ICIC Express Letters
Part B: Applications
Volume **15**, Number **12**, December **2024**

ICIC International ©2024 ISSN 2185-2766
pp. 1311–1318

SEMG-BASED MUSCULAR MOVEMENT RECOGNITION FOR HAND PROSTHESIS USING CNN-LSTM

Sakorn Mekruksavanich[1], Olarik Surinta[2] and Anuchit Jitpattanakul[3,4,*]

[1]Department of Computer Engineering
School of Information and Communication Technology
University of Phayao
19 Moo 2, Tambon Mae Ka, Amphur Mueang, Phayao 56000, Thailand
sakorn.me@up.ac.th

[2]Multi-agent Intelligent Simulation Laboratory (MISL) Research Unit
Department of Information Technology
Faculty of Informatics
Mahasarakham University
Khamriang Sub-District, Kantarawichai District, Mahasarakham 44150, Thailand
olarik.s@msu.ac.th

[3]Department of Mathematics
Faculty of Applied Science
[4]Intelligent and Nonlinear Dynamic Innovations Research Center
Science and Technology Research Institute
King Mongkut's University of Technology North Bangkok
1518 Pracharat 1 Road, Wongsawang, Bangsue, Bangkok 10800, Thailand
*Corresponding author: anuchit.j@sci.kmutnb.ac.th

Received February 2024; accepted April 2024

Abstract. *Recent advancements in sensing technology have enabled the development of more sophisticated assistive devices. Real-time myoelectric interfaces use surface electromyography (sEMG) to capture muscular activities. These signals can be utilized to create myoelectric prosthetic hands for individuals with physical disabilities. Accurate classification of the acquired sEMG signals is critical for effectively controlling external devices. This study introduces deep learning techniques for classifying muscular activities based on sEMG data. The methodology involves data acquisition, pre-processing, generation, and model training/testing. The Ninapro-DB1 dataset of sEMG signals from 27 healthy participants performing 53 hand motions was utilized. Multiple experiments compared various deep learning architectures – convolutional neural networks (CNN), long short-term memory networks (LSTM), bidirectional LSTM (BiLSTM), gated recurrent units (GRU), and bidirectional GRU (BiGRU). A novel hybrid CNN-LSTM model is proposed to automatically extract spatial and temporal features from the raw sEMG data. Experimental results demonstrate the hybrid model achieves 99.27% accuracy and F1-score, outperforming other deep learning models. Therefore, this study shows deep learning, specifically a CNN-LSTM hybrid, can effectively classify muscle movements from sEMG data for assistive technology applications.*
Keywords: Hand gesture recognition, Surface electromyography, Deep learning network, Classification, Wearable sensor

1. **Introduction.** The integration of wearable technologies for sensing has revolutionized the control of external prosthetic devices. This is achieved by analyzing surface electromyography (sEMG) signals, which correspond to different muscular actions [1]. This advancement has significantly improved quality of life for individuals with physical disabilities. Peripheral devices can be categorized as wearable or non-wearable, depending

DOI: 10.24507/icicelb.15.12.1311

on the specific needs of individuals with disabilities. Precise real-time classification of muscle movements is essential for accurately controlling external instruments and ensuring optimal functionality. Significant advancements have been made in prosthetic hand technology, enabling these devices to meet the fundamental needs of individuals who have undergone limb amputation. However, most existing approaches to prosthetic hand control fall into two main paradigms: non-pattern recognition and pattern recognition. Traditional non-pattern recognition techniques typically involve on/off control, threshold management, and proportional regulation [2]. In contrast, pattern recognition approaches based on sEMG have overcome non-pattern comprehension limitations and greatly improved prosthetic hand dexterity. In sEMG pattern identification, multi-dimensional features are extracted from sEMG signals, moving beyond reliance on EMG amplitude alone, as in non-pattern techniques. A sophisticated prosthetic hand design incorporates predefined movement patterns and trajectories. The control algorithm requires parameters including movement style, orientation, and kinematics. The sEMG pattern contains valuable information about the intended motion. By successfully classifying the sEMG signals for the desired movement, the prosthetic hand can be directed to perform the appropriate action. Implementing pattern recognition methods can therefore significantly improve control of prosthetic hands for individuals with limb loss.

The identification of sEMG patterns can be achieved through two main processing approaches: machine learning (ML) and deep learning (DL). ML uses feature engineering to acquire knowledge and perform tasks by automatically modeling input data. However, traditional ML algorithms have limitations when dealing with raw, inconsistent, noisy, high-dimensional and abstract data [3]. Several studies have shown the importance of hand-engineered features for effective sEMG pattern recognition. In contrast, DL, a subtype of ML, uses feature learning as its core principle. A key characteristic of DL is its multi-layered model architecture. Using multiple hidden layers enables extraction of high-level feature representations [4, 5]. Computational models perform feature extraction and model-building together, allowing automatic feature learning without manual effort. This makes DL convenient for complex gesture recognition [6]. In recent years, DL techniques have been increasingly adopted for analyzing and interpreting physiological signals [7]. This growing trend has sparked interest in exploring DL's potential for sEMG pattern recognition. Overall, DL shows promise as an automated approach to extracting discriminative features from complex sEMG datasets [8, 9].

This research presents an innovative hybrid DL approach called CNN-LSTM for recognizing muscle movements using sEMG sensor data. The proposed model can automatically extract distinctive features of muscle movements from raw, unprocessed sEMG data by leveraging 1-dimensional convolution layers and long short-term memory (LSTM) units. To evaluate the effectiveness of the detection model, comprehensive investigations were conducted under various conditions, employing deep learning networks such as convolutional neural networks (CNN), LSTM, bidirectional LSTM (BiLSTM), gated recurrent unit (GRU), and bidirectional GRU (BiGRU). The Ninapro-DB1 dataset, a widely available benchmark dataset, was utilized for training and testing these networks.

This paper is organized as follows. Section 2 offers a comprehensive review of the pertinent literature. Section 3 presents the framework utilized in this study for muscle movement recognition. The results of our analysis are presented and discussed in Section 4. Finally, Section 5 concludes the paper with closing remarks and suggestions for future research directions.

2. **Related Works.** In recent years, sEMG-based gesture recognition has transitioned from traditional ML to DL methodologies. DL integrates feature extraction into model architectures, eliminating manual engineering. DL structures can autonomously acquire effective features during training, reducing reliance on predefined extraction steps.

Previously, feature engineering was crucial for ML techniques with sEMG. Experts proposed various time-domain, frequency-domain, and time-frequency feature sets [10, 11]. Studies show conventional ML attained satisfactory accuracy for limited gestures using multichannel sEMG [12]. However, challenges emerge as complexity increases, including more motions or intricate actions. Atzori et al. [13] evaluated features on their Ninapro databases using ML. The highest accuracy for 50 gestures in DB2 was 75.27%. The DL shows promise for automated feature learning with sEMG, overcoming limitations in traditional ML approaches relying on manual feature engineering. Transitioning to DL can improve performance as gesture recognition tasks become more complex.

Recently, CNNs have been increasingly utilized for developing sEMG-based gesture detection systems [14]. Despite successes using raw HD-sEMG data [8], CNN effectiveness remains suboptimal for sparse multichannel sEMG signals. For example, on the Ninapro-DB2 database, Atzori et al. [15] achieved only 60.3% accuracy by inputting raw sEMG directly into a CNN. Similarly, Zhai et al. [9] obtained 78.7% accuracy by using sEMG spectrograms as CNN inputs. In contrast, Wei et al. [16] attained 83.7% accuracy by combining multiple CNN streams, each fed with selective features, within a multi-view deep learning framework. However, they reported a strong dependence between feature sets and datasets. One feature set yielded 83.7% accuracy on DB1 but only 8.1% on DB2. Therefore, while CNNs show promise for sEMG-based gesture detection, their performance depends heavily on factors like sEMG signal density and appropriate input features. Further research is required to optimize CNNs for sparse multichannel sEMG across different datasets. Careful feature selection as CNN inputs may improve effectiveness.

Despite extensive research on understanding movements using sEMG, opportunities remain for further advancement, especially concerning more comprehensive databases with greater diversity of individuals and motions. To enable real-time functionality for prosthetic hands, it is crucial to ensure accurate and prompt recognition and categorization of muscle movements while also minimizing computational costs and time. Expanding databases with diverse subjects and motions would allow development and testing of more robust sEMG-based movement classification systems. Meanwhile, optimizations in speed and efficiency are needed so these systems can analyze and classify complex muscle patterns in real time for seamless prosthetic control. Though progress has been made, there is still work to be done to advance sEMG-based movement recognition to a level that supports responsive, intuitive control of multifunctional prosthetic hands.

3. **Methodology.** This section outlines the methodology used in our investigation to evaluate the usefulness of sEMG data for identifying muscular movements. Figure 1 provides an overview of the muscle movement identification process utilized in this study. First, we acquired muscular activity data from the Ninapro-DB1 dataset. Next, we preprocessed the data to remove extraneous signals and standardize the information. We then applied a windowing technique to segmenting the data into relevant details. The segmented data was input into a deep learning network, which generated predicted labels for each segment. Finally, we evaluated the performance of the model using metrics like accuracy, loss, and F1-score.

3.1. **Ninapro-DB1 dataset.** The current study utilized the Ninapro-DB1 dataset [17], a publicly available repository of muscular activity data collected using OttoBock sEMG electrodes. The dataset was obtained using ten OttoBock MyoBock 13E200-50 electrodes that enabled amplification, bandpass filtering, and Root-Mean-Square (RMS) rectification of the raw sEMG signals. The Ninapro-DB1 dataset contains accurately categorized data comprising 10 repetitions of 52 distinct hand movements and 1 resting position. The data was collected from 27 healthy participants who were instructed to perform a series of 52 motions with their right hand, referred to as Exercises A, B, and C. Each motion was

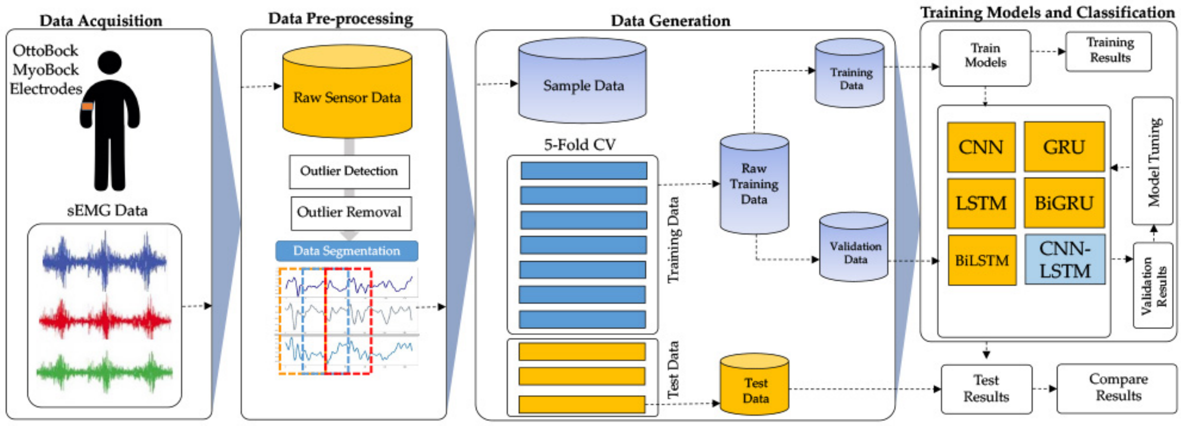

FIGURE 1. The muscular movement recognition framework used in this work

repeated for 5 seconds, followed by a 3-second rest period. The set of physical activities consists of the following motions.

- Exercise A consists of simple finger movements like stretching out the fingers and curling them into a fist.
- Exercise B covers 17 hand motions including 8 static and dynamic hand positions such as pinching, pointing, and grasping. It also includes 9 basic wrist actions like bending, extending, and pivoting the wrist.
- Exercise C encompasses 23 unique grasping activities and practical hand movements. These include making a fist, picking up objects with different shapes, and functional tasks like turning a key or opening a bottle.

3.2. **Data pre-processing.** Due to the large size of the Ninapro-DB1 dataset, it is impractical to use it for classification purposes due to the required processing time. Therefore, to train and test DL models effectively, applying techniques such as outlier elimination, data manipulation, and pre-processing to generating a refined dataset is necessary. These steps ensure the dataset is optimized and suitable for analysis and model development.

The methodology employed in this study involves utilizing outlier identification and elimination techniques to obtain the necessary data. A matching approach is employed to identify outliers to establish correspondences between the initial and modified labels. If any value of the initial labels aligns with the updated labels, the corresponding sEMG signals from the 10 electrodes and the previous label values remain unchanged. However, if the stimulus does not match the expected response, the corresponding sEMG signals and stimulus information are considered anomalous and excluded. This approach effectively removes irrelevant variables and outliers, allowing for the extraction of relevant sEMG and action label information for further analysis.

3.3. **The proposed DL model.** The present study introduces a novel DL architecture called CNN-LSTM, which combines CNN and LSTM to improve recognition accuracy. The CNN-LSTM architecture consists of two convolutional layers and a single LSTM layer. The architecture of the CNN-LSTM is depicted in Figure 2.

Convolutional layers are effective at extracting useful features from sensor data. This study uses 1D convolutional layers (ConV1D) to efficiently extract features. As shown in Figure 2, the sensor data has a vertical axis for time sequence and a horizontal axis for multi-channel features from multiple sensors. 2D convolution can compromise channel integrity when there are many sensors, since it is localized. 1D convolution operates along the sensor channels, combining all of them. So 1D convolution is preferable for extracting spatial features across all sensors, instead of standard 2D convolution. In 1D convolution,

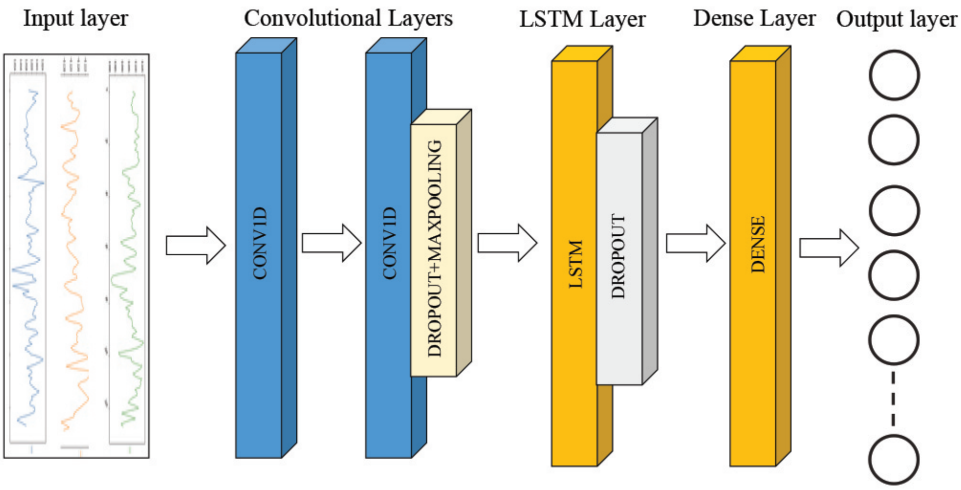

FIGURE 2. The proposed CNN-LSTM architecture

the input is convolved with each filter and then passed through a nonlinear activation function. The activation function is mathematically defined as

$$X_j = f\left(\sum_{i=1}^{n} \left(W^i \cdot x_j + b^i\right)\right) \tag{1}$$

where X_j is the activated output data, W^i is the weight of the ith filter, x_j is the sensing data convolved with W^i, b^i is the bias of the ith filter, n is the number of filters, and f is a nonlinear activation function.

LSTM networks are an extension of deep recurrent neural networks (RNNs) [18] that address the problems of vanishing and exploding gradients [19]. The LSTM network operates as follows: 1) Input data $X = \{x_0, x_1, x_2, \ldots, x_t, x_{t+1}, \ldots\}$ is transformed into a hidden layer $H = \{h_0, h_1, h_2, \ldots, h_t, h_{t+1}, \ldots\}$ using matrix transformations, and 2) an output layer $Y = \{y_0, y_1, y_2, \ldots, y_t, y_{t+1}, \ldots\}$ is activated using an activation function. LSTMs overcome issues with gradients that can prevent standard RNNs from learning long-term dependencies in sequence data as illustrated in Figure 3.

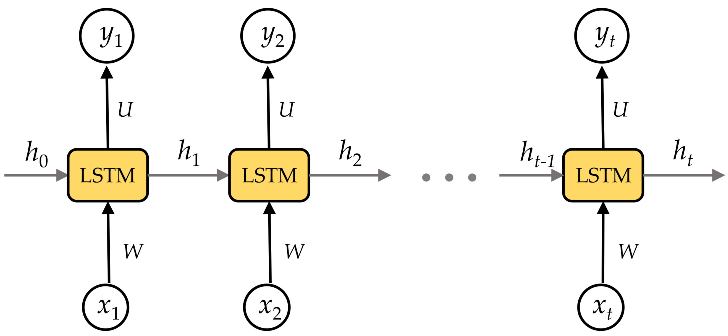

FIGURE 3. The unfold structure of one-layer baseline LSTM

This study proposes using a CNN-LSTM model to improve activity identification. The CNN-LSTM has two convolutional layers and one LSTM layer. The CNN-LSTM uses ConV1D to extract deep spatial features from the sensor data, with two layers having 3×1 kernels. The first and second ConV1D layers have 128 and 64 filters, respectively. The model uses rectified linear units for activation. After the second ConV1D layer, dropout is applied. A 128-unit LSTM layer comes before a dropout layer to extract temporal features. Finally, a SoftMax function is applied on the LSTM output using a fully connected dense layer. The CNN extracts spatial features from the sensor data, and the LSTM captures time dependencies, aiming to improve identification performance.

3.4. **Hyperparameter tuning by Bayesian optimization.** The present study utilized Bayesian optimization to automatically determine the parameters for the CNN-LSTM network. Bayesian optimization has proven to be a practical approach for addressing computationally demanding tasks that involve finding extrema. This method is beneficial for solving functions that lack a closed-form expression. Additionally, it can be applied to resource-intensive computations, situations where evaluating derivatives is challenging, or when dealing with non-convex functions. The objective of the optimization is to identify the highest possible value of an expression f at a given sampling point, even when the functions properties are not readily observable.

$$x^+ = \arg\max_{x \in A} f(x) \tag{2}$$

where A denotes the search space of x.

4. **Experiments and Findings.** This section provides an overview of the experimental setup and results used to evaluate the effectiveness of DL models in recognizing muscular movements from sEMG signals.

4.1. **Experiments.** The present study used the Google Colab Pro system equipped with a Tesla-V100 for all experimental procedures. Implementing the Python programming language involved Python, TensorFlow, Keras, Scikit-Learn, Numpy, and Pandas libraries.

Experimental investigations in this study were conducted on three different scenarios, each involving various muscle movements. The recommended CNN-LSTM network was trained and tested using separate sEMG data for each scenario, as detailed in Table 1.

TABLE 1. A list of scenarios employed in this work

Scenario	**Description**
I	Use sEMGs of exercise A (12 basic movements of the fingers)
II	Use sEMGs of exercise B (8 isometric and isotonic hand configurations and 9 fundamental movements of the wrist)
III	Use sEMGs of exercise C (23 grasping and functional movements)

4.2. **Experimental results.** In this study, the Ninapro-DB1 dataset was utilized for all experiments. DL models were developed and trained using this dataset, and their performance was assessed using the 5-fold cross-validation approach. The study's findings demonstrated the effectiveness of the CNN-LSTM algorithm and five other baseline DL models in terms of their ability to accurately recognize movements across the three scenarios specified in Tables 2, 3, and 4.

TABLE 2. Performance effectiveness of the proposed CNN-LSTM model and baseline DL models using sEMG data from Scenario I

Model	**Recognition effectiveness**		
	Accuracy	**Loss**	**F1-score**
CNN	90.47%(±0.74%)	0.33(±0.02)	90.42%(±0.77%)
LSTM	73.89%(±4.73%)	0.82(±0.15)	73.89%(±4.68%)
BiLSTM	94.49%(±1.20%)	0.18(±0.04)	94.47%(±1.20%)
GRU	92.06%(±1.41%)	0.27(±0.04)	92.01%(±1.41%)
BiGRU	95.59%(±0.14%)	0.15(±0.01)	95.56%(±0.14%)
CNN-LSTM	99.27%(±0.13%)	0.02(±0.01)	99.27%(±0.13%)

Table 3. Performance effectiveness of the proposed CNN-LSTM model and baseline DL models using sEMG data from Scenario II

Model	Recognition effectiveness		
	Accuracy	**Loss**	**F1-score**
CNN	85.79%(±0.76%)	0.48(±0.02)	85.85%(±0.76%)
LSTM	86.95%(±2.06%)	0.43(±0.07)	87.05%(±2.04%)
BiLSTM	95.70%(±0.88%)	0.15(±0.03)	95.70%(±0.88%)
GRU	92.36%(±2.65%)	0.28(±0.10)	92.37%(±2.64%)
BiGRU	96.18%(±0.63%)	0.14(±0.02)	96.19%(±0.63%)
CNN-LSTM	99.25%(±0.17%)	0.03(±0.01)	99.25%(±0.17%)

Table 4. Performance effectiveness of the proposed CNN-LSTM model and baseline DL models using sEMG data from Scenario III

Model	Recognition effectiveness		
	Accuracy	**Loss**	**F1-score**
CNN	70.18%(±1.07%)	0.99(±0.03)	70.05%(±1.06%)
LSTM	87.71%(±0.44%)	0.43(±0.02)	87.65%(±0.44%)
BiLSTM	94.85%(±1.10%)	0.19(±0.03)	94.82%(±1.09%)
GRU	90.72%(±0.79%)	0.34(±0.03)	90.67%(±0.79%)
BiGRU	95.27%(±0.41%)	0.18(±0.01)	95.23%(±0.41%)
CNN-LSTM	98.49%(±0.35%)	0.03(±0.00)	98.48%(±0.35%)

The CNN-LSTM model, as outlined in Table 2, was trained and evaluated using Scenario I. The investigation results reveal that the CNN-LSTM model achieved an impressive accuracy and F1-score of 99.27% in accurately classifying basic finger motions based on sEMG data, in line with the proposed methodology.

Table 3 presents the results of our investigations using surface sEMG data from Scenario II. The dataset comprises 8 isometric and isotonic hand arrangements and 9 fundamental wrist actions. The findings demonstrate that the CNN-LSTM model, developed in this study, outperforms other DL models used as baselines, achieving an impressive accuracy rate of 99.25% and the highest F1-score of 99.25%.

The investigations were conducted using sEMG data from Scenario III, including 23 grasping and practical actions, as shown in Table 4. The experimental findings reveal that the CNN-LSTM model proposed in this study demonstrates superior performance compared to other DL models used as benchmarks in the experiment. The suggested model achieved a remarkable accuracy of 98.49% and a maximum F1-score of 98.48%.

5. **Conclusion and Future Works.** This research uses sEMG sensor data and DL methodologies to identify muscle movements. The study evaluated the recognition capabilities of a hybrid CNN-LSTM model and five baseline DL models (CNN, LSTM, BiLSTM, GRU, and BiGRU) using the Ninapro-DB1 benchmark sEMG dataset. Multiple investigations were conducted across three distinct scenarios. The findings demonstrate that the CNN-LSTM approach developed in this study outperforms other standard models, achieving the highest accuracy and F1-score of 99.27%.

To enhance the quality and effectiveness of future research, it is recommended to gather sufficient datasets encompassing various categories of disabilities among amputee participants.

Acknowledgment. This project was funded by Thailand Science Research and Innovation Fund; the University of Phayao (Grant No. FF67-UoE-Sakorn); National Science, Research and Innovation Fund (NSRF); and King Mongkut's University of Technology North Bangkok with Contract no. KMUTNB-FF-66-07.

REFERENCES

[1] W. Budiharto, Low cost prosthetic hand based on 3-lead muscle/electromyography sensor and 1 channel EEG, *ICIC Express Letters*, vol.13, no.1, pp.77-82, 2019.
[2] H. Liu, D. Yang, S. Fan and H. Cai, On the development of intrinsically-actuated, multisensory dexterous robotic hands, *Robomech Journal*, vol.3, 2016.
[3] J. Wang, Y. Chen, S. Hao, X. Peng and L. Hu, Deep learning for sensor-based activity recognition: A survey, *Pattern Recognition Letters*, vol.119, pp.3-11, 2019.
[4] S. Mekruksavanich and A. Jitpattanakul, Hybrid convolution neural network with channel attention mechanism for sensor-based human activity recognition, *Scientific Reports*, vol.13, no.1, 2023.
[5] U. Côté-Allard, G. Gagnon-Turcotte, A. Phinyomark, K. Glette, E. J. Scheme, F. Laviolette and B. Gosselin, Unsupervised domain adversarial self-calibration for electromyography-based gesture recognition, *IEEE Access*, vol.8, pp.177941-177955, 2020.
[6] S. Mekruksavanich and A. Jitpattanakul, FallNeXt: A deep residual model based on multi-branch aggregation for sensor-based fall detection, *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol.16, no.4, pp.352-364, 2022.
[7] G. R. Naik, S. E. Selvan, M. Gobbo, A. Acharyya and H. T. Nguyen, Principal component analysis applied to surface electromyography: A comprehensive review, *IEEE Access*, vol.4, pp.4025-4037, 2016.
[8] Y. Du, W. Jin, W. Wei, Y. Hu and W. Geng, Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation, *Sensors*, vol.17, no.3, 2017.
[9] X. Zhai, B. Jelfs, R. H. M. Chan and C. Tin, Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network, *Frontiers in Neuroscience*, vol.11, 2017.
[10] R. N. Khushaba, A. H. Al-Timemy, A. Al-Ani and A. Al-Jumaily, A framework of temporal-spatial descriptors-based feature extraction for improved myoelectric pattern recognition, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol.25, no.10, pp.1821-1831, 2017.
[11] F. Duan, L. Dai, W. Chang, Z. Chen, C. Zhu and W. Li, sEMG-based identification of hand motion commands using wavelet neural network combined with discrete wavelet transform, *IEEE Transactions on Industrial Electronics*, vol.63, no.3, pp.1923-1934, 2016.
[12] Z. Li, X. Zhao, G. Liu, B. Zhang, D. Zhang and J. Han, Electrode shifts estimation and adaptive correction for improving robustness of sEMG-based recognition, *IEEE Journal of Biomedical and Health Informatics*, vol.25, no.4, pp.1101-1110, 2021.
[13] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. M. Hager, E. Simone, G. Giatsidis, F. Bassetto and H. Müller, Electromyography data for non-invasive naturally-controlled robotic hand prostheses, *Nature*, vol.1, 2014.
[14] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu and J. Li, Gesture recognition by instantaneous surface EMG images, *Scientific Reports*, vol.6, 36571, 2016.
[15] M. Atzori, M. Cognolato and H. Müller, Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands, *Frontiers in Neurorobotics*, vol.10, 2016.
[16] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli and W. Geng, Surface-electromyography-based gesture recognition by multi-view deep learning, *IEEE Transactions on Biomedical Engineering*, vol.66, no.10, pp.2964-2973, 2019.
[17] M. Atzori, A. Gijsberts, S. Heynen, A.-G. M. Hager, O. Deriaz, P. V. der Smagt, C. Castellini, B. Caputo and H. Müller, Building the Ninapro database: A resource for the biorobotics community, *Proc. of the IEEE International Conference on Biomedical Robotics and Biomechatronics*, Rome, Italy, pp.1258-1265, 2012.
[18] F. J. Ordóñez and D. Roggen, Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition, *Sensors*, vol.16, no.1, 2016.
[19] M. Devanne, P. Papadakis and S. M. Nguyen, Recognition of activities of daily living via hierarchical long-short term memory networks, *Proc. of 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy, pp.3318-3324, 2019.