Mamba-based Multi-Scale Classification Framework for sEMG Gesture Recognition

Yuhan Song[1,2], Anming Dong[1,2,*], Yubing Han[1,2], Jiguo Yu[3], You Zhou[4]
[1] Key Laboratory of Computing Power Network and Information Security, Ministry of Education,
Shandong Computer Science Center (National Supercomputer Center in Jinan),
Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[2] Shandong Provincial Key Laboratory of Industrial Network and Information System Security,
Shandong Fundamental Research Center for Computer Science, Jinan, China
[3] School of Information and Software Engineering,
University of Electronic Science and Technology of China, Chengdu, China
[4] Technology Department, Shandong HiCon New Media Institute Co., Ltd., Jinan, China

***Abstract*—Surface electromyography signals (sEMG) are widely used in gesture recognition and human-computer interaction due to their ability to reflect muscle activity patterns. However, the non-stationarity, multi-scale variations, and noise interference of the signals severely limit modeling and recognition performance. Existing methods based on traditional machine learning rely on manually designed temporal-frequency features, making them difficult to adapt to complex and variable signal environments. Although deep learning methods can improve recognition accuracy through end-to-end modeling, their robustness is still insufficient in complex scenarios such as small sample sizes, signal noise, and multi-scale variations, and they are inefficient in long-sequence modeling, which limits their performance in sEMG tasks. To address the aforementioned issues, this paper proposes a multi-path Mamba multi-scale classification framework (MaMSC). This method decomposes sEMG signals into high-frequency, low-frequency, and temporal residual components using hybrid frequency-temporal decomposition (HFTD), and models these multi-scale components using multi-path selective Mamba units within the MaMSC block. Subsequently, an inverse hybrid frequency-temporal decomposition (IHFTD) module fuses the features into a unified temporal representation. This design fully leverages Mamba's efficient long-sequence modeling capabilities and enhances robustness through symmetric residual paths. Experiments on the Ninapro DB2 dataset demonstrate that this method achieves accuracies of 99.18%, 92.82%, and 99.35% on gesture recognition Exercise B, C, and D, respectively. Cross-subject experiments further demonstrate that a generalization accuracy of 93.56% can be achieved with minimal fine-tuning of target user data, exhibiting extremely low calibration costs and excellent potential for real-world interaction adaptation. The code is available at: https://github.com/yu-han-song/MaMSC.**

***Index Terms*—Gesture recognition, surface electromyography, multi-scale, signal, Mamba.**

I. Introduction

As a non-invasive bioelectric signal, sEMG can reflect muscle activity patterns and has important application value in gesture recognition, motion control, and medical rehabilitation [1]–[6]. However, sEMG signals generally exhibit non-stationarity, multi-scale changes, and noise interference, posing challenges for feature extraction and sequence modeling.

Existing sEMG recognition methods are mainly divided into traditional machine learning and deep learning approaches [7]. Traditional methods rely on hand-crafted time-domain, frequency-domain, or time-frequency features, combined with classifiers such as Support Vector Machines (SVM) [8] and K-Nearest Neighbors (KNN) [9]. They have low computational costs and good real-time performance, but limited feature representation, making them unsuitable for complex gestures or large-scale data. Deep learning methods, in contrast, directly learn features from raw or preprocessed sEMG signals in an end-to-end manner, improving nonlinear modeling and recognition accuracy, and are becoming the mainstream. Representative models include Convolutional Neural Networks (CNN) [10], Recurrent Neural Networks (RNN) [11], and Transformers [12]. To capture spatial-temporal or global dependencies, hybrid models such as CNN-GRU [4], [13], CNN-Transformer [14], [15], and CNN-GRU-Transformer [5] have been proposed. However, CNN mainly captures local patterns and struggles with long-term dependencies. RNNs and GRUs handle temporal relations but suffer from vanishing or exploding gradients in long sequences. Hybrid models alleviate these issues but increase complexity and reduce efficiency. Transformers, while powerful, face $O(n^2)$ complexity in self-attention, leading to heavy GPU memory usage on long sequences, and show limited robustness with small samples and noisy data. The recently proposed Mamba architecture provides new opportunities for sequence modeling. Mamba is based on selective State Space Models (SSMs) [16], [17].Unlike Transformers that rely on attention, Mamba uses input-dependent state transitions for content-based reasoning, achieving linear scaling with sequence length, higher throughput, and strong cross-modal generalization. It combines the advantages of Transformers and RNNs: parallel training and global context capture, together with recurrent state updates that mitigate vanishing or exploding gradients. With $O(n)$ complexity, Mamba maintains efficiency and scalability on very long sequences. Mamba has achieved success in the fields of language modeling [18], speech or audio analysis [19], [20], and genomics [21], and has demonstrated wide applicability in the field of time series: NetMamba applies the unidirectional Mamba architecture to network traffic byte sequence classification, achieving a 60-fold inference acceleration while main-

*Corresponding author: anmingdong@qlu.edu.cn

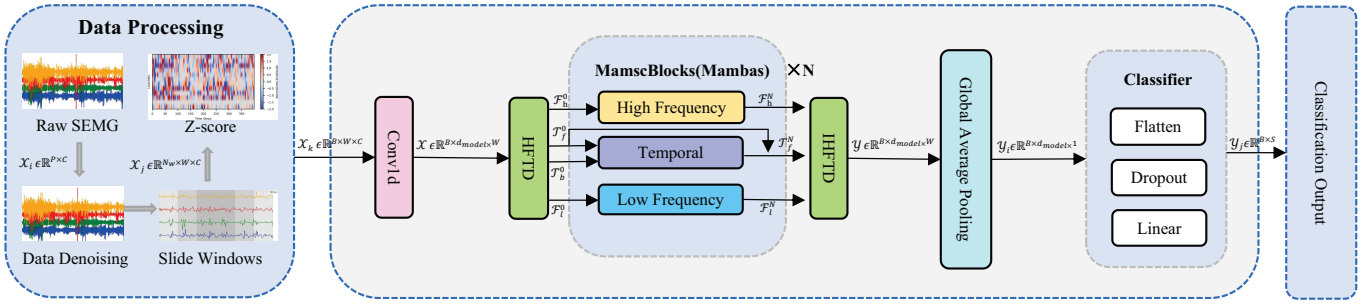

Fig. 1. The structure of model.

taining nearly 99% accuracy [22]; ActivityMamba designs a CNN-Mamba hybrid architecture to process multimodal sensor time series data, achieving SOTA performance on five human activity recognition datasets with significantly lower FLOPs than the Transformer baseline [23]; ECGMamba, targeting electrocardiogram, a typical physiological time series, proposes a multi-path Mamba block (MP-Mamba) to model the long-range temporal dependencies of 12-lead ECG, surpassing the performance of Transformer with only 1/17 of the computational complexity [24]. However, to our knowledge, Mamba's application in sEMG signals, which exhibit strong temporal dependence, non-stationarity, and individual variability, remains largely unexplored, and its modeling requirements are highly similar to those of physiological signals such as electrocardiograms (ECG). Therefore, we explore the application of Mamba in sEMG gesture recognition, leveraging its linear complexity and long-range dependency modeling capabilities to provide an efficient solution across various subject scenarios.

Based on this motivation, we propose a Multi-path MaMSC framework for sEMG analysis. The input sequence is first processed by a HFTD module [25], which applies first-order wavelet decomposition to obtain high-frequency, low-frequency, and residual components, reducing non-stationarity. Then, N MamscBlocks are built, where Selective Mamba units model sequences along high-frequency, low-frequency, and time-domain paths. A symmetric residual branch with channel flipping is added to the time-domain path. Finally, an IHFTD module fuses all path features into a unified representation for gesture recognition.

The main contributions of this paper are summarized as follows:

(1) We propose a novel multi-scale classification framework based on multipath Mamba for gesture recognition using sEMG signals. This framework achieves decoupling and fusion of multi-scale time-frequency features through a HFTD and IHFTD module.

(2) We design a multipath stacked Mamba encoder, termed MamscBlocks, to perform dedicated intra-scale modeling for high-frequency, low-frequency, and time-domain residual signals, enhancing discriminative sequence representation capabilities.

(3) Extensive experiments on the Ninapro DB2 dataset demonstrate that this method performs excellently on intra-subject gesture recognition tasks. With only a minimal amount of target user data for fine-tuning across subject tasks, the model achieves 93.56% accuracy, exhibiting low calibration costs and strong potential for real-world interaction adaptation.

II. sEMG Data and Preprocessing

The Ninapro DB2 dataset, widely used for sEMG-based gesture recognition, contains recordings from 40 subjects performing 49 hand and wrist movements plus rest, sampled at 2 kHz. DB2 uses the Delsys Trigno Wireless system to collect signals, with 12 sEMG channels arranged. The dataset categorizes gestures into three groups: (i) Exercise B, comprising 17 basic finger and wrist movements; (ii) Exercise C, comprising 23 grasping and functional movements; and (iii) Exercise D, comprising 9 force patterns. Due to its non-stationary and subject-dependent nature, DB2 requires models capable of capturing temporal and multi-scale frequency features while suppressing noise.

To address these challenges, we apply a preprocessing pipeline consisting of bandpass filtering, sliding window segmentation, and Z-score normalization. Specifically, a fourth-order Butterworth filter removes low-frequency motion artifacts and high-frequency noise. The signal is then divided into overlapping windows of length $W = 400$ [26], producing samples of shape $[N_w, W, C]$, where C is the number of channels and N_w the number of windows. Finally, Z-score normalization is applied across channels to standardize amplitudes and facilitate model convergence. The resulting data are used as input for the deep learning framework described below.

III. Proposed MaMSC sEMG Recognition Architechture

A. Proposed MAMBA-Based Recognition Framework

This paper proposes a sEMG gesture recognition framework based on the Mamsc architecture. Its overall structure is shown in Fig. 1. The framework primarily consists of data preprocessing, Conv1d, HFTD, MamscBlocks, IHFTD, global average pooling (GAP), and a classifier head.

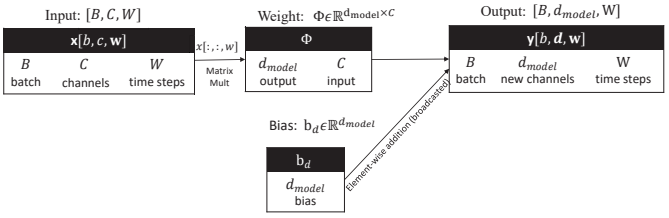

Fig. 2. One-dimensional convolution operation.

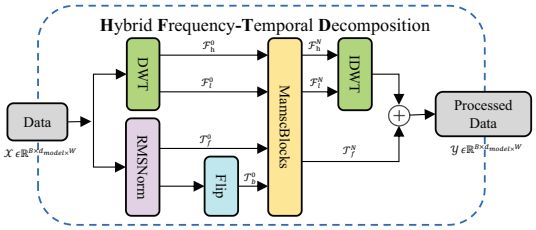

Fig. 3. HFTD: the input segment is decomposed into high-frequency, low-frequency, and temporal residual components.

First, the raw sEMG signal undergoes data preprocessing to generate uniformly scaled time series segments. Conv1d then maps the 12-channel input to the model's internal dimensions for subsequent modeling. The HFTD module uses Discrete Wavelet Transform (DWT) [27], [28] to obtain high- and low-frequency components and generates a temporal residual signal through RMSNorm and Flip operations. These three features are then fed into the MamscBlocks paths for parallel modeling. After stacking N layers, IHFTD fuses them to generate a unified feature representation. Finally, GAP aggregates temporal features, and a multi-layer classifier head outputs gesture categories.Below we will introduce these modules in detail.

B. Conv1d

After preprocessing, the dataset has a shape of $[N_w, W, C]$. During training, a mini-batch B is sampled from it, resulting in an input tensor $[B, W, C]$.

At the entry point of the feature extraction module, a one-dimensional convolutional layer (Conv1d) with a kernel size of 1 is used as a channel mapping layer. Linear projection is performed only on the channel dimension, mapping C channels to the model dimension d_{model}. The input $[B, W, C]$ is first transposed to $[B, C, W]$, and the output after Conv1d is $[B, d_{\text{model}}, W]$.The key point is that the transformation is applied independently at all time steps, allowing efficient parallel computation while preserving temporal structure.

The mathematical formula for a convolution operation with kernel size 1 is as follows, and this operation can also be visualized as shown in Fig. 2.

$$y[b, d, w] = \sum_{c=1}^{C} \Phi_{d,c} \cdot x[b, c, w] + b_d, \tag{1}$$

Where $x[b, c, w]$ is the input tensor of batch b, channel c, and time step w. $\Phi \in \mathbb{R}^{d_{\text{model}} \times C}$ is the convolution kernel's weight matrix, and $\Phi_{d,c}$ represents the weight parameter that maps the input channel c to the output channel d. $b_d \in \mathbb{R}^{d_{\text{model}}}$ is the bias vector, and $y[b, d, w]$ is the output tensor.

We denote the complete output tensor as $\mathcal{X} = \{y[b, d, w]\}$, where $\mathcal{X} \in \mathbb{R}^{B \times d_{\text{model}} \times W}$. This tensor serves as the input to the subsequent HFTD block with dimension d_{model}.

C. Hybrid Frequency-Temporal Decomposition

sEMG signals exhibit significant nonstationarity in the time domain and contain multi-scale dynamic characteristics related to movement in the frequency domain. Traditional models struggle to simultaneously model both types of information, limiting their ability to model EMG signals. To address this, this paper proposes an HFTD method, as shown in Fig. 3, which decouples and represents the multiscale features of sEMG signals in both the time and frequency domains. In this work, the Daubechies-4 wavelet is adopted for DWT. Given an input sample $\mathcal{X} \in \mathbb{R}^{B \times d_{\text{model}} \times W}$, where d_{model} is the latent feature dimension after channel mapping, and W is the temporal length of each window segment. To enhance the scale representation of features, we use the DWT to perform a channel-by-channel decomposition of $\mathcal{X}$:

$$[\mathcal{F}_h, \mathcal{F}_l] = \mathrm{DWT}(\mathcal{X}), \tag{2}$$

where $\mathcal{F}_h$ is the decomposed high-frequency component and $\mathcal{F}_l$ is the low-frequency component.

Thus, the decomposition produces two frequency-domain components:

$$\mathcal{X} \rightarrow \{\mathcal{F}_h \in \mathbb{R}^{B \times d_{\text{model}} \times M},\ \mathcal{F}_l \in \mathbb{R}^{B \times d_{\text{model}} \times M}\}.$$

M refers to the length of the time series obtained after DWT. This operation is performed independently on each feature channel, ensuring consistency of channel structure after decomposition.

Considering that frequency-domain transformation may weaken the direct expression of the original temporal structure, this paper introduces an RMSNorm-based residual path to preserve time-domain information and incorporate a symmetric modeling mechanism. For the input sample $\mathcal{X} \in \mathbb{R}^{B \times d_{\text{model}} \times W}$, two residual components are defined as

$$\mathcal{T}_f = \mathrm{RMSNorm}(\mathcal{X}), \quad \mathcal{T}_b = \mathrm{Flip}_W(\mathcal{T}_f). \tag{3}$$

Here, $\mathcal{T}_f$ is obtained by applying RMSNorm across the feature dimension d_{model} at each time step, which preserves the temporal structure of the sEMG sequence. $\mathcal{T}_b$ is generated by flipping $\mathcal{T}_f$ along the temporal axis W, forming a symmetric residual branch in the time domain.

D. MamscBlocks

Mamba is a general State-Space Model (SSM) architecture with linear complexity of $O(n)$ and excellent long-sequence modeling capabilities. The structure of Mamba is shown in Fig. 4.

Based on Mamba, we design task-specific MamscBlocks modules and extend them with a multi-path architecture. After

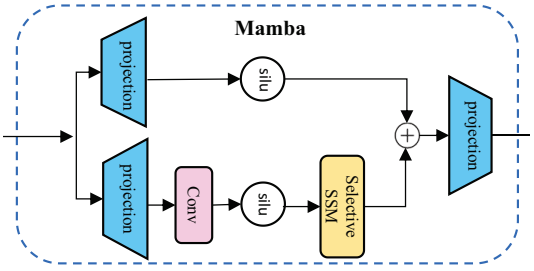

Fig. 4. Detailed design of Mamba.

HFTD decomposition, features are fed into the MamscBlocks for three-path modeling. MamscBlocks consist of three parallel branches: high-frequency, low-frequency, and temporal paths. Each branch is built with several Selective Mamba layers, combined with normalization and residual connections, to achieve multi-scale feature modeling. The update process is:

$$\begin{aligned} \mathcal{F}_h^n &= \text{Mamba}_\theta^{HF}(\mathcal{F}_h^{n-1}), \\ \mathcal{F}_l^n &= \text{Mamba}_\theta^{LF}(\mathcal{F}_l^{n-1}), \end{aligned} \tag{4}$$

$$\begin{aligned} \mathcal{T}_f^n &= \text{Mamba}_\theta^{T}(\text{RMSNorm}(\mathcal{T}_f^{n-1})) \\ &\quad + \text{Mamba}_\theta^{T}(\mathcal{T}_b^{n-1}) + \mathcal{T}_f^{n-1}, \\ \mathcal{T}_b^n &= \text{Flip}_W(\mathcal{T}_f^n). \end{aligned} \tag{5}$$

Here, where $n = 1, 2, \ldots, N$ denotes the n-th layer of the stacked MamscBlocks, and N is the total number of MamscBlocks layers. Mamba_θ^{HF}, Mamba_θ^{LF}, and Mamba_θ^{T} denote Mamba modules specialized for high-frequency, low-frequency, and temporal branches, respectively, each with independent parameters θ.

E. Inverse Hybrid Frequency-Temporal Decomposition

After processing through the three paths of MamscBlocks, the outputs are high-frequency features $\mathcal{F}_h^N$, low-frequency features $\mathcal{F}_l^N$, and the primary time-domain residual representation $\mathcal{T}_f^N$. These outputs are then fed into the IHFTD module for fusion, generating the final unified representation $\mathcal{Y}$:

$$\mathcal{Y} = \text{IDWT}([\mathcal{F}_h^N, \mathcal{F}_l^N]) + \mathcal{T}_f^N. \tag{6}$$

Finally, it is reconstructed into a unified spatiotemporal representation $\mathcal{Y} \in \mathbb{R}^{B \times d_{\text{model}} \times W}$.

F. Global Average Pooling and Classifier

In the final stage, we process the fused multi-scale feature tensor $\mathcal{Y} \in \mathbb{R}^{B \times d_{\text{model}} \times W}$ using a global average pooling (GAP) operation along the temporal dimension:

$$\mathcal{Y}_i = \text{GAP}(\mathcal{Y}) \in \mathbb{R}^{B \times d_{\text{model}} \times 1}. \tag{7}$$

This compact representation aggregates global discriminative information and reduces overfitting risks. Subsequently, the tensor is flattened to $[B, d_{\text{model}}]$, passed through a dropout layer, and projected by a fully connected layer into the category space $[B, S]$, where S denotes the number of gesture classes. Cross-entropy loss is used for training, where the logits are transformed into a probability distribution over gesture categories during optimization.

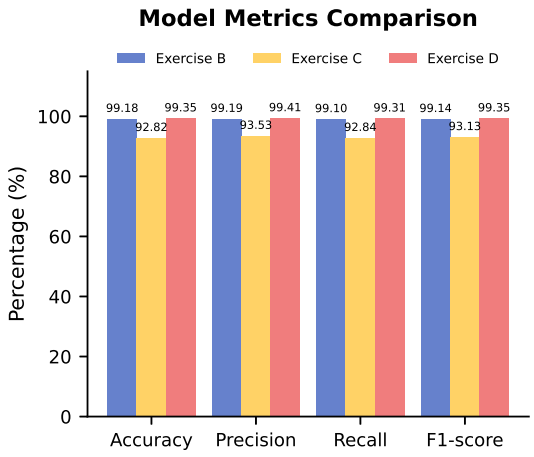

Fig. 5. Four-Metric Evaluation of Model on Exercise B, C, and D Datasets.

IV. Experiment and Analysis

A. Experiment Environment

The MaMSC framework proposed in this paper was trained using PyTorch 2.13 and the CUDA toolkit on an Ubuntu 22.04 system and an NVIDIA RTX 4090 (24GB) GPU. The Adam optimizer was used with early stopping and a dynamic learning rate decay strategy to accelerate convergence and prevent overfitting. The sEMG data was split into 70% training, 10% validation, and 20% testing. In cross-subject experiments, the data are divided into source domain subjects and target domain subjects: in the source domain pre-training phase, by jointly loading data from multiple subjects, general sEMG features are learned with an 80% training set and 20% validation set ratio; in the target domain few-shot adaptation phase, only a small amount of labeled data from the target subject is used for fine-tuning to verify the model's adaptive potential under low calibration costs. The model adopts the cross-entropy loss function. During training, gradient clipping is performed by limiting the norm of the gradient to 1.0 to prevent gradient explosion. The model hyperparameter configuration is as follows: the batch size is 32, and the maximum number of training epochs is 80. The Mamba backbone network has 256 channels, and the learning rate is 6×10^{-5}.

B. Performance on Datasets

To comprehensively assess its performance, we evaluate MaMSC on the Ninapro DB2 dataset using four metrics and confusion matrix analysis.

Four evaluation indicators, including Accuracy, Precision, Recall, and F1 score, are used to measure the performance of the MaMSC framework in the sEMG gesture classification task. As shown in Fig. 5, the proposed model achieves excellent performance across all exercises, achieving the highest accuracy in Exercise D, which has fewer gesture categories. While accuracy slightly decreases in Exercise C, which has more categories, it still maintains a high level of accuracy, validating the model's adaptability and robustness.

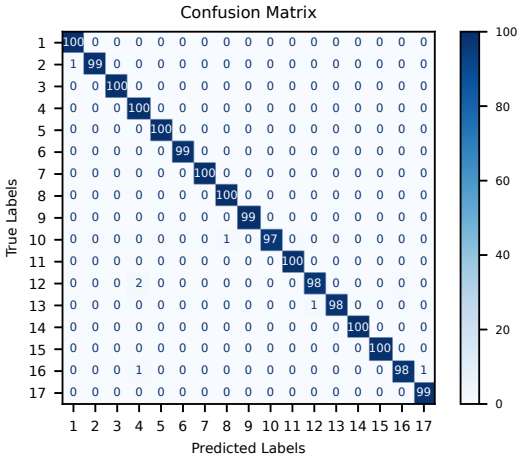

Fig. 6. The confusion matrix of this model on the Exercise B.

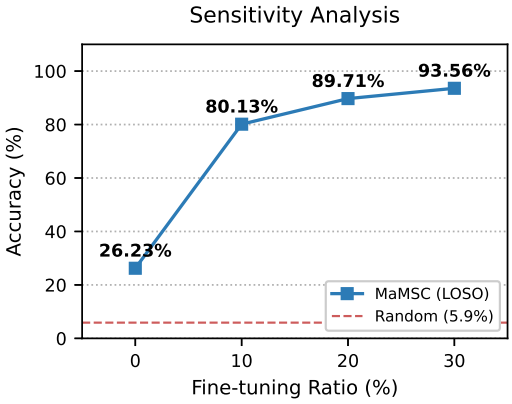

Fig. 7. The trend of changes in the model's classification accuracy when fine-tuning with different proportions of target subject data on the Exercise B dataset.

To further investigate classification performance across individual gesture categories, we analyze a confusion matrix constructed for the 17 categories in Exercise B. As shown in Fig. 6, most prediction counts on the main diagonal reach or approach 100%, while categories 10 have lower prediction counts, but still maintain at or above 97%. Furthermore, the values of the off-diagonal elements are generally low, indicating low misclassification probabilities. Overall, no obvious systematic confusion blocks appear in the matrix, demonstrating that MaMSC is able to effectively distinguish different gesture patterns while maintaining high robustness.

C. Comparison with Referenced Methods

As shown in Table I, the MaMSC framework demonstrates significant advantages in all three tasks: 99.18% accuracy in basic finger movements (Exercise B), 92.82% in grasping and functional movements (Exercise C), and 99.35% in force pattern recognition (Exercise D).

TABLE I
COMPARISON OF MODEL PERFORMANCE WITH DIFFERENT METHODS.

Model	ExerciseB	ExerciseC	ExerciseD
SVM [8]	82.18%	74.08%	91.47%
CNN [10]	91.65%	79.55%	97.67%
CNN+Transformer [14]	90.70%	74.66%	97.79%
CNN+GRU [4]	95.99%	92.75%	96.64%
TransGER [5]	96.44%	87.60%	99.26%
MaMSC	**99.18%**	**92.82%**	**99.35%**

These results consistently outperform existing methods, demonstrating the effectiveness of MaMSC's multi-scale dynamic modeling in handling diverse gesture recognition tasks with high accuracy and robustness.

D. Cross-Subject Generalization Analysis

To comprehensively evaluate the generalization ability of the MaMSC framework across different users, we divided the data in the Exercise B dataset into source domain subject datasets and target domain subject datasets. This study adopted a three-stage leave-one-subject-out validation scheme, which is as follows:

First, in the pre-training stage, the model learns subject-invariant representations on the source domain subjects. Subsequently, in the zero-shot inference stage, the completely isolated target subjects are directly tested. Affected by the covariate shift caused by physiological differences, the initial accuracy is at a low level of 26.23%, which objectively eliminates the risk of data leakage. Finally, in the few-shot adaptation stage, the model is fine-tuned using a very small amount of data from the target subjects. Fig. 7 shows the changing trend of the model's classification accuracy under fine-tuning with different proportions of target subject data. The results indicate that by fine-tuning with only 10% of the target subject's data, the accuracy rapidly climbs to 80.13%; when the proportion increases to 30%, the accuracy further breaks through to 93.56%.

E. Ablation Studies

To explore the role of each core component in the MaMSC framework, we conducted a series of ablation experiments on the Exercise B dataset of Ninapro DB2.

As shown in Table II, the complete MaMSC model achieved an accuracy of 99.18% and an F1 score of 99.14%, converging in only 60 epochs. Through comparison, it was found that: i) w/o HFTD: The accuracy drops to 98.84%, and the F1 score falls to 98.77%, which verifies the importance of time-domain feature extraction for capturing instantaneous gesture details. ii) w/o MamscBlocks: When using a single-path Mamba structure, the model accuracy fluctuates, and the number of training epochs required for convergence increases from 60 to 79, indicating that the multi-scale parallel structure can accelerate feature extraction. iii) w/o IHFTD: When simple feature concatenation is used instead of the physics-based reconstructed IHFTD, the performance is most severely impaired: the accuracy drops significantly to 97.70%, and the convergence time sharply doubles to 150 epochs.

TABLE II
ABLATION STUDY RESULTS

Setting	Accuracy	F1-score	Epochs
w/o HFTD	98.84%	98.77%	60
w/o MamscBlocks	99.06%	98.87%	79
w/o IHFTD	97.70%	97.63%	150
All	99.18%	99.14%	60

V. CONCLUSION

Based on research and experiments, the proposed multi-path Mamba encoding deep learning architecture demonstrates significant advantages in sEMG gesture recognition. This framework leverages HFTD and MamscBlocks to model the multi-scale temporal dependencies of the signal. By jointly modeling high- and low-frequency, as well as temporal residual paths, it captures fast and slow features while preserving temporal context and linking local and global dependencies. Experiments demonstrate superior classification performance on the Ninapro DB2 dataset, surpassing existing methods by achieving leading performance indicators.

VI. ACKNOWLEDGMENT

This work was supported in part by the Shandong Provincial Natural Science Foundation under Grant ZR2023MF040, the Jinan Science and Technology Plan Project under Grant 202527001, and the Jinan City-University Integration Project under Grant JNSX2024043. This research used artificial intelligence tools including ChatGPT and Gemini solely for language polishing. We has verified and revised all AI-assisted content and takes full responsibility for this article.

REFERENCES

[1] S. Sharma, K. N. Faisal, and R. R. Sharma, "Automated grasp recognition using semg: Recent advances, challenges and future developments," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[2] L. Guo, Z. Lu, and L. Yao, "Human-machine interaction sensing technology based on hand gesture recognition: A review," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 4, pp. 300–309, 2021.

[3] A. Fatayer, W. Gao, and Y. Fu, "semg-based gesture recognition using deep learning from noisy labels," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4462–4473, 2022.

[4] S. Song, A. Dong, J. Yu, Y. Han, and Y. Zhou, "A multichannel cnn-gru hybrid architecture for semg gesture recognition," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 4132–4139.

[5] Y. Yuan, A. Dong, W. Xu, Y. Han, J. Yu, and Y. Zhou, "Transger: Transformer-based cnn-bigru architecture for semg gesture recognition in time-frequency domain," in *International Conference on Wireless Artificial Intelligent Computing Systems and Applications*. Springer, 2025, pp. 297–306.

[6] B. Gao, Y. Han, Y. Zhou, J. Yu, S. Li, and A. Dong, "Wireless portable dry electrode multi-channel semg acquisition system," in *International Conference on Wireless Artificial Intelligent Computing Systems and Applications*. Springer, 2024, pp. 124–135.

[7] H. Sharif, S. B. Seo, and T. K. Kesavadas, "Hand gesture recognition using surface electromyography," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 682–685.

[8] T. Prabhavathy, V. K. Elumalai, and E. Balaji, "Hand gesture classification framework leveraging the entropy features from semg signals and vmd augmented multi-class svm," *Expert Systems with Applications*, vol. 238, p. 121972, 2024.

[9] N. K. Karnam, A. C. Turlapaty, S. R. Dubey, and B. Gokaraju, "Classification of semg signals of hand gestures based on energy features," *Biomedical Signal Processing and Control*, vol. 70, p. 102948, 2021.

[10] Z. Xu, J. Yu, W. Xiang, S. Zhu, B. Liu, J. Li *et al.*, "A novel se-cnn attention architecture for semg-based hand gesture recognition." *CMES-Computer Modeling in Engineering & Sciences*, vol. 134, no. 1, 2023.

[11] T. M. Bittibssi, M. A. Genedy, S. A. Maged *et al.*, "semg pattern recognition based on recurrent neural network," *Biomedical signal processing and control*, vol. 70, p. 103048, 2021.

[12] Z. Wang, J. Yao, M. Xu, M. Jiang, and J. Su, "Transformer-based network with temporal depthwise convolutions for semg recognition," *Pattern Recognition*, vol. 145, p. 109967, 2024.

[13] J. Li, B. Zhang, Z. Wang, J. Feng, D. Ye, and X. Zhao, "Optimizing semg-based gesture recognition under non-ideal conditions through a robust deep learning approach," *IEEE Transactions on Instrumentation and Measurement*, 2025.

[14] S. Blade, Z. Yao, Y. Hou, Y. Li, S. Zhou, Y. Wang, and X. Liu, "An semg-controlled prosthetic hand featuring a tiny cnn-transformer model and force feedback," in *2023 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2023, pp. 1–5.

[15] Y. Liu, X. Li, L. Yang, G. Bian, and H. Yu, "A cnn-transformer hybrid recognition approach for semg-based dynamic gesture prediction," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–16, 2023.

[16] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[17] T. Dao and A. Gu, "Transformers are ssms: Generalized models and efficient algorithms through structured state space duality," *arXiv preprint arXiv:2405.21060*, 2024.

[18] H. Zhao, M. Zhang, W. Zhao, P. Ding, S. Huang, and D. Wang, "Cobra: Extending mamba to multi-modal large language model for efficient inference," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 10, 2025, pp. 10 421–10 429.

[19] X. Zhang, Q. Zhang, H. Liu, T. Xiao, X. Qian, B. Ahmed, E. Ambikairajah, H. Li, and J. Epps, "Mamba in speech: Towards an alternative to self-attention," *IEEE Transactions on Audio, Speech and Language Processing*, 2025.

[20] S. Gong, Y. Zhuge, L. Zhang, Y. Wang, P. Zhang, L. Wang, and H. Lu, "Avs-mamba: Exploring temporal and multi-modal mamba for audio-visual segmentation," *IEEE Transactions on Multimedia*, vol. 27, pp. 5413–5425, 2025.

[21] Q. Zhao, Z. Li, Q. Mao, T. Chen, Y. Zhang, B. Li, Z. Zhao, and X. Fan, "Mambacpg: an accurate model for single-cell dna methylation status imputation using mamba," *Briefings in Bioinformatics*, vol. 26, no. 4, p. bbaf360, 2025.

[22] T. Wang, X. Xie, W. Wang, C. Wang, Y. Zhao, and Y. Cui, "Netmamba: Efficient network traffic classification via pre-training unidirectional mamba," in *2024 IEEE 32nd International Conference on Network Protocols (ICNP)*. IEEE, 2024, pp. 1–11.

[23] F. Luo, A. Li, B. Jiang, S. Khan, K. Wu, and L. Wang, "Activitymamba: a cnn-mamba hybrid neural network for efficient human activity recognition," *IEEE Transactions on Mobile Computing*, 2025.

[24] Y. Qiang, X. Dong, X. Liu, Y. Yang, F. Hu, and R. Wang, "Ecgmamba: Towards ecg classification with state space models," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2024, pp. 6498–6505.

[25] H. Ye, G. Duan, H. Zeng, Y. Zhu, L. Meng, X. Zheng, and Y. Zhu, "Karma: A multilevel decomposition hybrid mamba framework for multivariate long-term time series forecasting," in *International Conference on Wireless Artificial Intelligent Computing Systems and Applications*. Springer, 2025, pp. 266–276.

[26] T. Tanaka, I. Nambu, and Y. Wada, "Mitigating the impact of electrode shift on classification performance in electromyography applications using sliding-window normalization," *Sensors*, vol. 25, no. 13, p. 4119, 2025.

[27] 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, 2015.

[28] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE transactions on information theory*, vol. 36, no. 5, pp. 961–1005, 2002.