Advances in Artificial Intelligence and Machine Learning; Research 5 (3) 4094-4114 Received 12-05-2025; Accepted 24-07-2025; Published 31-07-2025

sEMG-based Hand Gesture Combination Detection via Decision Fusion with CNN/GRU Based Models and Random Forest Classifier

H.M.P. Priyanga priyanjithpr@gmail.com
Information and Communication Technology Department
Faculty of Technology, University of Sri Jayewardenepura
Sri Lanka

A.K.S. Srinath sumi1999srinath@gmail.com
Information and Communication Technology Department
Faculty of Technology, University of Sri Jayewardenepura
Sri Lanka

J.B.P. Perera priyanjanjb@gmail.com
Information and Communication Technology Department
Faculty of Technology, University of Sri Jayewardenepura
Sri Lanka

B.N.S. Lankasena nalaka@sjp.ac.lk
Information and Communication Technology Department
Faculty of Technology, University of Sri Jayewardenepura
Sri Lanka

B.M. Seneviratne bathiyaseneviratne@sjp.ac.lk
Information and Communication Technology Department
Faculty of Technology, University of Sri Jayewardenepura
Sri Lanka

M.H.Paul hansamalipaul@sjp.ac.lk
Information and Communication Technology Department, Faculty of Technology
University of Sri Jayewardenepura
Sri Lanka

Corresponding Author: B.N.S. Lankasena

Abstract

This research investigates the development of predictive models for recognizing dual-hand gestures using surface electromyography (sEMG) signals. The main focus was put on dual-hand gestures, an area that remains a gap in research and is essential to advance technologies in human-computer interaction. Most of the previous systems are designed for single-hand gestures, leading to low accuracy for more complex dual-hand gestures. The basic objective of this research is to develop a robust predictive model which enhances the recognition of dual-hand gestures significantly through analysis of the NinaPro DB1 database. We used two approaches to explore into the problem: one using pre-trained models and the other with

Citation: H.M.P. Priyanga et al. sEMG-based Hand Gesture Combination Detection via Decision Fusion with CNN/GRU Based Models and Random Forest Classifier. **Advances in Artificial Intelligence and Machine Learning**. 2025;5(3):230.

our ensemble learning method. The second approach is a novel hybrid model that consists of a Convolutional Neural Network (CNN) and a Gated Recurrent Unit (GRU). In that, the spatial and temporal features from the data communicated by the sEMG signals were captured for the description of the complex dynamics of dual hand gestures. The application of the models provided different levels of success. The pre-trained model of VGG16 resulted in an accuracy of 60%, illustrating the complexity of adapting sEMG signals for image-based neural networks. The hybrid CNN-GRU model yielded better accuracy, with a first set of gestures achieving 83% and a second set achieving 85% over the dataset. These two low-level models were combined using a higher model that utilizes random forest, which can be used through action mapping to support various operations. The higher model achieved an impressive 99% accuracy, indicating the success of combining CNN and GRU for this type of data. The high classification performance of the hybrid model infers success in effectively handling the spatial-temporal complexities of dual-hand gestures.

Keywords: Human-Computer Interaction (HCI), Surface Electromyography (sEMG), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Hand Gestures, Hybrid models.

1. INTRODUCTION

Human-computer interaction (HCI) and biomedical engineering are broad domains where sEMG-based hand gesture recognition is emerging. sEMG-based technology uses information in an individual's muscles' electrical activity to allow a non-invasive approach to acquiring and decoding hand movements. The advantages are very promising for fields in rehabilitation, prostheses control, gaming, and virtual reality. This becomes even more important as it can affect the quality of life for people who have motor impairments and the user experience in interactive systems. sEMG hand gesture recognition is important because it helps design more intuitive and natural interfaces between humans and machines. Unlike classical input devices, such as keyboards and mice, gesture-based systems can easily lend themselves to immersive interaction, which is particularly beneficial for contexts in which physical limitations or the need for hands-free operation are considerations. Moreover, advances in assistive technologies that can arise from correctly recognizing and interpreting a wide range of hand gestures go a long way toward user access, control, and manipulation.

This paper focuses on creating a model for detecting combinations of hand gestures using sEMG. The new model combines CNN and GRU layers to analyze and interpret the complex sEMG signals. The CNN layers extract spatial features from sEMG data while the GRU layers extract the temporal features. This research is done with Ninapro database (DB1), which is a very rich database with the most variations of combined sEMG recordings taken from different participants [1]. Such data will make it possible to develop an accurate, robust, and generalizable model that can predict an extensive range of combined hand gestures. The next generation of gesture recognition systems will likely continue to evolve towards utilizing sEMG for dual-hand coordination, driven by ongoing advancements in deep learning methodologies and enhanced signal processing techniques, thereby expanding the potential applications in both assistive devices and interactive interfaces [2, 3].

Surface electromyography (sEMG) has a key advantage over motion sensors and cameras; it shows movement intentions in real time, making it very useful for gesture recognition systems [4–6].

Recent work focuses on recognizing gestures from both hands, aiming for more natural and flexible control in prosthetics and HCI [2, 7]. Earlier methods mostly focused on one-hand and used ML models like SVM and neural networks [8, 9]. Recognizing two-hand gestures adds complexity and noise to the sEMG signals [10], but it also opens up big possibilities for both able-bodied and disabled users [2, 6]. To succeed, it needs advanced methods for handling more gestures [11], and models that combine CNNs and LSTM networks look promising for this task [12]. CNN-LSTM models improve dual-hand gesture recognition by up to 10% over simpler methods [8, 13]. Though SVMs with PCA perform well, neural networks handle complexity better due to their flexibility [14, 15]. Neural networks offer strong accuracy and adaptability but require lots of data, are resource-heavy, and can overfit or lack transparency [6, 16]. These challenges are critical to solve as dual-hand gestures make sEMG signals more complex and noisy. Such complexity requires very complex methodologies in signal processing and pattern recognition. Most recent works in this area focus on integrated models, combining CNNs for spatial feature extraction with an LSTM for capturing temporal dependencies [17]. Such an integrated model approach proved to improve the recognition accuracy of dual hand gestures, where the patterns of muscle activation across various regions of the forearm are accurately localized and identified while underpinning the transitions of the gesture sequence across time [6]. This attention has been designed in the neural network to help the network focus on relevant information available in the sEMG data [13, 18].

Existing models lack training on diverse datasets, which requires new models to use extensive and diverse datasets to improve research in the field [8, 19, 20]. Real-time processing capability is identified as a critical problem within this field because it plays a vital role, especially for applications that include robotics and advanced prostheses. The present technologies demonstrate both latency issues and computational inefficiencies, which impact the user experience and the ability to develop and deploy practical systems [4, 21]. It claims that research in the future needs to concentrate on creating improved neural network structures and implementing model pruning and quantization while making use of neural processing units to enhance performance in real-time operations [20, 22]. The advancement of robustness and effectiveness in sEMG-based gesture recognition systems has been achieved through enhanced signal processing techniques alongside ML and AI applications, hybrid system deployment and fusion techniques [8, 20]. These sEMG-based gesture recognition systems can significantly enhance user experience and accessibility in diverse domains such as prosthetic arms for rehabilitation, virtual reality and robotics by being more accurate and getting closer to how human hands function [21].

Current research mainly examines movements performed by one-hand whereas typical human activities require both hands to operate simultaneously. Therefore, single-hand recognition systems limit the functionality of assistive technologies. While recent models have attempted to solve this problem with systems comprising machine learning models designed to increase recognition accuracy, the problem persists given the high variability and similarity in the activation patterns of muscles when it comes to dual-hand gestures [8, 23]. Considering the real-time processing of sEMG signals and the need for real-time operation in interaction applications like robotic control and advanced prostheses, the problem becomes even more complex [24]. Therefore, we propose that a strong predictive model formulated with advanced signal processing techniques and machine learning algorithms will remarkably enhance the accuracy and efficiency of recognizing simultaneous dual-hand gestures from sEMG data. This will improve sEMG-based systems' functional capabilities and make them more practical for areas of application.

The contribution made in this study falls under critical human-computer interaction (HCI). It fills the existing gap that exists in models of dual-handed gesture recognition. The predictive model developed is in a position to grab and interpret the combinations of gestures from both hands at the same time; the effective advancement of this investigation gives us a big jump in our conceptual understanding of sEMG-based gesture recognition. The hybrid CNN-GRU model, with the functions of integrating and representing spatial and temporal information, exhibits better performance in the subtask of gesture recognition, making the dual-hand detection model both superior and theoretically enriched. This study significantly extends the theoretical and practical applications of gesture recognition technology, considering that this model can process complex dynamics of muscle activities from dual-hand gestures. The development allows for deeper decoding of patterns in muscular activity, which is critical to implementing intuitive and effective HCI systems.

In addition, the research in this work proposes new methods and processing techniques on how to manage sEMG signals, advancing the theoretical bases that consequently become the main building blocks for both robust machine learning and gesture recognition algorithms. This study has thus gone far in providing a strong theoretical framework that can support related research in other disciplines, thereby pushing the frontier beyond what might have been previously considered possible in HCI and technology dependent on sEMG. However, the practical implications of these findings go beyond theoretical improvements. The developed models' numerous capabilities can be used to revolutionize applications in many areas, for instance: virtual reality, gaming, healthcare, rehabilitation, manufacturing, entertainment, and education, among others. Such research aids in achieving technologies that naturally support human-machine interfaces with improved user experience, boosting productivity, and raising access to digital services in the fast-changing technological landscape. Overall, this study not only fills a critical loophole in the existing models but also acts as a cornerstone for future explorations and sets the stage for a series of innovative applications that leverage new technologies, making advanced gesture recognition power more engaging through the development of intuitive interactions between humans and computers.

2. DATA & METHODS

The process for creating a predictive model to identify integrated dual hand movements using sEMG data is described in this section. The procedure consists of multiple steps, such as creating the dataset, gathering and preparing the data, filtering and preprocessing the data, extracting features, and building a neural network model.

2.1 System Design Overview

As for the system overview, we addressed the objectives parallelly with two approaches. One approach used a pre-trained model trained with the Continuous Wavelet Transform (CWT) image dataset. The other approach involved a hybrid model, consisting of a collection of three models. Two base models processed the sEMG signals from each hand, and the higher model combined these to predict the combination of gestures.

2.2 Dataset and Processing

The dataset: Ninapro DB1, is a thorough collection of synchronized sEMG and kinematic data collected from 27 healthy subjects. This dataset has 52 hand movements, as well as a rest position, which are repeated across multiple experiments. It has been created with high-precision instruments to capture the subtle dynamics of hand movements to support research in prosthetics, rehabilitation, and HCI. sEMG data are collected by utilizing 10 Otto Bock MyoBock 13E200 electrodes that were positioned optimally for muscle activity detection. Further, Cyberglove 2 data glove utilizes its 22 sensors to collect kinematic data through sensor attachments on the glove.

Ninapro DB1 dataset utilization

Segmentation of Individual Gestures: Individual gesture signals were divided into 7800 distinct CSV files, each of which corresponded to a different gesture, using the Ninapro DB1 dataset. *Combined Gesture Formation*: 3900 CSV files representing combined gestures were created by combining signal pairs from the separated CSV files to replicate dual-hand interactions. A specialized CWT script was then used to transform these files into CWT images.

Data preparation: The dataset includes sEMG data for 17 different hand gestures, and options ranging from simple finger movements to more intricate gestures. Ten distinct gestures were chosen from the original set by their applicability to the objectives of the study. These motions cover a wide variety of motions that are necessary to investigate efficient two-handed interactions.

Since the dataset consisted of both sEMG and kinematic data in MAT format, the first step was to extract only the sEMG into CSV files. Each subject has done three sessions, and each CSV file contains 10 hand gestures. The next step was to extract each hand gesture into separate CSV files according to the sampling rate.

The Ninapro DB1 dataset undergoes comprehensive preprocessing by the original database creators, including bandpass filtering (20-450 Hz), notch filtering (50 Hz) for power line interference removal, signal rectification, and amplitude normalization. Additionally, the dataset provides presegmented movement windows (5-second trials with 3-second rest periods), eliminating the need for manual segmentation. Our methodology builds upon this foundation by applying CWT transformation directly to the preprocessed signals, as the time-frequency analysis inherently handles any remaining noise through its multi-resolution decomposition. The robust preprocessing pipeline of Ninapro DB1 is well-documented in [1] and widely accepted in the sEMG research community.

Choosing 10 hand gestures: Out of the 52 hand gestures, we chose 10 hand gestures, 5 for each hand to have 25 combinations of hand gestures (FIGURE 1), and different hand gestures were assigned to each hand rather than using the same gesture for both. This decision was made to broaden the expressive scope of the gesture set and to enable more diverse and meaningful interactions. Assigning distinct gestures to each hand allows the system to capture a wider range of combinations and improve its potential for complex gesture recognition tasks.

Left Hand	Right hand
 Thumbs up	 Extension of the index and middle fingers
 Flexion of the ring and little finger, extension of the others	 Thumb opposing the base of the finger
 Abduction of all fingers	 Fingers flex together in the fist
 Pointing the index finger	 Adduction of extended fingers
 Wrist supination	 Wrist pronation

Figure 1: Chosen hand gestures

2.3 Model Development

A combination of recurrent and CNNs was used in the predictive model. RNNs examine the temporal dependencies in the gesture sequences, whereas CNNs extract spatial characteristics from the sEMG data.

Pre-trained approach: As the first approach, we experimented with the pre-trained CNN models. We chose to experiment with a few pre-trained models. The selected pre-trained models were VGG-19, VGG16, EfficientNet, and Mobilenet. The VGG-16 model was the one for which we were able to get the highest accuracy out of all of the models. The dataset was converted into CWT pictures so that the VGG-16 architecture could be used for training to modify this model for sEMG data. In this step, the distinct time-frequency properties of sEMG data are integrated with the powerful feature extraction capabilities of VGG-16.

System overview for the pre-trained model approach: As shown in FIGURE 2, the first step was to merge the two relevant gesture combinations into one CSV file, resulting in each gesture consisting of 10 electrode signals and forming a total of 20 electrode signals. The CSV files were

horizontally aligned so that each contained 20 columns. In the next step, this CSV file was converted into a CWT image. These CWT images were then used to train the pre-trained models.

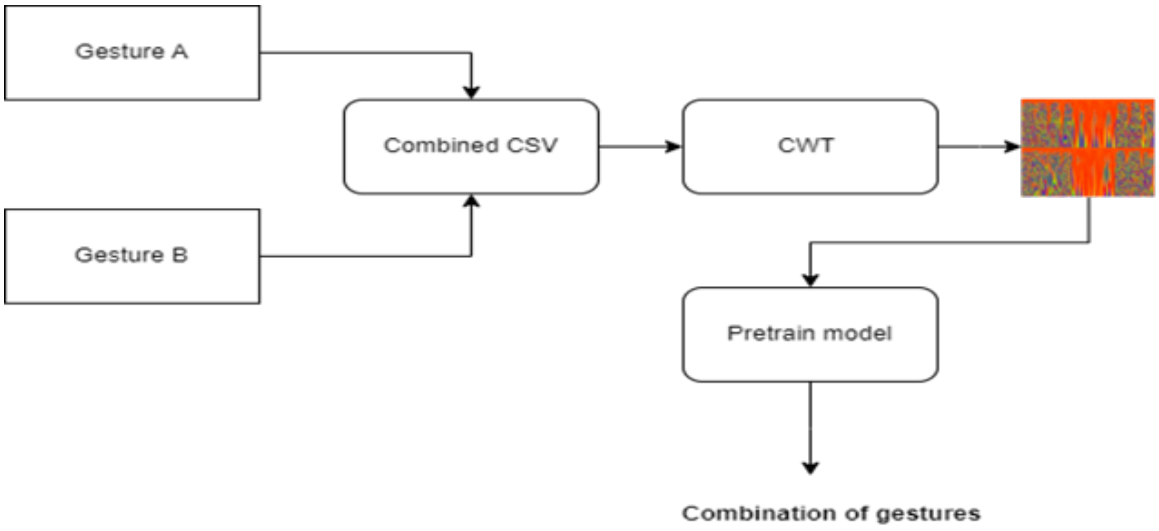


Figure 2: Pre-train model overview.

CWT images and the dataset: At this step, sEMG data was transformed into a time-frequency representation using CWT. It transforms a signal into tiny waves that are localized in both time and frequency. This creates an image that has captured both the spatial and temporal dynamics of the signal.

The continuous wavelet transform $CWT(a,b)$ is defined as [25]:

$$CWT(a,b) = |a|^{-1} \int_{\infty}^{\infty} x(t)\psi^{*}(at-b)dt \tag{1}$$

In this equation, CWT(a,b) denotes the wavelet coefficient at scale a and translation b, x(t) represents the input signal, and ψ(t) is the mother wavelet. Parameters a and b control the scale and translation of the wavelet, respectively, allowing the CWT to analyze signals over various resolutions and positions. This mathematical framework enables the extraction of time-frequency information from signals, making CWT a powerful tool in signal processing, feature extraction, and data analysis across diverse fields such as neuroscience, engineering, and finance.

We were able to create 4000 images by combining two gestures into CSV files. We divided each class's 800 images into two datasets: 720 for testing and 80 for validation.

CWT parameters: The continuous wavelet transform was applied to sEMG signals using the following parameters:

Mother wavelet: Morlet wavelet for optimal time-frequency localization

Scales: 128 scales (logarithmic spacing)

Frequency range: 10-500 Hz

Sampling frequency: 2000 Hz

Window length: 256 ms with 75% overlap

Output image size: 128 × 128 pixels

Normalization: Min-max scaling [0, 255]

Color format: Color scale (RGB) for enhanced feature visualization

These parameters generated 4000 time-frequency images from the preprocessed sEMG data for gesture classification.

Hybrid approach/decision fusion: In the next step, we explored hybrid approaches or ensemble learning. As shown in FIGURE 3, this approach entails the use of two low-level models, one for each hand, both of which are 13-layer CNN-based. To consider the sequence dependencies, the GRU layer was included. All the models were trained with five different hand gestures that were not shared with the other models.

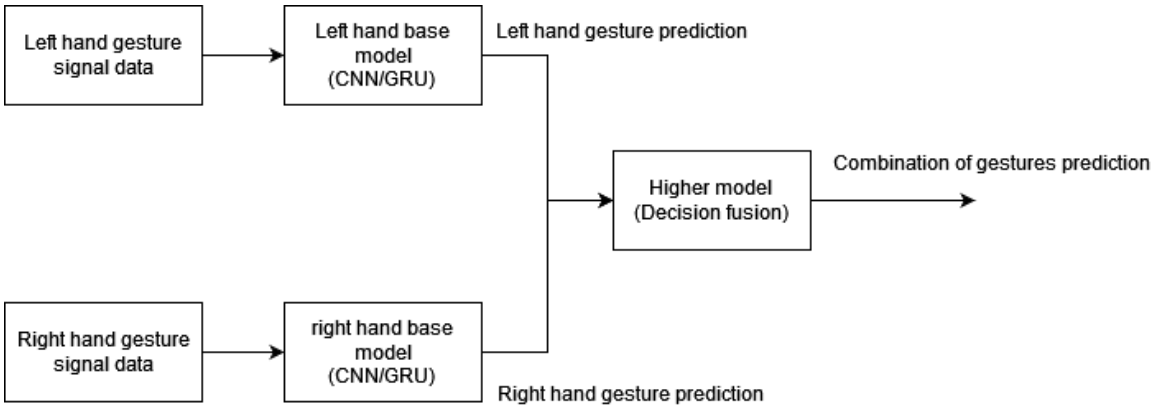


Figure 3: Hybrid model overview.

Both models had the same architecture. Subsequently, the predictions from these two models were used to train a higher-level model. The purpose of the higher-level model is to predict the occurrence of some pairs of gestures. Our selected high-level model is a random forest model that predicts 25 different gesture combinations. Users can modify and personalize the higher-level model to serve multiple applications.

Hybrid model architecture: The proposed hybrid model follows a stacked ensemble architecture optimized for SMG-based gesture recognition. It consists of two independently trained base models with identical CNN-GRU architectures and a Random Forest classifier acting as a meta-learner. Each base model processes time-series input shaped (T,10)(T, 10)(T,10), beginning with two 1D convolutional layers (64 filters, kernel size 5) and batch normalization, followed by max pooling, then two additional convolutional layers with 128 filters. A global average pooling layer reduces

dimensionality before reshaping for temporal modelling via a 256-unit GRU layer. This is followed by a fully connected layer (256 units, ReLU), dropout (**??**), and a softmax output layer.

The models are optimized independently using Adam optimiser with exponential learning rate decay, promoting diversity in learned representations. After training, the softmax outputs from both models are concatenated and passed to a Random Forest classifier, which serves as the final decision layer. This ensemble structure balances deep feature extraction with robust aggregation, improving performance in classifying single and compound gesture patterns.

Decision fusion: Decision fusion was used in the proposed system to improve hand gesture recognition accuracy and reliability. The system combines outputs from two separate models where CNN and GRU architectures process sEMG signals from each hand individually.

Individual model architecture: The system assigned each hand its own CNN/GRU model, which was designed to process the sEMG signals from CSV files for accurate hand gesture prediction. The spatial features from the sEMG data are extracted by CNN layers, and temporal dependencies are extracted through GRU layers which together allow models to accurately interpret hand movements.

Decision Fusion Process: The individual models process sEMG signal data from CSV files to generate hand gesture outputs before sending these outputs to the higher-level model for decision fusion. The fusion process merges the results from both hands to determine which gestures the user performed.

CSV dataset: For the hybrid approach, the dataset was split into individual gestures, and each gesture was sampled at a rate of 512. Every single gesture was saved in a separate CSV file, which led to the creation of 7800 CSV files. We used an 80/20 split for training and validation. For each class, 624 CSV files were assigned for training and 154 were for validation. The model was trained using a split of the data into training, validation, and testing sets. This approach ensures that the model is not only accurate but also generalizable to new, unseen data. The model's performance was evaluated based on accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the model's ability to recognize the combined gestures accurately.

3. RESULTS AND ANALYSIS

TABLE 1 compares two distinct modeling strategies evaluated in this study. Approach 1 employs Continuous Wavelet Transform (CWT) to convert raw input signals into time-frequency spectrograms, which are then fed into well-established pretrained CNN architectures such as VGG-16 and VGG-19. This approach leverages transfer learning by fine-tuning these pretrained models for gesture classification tasks. In contrast, Approach 2 is a hybrid model developed from the ground up. It integrates two independently trained CNN-GRU models as base learners, which are specifically tailored to capture both spatial and temporal dynamics in the gesture data. The outputs of these base models are then aggregated using a higher-level Random Forest classifier, which serves as a meta-learner to predict the combination of gestures being performed.

The first four entries of the table show the results of the accuracies of models whose bases are pre-trained CNNs; they are progressive in increase: EfficientNetB0, MobileNet, VGG-19, and VGG-16, with the last one in the group attaining the best accuracy of 60%. Notably, the embedding of GRUs into the CNNs had a dramatic effect on greatly improving the accuracy by 85% with the 'Right-hand' model and surpassed this with an even higher mark of 83% in the 'Left-hand' model.

Table 1: Architectures and accuracies

Model	Accuracy	F1 Score	Architecture
EfficientNetB0	21%	0.20	CNN (Pre-trained)
Mobilenet	29%	0.20	CNN (Pre-trained)
VGG-19	29%	0.21	CNN (Pre-trained)
VGG-16	60%	0.24	CNN (Pre-trained)
Left Hand	83%	0.83	CNN/GRU
Right Hand	85%	0.84	CNN/GRU
Higher Model	99%	0.93	Random Forest

3.1 Pre-train Model Training

The pre-trained models showed signs of overfitting. It indicated that the CWT images are quite dissimilar to the natural images that the pre-trained models were trained on. We attempted fine-tuning by unfreezing the lower layers, but it was unsuccessful. Out of the pre-trained models of the study, VGG-16 produced the best performance.

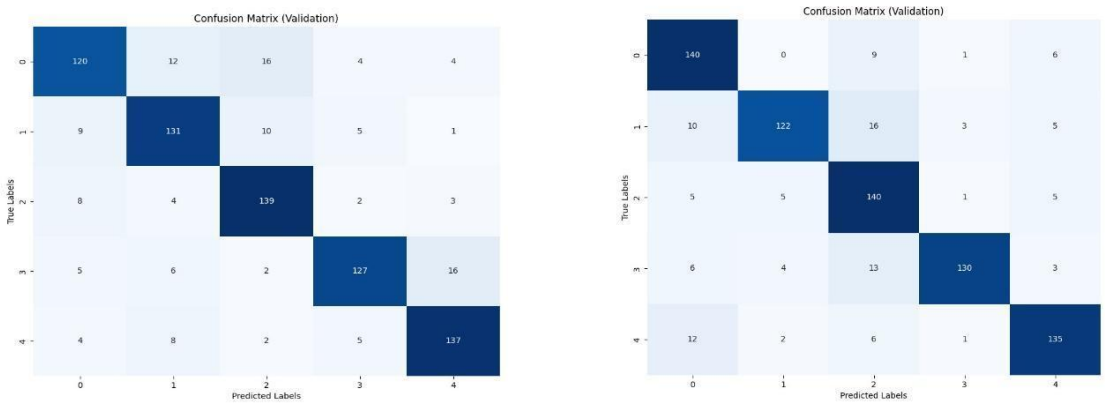
3.2 Hybrid Model/Decision Fusion

3.2.1 Left- and right-hand models

The low-level models were trained using the individual gesture CSV files that we derived from the Ninapro dataset. Each low-level model was trained using 150 epochs. After 150 epochs, the first model achieved 84% validation accuracy, and the validation loss was 0.4844 while the precision was 0.8385 (FIGURE 4(a)). The Right-hand model also achieved 85% validation accuracy, similar to the left-hand model (validation accuracy: 0.8551 and precision: 0.8551) (FIGURE 4(b)).

The confusion matrix (FIGURE 4 (a) and (b)) illustrates the distribution of predicted versus true labels across all gesture classes. Strong diagonal dominance indicates high classification accuracy, though certain off-diagonal entries (e.g., confusion between classes 2 and 5) suggest the need for further feature refinement or additional training samples for those gestures.

The ROC Curves (FIGURE 5(a) and (b) and classification report (FIGURE 6 (a) and FIGURE 6 (b)) show high precision and recall for most gesture classes from both left hand and right hand, with average F1-scores above 0.83 and 0.80. Lower recall in certain classes (e.g., class 7) reflects difficulties in detecting those gestures reliably, aligning with observations from the confusion matrix.



(a) Left-hand (b) Right-hand

Figure 4: Model's Confusion Matrix.

3.2.2 Higher model

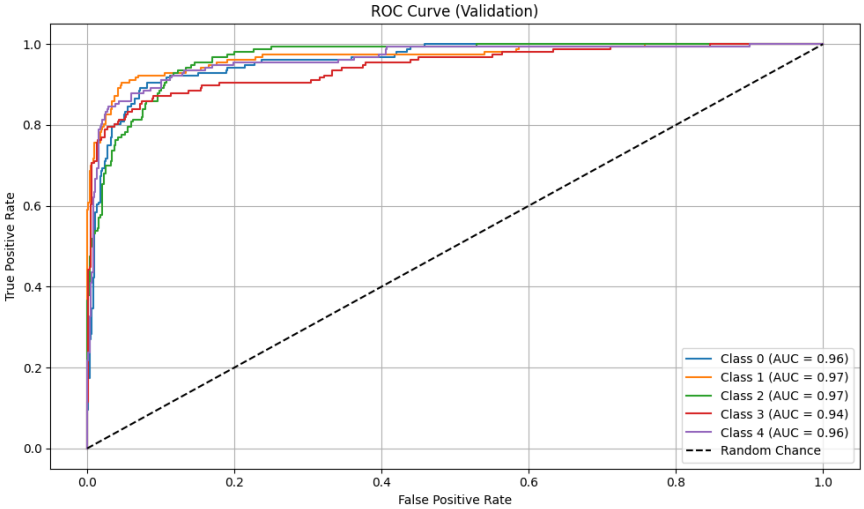
After training both the left- and right-hand models, the predictions were used to train a random forest model as a higher model to predict the combination of 25 hand gestures. This model was trained on the outputs of left- and right-hand models. The fusion process entails the training of the low-level models for the left- and right-hand, and their outputs are used to train the high-level models for decision fusion. In particular, a random forest model is used to predict all possible 25 hand gestures taking into account predictions from both left- and right-hand models.

Integration of left- and right-hand predictions: The fusion process involves combining the predictions from the left- and right-hand models to make a final decision regarding the combination of gestures performed by the user. Each low-level model produces probabilities or class labels for the respective hand gestures based on the sEMG signals processed through the CNN/GRU architectures.

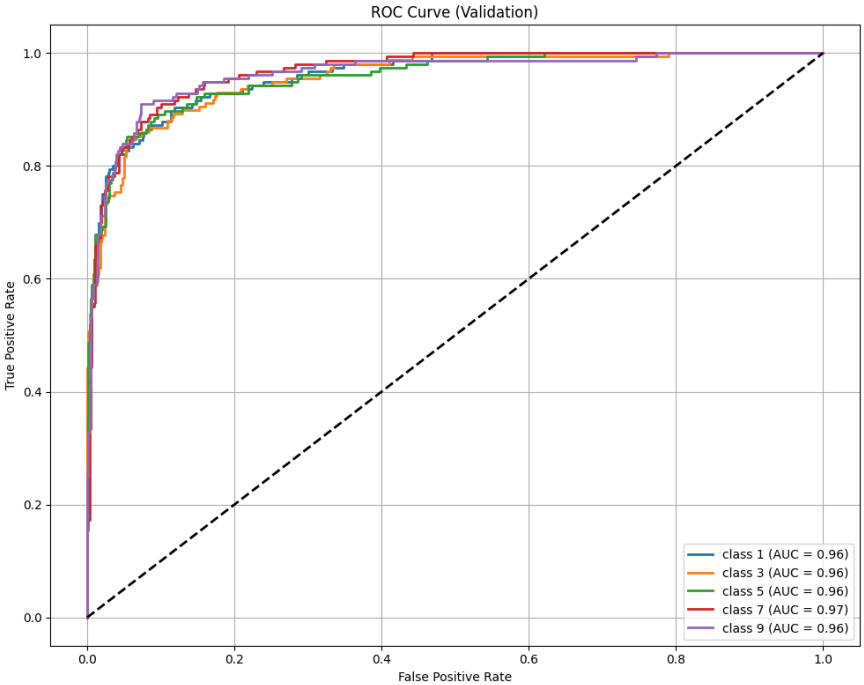
The random forest model was trained using the outputs of the left- and right-hand models as input features. These features represent the predicted probabilities or class labels for each of the 25 hand gestures. By incorporating the predictions from both hands, the random forest model learns to identify patterns and relationships between the gestures performed by each hand.

Prediction of gesture combination: Once the random forest model is trained on the integrated outputs of the left- and right-hand models, it is capable of predicting the combination of gestures performed by the user. By considering the joint probabilities or class labels provided by both low-level models, the higher-level model can make informed decisions regarding the simultaneous gestures executed by each hand.

The fusion model, combining predictions from both the left- and right-hand models, achieves a validation accuracy rate of 99%, surpassing the individual base models' accuracies of 84% for the left-hand model and 85% for the right-hand model. This integration allows the fusion model to



(a): Right-hand



(b): Left-hand

Figure 5: Model's ROC curve.

gain a comprehensive understanding of the user's gestures, enabling precise prediction of complex combinations of hand movements.

The ROC curve (FIGURE 7) and confusion matrix (FIGURE 8) for each class were computed using a one-vs-all strategy. The average AUC score was 0.99, demonstrating the model's strong ability to

Classification Report:

	precision	recall	f1-score	support
class 0	0.78	0.85	0.81	156
class 1	0.83	0.88	0.85	156
class 2	0.75	0.82	0.79	156
class 3	0.91	0.79	0.85	156
class 4	0.90	0.80	0.85	156
accuracy			0.83	780
macro avg	0.83	0.83	0.83	780
weighted avg	0.83	0.83	0.83	780

(a): Left-hand

Classification Report:

	precision	recall	f1-score	support
class 1	0.83	0.78	0.81	156
class 3	0.80	0.80	0.80	158
class 5	0.82	0.78	0.80	156
class 7	0.72	0.91	0.81	156
class 9	0.88	0.74	0.81	156
accuracy			0.80	782
macro avg	0.81	0.80	0.80	782
weighted avg	0.81	0.80	0.80	782

(b): Right-hand

Figure 6: Classification reports

distinguish between different gesture classes. Notably, classes with lower AUC values corresponded to those with more confusion in the confusion matrix.

The classification report (FIGURE 9) shows high precision and recall for most gesture classes, with average F1-scores above 0.99. Lower recall in certain classes (e.g., class 1-8) reflects difficulties in detecting those gestures reliably, aligning with observations from the confusion matrix.

3.3 User Interface for the Predictive Model

A user interface was created using Flask to incorporate all three hybrid model architectures. As shown in FIGURE 10, once the CSV file is uploaded to its corresponding model, the predictions of these models are then fed into the higher-level model to get the final output of the hand gestures. The final combination of gestures is then presented to the user on the front end of the application. The coding, model development, and system implementation were carried out in accordance with industry standards and best practices, with a focus on future advancements and evolution [26–28].

4. DISCUSSION

This study improves gesture recognition by creating a prediction model that can accurately recognize integrated dual-hand movements using sEMG data. By combining CNNs for spatial feature extraction and GRUs for temporal sequence modeling, the study not only achieved its primary goal but also enhanced the field of HCI. Unlike traditional approaches, which struggle with the complexity and temporal dependencies of such gestures, our model effectively integrates spatial and sequential learning to achieve a remarkable 99% accuracy. By leveraging the strengths of CNNs for feature extraction and GRUs for temporal modeling, this research not only surpasses previous methodologies but also sets a new benchmark for real-time human-computer interaction. These findings underscore the potential of hybrid neural networks in refining gesture recognition systems, paving the way for more responsive and efficient applications in prosthetics, rehabilitation, and human-machine interaction.

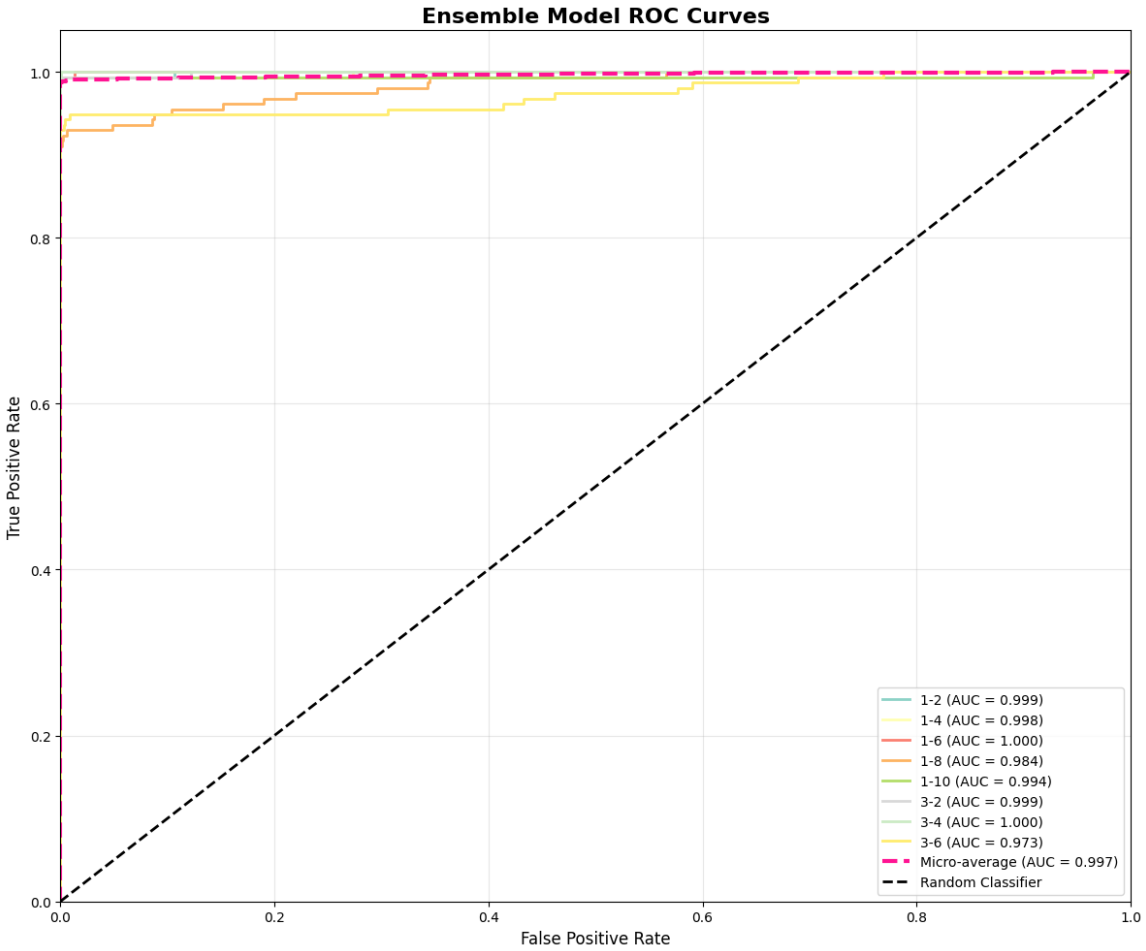


Figure 7: ROC Curve

The literature increasingly recognizes the integration of CNNs and GRUs as a robust method for gesture recognition tasks, showcasing high performance in various studies [10, 29]. The study in [8] highlighted the challenges posed by the non-linear relationship between sEMG signals and hand gestures. To address these issues, this study's use of combinations of CNNs and GRUs proves helpful. CNNs efficiently capture the spatial information from sEMG signals, while the GRUs handle and capture the temporal dependencies, providing a complete approach to understanding complex gesture signals. Also, this method differs from the approach in [4], which investigated SVM classifiers for similar tasks but did not include the sequential processing capabilities of RNN-based models like GRUs. Other approaches, such as the one presented in [6], used EMD with RF classifiers to achieve high in-hand motion recognition rates because it dealt with the non-linear and non-stationary nature of sEMG signals. However, this method fails to employ CNNs' feature extraction power together with GRUs sequence learning abilities, which the current study does. For example, traditional sEMG-based gesture recognition techniques have used various neural network configurations, including CNNs, to classify hand gestures, with an approximate accuracy of about 80.40% using multi-sensor data fusion techniques [30]. Still, such kinds of models are reasonably

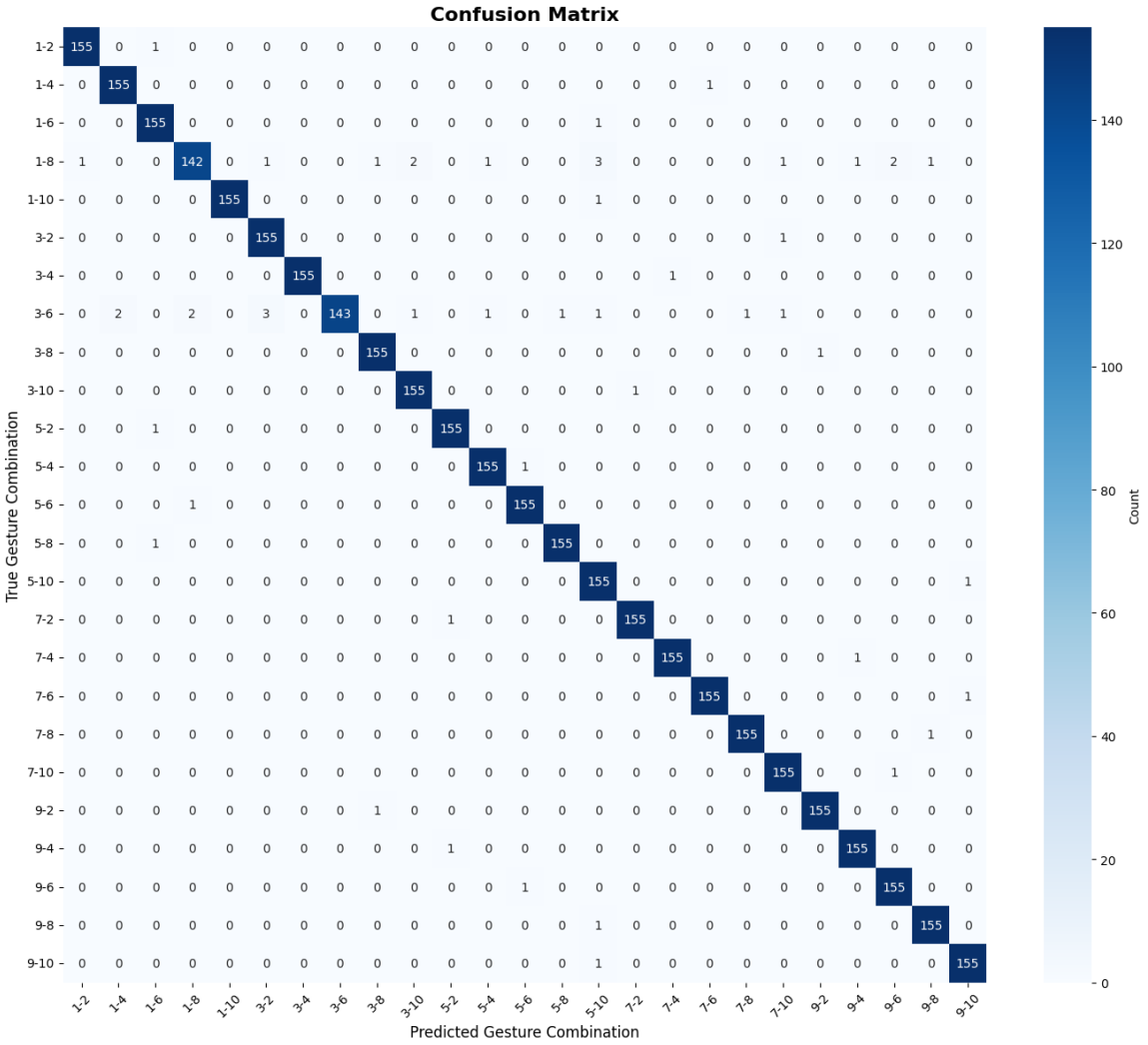


Figure 8: Confusion Matrix

unsuitable and insufficient to perform with the complexity and subtlety of dual-hand gestures due to having the minimum capability to temporal dependencies that are implicated in sEMG signals. This, as compared to the hybrid model of CNN-GRU which is being developed in the current research work and will effectively capture spatial as well as temporal features, will make a difference in the phenomenal accuracy of recognition of these gestures.

In [7], average classification accuracies of 80.88% and 82.64% using Slow Fusion and Inception models. However, these models struggled with the complexity of dual-hand gestures, particularly in capturing temporal dependencies in sEMG signals. To address this, we used a hybrid CNN-GRU model that effectively captures both spatial and temporal features, achieving a much higher accuracy of 99%. Similarly, a CNN-GRU model was developed in [8], which successfully extracts complex patterns from sEMG data, showing its strength in recognizing both spatial and temporal

	precision	recall	f1-score	support
1-2	0.99	0.99	0.99	156
1-4	0.99	0.99	0.99	156
1-6	0.98	0.99	0.99	156
1-8	0.98	0.91	0.94	156
1-10	1.00	0.99	1.00	156
3-2	0.97	0.99	0.98	156
3-4	1.00	0.99	1.00	156
3-6	1.00	0.92	0.96	156
3-8	0.99	0.99	0.99	156
3-10	0.98	0.99	0.99	156
5-2	0.99	0.99	0.99	156
5-4	0.99	0.99	0.99	156
5-6	0.99	0.99	0.99	156
5-8	0.99	0.99	0.99	156
5-10	0.95	0.99	0.97	156
7-2	0.99	0.99	0.99	156
7-4	0.99	0.99	0.99	156
7-6	0.99	0.99	0.99	156
7-8	0.99	0.99	0.99	156
7-10	0.98	0.99	0.99	156
9-2	0.99	0.99	0.99	156
9-4	0.99	0.99	0.99	156
9-6	0.98	0.99	0.99	156
9-8	0.99	0.99	0.99	156
9-10	0.99	0.99	0.99	156
accuracy			0.99	3900
macro avg	0.99	0.99	0.99	3900
weighted avg	0.99	0.99	0.99	3900

Figure 9: Higher model's classification report.

Figure 10: User interface.

characteristics. In contrast, [4] used SVMs without deep learning, which were less effective in handling temporal information, achieving 83% and 85% accuracy for left- and right-hand gestures. This is comparable to the average 83.97% accuracy reported for hybrid Slow Fusion and Inception models in [7], further highlighting the advantage of using GRU-based approaches.

The combination of EMD and RF classifiers presented in [6] achieved an average recognition performance rate of 91.67% for ten in-hand motions. Though their approach demonstrates high accuracy, it primarily captures nonlinear characteristics using EMD. However, on the other hand, the CNN-GRU approach adopted in the current study leverages the strength of deep learning in handling complex temporal sequences and spatial features, which are of primary importance in the recognition of dual hand gestures. An approach implemented in the current study adapts the pre-trained VGG16 model for sEMG data, although its accuracy is only 60%. This result is relatively low compared to the model in [7], which noted that most general-purpose models do not perform well with sEMG signals, particularly in specialized datasets. Since the hybrid CNN-GRU model outperforms the VGG16 model there is evidence that sEMG data possess unusual properties that need specialized modeling approaches.

The hybrid model demonstrated performance results that matched those reported in [22], which was 89.76% using Temporal Convolutional Networks for sEMG-based gesture recognition. The latter model improves the accuracy of the state-of-the-art by approximately 5% and accentuates the effectiveness of TDA in gesture recognition. Similarly, the study in [19], validated CNN performance for decoding hand gestures through sEMG signals was tested offline with an accuracy of over 95% for some hand gestures. Achieving high performance in real-time applications presents significant challenges which the present study's CNN-GRU model successfully addressed. Further, the study in [30] demonstrated the benefits of the multi-sensor fusion technique in gesture recognition, where a higher recognition rate is attained based on the fusion of Kinect and sEMG signals compared to a single-sensor-based method. However, their approach dealt with various types of sensors and integration at the sensor level. In contrast, the hybrid CNN-GRU model applied in the current study is being adopted to exploit the benefits of data acquired from multi-type information (spatial and temporal features) of sEMG signals to improve recognition.

Limitations and challenges: The difficulties and limitations encountered in this study reflect common issues in the field of sEMG-based gesture recognition, as well as specific problems with the application of advanced machine learning models. GRUs for sequence learning directly address these nonlinearities, resulting in a more comprehensive model of muscle activity. The current novel hybrid model of CNN-GRU has some distinct advantages over traditional methodologies. This should permit effective extraction of the features of spatial characteristics of the sEMG signals, which becomes very critical if the minor patterns for dual hand gestures are to be recognized. Indeed, the study in [7] demonstrated that CNN models like Inception will lead to great classification accuracy due to the potential to capture very fine details of the input data. The proposed model overcomes the limitation of purely spatial models with the incorporation of GRUs capable of handling the temporal sequence of features. However, in our method, the integration of the GRU layer further enhances the capability of the model to learn the dynamic temporal pattern. As a result, we achieve better recognition accuracy for complex dual-hand gestures.

According to the study in [19], hand gestures can be decoded accurately from the sEMG data using the CNN architecture. However, its ability to handle real-time applications could not be guaranteed without RNNs like GRU and LSTM. The hybrid CNN-GRU model that we proposed

in this work ensures high performance in both static and real-time situations. The application of the pre-trained VGG16 model to CWT images of sEMG signals resulted in a poor accuracy of 60%. The existing feature acceleration capability from pre-trained models struggles to match specialized dataset requirements such as sEMG. A misalignment occurs because VGG16 models achieve optimal performance with natural images, yet these measurements differ substantially from the sEMG image characteristics. The research shows that sEMG signals present elaborate patterns, which establish non-linear correlations with various hand gestures. Due to its complex structure, sEMG signals present challenges for standard image recognition algorithms to successfully decode this complexity. The pre-trained VGG16 model shows lower accuracy in this work, which points to an ongoing principal problem. The features pre-trained models learn in their original training phase do not perform well when they try to process specialized datasets like sEMG signals. Studies revealed that existing models faced similar challenges which prompted the suggestion to develop particular sEMG architectures [7, 31].

A key limitation of this study is the use of only 10 hand gestures out of a possible 52. This reduction was necessary due to constraints in time and computational resources. Expanding the gesture set would have significantly increased the demands on model training and validation, which were already resource-intensive processes. Another factor to consider is the low accuracy observed with pretrained models utilizing CWT-transformed sEMG data. These models, such as VGG-16 and VGG-19, are originally trained on natural image datasets consisting of physical and well-structured objects, and are not inherently suited to handle the irregular, noise-prone patterns typical of SMG signals. Furthermore, the absence of noise removal preprocessing, an established step in most sEMG gesture recognition pipelines, likely contributed to the degraded performance. This omission weakens the reliability of the CWT-based approach and should be addressed in future work to ensure a more effective application of pretrained architectures in this domain.

Future research: Future work can focus on optimizing models to reduce computational requirements while maintaining high performance. This can be achieved through efficient neural network designs or model pruning techniques. Advanced methods for collecting high-quality sEMG data are also essential to address variability and improve data reliability. Research should explore how models can generalize across diverse user populations by incorporating adaptive algorithms. Integrating sEMG data with other modalities, such as vision or inertial sensors, may enhance model accuracy and robustness. Additionally, future work will focus on expanding the gesture set from the initial 10 to all 52 gestures, enabling a more comprehensive evaluation of the model's performance and enhancing its applicability in real-world scenarios. Finally, large-scale real-world testing is essential to assess practical usability and uncover further development opportunities.

5. CONCLUSION

This research makes significant contributions to sEMG-based dual-hand gesture recognition for several specific reasons. The model succeeds in resolving numerous research deficiencies related to gesture recognition while establishing an effective system for hand gesture combination prediction. The primary source of sEMG data for our research was the NinaPro DB1 database, which supplied comprehensive details for creating and testing our models. We have managed to predict the effectiveness of a pre-existing VGG16 module together with a hybrid CNNGRU model through our work. The VGG16 model demonstrated poor performance with only 60% accuracy on sEMG data,

which indicates that the characteristics of these signals are very different from natural image data causing complex problems. The custom-designed hybrid model, which combines CNN with GRUs yielded significant performance improvements with an 83% accuracy for left-hand gestures and 85% for right-hand gestures during our research. The hybrid model processed dual-hand gestures with success because it effectively used both spatial and temporal features from sEMG data. The Higher model represents a framework that allows generalization through the application of action mapping.

This research not only demonstrates the effectiveness of custom machine-learning strategies beyond the capabilities of pre-trained systems but also shows that complex neural network architectures are capable of deciphering high-level patterns in muscle activity. It leads to new research directions, especially in terms of making these models more generalizable over diverse datasets and real-world scenarios. Gesture recognition technologies that improve in the prosthetics and rehabilitation sector lead to better patient life quality. However, it does have its limitations. The computational demands of our model arise from its combination of CNNs and RNNs which makes it difficult to deploy in environments where resources are limited. The performance of this approach requires high-quality and consistent sEMG data collection to ensure effectiveness. Factors like electrode shifts or variations between users can impact accuracy and reduce the model's robustness in gesture recognition. This study paves the way for further advancements in human-computer interaction technologies. Future studies need to combine efforts on creating superior model structures with enhancing system adaptability to data variations and growing the practical applications of sEMG-based gesture control. New progress in this field will produce major improvements in assistive technologies that will result in more intuitive and effective human-machine interface designs.

References

[1] Atzori M, Gijsberts A, Heynen S, Hager AG, Deriaz O, et al. Building the Ninapro Database: A Resource for the Biorobotics Community. In: 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob). IEEE. 2012:1258-1265.

[2] Wang Z, Huang W, Qi Z, Yin S. MS-CLSTM: Myoelectric Manipulator Gesture Recognition Based on Multi-Scale Feature Fusion CNN-LSTM Network. Biomimetics. 2024;9:784.

[3] Cote-Allard U, Fall CL, Campeau-Lecours A, Gosselin C, Laviolette F, et al. Transfer Learning for sEMG Hand Gestures Recognition Using Convolutional Neural Networks. In: IEEE International Conference on Systems Man and Cybernetics (SMC). IEEE. 2017:1663-1668.

[4] Bian F, Li R, Liang P. SVM Based Simultaneous Hand Movements Classification Using sEMG Signals. In: IEEE International Conference on Mechatronics and Automation (ICMA). IEEE. 2017:427-432.

[5] Simão M, Mendes N, Gibaru O, Neto P. A Review on Electromyography Decoding and Pattern Recognition for Human-Machine Interaction. IEEE Access. 2019;7:39564–39582.

[6] Xue Y, Ji X, Zhou D, Li J, Ju Z. SEMG-Based Human In-Hand Motion Recognition Using Nonlinear Time Series Analysis and Random Forest. IEEE Access. 2019;7:176448-176457.

[7] Erözen AT. A New CNN Approach for Hand Gesture Classification Using sEMG Data. J Innov Sci Eng. 2020:44-55.

[8] Quivira F, Koike-Akino T, Wang Y, Erdogmus D. Translating sEMG Signals to Continuous Hand Poses Using Recurrent Neural Networks. In: IEEE EMBS International Conference on Biomedical Health Informatics. IEEE. 2018:166-169.

[9] Cheng Y, Li G, Yu M, Jiang D, Yun J, et al. Gesture Recognition Based on Surface Electromyography-Feature Image. Concurrency Comput Pract Experience. 2020;33:e6051.

[10] Shen S, Gu K, Chen XR, Yang M, Wang RC. Movements Classification of Multi-Channel sEMG Based on CNN and Stacking Ensemble Learning. IEEE Access. 2019;7:137489-137500.

[11] Yin Z. Research on Hand Action Pattern Recognition of Bionic Limb Based on Surface Electromyography. 2021 2nd International Academic Conference on Energy Conservation, Environmental Protection and Energy Science (ICEPE 2021). E3S Web Conf. 2021;271:01030.

[12] Zhang Y, Yu J, Zhou D, Liu H. SEMG-Based Hand Gesture Classification With Transient Signal. Communications in Computer and Information Science. Springer. 2020:401-412.

[13] Yu G, Deng Z, Bao Z, Zhang Y, He B. Gesture Classification in Electromyography Signals for Real-Time Prosthetic Hand Control Using a Convolutional Neural Network-Enhanced Channel Attention Model. Bioengineering. 2023;10:1324.

[14] Suri K, Gupta R. Transfer Learning for SEMG-Based Hand Gesture Classification Using Deep Learning in a Master-Slave Architecture. In: 3rd International Conference on Contemporary Computing and Informatics (IC3I). IEEE. 2018:178-183.

[15] Bao T, Xie SQ, Yang P, Zhou P, Zhang ZQ. Toward Robust, Adaptive and Reliable Upper-Limb Motion Estimation Using Machine Learning and Deep Learning–a Survey in Myoelectric Control. IEEE J Biomed Health Inform. 2022;26:3822-3835.

[16] Wu C, Yan Y, Cao Q, Fei F, Yang D, et al. sEMG Measurement Position and Feature Optimization Strategy for Gesture Recognition Based on Anova and Neural Networks. IEEE Access. 2020;8:56290–56299.

[17] Chen X, Zhang X, Chen X, Chen X. Decoding Silent Speech Based on High-Density Surface Electromyogram Using Spatiotemporal Neural Network. IEEE Trans Neural Syst Rehabil Eng. 2023;31:2069-2078.

[18] Khezri M, Jahed M. A Novel Approach to Recognize Hand Movements via SEMG Patterns. 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2007:4907-4910.

[19] Asif AR, Waris A, Gilani SO, Jamil M, Ashraf H, et al. Performance Evaluation of Convolutional Neural Network for Hand Gesture Recognition Using EMG. Sensors. 2020;20:1642.

[20] Al-nagashi F, Rahim N, Shukor S, Hamid M. Mitigating Overfitting in Extreme Learning Machine Classifier Through Dropout Regularization. Appl Math Comput Intell. 2024;13:26-35.

[21] Gong Q, Jiang X, Liu Y, Yu M, Hu Y. A Flexible Wireless sEMG System for Wearable Muscle Strength and Fatigue Monitoring in Real Time. Adv Electron Mater. 2023;9: 2200916.

[22] Tsinganos P, Cornelis B, Cornelis J, Jansen B, Skodras A. Improved Gesture Recognition Based on sEMG Signals and TCN. In: ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019:1169-1173.

[23] Rabbi MF, Pizzolato C, Lloyd DG, Carty CP, Devaprakash D, et al. Non-negative Matrix Factorisation Is the Most Appropriate Method for Extraction of Muscle Synergies in Walking and Running. Sci Rep. 2020;10:8266.

[24] Li W, Shi P, Yu H. Gesture Recognition Using Surface Electromyography and Deep Learning for Prostheses Hand: State-Of-The-Art Challenges and Future. Front Neurosci. 2021;15:621885.

[25] Jo YU, Oh DC. Real-Time Hand Gesture Classification Using CRNN With Scale Average Wavelet Transform. J Mech Med Biol. 2020;20:2040028.

[26] http://ir.kdu.ac.lk/bitstream/handle/345/7407/FOC_IRC2023_Proceeding-Book-164-172.pdf?sequence=1&isAllowed=y

[27] http://ir.kdu.ac.lk/bitstream/handle/345/7425/FOC_IRC2023_Proceeding-Book-292-296.pdf?sequence=1&isAllowed=y

[28] N. Wisidagama, M. Karunarathne, P. Paranagama, R. Rathnayake, and B. Lankasena, "A comprehensive study on software evolution in plan driven and agile methodologies," 2023, available at: http://ir.kdu.ac.lk/bitstream/handle/345/7418/FOC_IRC2023_Proceeding-Book-237-245.pdf?sequence=1&isAllowed=y

[29] Chen J, Meng J, Wang X, Yuan J. Dynamic Graph CNN for Event Camera Based Gesture Recognition. In: IEEE International Symposium on Circuits and Systems (ISCAS). IEEE. 2020:1-5.

[30] Sun Y, Li C, Li G, Jiang G, Jiang D, et al. Gesture Recognition Based on Kinect and SEMG Signal Fusion. Mob Netw Appl. 2018;23:797-805.

[31] Pizzolato S, Tagliapietra L, Cognolato M, Reggiani M, Müller H, et al. Comparison of Six Electromyography Acquisition Setups on Hand Movement Classification Tasks. PLOS One. 2017;12:e0186132.