EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

Human Muscle sEMG Signal and Gesture Recognition Technology Based on Multi-Stream Feature Fusion Network

Xiaoyun Wang[1*]

[1]Anhui Vocational and Technical College, School of Intelligent Manufacturing, Hefei, 230011, China

Abstract

Surface electromyography signals have significant value in gesture recognition due to their ability to reflect muscle activity in real time. However, existing gesture recognition technologies have not fully utilized surface electromyography signals, resulting in unsatisfactory recognition results. To this end, firstly, a Butterworth filter was adopted to remove high-frequency noise from the signal. A combined method of moving translation threshold was introduced to extract effective signals. Then, a gesture recognition model based on multi-stream feature fusion network was constructed. Feature extraction and fusion were carried out through multiple parallel feature extraction paths, combined with convolutional neural networks and residual attention mechanisms. Compared to popular methods of the same type, this new recognition method had the highest recognition accuracy of 92.1% and the lowest recognition error of 5%. Its recognition time for a single-gesture image was as short as 4s, with a maximum Kappa coefficient of 0.92. Therefore, this method combining multi-stream feature fusion networks can effectively improve the recognition accuracy and robustness of gestures and has high practical value.

Keywords: Multi-stream Characteristics, Convolutional Neural Networks, Surface Electromyography Signal, Gestures, Recognition

Received on 24 June 2024, accepted on 02 September 2024, published on 09 September 2024

doi:10.4108/eetpht.10.7230

1. Introduction

Surface Electromyography Signal (sEMG) records the electrophysiological signals caused by muscle activity. When muscle activity occurs, the action potential generated by muscle fibers is transmitted through the skin to the electrode, which transmits these electrical signals to amplifiers and recording devices [1-2]. In rehabilitation medicine, sEMG is utilized to evaluate and monitor muscle function, helping to design rehabilitation training programs. In sports science, sEMG is utilized to analyze the muscle activity of athletes in different sports, optimize training, and improve athletic performance. In human-machine interfaces, sEMG is utilized to control prosthetics, wheelchairs, and other assistive devices, achieving more natural and precise motion control [3]. In recent years, with the advancement of biomedical engineering and computing technology, the application of sEMG in Gesture Recognition (GR) has rapidly developed, promoting the development of health technology. The current GR technology utilizes machine learning and other methods to classify and analyze sEMGs, achieving high recognition accuracy and real-time performance [4]. Common methods include Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) [5]. However, there are still some shortcomings in the application of sEMG in GR. Firstly, sEMG is susceptible to factors such as muscle fatigue, electrode displacement, and

*Corresponding author. Email: wangxy@uta.edu.cn

skin resistance changes, leading to unstable signal quality and thus affecting recognition accuracy. Secondly, existing recognition techniques only consider individual sEMG data features, resulting in poor performance of GR. To this end, innovative attempts are made to preprocess and optimize gesture images, enhancing the recognition of image data features. Meanwhile, gesture feature recognition in various dimensions is achieved through a multi-stream feature fusion network. This study aims to enhance the application effectiveness and recognition robustness of sEMG in GR, and its contribution lies in providing effective support for subsequent GR.

2. Related Works

By capturing and understanding hand movements, GR can achieve natural and efficient interaction. GR has broad application prospects in fields such as virtual reality, augmented reality, smart homes, and medical rehabilitation [6]. To enhance the level of existing GR technology in visual driving, Sahoo J P et al. combined AlexNet and VGG-16 to propose a novel composite convolutional GR algorithm. This algorithm had better recognition stability and accuracy than traditional methods on two conventional datasets [7]. Bhushan S et al. constructed a GR model combining Random Forest (RF) to improve the GR efficiency and verified its effectiveness. This new model's highest recognition accuracy was 97.89%, which adapted to various GRs in different environments [8]. Gao R et al. found that previous work proposed using gesture position independent features to represent gestures, rather than directly matching signal change patterns. To this end, they constructed a new dynamic GR method by combining WiFi signals. This method significantly improved the feature strength of GR signals, with a maximum testing accuracy of 96.7% [9]. Faisal MA et al. believed that due to the availability of hardware and deep learning algorithms, GR research gained new momentum. To this end, they developed a cost-effective identification model using five flexible sensors, an inertial measurement unit, and a powerful microcontroller. After nearly 30 static and dynamic tests on 25 subjects, this model's F1 improved by at least 10% compared to traditional methods [10].

SEMG is an electrophysiological signal caused by muscle activity, which is recorded from the surface of the skin through non-invasive electrodes. In recent years, the application of sEMG in GR has received widespread attention. Fateyer A et al. proposed a novel GR method using sEMG spectral signals to enhance the GR of sparse multi-channel sEMG controlled electromyographic implants. The average classification accuracy of this method on large public databases was 95.5% [11]. Wang H et al. proposed a novel GR method combining deformable convolutional networks to enhance the effectiveness of peripheral device interfaces for prosthetic hands. This method extracted implicit correlations between different channels from sparse multi-channel sEMG, demonstrating strong robustness and feasibility [12]. Lv X et al. proposed a remote GR system based on multi-attention mechanism CNN to enhance the signal extraction and prosthetic control effects of sEMG. This system significantly enhanced the relevant features of sEMG, with a recognition accuracy of up to 97.86% [13]. Jiang Y et al. found that the existing GR using sEMG lacked a dataset for multi-class gestures. To this end, they constructed a rich dataset by combining inertial measurement units and conducted training tests using recurrent neural networks. Optimizing the training data of GR was indeed a unique method. However, the goal of improving GR accuracy could still be achieved [14].

In summary, existing GR research mainly focuses on two aspects: visual driving and signal feature extraction, such as composite convolutional GR, RF, sEMG spectrum signal recognition methods, etc. Although these methods can achieve GR, they are susceptible to noise interference and affect recognition accuracy. At the same time, a single feature extraction path is difficult to fully capture complex changes in electromyographic signals. To this end, an innovative GR method based on multi-stream feature fusion networks is proposed. The effectiveness of GR is enhanced through preprocessing optimization of sEMG and multi-dimensional convolutional feature extraction.

3. Materials and Methods

The existing GR technology has drawbacks, such as poor utilization of sEMG and slow recognition speed. Firstly, the preprocessing optimization of sEMG collection is studied, followed by the construction of a multi-stream feature fusion network. A GR model utilizing the flow feature fusion network is proposed.

3.1 sEMG preprocessing and extraction

SEMG is a technique that records and analyzes electrophysiological signals caused by muscle activity by placing electrodes on the skin surface. The basic principle is based on the transmission of action potentials emitted by motor neurons to muscle fibers during muscle contraction, leading to changes in the internal and external potentials of muscle fibers [15-16]. This potential change generates an electric current in muscle fibers, which in turn forms a potential signal on the surface of the skin. Figure 1 shows the changes in sEMG point sequence and the principle of signal generation.

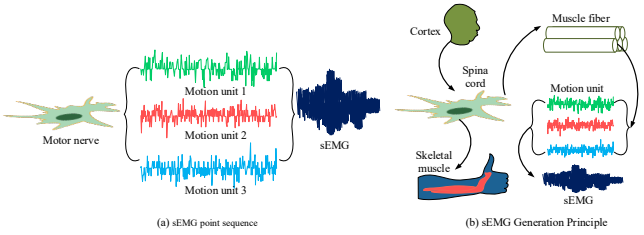

Figure 1. sEMG point sequence variations and the principle mode of generation

Figure 1 (a) is a schematic diagram of the changes in sEMG point sequence. Figure 1 (b) shows the principle of sEMG generation. The motor nerve units in the spinal cord generate various electromyographic signals. The signal stimulates the presynaptic membrane through neural impulses to produce acetylcholine, which then binds to the motor endplate to generate a potential. The sEMG electrode can capture these potential changes, amplify weak electrical signals through an amplifier, and record and analyze them through a data acquisition system [17]. In addition, after being transmitted through multiple layers of muscle tissue, sEMG is inevitably mixed with noise and interference due to physiological characteristics such as arm shaking and the influence of the acquisition environment. This reduces the signal strength of sEMG and weakens the features. For this study, Butterworth Filter (BF) is introduced to filter the original sEMG. Compared to other filtering methods, BF has smooth frequency response characteristics and can effectively remove high-frequency noise in the signal. Meanwhile, BF can maintain the main components of the signal, thereby improving the quality and accuracy of the signal. Specifically, BF determines the filter characteristics through transfer functions, uses differential equations to achieve digital filtering of signals, and analyzes the filtering effect through frequency response [18]. The transfer function is represented by equation (1).

$$H(s) = -\frac{G}{\sqrt{1+(\frac{s}{\omega_c})^{2n}}} (1)$$

In equation (1), $H(s)$ refers to a filter's transfer function. G refers to the filtering gain. s refers to complex frequency variables. ω_c refers to a filter's cutoff frequency. n refers to a filter's order. The difference equation is represented by equation (2).

$$y(n) = \frac{1}{a_0}(b_0 x(n) + b_M x(n) - a_1 y(n-1) - a_N y(n-1) (2)$$

In equation (2), (n) refers to the current sample value of the filtered output signal. $x(n)$ refers to the current sample value of the input signal. b_M refers to the molecular coefficient. a_y refers to the denominator coefficient. M and N correspond to the numerator and denominator parts' orders. Figure 2 shows a comparison of time-domain waveforms after sEMG filtering.

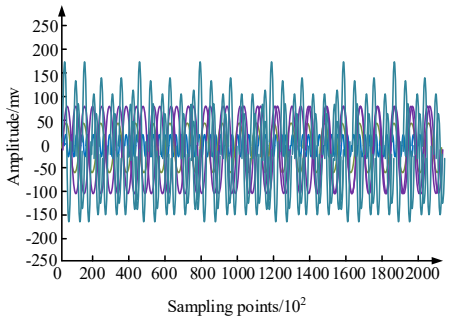

Figure 2. Time-domain waveforms of surface EMG after BF filtering

In Figure 2, after BF filtering optimization, the original sEMG time-domain waveform's amplitude remained stable in [-150, 150]. This eliminates power frequency noise above 150 and below -150, effectively enhancing the characteristic strength of sEMG time-domain signals. General gesture acquisition includes both stationary and motion states. When the arm changes from stationary to moving, sEMG also follows the change to active state. After considering the computational complexity of sEMG, a Moving Panning Threshold Combination Method (MPTCM) is proposed. This method sets an effective threshold by calculating the average instantaneous energy, thereby constraining and extracting the starting and active points. Firstly, the filtered sequence of sEMG is obtained by using the difference squared method to obtain the instantaneous energy average sequence. Secondly, energy extraction is performed by sliding through a fixed window. The energy values of each window are calculated sequentially. Finally, an amplitude threshold is set to identify and filter sEMG sequence data with energy values greater than the threshold. The average instantaneous energy is represented by equation (3).

$$E_{ave}(i) = \frac{1}{I}\sum_{k=1}^{I} X_k(i+1) + X(i)^2 (3)$$

In equation (3), X_k represents the sEMG sequence value. $E_{ave}(i)$ represents the average energy value of the electromyographic signal. I represents the total number of signal channels. Assuming a window length of 128 is taken for instantaneous energy extraction, the average energy of each window is calculated one by one. The relevant calculations are represented by equation (4).

$$E_{mean}(i) = \frac{1}{\varphi}\sum_{j=1}^{i+\varphi-1} E_{ave}(j) (4)$$

In equation (4), $E_{mean}(i)$ refers to the average energy of each window. φ refers to the number of windows. i refers to the sequence value of the electromyographic signal. j refers to the signal sampling point. Figure 3 shows the average energy of the sEMG active segment after MPTCM processing with an amplitude threshold of 120.

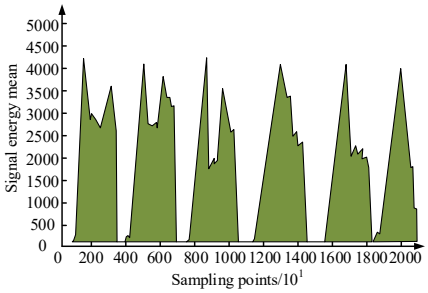

Figure 3. Representation of the mean energy of the sEMG active segment after MPTCM optimization

In Figure 3, signals below the threshold are marked as 0. At this point, it is easier to distinguish the sEMG energy fluctuations under the data volume of the two related sampling points. Using 2000 samples as a quantity interval, the optimized sEMG activity segment is clearer and better displays sEMG sequence information under different actions compared to before optimization.

3.2 Construction of gesture recognition model based on multi-stream feature fusion network

After completing the preprocessing and extraction optimization of sEMG, the study attempts to construct a novel GR model. CNN is a commonly utilized data signal feature extraction model in deep learning methods. Firstly, the collected sEMG is preprocessed to remove noise and normalize the signal. Then, the preprocessed signals are converted into input formats suitable for CNN processing. The input data are processed through multiple layers of convolutional layers. Spatial features are extracted through a mechanism of local receptive fields and weight sharing. Each convolutional layer is followed by a non-linear activation function to enhance the network's non-linear ability [19-20]. Subsequently, the pooling layer downsamples the feature map, reducing data dimensions and maintaining important features. After multi-layer convolution and pooling operations, the high-level feature map obtained is integrated through fully connected layers. Finally, the Softmax layer is utilized for classification. The probability of each gesture category is output. CNN convolution is represented by equation (5).

$$C_{cn} = f(X \cdot W_{cn} + b_{cn}) (5)$$

In equation (5), f is a convolutional activation function. X is the input data. W_{cn} is the convolutional kernel's weight. b_{cn} is the bias. The pooling layer calculation is represented by equation (6).

$$a = \begin{bmatrix} 2 & 4 & 3 \\ 4 & 6 & 5 \\ 1 & 0 & 3 \end{bmatrix},$$

$$a_{max} = \begin{bmatrix} 4 & 4 \\ 6 & 4 \end{bmatrix}, a_{ave} = \begin{bmatrix} 2.5 & 2.25 \\ 3 & 2 \end{bmatrix} (6)$$

In equation (6), a_{max} represents the maximum pooling value, which means dividing the listed numbers in equation a into four equal parts: top, bottom, left, right, and then selecting the maximum value from each part in order to obtain equation a_{max}. The pooling layer only retains the most significant one in each output data, thereby reducing the difficulty of the entire filtering, represented by equation (7).

$$y = f(W \cdot x + z) (7)$$

In equation (7), W refers to the weight matrix. x refers to the input value. z refers to bias. y refers to output. However, CNN mainly relies on feature extraction and classification of single-stream signals. This often leads to insufficient feature representation and inability to fully capture complex changes in electromyographic signals, thereby affecting accurate GR. Therefore, the study attempts to use a multi-stream feature fusion approach, which decomposes and processes sEMG from different muscle parts through multiple parallel feature extraction paths. Finally, these features are integrated to construct a novel GR network in Figure 4.

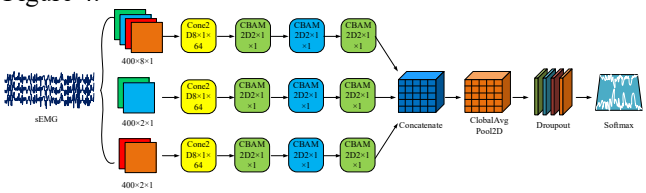

Figure 4. Multi-stream feature fusion network model structure

In Figure 4, the entire network model can be divided into input module, multi-stream convolution module, and multi-stream feature fusion module. The input module contains sEMG fragments from different muscle groups in the forearm and upper arm, with each fragment size of 400×12×1. The multi-stream convolution module processes the input data through multiple parallel feature extraction paths. Each path extracts features through multiple 2D convolutional layers and residual attention mechanism modules. Each feature extraction path consists of a three-layer convolutional structure. The first layer consists of 8 1×64 convolution kernels, followed by 2 1×64 convolution kernels. The last layer includes 1 8×128 convolution kernel and 1 2×128 convolution kernel. Finally, the outputs of multiple feature extraction paths are fused through concatenation operations. The fused features are dimensionally reduced through a global average pooling layer. At this point, the convolution operation is represented by equation (8).

$$y_{e,u,p} = \sum_{t=1}^{T} \sum_{v=1}^{V} x_{e+t-1,u+v-1} w_{t,v,p} + b_p (8)$$

In equation (8), $y_{e,u,p}$ refers to the output feature map's value at position (e, u) and channel I. $X_{e+t-1,u+v-1}$ refers to this input feature map's position at position (e + t − 1, u +

$v-1$). $w_{t,v,p}$ refers to the convolutional kernel's weight at position $w_{t,v,p}$ and channel p. b_p refers to the bias of channel p. The calculation of residual connections is represented by equation (9).

$$y *= F(x^*, W_e) + x^*(9)$$

In equation (9), y^* refers to the residual block output. (x^*, W_e) refers to the convolution operation inside the residual block. x^* refers to input features. The calculation of attention mechanism is represented by equation (10).

$$Mc(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))(10)$$

In equation (10), Mc refers to channel attention mapping. σ refers to the activation function, such as sigmoid. MLP refers to multi-layer perceptrons. AvgPool(F) and MaxPool(F) correspond to global average pooling and global maximum pooling operations. In summary, Figure 5 shows the GR model combined with sEMG preprocessing optimization and multi-stream feature fusion network.

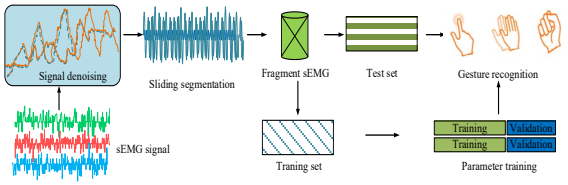

Figure 5. Novel gesture recognition model computational flow

In Figure 5, first, the arm muscle electrode is inserted using an invasive sensor to export sEMG. Secondly, the sEMG is transmitted to the preprocessing module and filtered by BF to remove high-frequency noise while preserving signal features. After completion, MPTCM extraction is performed on each signal data. Data filtering is performed by calculating instantaneous energy and threshold judgment. Then, the filtered sEMG fragments are constructed into a database and divided into a training set and a testing set. In the training section, a multi-stream feature fusion network model is trained multiple times to effectively provide a recognition model with the best parameter value performance. Finally, the sEMG GR is performed using the best performing model.

4. Results

The CPU is Intel Core i7. The GPU is NVIDIA GeForce GTX 1060 with 16GB of memory. The operating system is Windows 10 and adopts the Python 3.8 framework. The GR model was trained using the Adam optimizer with a training period of 40 and a learning rate of 0.001. The NinaPro dataset was used as the experimental data source. NinaPro is a publicly available dataset specifically designed for GR and sEMG research, containing sEMG gestures and movements from healthy volunteers and amputees, covering a variety of gestures and movements. The study divided the dataset data into training and testing sets in an 8:2 ratio. Firstly, using recognition accuracy as an indicator, the proposed GR model was subjected to ablation testing in Figure 6.

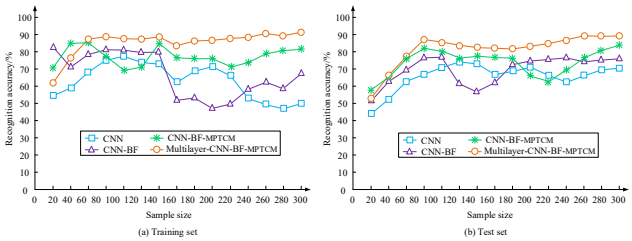

Figure 6. Novel gesture recognition model ablation test results

Figure 6 (a) shows the ablation test results of the new GR model on the training set. Figure 6 (b) shows the ablation test results of the new GR model under the test set. As the training samples increased, the recognition accuracy of each module generally improved. However, in the later stage, there was a slight decrease in CNN-BF and CNN-BF-MPTCM. The reason is that although BF and MPTCM have preprocessed and optimized sEMG, the inputted single stream features still affect this model under complex samples. This proposed model greatly optimized the image convolution process through multi-stream feature fusion, reducing the dimensionality of complex data processing. Its recognition accuracy in the training set and testing was the highest at 92.1% and 90.8%, respectively. The study compared popular gesture detection algorithms of the same type, such as RF, Deep Belief Network (DBN), and Support Vector Machine (SVM). The above methods were tested 15 times each based on detection error. Figure 7 is a test box diagram.

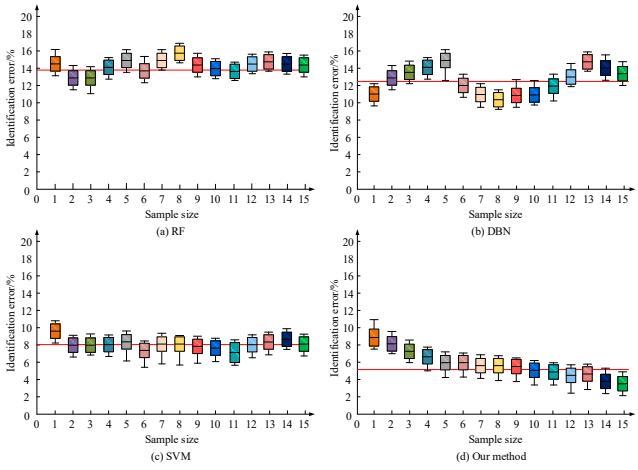

Figure 7. 15 Recognition error tests with different algorithms

Figure 7 (a) shows the GR results of 15 repetitions of RF. Figure 7 (b) shows the 15 repeated GR results of DBN. Figure 7 (c) shows the 15 repeated GR results of the SVM. Figure 7 (d) shows the 15 repeated GR results of the proposed algorithm. The testing errors of RF and DBN generally tended to be over 12%, demonstrating poor recognition performance. SVM, relying on its unique linear and nonlinear classification capabilities, reduced recognition errors by 4%. This proposed method's recognition error significantly decreased after multiple repeated tests, and the error range gradually narrowed to 5%. The study randomly selected 4 types of gesture actions from NinaPro to verify this proposed method's practical application effect in Figure 8.

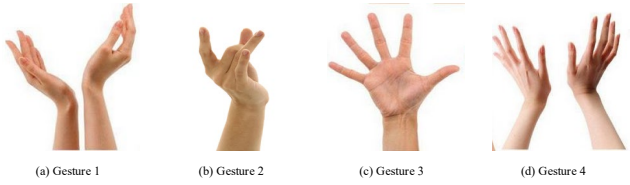

Figure 8. Four types of gestural movements

The study continued to introduce more advanced GR techniques for comparison, such as Deep Convolutional Generative Adversarial Network (DCGAN), Variational Autoencoder (VAE), and Spatio-Temporal Graph Convolutional Network (ST-GCN). Figure 9 shows the test results.

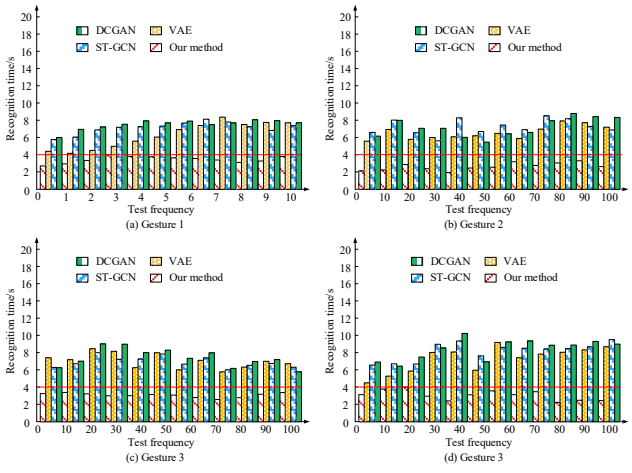

Figure 9. Gesture recognition results for four recognition models

Figure 9 (a) shows the 10 recognition results of gesture 1 by four models. Figure 9 (b) shows the 10 recognition results of gesture 2 by four models. Figure 9 (c) shows the 10 recognition results of gesture 3 by four models. Figure 9 (d) shows the 10 recognition results of gesture 4 by four models. The GR time of DCGAN and ST-GCN was generally greater than 6s, and the GR performance was relatively average. Early recognition of VAE was better. However, as this model's fatigue level increased, its recognition and detection efficiency decreased. The proposed method had the shortest GR time, especially in gestures 2 and 3, with an average GR time of 4s. Therefore, this proposed method had significant advantages among many existing methods. Tests were conducted using Precision (P), Recall (R), F1 value, and Kappa coefficient as indicators. The Kappa coefficient took a value of [-1,1], and a large value indicates good recognition and prediction performance. Table 1 shows the results.

Table 1. Indicator test results for different models

Method	P/%	R/%	F1/%	Kappa
RF	87.47	89.67	88.57	0.78
DBN	89.64	90.54	90.09	0.84
SVM	93.47	92.11	92.79	0.88
DCGAN	90.26	91.27	90.77	0.86
VAE	92.44	93.68	93.06	0.83
ST-GCN	92.87	94.12	93.50	0.89
Our model	96.73	95.43	96.08	0.92

In Table 1, the proposed method performed the best in all four indicators, including P, R, F1 values, and Kappa coefficient. Its P was 96.73%, R was 95.43%, F1 value was 96.08%, and Kappa coefficient was 0.92, which is significantly better than other models. These results confirmed that the method had higher accuracy and consistency in GR tasks. Among other models, ST-GCN and SVM also performed well, but the overall performance was not as good as the proposed method.

Conclusion

There are issues with singularity, real-time performance, and accuracy in the application of sEMG in the GR field. In this regard, the study conducted in-depth analysis and optimization of sEMG preprocessing and feature extraction by combining BF and MPTCM. Subsequently, based on CNN and residual attention mechanism, a GR model combining multi-stream feature fusion network was constructed. The proposed method achieved the highest recognition accuracy of 92.1% and 90.8% in the training and testing sets, respectively. Compared to RF, DBN, and SVM, the proposed method had a minimum GR error of 5% after 15 repeated tests, which was significantly better than the 12% error rate of RF and DBN. After conducting comparative tests on the recognition of four random gestures, this new method's average recognition time was 4s. Other methods' shortest recognition time was 6s. This indicated that the proposed method had significant advantages in recognition efficiency. The new model had a P value of 96.73%, R value of 95.43%, F1 value of 96.08%, and Kappa coefficient of 0.92. In summary, the multi-stream

feature fusion network can significantly improve the GR performance of sEMG and has high practical value. Although the proposed method has achieved certain results, it has not yet taken into account the GR effects in other complex environments, such as light changes and noise interference. Therefore, subsequent research can explore more efficient feature extraction methods and more optimized network structures to further enhance the GR capability of sEMG.

Acknowledgements

This research was supported by Natural Science Research Project of Anhui Educational Committee（2022AH040280.

References

[1] F. A. Farid, N. Hashim J. Abdullah, R. Bhuiyan, J. Uddin, M. A. Haque and M. N. Huse,“A structured and methodological review on vision-based hand gesture recognition system,” *J. Imaging*,vol. 8, no. 6, pp. 153-154, May. 2022, doi:10.3390/jimaging8060153.

[2] J. Qi, L. Ma, Z. Cui, and Y. Yu, “Computer vision-based hand gesture recognition for human-robot interaction: a review,"*Complex Intell. Syst*, vol. 10, no. 1, pp. 1581-1606, June. 2024, doi: 10.1007/s40747-023-01173-6.

[3] D. Ryumin, D. Ivanko, and E. Ryumina, “Audio-visual speech and gesture recognition by sensors of mobile devices,” *Sens*, vol. 23, no. 4, pp. 2284-2285, November. 2023, doi: 10.3390/s23042284.

[4] J. Yu, M. Qin, and S. Zhou, “Dynamic gesture recognition based on 2D convolutional neural network and feature fusion,”*Sci. Rep*, vol. 12, no. 1, pp. 4345-4346, March. 2022, doi: 10.1038/s41598-022-08133-z.

[5] A. Calado, P. Roselli, and V. Errico, “A geometric model-based approach to hand gesture recognition,” *IEEE Trans. Syst. Man Cybern. Part B Cybern*, vol. 52, no. 10, pp. 6151-6161, January. 2022, doi: 10.1109/TSMC.2021.3138589.

[6] A. S. M. Miah, M. A. M. Hasan, and J. Shin, “Multistage spatial attention-based neural network for hand gesture recognition,” *Comput*, vol. 12, no. 1, pp. 13-14, December. 2023, doi: 10.3390/computers12010013.

[7] J. P. Sahoo, A. J. Prakash, P. Pławiak, and S. Samantray, “Real-time hand gesture recognition using fine-tuned convolutional neural network,” *Sens*, vol. 22, no. 3, pp. 706-707, January. 2022, doi: 10.3390/s22030706.

[8] S. Bhushan, M. Alshehri, I. Keshta, A. K. Chakraverti, J. Rajpurohit, and A. Abugabah, “An experimental analysis of various machine learning algorithms for hand gesture recognition,” *Electron*, vol. 11, no. 6, pp. 968-967, March. 2022, doi: 10.3390/electronics11060968.

[9] R. Gao, W. Li, Y. Xie, E. Yi, L. Wang, D. Wu, and D. Zhang, “Towards robust gesture recognition by characterizing the sensing quality of WiFi signals,” Proc. *ACM Interact. Mobile Wearable Ubiquitous Technol*, vol. 6, no. 1, pp. 1-26, March. 2022, doi: 10.1145/3517241.

[10] M. A. A. Faisal, F. F. Abir, and M. U. Ahmed, “Exploiting domain transformation and deep learning for hand gesture recognition using a low-cost dataglove,”*Sci. Rep*, vol. 12, no. 1, p. 21446, December. 2022, doi: 10.1038/s41598-022-25108-2.

[11] A. Fatayer, W. Gao, and Y. Fu, “sEMG-based gesture recognition using deep learning from noisy labels,"*IEEE J. Biomed. Health. Inf*, vol. 26, no. 9, pp. 4462-4473, June. 2022, doi: 10.1109/JBHI.2022.3179630.

[12] H. Wang, Y. Zhang, C. Liu, and H. Liu, “sEMG based hand gesture recognition with deformable convolutional network,” *Int. J. Mach. Learn. Cybern*, vol. 13, no. 6, pp. 1729-1738, March. 2022, doi: 10.1007/s13042-021-01482-7.

[13] X. Lv, C. Dai, H. Liu, Y. Tian, L. Chen, Y. Lang, R. Tang, and J. He, “Gesture recognition based on sEMG using multi-attention mechanism for remote control,” *Neural Comput.Appl*., vol. 35, no. 19, pp. 13839-13849, July. 2023, doi: 10.1007/s00521-021-06729-6.

[14] Y. Jiang, L. Song, J. Zhang, Y. Song, and M. Yan, “Multi-category gesture recognition modeling based on sEMG and IMU signals,” *Sens*, vol. 22, no. 15, pp. 5855-5856, August. 2022, doi: 10.3390/s22155855.

[15] R. Zhang, C. Jiang, and S. Wu, “Wi-Fi sensing for joint gesture recognition and human identification from few samples in human-computer interaction,” IEEE J. Sel. Areas Commun, vol. 40, no. 7, pp. 2193-2205, March. 2022, doi: 10.1109/JSAC.2022.3155526.

[16] J. Yang, S. Liu, and Y. Meng, “Self-powered tactile sensor for gesture recognition using deep learning algorithms,” *ACS Appl. Mater. Interfaces*, vol. 14, no. 22, pp. 25629-25637, May. 2022, doi: 10.1021/acsami.2c01730.

[17] L. Liu, W. Xu, and Y. Ni, “Stretchable neuromorphic transistor that combines multisensing and information processing for epidermal gesture recognition,”*ACS Nano*, vol. 16, no. 2, pp. 2282-2291, January. 2022, doi: 10.1021/acsnano.1c08482.

[18] L. Guo, Z. Lu, and L. Yao, “A gesture recognition strategy based on A-mode ultrasound for identifying known and unknown gestures,”*IEEE Sens. J*, vol. 22, no. 11, pp. 10730-10739, April. 2022, doi: 10.1109/JSEN.2022.3167696.