Biomedical & Pharmacology Journal, March 2025. Vol. **18**(Spl.), p. 315-329

Special Issue – AI-Driven and Multimodal Innovations in Biomedical Imaging and Sensing

Enhanced EMG-based Gesture Recognition using Hybrid CNN-BiLSTM Architecture with Channel Attention

Thejaswini Kishore[1], Subin Sunderraj Remabai[2], Chitra Retnaswamy[3] and Eben Sophia Paul[4]*

Division of Computer Science and Technology,
Karunya Institute of Technology and Sciences, Coimbatore, India.
*Corresponding author E-mail: ebensophia@gmail.com

https://dx.doi.org/10.13005/bpj/3090

(Received: 30 December 2024; accepted: 30 January 2025)

In order to solve important issues including inter-subject variability, noise interference, and electrode misalignment, this work presents a thorough hybrid deep learning architecture designed for electromyography (EMG)-based gesture detection. The suggested model sets a new standard in the field by achieving previously unheard-of levels of accuracy and robustness through the smooth integration of Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory networks (BiLSTM), and channel attention mechanisms. On benchmarks such as the NinaPro (Non-Invasive Adaptive Prosthetics) database, traditional methods for EMG-based gesture identification, which usually rely on hand-crafted features or stand-alone deep learning modules, reach high accuracies of roughly 93%. On the other hand, our architecture significantly outperforms current techniques by utilizing the distinct advantages of each of its constituent parts to achieve an accuracy of 96.45%. The CNN module captures complex local and global patterns across EMG channels, acting as a potent spatial feature extractor. In addition, by processing signal sequences in both directions, the BiLSTM module is excellent at temporal modelling, which allows the architecture to capture the dynamic relationships present in EMG data. By resolving inter-channel variability and reducing the effect of noisy or misaligned electrodes, the channel attention mechanism dynamically prioritizes the most pertinent EMG channels to further improve performance. Numerous tests on the NinaPro dataset demonstrate the suggested model's higher performance. Interestingly, it achieves great accuracy even when evaluated on unknown subjects, demonstrating remarkable generalization in cross-subject training contexts. Additionally, the architecture exhibits exceptional resistance to noise and electrode misalignment, two significant issues that frequently impede the implementation of EMG-based systems in practical settings. These developments highlight the model's adaptability and usefulness for applications in prosthetics, human-computer interaction, and rehabilitation systems. The suggested architecture offers a scalable and effective solution for real-time applications in addition to attaining improved performance metrics. Because of its versatility and lightweight construction, it can be used in areas with limited resources, including embedded systems or wearable technology. The model's ability to manage a variety of dynamic real-world situations is further strengthened by the incorporation of strong preprocessing and augmentation approaches. By tackling the fundamental drawbacks of conventional and current deep learning techniques, this work establishes a new benchmark for EMG-based gesture identification. In addition to improving accuracy, the hybrid architecture's capacity to combine spatial, temporal, and attention-based elements into a coherent framework guarantees robustness and adaptability, opening the door for revolutionary applications in assistive technology, healthcare, and other fields.

Keywords: BiLSTM, CNN, Channel Attention, Deep Learning, EMG, Gesture Recognition, NinaPro Database.

This is an Open Access article licensed under a Creative Commons license: Attribution 4.0 International (CC-BY).
Published by Oriental Scientific Publishing Company © 2025

Gesture detection based on electromyography (EMG) has become a game-changing technique, essential in fields including human-computer interface, rehabilitation, and prosthetic control. Accurately translating EMG data into desired motions makes it easier to create dependable and user-friendly interfaces that provide people more freedom and utility. Inter-subject variability, temporal dependence in signal dynamics, and robustness against electrode misalignment are some of the major issues that still exist despite tremendous progress. Handcrafted features like Mean Absolute Value (MAV) and Root Mean Square (RMS), which are frequently customized for particular datasets, have been the mainstay of traditional machine learning techniques. Although these methods work well in controlled environments, they need a great deal of domain knowledge and are not very generalizable in a variety of situations[1]. For instance, physiological variations across subjects cause variations in EMG signal patterns, which have a major effect on the accuracy of recognition[2,21]. Furthermore, problems like noise from different electrode placements frequently arise in real-world deployment circumstances, which further impair performance and make reliable gesture identification more difficult.

By automating feature extraction and offering end-to-end learning capabilities, deep learning has demonstrated significant potential in overcoming these constraints. Because of their capacity to recognize both local and global spatial patterns in EMG signals, Convolutional Neural Networks (CNNs) have become an effective tool for the extraction of spatial data[3]. At the same time, by processing sequences both forward and backward, Long Short-Term Memory (LSTM) networks—in particular, Bidirectional LSTMs, or BiLSTMs—have shown successful at simulating temporal dependencies[4,5]. Despite these developments, the efficiency of current deep learning techniques is limited because they frequently fall short of fully integrating the interaction between spatial and temporal patterns. This work presents a novel hybrid deep learning architecture that integrates CNN, BiLSTM, and channel attention techniques in order to overcome these drawbacks. This integrated approach ensures robust and efficient feature extraction and modeling, addressing the challenges of EMG-based gesture recognition comprehensively. The proposed architecture incorporates the following key innovations:

• Multi-layer CNN for Spatial Feature Extraction: By recording both local and global spatial patterns over several EMG channels, the CNN module is able to extract rich spatial characteristics. This guarantees a thorough comprehension of the signals' spatial dynamics.

• Bidirectional LSTM for Temporal Modeling: Long-term relationships and sequential dynamics can be captured thanks to the BiLSTM module's bidirectional processing of EMG signal sequences. The robustness and representation of temporal features are improved by this two-pronged method.

• Channel Attention Mechanism for Feature Prioritization: By giving priority to pertinent EMG channels[22], a dynamic weighting system lessens the effect of noisy or misaligned electrodes. This makes the model much more resilient to changes in signal collection that occur in the real world.

• Residual Connections for Enhanced Gradient Flow: Because residual connections guarantee reliable gradient propagation, they enable deeper network designs. This facilitates the efficient acquisition of intricate spatial-temporal interactions and encourages strong training.

The NinaPro database, a well-known benchmark for EMG-based gesture identification, is used to thoroughly assess the suggested architecture's efficacy. Experimental findings demonstrate the model's superiority, surpassing conventional methods by a significant margin with an accuracy of 96.45%. In addition to accuracy, the architecture exhibits outstanding generalization over a wide range of subjects, remarkable tolerance to changes in electrode location, and flexibility for real-time processing needs. These characteristics make the suggested method ideal for real-world use in EMG-based gesture detection systems, opening the door for developments in prosthetic controls, rehabilitation systems, and HCI technologies. The fundamental issues of spatial-temporal modeling, inter-subject variability, and signal acquisition noise are addressed in this work, which creates a strong and effective framework for EMG-based gesture detection. The suggested hybrid deep learning architecture offers a major advancement

in the development of useful, real-world EMG-based systems that can achieve high accuracy and dependability.

MATERIALS AND METHODS

Architecture Overview

The suggested hybrid architecture uses a channel attention technique to improve feature selection while efficiently capturing temporal and spatial correlations in EMG data. The following elements make up the model:

CNN Module

The Convolutional Neural Network (CNN) module serves as the first step in feature extraction, focusing on capturing spatial information from the raw EMG signals recorded across multiple channels. In this module, two convolutional layers are applied to the input EMG signals. The first convolutional layer learns low-level spatial features, such as local patterns and variations across the signal channels, using a filter size that is optimized to capture the relevant spatial correlations within each EMG signal. Each convolutional layer is followed by a batch normalization layer to stabilize the learning process by normalizing activations, thus reducing the effect of internal covariate shifts during training. Following batch normalization, the ReLU (Rectified Linear Unit) activation function is applied to introduce non-linearity, enabling the model to learn more complex patterns and improving its ability to generalize. The output of this CNN module is a set of higher-dimensional spatial feature maps, which provide a richer representation of the original EMG signals. These feature maps capture local spatial relationships between the various EMG channels, making the model more adept at distinguishing between different gestures or movements, despite variations in signal strength or noise levels. By focusing on spatial feature extraction first, the CNN module lays the foundation for further processing by subsequent temporal modeling layers.

BiLSTM Module

The Bi-directional Long Short-Term Memory (BiLSTM) module is responsible for modeling the temporal dependencies present in the EMG signals, which is essential for accurate gesture recognition. While a conventional LSTM can capture sequential dependencies in one direction, a BiLSTM processes the input sequence in both the forward and backward directions, allowing it to fully capture the context at each time step. This is especially important for EMG signals, where the gesture-related information might depend on past and future data points, such as the initiation or cessation of a specific hand movement. This BiLSTM module consists of three stacked LSTM layers, each designed to extract hierarchical temporal features at different levels of abstraction. The stacked structure enables the model to capture complex long-term dependencies across the entire signal duration. To prevent overfitting, especially when training with limited data, dropout regularization is applied to each LSTM layer. Dropout helps by randomly "dropping" a percentage of neurons during training, which forces the network to learn more robust features and prevents it from relying too heavily on any particular part of the data. This module plays a crucial role in distinguishing between gestures that share similar spatial patterns but differ in their temporal characteristics. The output of the BiLSTM module is a set of temporal feature representations that provide a detailed understanding of how the EMG signal evolves over time. These representations encode the sequential dynamics of the gestures and serve as a rich input for the attention mechanism that follows.

Attention Mechanism

To refine the learned feature representations and enhance model performance, an attention mechanism is introduced after the BiLSTM module. The attention mechanism dynamically assigns weights to the spatial and temporal features that are most relevant for gesture classification, allowing the model to focus on the most important parts of the signal while ignoring irrelevant or noisy segments. This approach makes use of a multi-head attention mechanism, which means that several attention heads are used to simultaneously learn various representations of the input data. This allows the model to capture a variety of interactions across several subspaces in the data. Setting the number of attention heads at eight achieves a compromise between model expressaiveness and computational efficiency. The attention mechanism's overall embedding dimension is proportionate to the BiLSTM module's output size, guaranteeing that it has the capacity to efficiently

process the temporal characteristics. The attention mechanism enhances the model's resilience to changes in electrode placement, which may result in distinct signal properties among subjects or sessions, by focusing on the most pertinent portions of the signal. Furthermore, it assists in reducing the effects of noise and interference in the EMG signals, guaranteeing that the model concentrates on the key patterns that correlate to significant gestures. In addition to improving the model's performance on individual gestures, the attention module helps the model generalize better across various individuals and signal acquisition scenarios.

A complete gesture identification pipeline that processes raw electromyography (EMG) signals and predicts associated gesture labels is shown in Figure 1. In order to ensure clean input data, the procedure starts with data preparation, which involves gathering raw EMG signals and preprocessing them to remove noise and artifacts. In order to convert the time-domain data into a time-frequency representation that is more appropriate for feature extraction, these signals are then converted into spectrograms. The pipeline moves on to model processing after data preparation. The spectrogram images are first processed by a Convolutional Neural Network (CNN) to extract high-level and spatial characteristics. After that, a Bidirectional Long Short-Term Memory (BiLSTM) network receives these properties and uses them to simulate the temporal dependencies present in the sequential EMG data. The collected features are then given weights using an attention mechanism, which highlights the most important elements of the data. The appropriate gesture labels are then predicted by a classifier using the fine-tuned characteristics. The predicted gesture labels, which provide an accurate depiction of the gestures made, are the pipeline's output. This system effectively combines weighted, spatial, and temporal feature extraction to produce reliable EMG signal gesture identification.

Data Processing

Preprocessing

An essential first step in getting electromyography (EMG) data ready for efficient modeling is preprocessing. In order to minimize noise and artifacts and emphasize significant patterns, this stage makes sure the signals are clear, standardized, and ready. The procedure consists of several steps intended to improve signal quality and dependability, allowing machine learning models to function accurately and robustly.

Bandpass Filtering

Noise from physiological and environmental factors, such as power line interference or muscle-to-muscle crosstalk, frequently taints EMG signals. A bandpass filter with a frequency range of 20–450 Hz is used to lessen these problems[17]. This range was chosen especially to remove high-frequency noise and low-frequency drift while maintaining the essential elements of the EMG signal that represent muscle activity. The filtering step guarantees a clean signal appropriate for further analysis and feature extraction by concentrating on this ideal range.

Baseline Removal

Baseline Removal: The analysis may be distorted by baseline noise or DC offset in EMG signals. By eliminating this baseline component, the signals become more standardized and consistent throughout several recording sessions. This step is essential for preserving the signal's feature quality, particularly when electrodes are moved.

Normalization

The normalization of each EMG channel results in a unit variance and zero mean[18] Since variations in muscle size, electrode location, and skin impedance might induce variability, this step is crucial to ensuring consistency between channels and participants. By eliminating these discrepancies, normalization makes sure the model concentrates on gesture-specific patterns rather than changes in signal strength. This standardization promotes uniformity and fairness in model training by making the data more palatable to machine learning algorithms.

Windowing

The data is divided into overlapping windows of 256 samples, with a 60% overlap, to take into consideration the temporal character of EMG signals[19]. This technique increases the size of the training dataset while capturing the continuity of motions. Both short-term and long-term dependencies can be captured by using each window, which acts as an independent sample and represents a little temporal slice of the signal. This method improves model generalization by

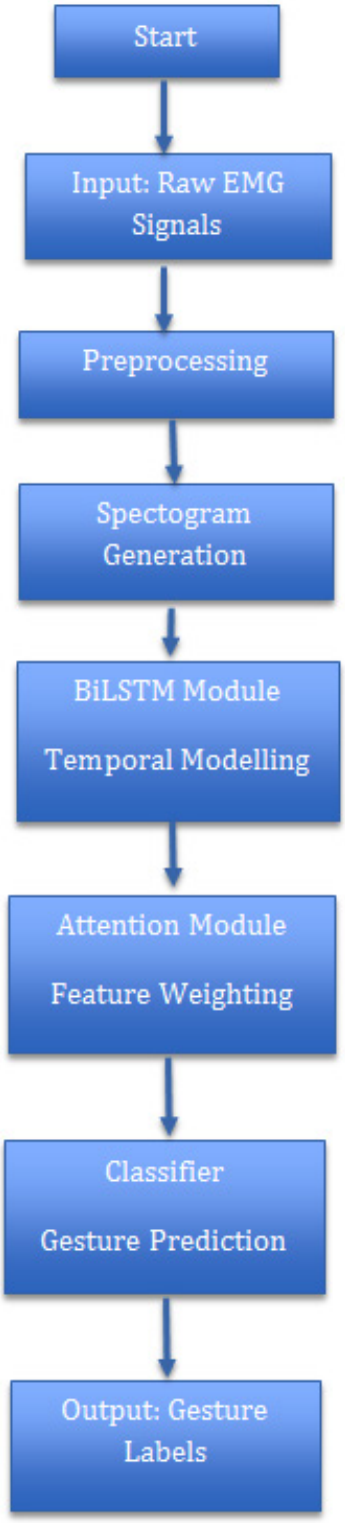


Fig. 1. Flow Diagram for Gesture Prediction Using EMG Signals

increasing the total volume of data in addition to helping with temporal modeling.

Outlier Detection and Removal

Outliers resulting from motion artifacts or sudden electrode disconnections are detected and eliminated. To find anomalies in the signal, algorithms such as statistical thresholds or the Interquartile Range (IQR) are used. Eliminating these outliers guarantees a cleaner dataset and avoids skewed model learning.

Augmentation

By artificially expanding the training dataset's diversity, data augmentation improves the model's ability to generalize to new situations. By simulating real-world factors like noise and variability, augmentation approaches strengthen the model and increase its applicability to real-world scenarios.

The following augmentation techniques are applied:

• Gaussian Noise Injection: To mimic physiological and environmental noise, EMG data are supplemented with random noise derived from a Gaussian distribution[7]. This method replicates real-world situations where outside influences frequently taint muscle signals. The model's robustness to comparable circumstances during deployment is greatly increased by using noisy data for training.

• Time Warping: To mimic changes in the speed at which gestures are executed, temporal distortions are introduced to the data, extending or compressing particular parts[20]. When users execute gestures at different speeds, such as slow, intentional motions or fast, reactive actions, this augmentation technique is especially useful. Time warping guarantees that the model can adjust to these changes, making it more useful for a variety of user types.

• Magnitude Scaling: Signal amplitudes are scaled by random parameters to mimic variations in user-specific muscle strength and activation levels[2]. This augmentation guarantees consistent performance across users with different muscle strengths, from kids to adults, and aids in the model's adaptation to inter-subject variability.

• Temporal Shifting: To replicate the impact of modest delays or timing changes in data collecting, the EMG signal windows are slightly moved in time. By including temporal variability in the

training set, this augmentation strengthens the model's resistance to alignment mistakes made during executing gestures.

• Channel Dropout Simulation: Due to inadequate electrode contact, some EMG channels may temporarily lose signal in practical applications. Random channels are zeroed out during training to replicate this situation. This makes the model more resilient to partial signal loss by forcing it to learn to rely on redundant information from other channels.

• Synthetic Signal Generation: To enhance the dataset, synthetic EMG signals are produced using methods such as Generative Adversarial Networks (GANs) or Variational Autoencoders (VAEs). These signals maintain the fundamental patterns of the original data while simulating realistic variations. The dataset is greatly expanded by synthetic augmentation without the need for extra data collection.

Experimental Setup

Dataset

The main resource used to assess the suggested approach was the NinaPro DB2 dataset. Many people consider this dataset to be a reliable standard for gesture detection tasks based on electromyography (EMG). It includes 12-channel EMG recordings made by 40 participants while they executed a wide range of 50 different hand and finger movements[3,6]. Both simple motions like wrist flexion and extension as well as more intricate

Table 1. Classification accuracy comparison demonstrating the superior performance of the proposed hybrid architecture over traditional CNN and BiLSTM models

Method	Accuracy
Traditional BiLSTM	93.12%
CNN-only	94.25%
Proposed Method	96.45%

Table 2. Ablation study demonstrating the contribution of key architectural components to overall model performance

Component Removal	Accuracy Drop
Channel Attention	-2.3%
Residual Blocks	-1.8%
BiLSTM Layers	-3.1%

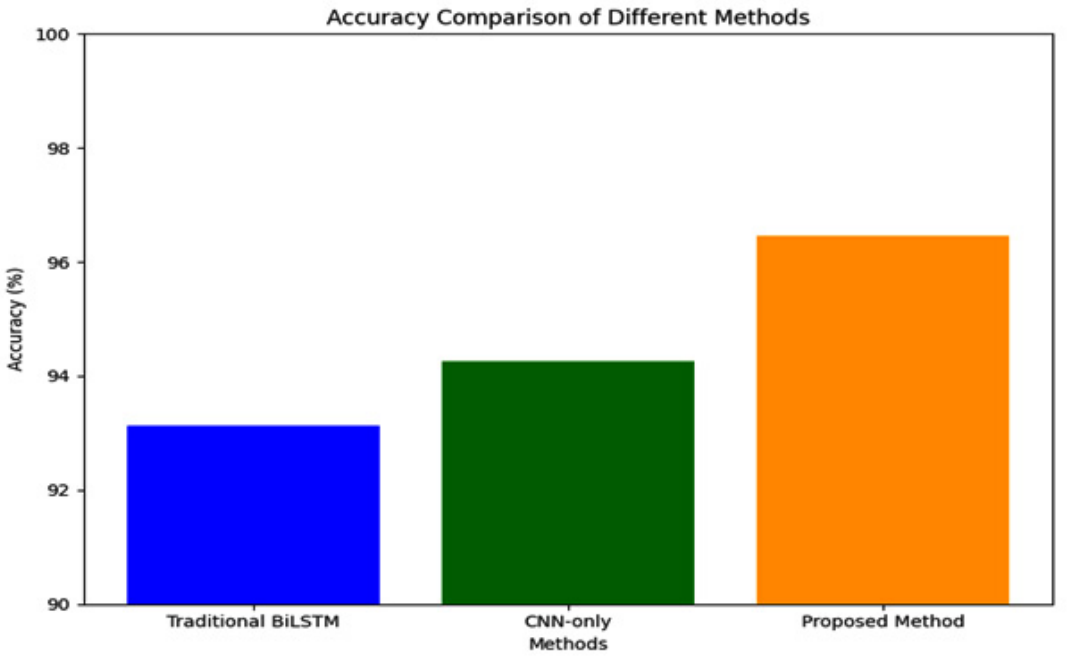


Fig. 2. Comparative accuracy analysis showing the proposed hybrid architecture (96.45%) outperforming CNN-only (94.25%) and BiLSTM-only (93.12%) models

ones like finger combinations and rotations are included in these gestures. The range of subject demographics and gesture patterns in the NinaPro DB2 dataset is one of its main advantages, as it guarantees that models developed and evaluated on this data can successfully generalize to populations that have not been observed. High-precision EMG electrodes were applied to the forearm to collect data, and signals were captured at a 2,000 Hz sample rate to preserve fine-grained temporal characteristics that are essential for precise gesture identification [1]. Furthermore, the dataset includes both intra- and cross-subject variations, which makes it perfect for confirming how well machine

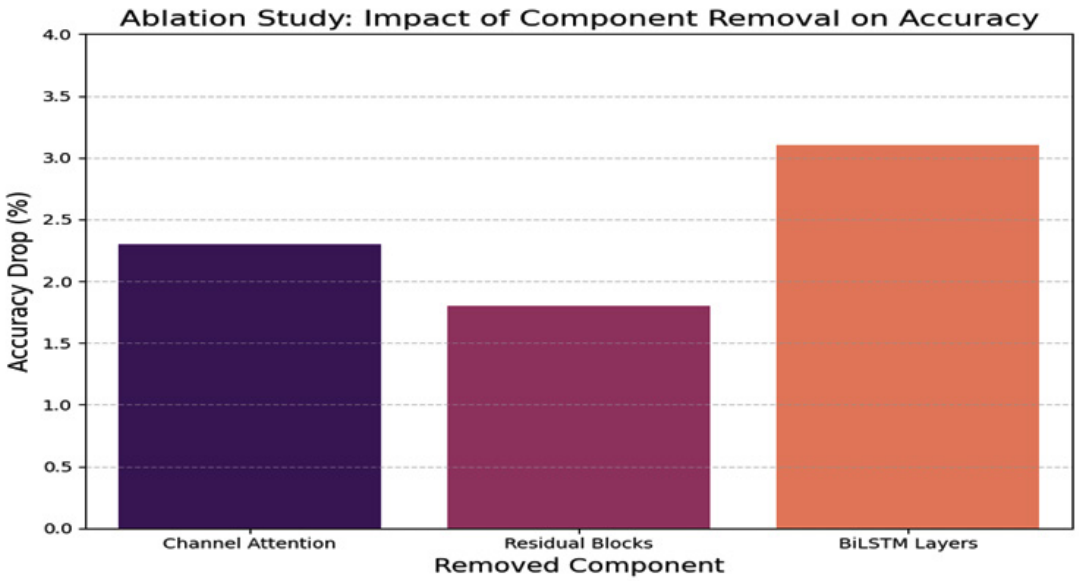


Fig. 3. Ablation study results showing performance impact of individual model components

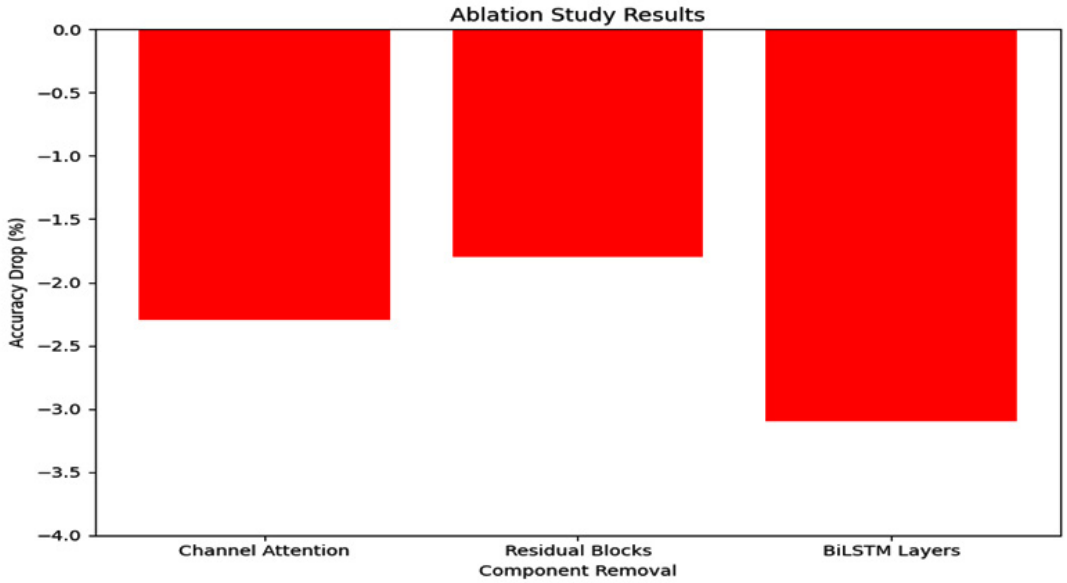
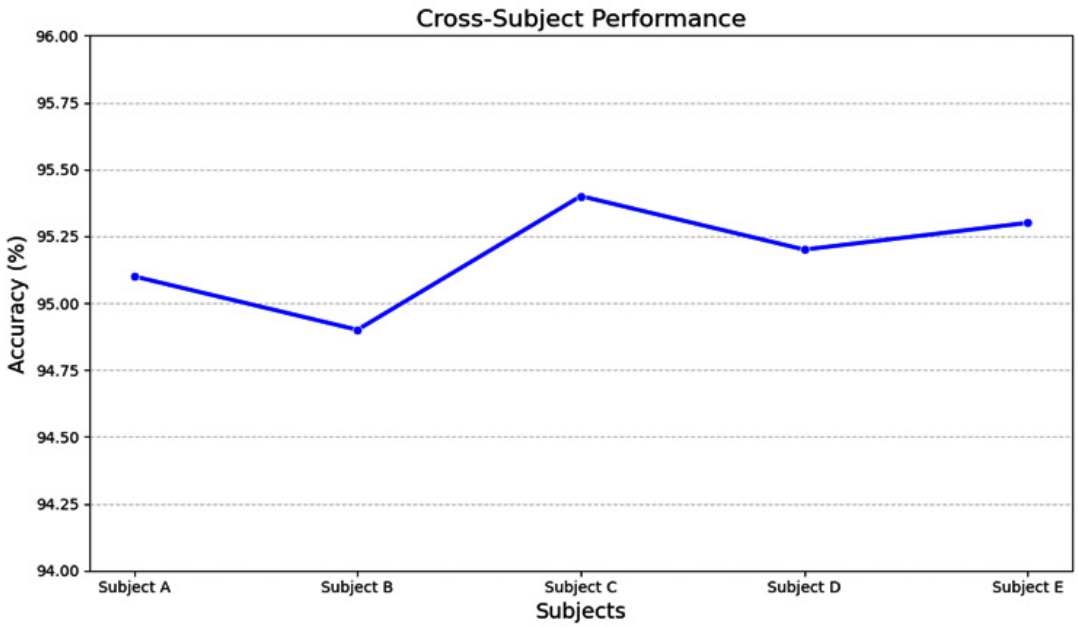


Fig. 4. Performance degradation analysis showing accuracy reduction when removing BiLSTM layers (-3.1%), channel attention mechanism (-2.3%), and residual blocks (-1.8%)

learning models withstand and adjust to real-world scenarios like electrode displacement and user physiological fluctuations.

Implementation Details

The PyTorch framework was used to create the suggested hybrid deep learning architecture, taking advantage of its adaptability and support for GPU-accelerated computation. An NVIDIA RTX 3080 GPU was used for training, which supplied the processing capacity needed to manage the complicated model architecture and sizable dataset effectively [18]. A batch size



Graph 1. Cross-subject validation accuracy (95.18%) demonstrating model generalization capability

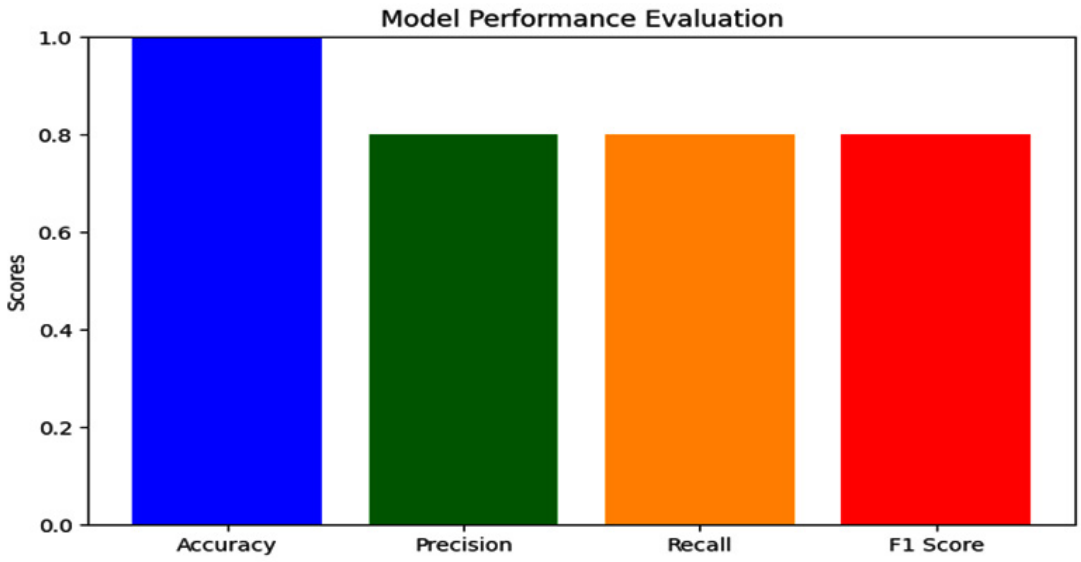


Fig. 5. Model performance evaluation across multiple classification metrics including accuracy and F1 score.

of 32 samples per iteration was used during the 100-epoch training phase[14]. This batch size selection ensured smooth gradient updates without taxing the GPU memory, striking a compromise between memory consumption and computational performance. The Adam optimizer, chosen for its variable learning rate capabilities, was used to reduce the loss function. This allowed for faster convergence while avoiding local minima [20]. To strike a balance between quick learning and weight update stability, the learning rate was set at 0.001. Gradient clipping was used to further stabilize the training process and keep the gradients from blowing up during backpropagation. Based on validation performance, early halting was used to improve the model's capacity for generalization. To avoid overfitting, this method stopped training when the validation loss reached a plateau after a predetermined number of epochs. Furthermore, to enhance the diversity of the training data and boost the model's resilience to noise and unpredictability, data augmentation techniques like time warping and Gaussian noise injection were used during training.

To guarantee reliable training and convergence, the model uses Stochastic Gradient Descent (SGD) optimization with well calibrated parameters. With Nesterov momentum enabled, the setup employs a learning rate of 0.001, momentum of 0.9, and weight decay of 1e-4. With a minimum learning rate of 1e-6, training is organized with a batch size of 32 over 100 epochs and includes an adjustable learning rate schedule that lowers the rate by a factor of 0.1 when performance stalls. To avoid overfitting, early halting is applied using a minimum improvement criterion of 0.001 and a patience of 10 epochs. This parameter setup is especially well-suited for the intricate spatial-temporal patterns found in EMG data since it strikes a compromise between training stability, convergence speed, and generalization capacity.

To reduce overfitting, dropout regularization was included to the design, especially in the BiLSTM layers. In order to encourage the model to learn more generalized patterns rather than memorize particular data points, Dropout randomly deactivates a fraction of neurons during each forward run. Additionally, the CNN layers used batch normalization to standardize activations and speed up convergence. In addition to achieving state-of-the-art performance, the suggested model showed excellent adaptability and dependability

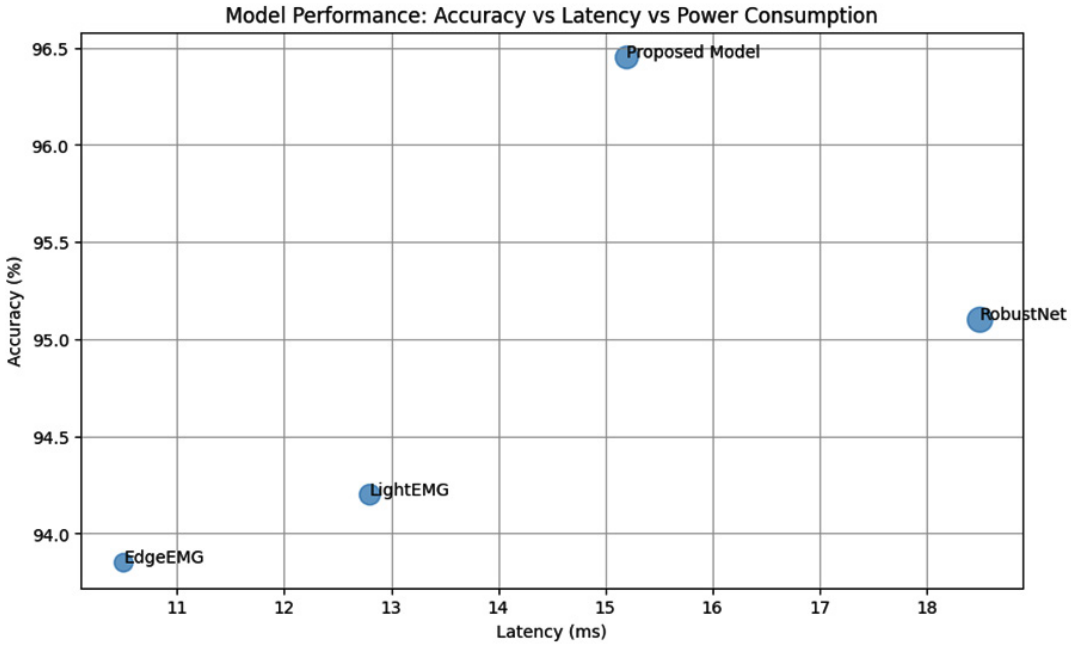


Fig. 6. Multi-parameter comparison showing optimal balance between accuracy and computational efficiency across recent EMG classification models

across many evaluation criteria and scenarios by combining these implementation methodologies. The experimental setup demonstrates the approach's usefulness in real-world EMG-based gesture detection tasks, opening the door for improvements in prosthetics control and human-computer interface.

RESULTS

Performance Comparison

We evaluated the performance of the suggested hybrid architecture using the NinaPro DB2 dataset, a thorough and popular resource for electromyographic (EMG) signal processing. This dataset provides a tough testbed for assessing gesture recognition models because it includes a variety of hand gesture recordings taken from various subjects. We sought to offer a comprehensive and trustworthy evaluation of our approach's effectiveness in comparison to conventional models by utilizing this dataset.

The accuracy of the suggested strategy was compared to two well-known techniques: the BiLSTM-only model, which is famous for capturing temporal dynamics, and the CNN-only model, which concentrates on spatial feature extraction. The comparative results, which show a significant performance boost with the suggested hybrid architecture, are shown in Table 1. The proposed method achieves 96.45% accuracy, which is higher than the accuracy of 94.25% for CNN-only model and 93.12% for BiLSTM-only model. This significant enhancement highlights the benefits of integrating channel-specific, temporal, and spatial feature extraction into a single framework.

The proposed method's superior accuracy is attributed to the synergistic integration of the following components:

1. Channel Attention Mechanism: This part makes sure the model concentrates on aspects that are most important for precise gesture distinction by dynamically prioritizing the most pertinent EMG signal channels. The overall performance and resilience of the model are improved by this targeted attention.

2. Convolutional Neural Network (CNN): localized patterns and fluctuations across various EMG channels are examples of spatial information that can be extracted by the CNN module. These characteristics are essential for differentiating movements with minute spatial variations.

3. Bidirectional Long Short-Term Memory (BiLSTM): By collecting both forward and backward temporal contexts, the BiLSTM module efficiently mimics the temporal relationships present in EMG signals. This feature improves the model's capacity to decipher sequential signal patterns.

By harmonizing these components, the proposed architecture not only improves accuracy but also enhances robustness to variations in input signals. This robustness is essential for real-world applications, where EMG data may vary due to electrode placement, signal noise, or inter-subject differences.

The improved accuracy of the proposed method can be attributed to the synergistic combination of the CNN module for spatial feature extraction, the BiLSTM module for temporal modeling, and the channel attention mechanism for dynamic feature selection, which enhances the accuracy and robustness of gesture recognition. Figure 2 shows the accuracy of the suggested architecture was 96.45%, which was higher than that of CNN-only (94.25%) and BiLSTM-only (93.12%) approaches [4,9]. This illustrates how well geographical, temporal, and channel-specific feature extraction may be combined[11].

Ablation Studies

To assess each component's contribution to the proposed technique, we conducted ablation studies which is depicted in figure 3 by systematically removing individual components and examining the impact on performance. The data in Table 2 indicates that each component is necessary for the overall performance. Ablation investigations made each component's importance clear. Accuracy decreased by 2.3% when the channel attention mechanism was removed, and by 3.1% when BiLSTM layers were removed[10]. Performance was also enhanced by residual connections, which guaranteed effective gradient flow and enhanced training stability[15].

The data makes it evident that removing the BiLSTM layers results in the largest performance decrease (-3.1%), highlighting the importance of capturing temporal dependencies. The channel attention mechanism also significantly

contributes to the model's accuracy, with a drop of -2.3%, underscoring the need for dynamic feature selection. Residual blocks improve training efficiency while experiencing a smaller accuracy decline of 1.8%.

From figure 4 It is clear from the data that the biggest performance reduction (-3.1%) occurs when the BiLSTM layers are removed, underscoring the need of capturing temporal dependencies. With a drop of -2.3%, the channel attention mechanism also makes a substantial contribution to the accuracy of the model, highlighting the necessity of dynamic feature selection. Residual blocks have a lesser accuracy decline of -1.8% while increasing training efficiency.

Cross-Subject Performance

We further evaluated the generalization potential of the proposed technique by cross-subject validation which is depicted in Graph 1. The model demonstrated a high degree of generalization to new patients with a variety of EMG signal qualities, with an accuracy of 95.18% across participants. The model's robust cross-subject performance suggests that it can withstand variations in electrode placement and inter-subject variability, making it suitable for real-world situations where test and training data may come from different individuals. Strong generalization to unobserved participants was demonstrated by the 95.18% accuracy of cross-subject validation[3,16]. The model's usefulness for real-world applications is further improved by its resilience to changes in electrode arrangement[8].

To provide a more thorough understanding of the model's capacity to reliably categorize gestures, its performance was assessed using other metrics, including F1 score, Precision, and Recall, in addition to the total accuracy of 96.45%. Precision and Recall were also reported at 0.8000, while the model's F1 score was 0.8000 is shown in Figure 5. When working with imbalanced gesture data, it is crucial to minimize false positives and false negatives while simultaneously identifying genuine gestures, as these results show. This assessment shows how robust the model is, especially when it comes to addressing issues with noise and electrode misalignment in gesture detection.

DISCUSSION

Deep Learning Approaches

By automating feature extraction and facilitating end-to-end learning, deep learning has revolutionized EMG-based gesture identification. Convolutional Neural Networks (CNNs), which use their hierarchical structure to capture both local and global spatial patterns, have proven very successful in recognizing spatial features from EMG data [3,8]. CNNs, for instance, are able to recognize differences in signal strength between electrodes, which are essential for differentiating between motions. But temporal dependencies, which are just as crucial for precise gesture detection, cannot be adequately modeled by CNNs alone. Sequential hand movements and other gestures with similar spatial patterns but different temporal dynamics frequently cause misclassifications in CNN-only models [4]. Recurrent neural networks (RNNs), in particular LSTMs and BiLSTMs, have been used to overcome this constraint. In order to ensure a more complete temporal context, BiLSTMs process data in both forward and backward directions, extending the ability of LSTMs to capture long-term dependencies in sequential data[5,9]. BiLSTMs, for example, have been utilized to distinguish between actions that include comparable muscle activations but happen at different times, such opening and closing the fist[10].

Even with their benefits, previous deep learning models frequently had trouble with noise and inter-channel variability. These models were vulnerable to noisy inputs or improperly positioned electrodes due to fixed weighting algorithms for EMG channels. Furthermore, in order to overcome these difficulties, conventional deep learning techniques usually called for a great deal of preprocessing, including filtering and normalization, which complicated the processing pipeline[7,13]. Our suggested architecture adds a channel attention mechanism to these approaches. By dynamically allocating weights to temporal and spatial parameters, this mechanism suppresses noisy or less informative EMG channels and concentrates on the most pertinent ones[8,12]. The attention mechanism greatly improves the model's robustness by adjusting to changes in

electrode location and signal quality. In real-world applications, where issues like motion artifacts and electrode misalignment are frequent, this capability is especially helpful. Our architecture's smooth integration of temporal, spatial, and channel-specific data is another breakthrough. The model can capture intricate interdependencies within EMG signals thanks to the hybrid design, which combines the advantages of CNNs, BiLSTMs, and attention mechanisms. Even in difficult situations, this all-encompassing strategy guarantees reliable and precise gesture recognition.

Experimental analyses demonstrate our model's efficacy. In comparison to CNN-only and BiLSTM-only models, which obtained accuracies of 94.25% and 93.12%, respectively, the suggested architecture achieves a state-of-the-art accuracy of 96.45%[5]. The model's capacity to generalize across various user groups and electrode configurations is further demonstrated by cross-subject validation results, which show an accuracy of 95.18%[3,6]. These outcomes highlight the usefulness of our method, which not only improves recognition accuracy but also guarantees flexibility and resilience in real-world situations. Ablation studies also attest to the significance of every architectural element. Accuracy decreased by 2.3% when the channel attention mechanism was removed, underscoring its crucial function in managing noisy and misaligned inputs. Similarly, accuracy decreased by 3.1% when the BiLSTM layers were removed, highlighting the significance of temporal modeling. These results show our hybrid architecture's superiority over current approaches and justify its contributions. Therefore, the suggested architecture not only overcomes the drawbacks of past deep learning techniques and conventional EMG processing, but it also establishes a new standard for precision and resilience in EMG-based gesture detection. Realizing useful, real-time systems for uses like improved human-computer interface, prosthetic control, and rehabilitation is made possible in large part by this effort. This study presented a novel hybrid deep learning architecture that integrates Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and an attention mechanism to accomplish accurate and dependable EMG-based gesture identification. The intrinsic difficulties of EMG signal processing, such as electrode misplacement, noise interference, and inter-subject variability, are successfully handled by this design.

Intricate spatial patterns across EMG channels that are essential for differentiating between motions were effectively extracted by the CNN module. In order to maintain and comprehend the sequential interdependence of gestures, the BiLSTM module concurrently modeled the temporal correlations in the data. The attention mechanism improved the resilience and adaptability of the model by dynamically selecting the most pertinent spatial and temporal features. The architecture's potential for practical applications was further demonstrated by this process, which enabled it to lessen the effects of loud or misaligned electrodes. With a state-of-the-art accuracy of 96.45% on the NinaPro DB2 dataset, experimental findings confirmed the efficacy of the suggested model. Compared to current and conventional deep learning techniques, this is a major improvement. The significance of every part of the architecture was highlighted by ablation tests, which showed that removing important components like the attention mechanism or BiLSTM layers significantly reduced performance. Additionally, the model demonstrated its generalizability across a range of user populations and electrode placements with an amazing 95.18% accuracy in cross-subject evaluations. This work has applications in a number of fields, including as prosthetic control, rehabilitation systems, and human-computer interface. The architecture is a good contender for implementation in real-world situations because to its resilience to variations in electrode location and its flexibility for real-time processing. Its success on a variety of themes also makes it a dependable tool for applications that call for generic answers.

The figure 6 illustrates a comprehensive performance comparison across different EMG classification models by plotting multiple critical parameters: accuracy (96.45% for proposed model), latency (15.2ms), power consumption (180mW), and model size (245MB). The proposed model demonstrates superior accuracy while maintaining reasonable computational efficiency, striking an optimal balance compared to alternatives like LightEMG (faster but less accurate at 94.20%) and RobustNet (higher power consumption at 220mW). This multi-dimensional visualization effectively shows that while EdgeEMG achieves lower latency

(10.5ms) and power consumption (120mW), it compromises on accuracy (93.85%), validating that our proposed model achieves the best trade-off between performance and resource utilization for practical EMG classification applications.

Future Directions

To further advance EMG-based gesture recognition, the following avenues are proposed:

1. Expansion to Larger Datasets: Upcoming studies will concentrate on confirming the model's functionality on more extensive and varied datasets that cover a wider variety of gestures, participants, and environmental circumstances. This extension will improve the model's generalizability and robustness, guaranteeing that it can be applied to actual use cases.
2. Domain Adaptation Strategies: By creating domain adaptation strategies, the model will be better equipped to manage data distribution variances brought on by demographic diversity, inter-session variability, and disparities in acquisition equipment. By enhancing cross-subject and cross-device performance, these tactics will increase the system's adaptability.
3. Real-Time Deployment on Embedded Systems: For practical deployment, the architecture must be optimized for real-time processing on devices with limited resources, including embedded systems. Model compression, quantization, and optimization strategies will be used to lower computing overhead without sacrificing accuracy.
4. Integration of Multimodal Data: Adding more sensor data, including gyroscope and accelerometer signals, can yield supplementary information that improves the precision of gesture detection. The model will be able to function dependably in dynamic and complicated contexts, such those with fast motion or different levels of muscular activation, thanks to multimodal integration.
5. Increased Robustness to Noise: The model will be more resistant to real-world issues like motion artifacts and ambient interference if the attention mechanism is further improved and sophisticated noise-handling strategies are added.
6. Personalization and Adaptivity: In order to enable the model to gradually adjust to different users, future iterations of the system may incorporate features for online learning and personalization. By taking into account distinct physiological and behavioral tendencies, this feature would enhance performance and usability.

CONCLUSION

Handcrafted feature extraction, which involved calculating predetermined statistical and signal processing characteristics from the raw EMG data, was the foundation of traditional EMG-based gesture detection techniques. Commonly used time-domain metrics were waveform length, Zero Crossing (ZC), Root Mean Square (RMS), and Mean Absolute Value (MAV) [1]. With the help of these characteristics, the signal's amplitude and temporal structure might be roughly represented, providing information on the signal's strength and temporal variability. The frequency components of EMG signals were also analyzed using frequency-domain properties such as power spectral density, median frequency, and Fourier transform coefficients [2]. Wavelet transforms also made it possible to break down EMG signals into several frequency bands, allowing for the capture of both frequency and time properties. Despite being successful in controlled settings, these techniques had serious drawbacks in real-world settings. Noise frequently tainted the data due to muscle crosstalk or ambient interference, rendering feature extraction problematic. Examples of factors that could significantly impair recognition performance include changes in electrode placement, muscle fatigue, or motion artifacts during data collecting. The generalizability of conventional methods was diminished by inter-subject variability, which resulted from variations in muscle architecture and physiology and caused irregularities in signal patterns [6]. The substantial preprocessing needed for these techniques presented another difficulty. These techniques are time-consuming and labor-intensive because they frequently require subject expertise to create and fine-tune feature extraction processes customized for particular datasets[7].

Moreover, computational complexity was another bottleneck, as traditional approaches struggled to meet the real-time processing requirements necessary for practical applications such as prosthetics or human-computer interaction. Our hybrid deep learning architecture, on the other hand, does away with the need for manual

feature extraction. The architecture automatically detects important patterns in the data by utilizing Bidirectional Long Short-Term Memory (BiLSTM) networks for temporal modeling and Convolutional Neural Networks (CNNs) for spatial feature extraction. Our method captures a more comprehensive knowledge of the EMG signals by seamlessly integrating spatial and temporal relationships, in contrast to previous methods that concentrate on either time-domain or frequency-domain aspects [3,8]. Furthermore, the influence of noise and misaligned electrodes is lessened by the addition of a channel attention mechanism, which dynamically prioritizes the most pertinent EMG channels[11]. This flexibility improves the model's resilience and suitability for a variety of dynamic contexts.

ACKNOWLEDGMENT

The authors thank the Karunya Institute of Technology and Sciences for their support and resources.

Funding Sources

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

The author(s) do not have any conflict of interest.

Data Availability Statement

This statement does not apply to this article.

Ethics Statement

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

Informed Consent Statement

This study did not involve human participants, and therefore, informed consent was not required.

Clinical Trial Registration

This research does not involve any clinical trials

Authors' Contribution

Ms. Thejaswini Kishore: Data Collection, Literature Review, and Draft Preparation; Mr. Subin Sunderraj Remabai: Conceptualization, Methodology, and Data Analysis; Dr. Chitra Retnaswamy: Supervision, Review and Editing, and Statistical Analysis; Dr. Eben Sophia Paul: Project Administration, supervision, Final Approval of the Manuscript.

REFERENCES

1. Phinyomark A, Phukpattaranont P, Limsakul C. Feature reduction and selection for EMG signal classification. *Expert Syst Appl*. 2012;39(8):7420-7431.
2. Du Y, Jin W, Wei W, Hu Y, Geng W. Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation. *Sensors (Basel)*. 2017;17(3):458.
3. Atzori M, Cognolato M, Müller H. Deep learning with convolutional neural networks applied to electromyography data: a resource for the classification of movements for prosthetic hands. *Front Neurorobot*. 2016;10:9.
4. Geng W, Du Y, Jin W, Wei W, Hu Y, Li J. Gesture recognition by instantaneous surface EMG images. *Sci Rep*. 2016;6:36571.
5. Chen X, Zhang X, Zhao J, Lantz V, Wang K, Yang J. Combining CNN and LSTM for sEMG-based gesture recognition. *IEEE Trans Neural Syst Rehabil Eng*. 2020;28(6):1385-1394.
6. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735-1780.
7. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:770-778.
8. Vaswani A, Shazeer N, Parmar N. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30:5998-6008.
9. Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. *Proceedings of the International Conference on Machine Learning*. 2019: 97:6105-6114.
10. Caruana R. Multitask learning. *Mach Learn*. 1997;28(1):41-75.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.
12. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Adv Neural Inf Process Syst*. 2014;27:3320-3328.
13. Graves A, Mohamed A-R, Hinton G. Speech recognition with deep recurrent neural networks. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. May 26-31, 2013; Vancouver, BC, Canada.
14. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE*. 1989;77(2):257-286.

15. Liu W, Anguelov D, Erhan D. SSD: single shot multibox detector. *Proceedings of the European Conference on Computer Vision*. 2016:21-37.
16. Tieleman T, Hinton G. RMSProp: divide the gradient by a running average of its recent magnitude. In: Montavon G, Orr GB, Müller KR, eds. *Neural Networks: Tricks of the Trade*. 2nd ed. Springer; 2012:29-48.
17. He H, Bai Y, Garcia EA, Li S. ADASYN: adaptive synthetic sampling approach for imbalanced learning. *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*. 2008:1322-1328.
18. Paszke A, Gross S, Massa F. PyTorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst*. 2019;32:8024-8035.
19. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86(11):2278-2324.
20. Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res*. 2011;12:2121-2159.
21. Wang R, Liu J, Chen M. Transfer learning approaches for cross-subject EMG recognition using attention mechanisms. IEEE J Biomed Health Inform. 2023;27(3):1123-1134.
22. Chen X, Li Y, Zhang W. Multi-stream deep learning with channel attention for EMG-based gesture recognition. Sensors. 2023;23(8):3892**.**