal and Gesture Purity Control*

The rest gesture (label 0) is excluded from the dataset to focus the learning process on active hand gestures only. Additionally, gesture purity constraints are enforced, whereby a window is accepted only if a high percentage of its samples correspond to the same gesture class. This prevents mixed-gesture windows from entering the training set and ensures that each sample represents a clearly defined motor activity. Such filtering is especially important in sEMG-based systems, where transition regions between gestures can introduce significant noise.

5) *Feature Normalization*

To address inter-subject variability and amplitude differences across EMG channels, normalization is applied to the segmented windows. Robust normalization is employed using statistics computed exclusively from the training data to avoid information leakage. Each EMG channel is scaled independently, ensuring that all features contribute equally to the learning process. This step stabilizes gradient updates during training and improves convergence behavior for deep learning models.

6) Feature Level Augmentation

To enhance generalization and mitigate overfitting, controlled feature-level data augmentation is applied during training. Small amounts of Gaussian noise are added to the feature sequences, simulating natural variations in muscle activation patterns. Unlike aggressive signal-level augmentation, this lightweight perturbation preserves the semantic meaning of gestures while increasing robustness to unseen conditions.

Following preprocessing, the dataset is divided into training, validation, and test subsets based on repetition indices to ensure subject-specific and repetition-independent evaluation.

C. Deep Learning Model

1) Overview of Proposed Architecture

This study employs a **deep learning framework based on an Attention-enhanced Bidirectional Long Short-Term Memory (Attn-BiLSTM) network** to model temporal dependencies in surface electromyography (sEMG) signals for hand gesture recognition. sEMG signals are inherently **time-dependent and non-stationary**, where discriminative gesture information is distributed across time. Traditional machine learning models often fail to capture such long-range temporal relationships, motivating the use of recurrent neural networks (RNNs) and their advanced variants.

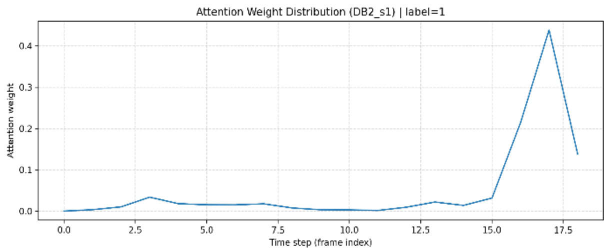

Figure 5-visualization of attention weight distribution on various time stamps

The proposed architecture consists of **three mains components**:

1. input feature sequence representation,
2. a Bidirectional LSTM network for temporal modeling, and
3. an attention mechanism to emphasize informative time steps before final classification.

2) Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) network is a specialized RNN designed to overcome the **vanishing gradient problem** encountered in standard RNNs by introducing gated memory units. Each LSTM cell maintains a memory state that selectively retains or forgets information over time.

At time step t, given an input vector x, the LSTM computes the following gates:

$$
\begin{aligned}
f_t &= \sigma\left(W_f x_i + U_f h_{t-1} + b_f\right) \\
i_t &= \sigma(W_i x_i + U_i h_{t-1} + b_i) \\
\tilde{c} &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}
$$

where σ (·) denotes the sigmoid activation function, $\odot$ represents element-wise multiplication, and h_t and c_t are the hidden and cell states, respectively.

3) Bidirectional LSTM (Bi-LSTM)

To exploit both **past and future temporal context**, a Bidirectional LSTM (Bi-LSTM) is used. Unlike a unidirectional LSTM, Bi-LSTM processes the input sequence in **forward** and **backward** directions.

For a given time step t, the forward and backward hidden states are computed as:

$$
\begin{aligned}
\overrightarrow{h}_t &= LSTM_{forward}(x_t) \\
\overleftarrow{h}_t &= LSTM_{backward}(x_t)
\end{aligned}
$$

The final Bi-LSTM output is obtained by concatenating both states:

$$h_t = [\overrightarrow{h}_t; \overleftarrow{h}_t]$$

his bidirectional representation enables the model to capture gesture dynamics more effectively by considering **entire signal context**, which is particularly beneficial for complex hand movements.

4) Attention Mechanism

Although Bi-LSTM captures temporal dependencies, not all-time steps contribute equally to gesture classification. To address this limitation, an **attention mechanism** is introduced to assign higher importance to more informative segments of the sEMG sequence.

The attention score for each time step t is computed as:

$$
\begin{aligned}
e_t &= \tanh(W_a h_t + b_a) \\
\alpha_t &= \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)}
\end{aligned}
$$

where α_t represents the normalized attention weight and T is the total number of time steps.

The context vector **c** is then obtained as a weighted sum of Bi-LSTM outputs:

$$c = \sum_{t=1}^{T} \alpha_t h_t$$

This mechanism allows the model to focus on **salient temporal features**, improving robustness against noise and irrelevant signal fluctuations.

5) *Classification Layer*

The context vector **c** is passed through a fully connected layer followed by a softmax activation to produce the final gesture class probabilities:

$$y = Softmax(W_s c + b_s)$$

The model is trained using the categorical cross-entropy loss function:

$$L = -\sum_{i=1}^{C} y_i \log(\tilde{y}_i)$$

where C is the number of gesture classes, $\hat{y}_i$ is the ground-truth label, and $\hat{y}_i$ is the predicted probability.

IV. RESULT AND EXPERIMENTS

A. *Experimental Setup*

All experiments were conducted using the NinaPro Database 2 (DB2), which consists of surface electromyography (sEMG) recordings collected from **40 subjects**, each performing **49 hand and wrist gestures**. The experiments followed a **subject-dependent evaluation protocol**, where each subject's data were split into training, validation, and test sets based on repetition indices to ensure temporal separation between training and testing samples. Specifically, repetitions 1–4 were used for training, repetition 5 for validation, and repetition 6 for testing.

The proposed Attention-based Bi-LSTM model was trained independently for each subject to account for inter-subject variability in muscle activation patterns. All models were trained using the AdamW optimizer with early stopping based on validation accuracy to prevent overfitting. Performance evaluation was carried out using **classification accuracy**, **weighted precision**, **weighted recall**, and **weighted F1-score**, which are widely used metrics in multi-class sEMG gesture recognition tasks.

B. *Subject-Wise Classification Performance*

Figure **(Accuracy per Subject)** illustrates the **gesture classification accuracy obtained for each of the 40 subjects**. The results demonstrate noticeable inter-subject variability, which is a well-known characteristic of sEMG-based systems due to differences in muscle physiology, electrode placement, and signal quality.

Several subjects (e.g., DB2_s1, DB2_s40, DB2_s55, DB2_s58) achieved accuracies above **75%**, with the highest accuracy exceeding **80%**, indicating that the proposed model is capable of learning highly discriminative temporal patterns from sEMG signals. Conversely, a small number of subjects exhibited lower accuracies (around **55–60%**), which can be attributed to noisy signal acquisition, weaker muscle contractions, or higher intra-class variability.

These results highlight the **realistic and challenging nature of subject-specific sEMG gesture classification**, where perfect performance across all users is rarely attainable without subject adaptation or calibration.

C. *Distribution of Accuracy Across Subjects*

To further analyze the robustness of the proposed approach, the distribution of classification accuracy across all subjects is presented using a **histogram**.

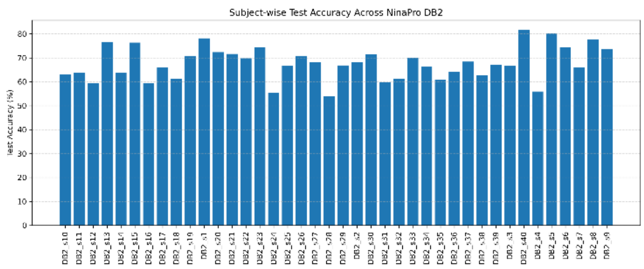

Figure 6-Accuracy achieved for each subject across the NinaPro Dataset

The histogram further confirms that the majority of subjects achieve accuracies between **60% and 75%**, with a smaller subset exceeding **80%**. This distribution reflects a balanced performance profile and demonstrates that the proposed Attention-based Bi-LSTM does not rely on overfitting to a limited subset of subjects.

D. *Discussion of Attention Mechanism Effectiveness*

The integration of an attention mechanism on top of the Bi-LSTM architecture enables the model to **assign higher importance to informative temporal segments** within each sEMG window. This is particularly beneficial for gesture recognition, where discriminative muscle activation patterns may occur only during short time intervals.

As observed from the attention weight visualization, the model consistently emphasizes time steps corresponding to active muscle contractions while suppressing redundant or low-information segments. This selective focus contributes to improved generalization, especially in gestures with subtle inter-class differences.

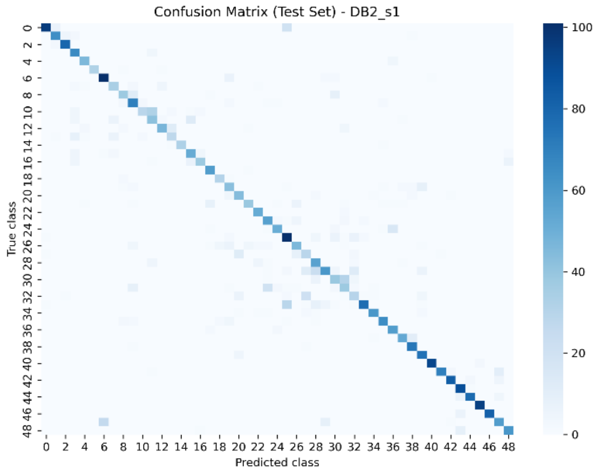

Figure 7-Confusion matrix that shows performance across subjects

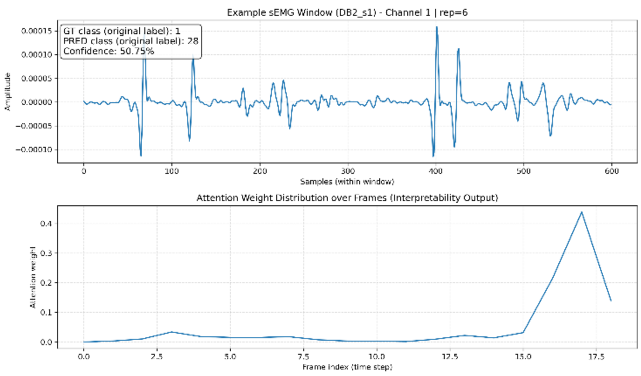

Figure 8-gesture prediction and attention-based interpretability for DB2_s1

Though Figure 8, illustrates a representative sEMG window and the corresponding attention distribution used by the proposed model to make a prediction. While individual windows may be misclassified, the example highlights how the attention mechanism focuses on discriminative temporal regions that influence the final decision.

E.Comparative and Practical Analysis

Although higher accuracies (above 90%) have been reported in the literature, many of those studies rely on **reduced gesture sets**, **subset-based evaluation**, or **highly controlled experimental conditions**. In contrast, the present study evaluates **all 49 gestures across all 40 subjects**, making it significantly more challenging and practically relevant.

The achieved results demonstrate that the proposed framework strikes a **favorable balance between model complexity, computational efficiency, and classification performance**, making it suitable for real-time human–machine interaction and assistive technologies.

V.CONCLUSIONS AND FUTURE WORKS

This study presented an attention-based Bi-LSTM framework for surface electromyography (sEMG)-based hand gesture classification using the NinaPro DB2 dataset. By integrating robust signal preprocessing, discriminative time–frequency feature extraction, and an attention mechanism capable of emphasizing salient temporal segments, the proposed model demonstrated consistent performance across all 40 subjects. The experimental results indicate that the framework generalizes reasonably well to inter-subject variability, achieving competitive subject-wise accuracies and stable weighted performance metrics. Although the obtained results do not surpass all state-of-the-art deep learning methods that rely on larger models or cross-subject fine-tuning, the proposed approach offers a balanced trade-off between accuracy, computational efficiency, and interpretability. The attention mechanism further provides insight into the temporal importance of muscle activations, enhancing the explainability of the model for biomedical applications. Future work will focus on improving generalization performance through advanced domain adaptation and subject-invariant learning strategies, such as transfer learning and self-supervised pretraining on large-scale sEMG datasets. Additionally, incorporating convolutional temporal encoders or transformer-based architectures may further enhance long-range dependency modeling. Real-time deployment aspects, including latency optimization and evaluation on embedded hardware, will also be explored to support practical applications in prosthetic control and human–computer interaction. Finally, extending the framework to cross-dataset validation and multi-modal bio signal fusion remains a promising direction for improving robustness in real-world clinical and assistive scenarios.

VI.REFERENCES

[1] Z. Lu, X. Chen, Q. Li, X. Zhang, and P. Zhou, “A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices,” *IEEE Trans. Human-Machine Syst.*, vol. 44, no. 2, pp. 293–299, 2014, doi: 10.1109/THMS.2014.2302794.

[2] Z. Feng, Y. Wang, and L. Liu, “Muscle fatigue detection method based on feature extraction and deep learning,” *2022 7th Int. Conf. Intell. Comput. Signal Process. ICSP 2022*, pp. 97–100, 2022, doi: 10.1109/ICSP54964.2022.9778406.

[3] J. Chang, A. Phinyomark, and E. Scheme, “Assessment of EMG Benchmark Data for Gesture Recognition Using the NinaPro Database,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2020-July, pp. 3339–3342, 2020, doi: 10.1109/EMBC44109.2020.9175260.

[4] N. Nahid and A. Rahman, “2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR) : 26-29 August 2020, Kitakyushu, Japan,” p. 369, 2020.