

Continuous Prediction of Wrist Joint Kinematics Using Surface Electromyography from the Perspective of Muscle Anatomy and Muscle Synergy Feature Extraction

Zijun Wei, Meiju Li, Zhi-Qiang Zhang, *Member, IEEE*, Sheng Quan Xie, *Fellow, IEEE*

Abstract—Post-stroke upper limb dysfunction severely impacts patients' daily life quality. Utilizing sEMG signals to predict patients' motion intentions enables more effective rehabilitation by precisely adjusting the assistance level of rehabilitation robots. Employing the muscle synergy (MS) features can establish more accurate and robust mappings between sEMG and motion intentions. However, traditional matrix factorization algorithms based on blind source separation still exhibit certain limitations in extracting MS features. This paper proposes four deep learning models to extract MS features from four distinct perspectives: spatiotemporal convolutional kernels, compression and reconstruction of sEMG, graph topological structure, and the anatomy of target muscles. Among these models, the one based on 3DCNN predicts motion intentions from the muscle anatomy perspective for the first time. It reconstructs 1D sEMG samples collected at each time point into 2D sEMG frames based on the anatomical distribution of target muscles and sEMG electrode placement. These 2D frames are then stacked as video segments and input into 3DCNN for MS feature extraction. Experimental results on both our wrist motion dataset and public Ninapro DB2 dataset demonstrate that the proposed 3DCNN model outperforms other models in terms of prediction accuracy, robustness, training efficiency, and MS feature extraction for continuous prediction of wrist flexion/extension angles. Specifically, the average nRMSE and R^2 values of 3DCNN on these two datasets are (0.14/0.93) and (0.04/0.95), respectively. Furthermore, compared to existing studies, the 3DCNN outperforms musculoskeletal models based on direct collocation optimization, physics-informed GANs, and CNN-LSTM-based deep Kalman filter models when evaluated on our dataset.

Index Terms—Surface electromyography (sEMG), upper-limb rehabilitation, deep learning, muscle synergy, muscle anatomy, continuous joint kinematics estimation methods.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the MaPS and Engineering Joint Faculty Research Ethics Committee of the University of Leeds under Application No. MEEC18-002.

Zijun Wei, Zhi-Qiang Zhang, and Sheng Quan Xie are with the School of Electronic and Electrical Engineering, University of Leeds, LS2 9JT Leeds, U.K. (e-mail: elzw@leeds.ac.uk; z.zhang3@leeds.ac.uk; s.q.xie@leeds.ac.uk).

Meiju Li is with the School of Information Science and Engineering, Zaozhuang University, 277100, Zaozhuang, China. (e-mail: meiju683@gmail.com).

I. INTRODUCTION

Post-stroke hemiplegia significantly impacts patients' basic activities of daily living (ADLs), with 60% of patients still exhibiting upper limb impairments, notably in finger and wrist

joints, following the stroke rehabilitation [1]. Consequently, the rapid rehabilitation of hand and wrist joints is crucial. With the advent of the neuroplasticity theory and rapid advancements in rehabilitation robotics, using these robots to accelerate upper limb recovery has become increasingly popular. Research [2] indicates that active patient participation and training intensity are key factors in improving the therapeutic outcomes of rehabilitation robots. However, most robots based on traditional impedance and admittance control primarily focus on passive rehabilitation, failing to maximize patient participation, thereby limiting recovery outcomes. Hence, it is essential to develop the "Assistance-as-Needed" (AAN) control strategies based on motion intention prediction, which can precisely adjust the assistance level of robots in real-time according to the patient's effort level, thereby maximizing the active participation [2]. Previous studies extensively utilized surface electromyography (sEMG) for motion intention prediction, and these studies are generally categorized into two types: pattern recognition based on discrete classification, and continuous regression for joint kinematics and dynamics estimation [3], [4], [5]. Considering that the AAN control strategy pertains to continuous control, this study will further explore continuous estimation methods of joint kinematics based on sEMG.

Based on our previous systematic review [6], which analyzed 186 studies over the past decade on continuous estimation of upper limb kinematics and dynamics using sEMG, it revealed that extracting muscle synergy (MS) features is crucial for establishing the precise, robust, and long-duration mappings between sEMG and motion intentions. MS theory posits that the central nervous system coordinates the activity of muscle groups rather than controlling individual muscles [7], [8]. MS co-activates specific muscle groups as synergistic units. These units are modularly combined with varying weights to simplify the complexity and the high dimensionality of motor control, thereby addressing the redundancy in degrees of freedom (DoFs) in motor control [7], [9]. From the perspective of time-varying model, sEMG can be represented as a linear combination of the MS matrix (spatial component) multiplied by the corresponding muscle activation coefficient matrix (temporal component), plus the residual error matrix [8], [9], [10], [11], [12], [13]. Therefore, most studies on MS feature extraction focus on matrix factorization algorithms (MFAs) based on Blind Source Separation (BSS) and methods based on deep learning (DL). Common MFAs include Non-negative Matrix Factorization (NMF), which is the most widely used, as well as Principal Component Analysis (PCA), Factor Analysis (FA), and Independent Component Analysis (ICA), whereas DL-based methods typically involve Autoencoders (AE) [8], [9], [14].

However, common MFAs still have flaws. Given the non-orthogonality of actual muscle activation patterns and the non-Gaussian nature of sEMG signals, PCA imposes additional constraints on muscle activation and, like FA, performs worse than NMF and ICA which suited for non-Gaussian data [9]. Furthermore, PCA lacks non-negativity constraints and only performs linear transformations, leading to extract MS features with poor consistency and are susceptible to noise and outliers [10], [11]. Similarly, ICA is also sensitive to noise and outliers [9]. As for NMF, although its non-negativity constraint aligns with the physiological nature of neural commands and muscle activation, the non-sparse MSs resulting from multi-joint and multi-DoF movements, as well as abnormal muscle activation in stroke patients, can trap NMF in local optima [9], [10], [11]. Moreover, due to the sparsity of NMF and the lack of constraints on the arrangement of decomposed MSs, the stability and reproducibility of MS features extracted by NMF are poor [10], [11], [15]. Regarding the common flaws, all MFAs perform better with an increasing number of sEMG channels and require manual setting of thresholds, such as the Variance Accounted For (VAF), to adjust the number of MSs and reconstruction errors, while the ignored higher-order MSs representing fine motor patterns may still improve predictive performance [8], [10], [11], [13], [16], [17]. In contrast to linear MFAs, the non-linear nature of AEs makes them more suitable for analyzing non-linear and non-stationary sEMG signals, thus outperforming the PCA and NMF in extracting MSs [9]. For instance, study [18] have demonstrated that the Constrained Autoencoder Network (CAEN) outperforms NMF and NMF-HP (Hadamard product) in offline and online prediction of finger force. Studies [19], [20] also indicated that although NMF can reconstruct muscle activation signals better than AEs, AEs achieve the best balance between muscle activation reconstruction and motion intention estimation, thus making AEs superior in predicting shoulder and elbow joint torques.

Given that the focus of this paper is on predicting motion intentions rather than reconstructing muscle activations and considering the inherent limitations of MFAs in extracting MS features, this paper will further explore other DL-based MS feature extraction algorithms beyond MFAs. Since the graph-based topological structure can effectively capture the dynamic interaction relationship between nodes, Graph Neural Networks (GNNs) also have great potential in extracting MS features. Various GNN-based algorithms such as Graph Convolutional Network (GCN), GraphSAGE (Graph Sample and aggregate), and Graph Attention Network (GAT) have been extensively employed in the gesture pattern recognition and continuous estimation of wrist joint angles and grip force [21], [22], [23], [24], [25], [26], [27], [28]. However, only study [28] has addressed continuous estimation of wrist joint multi-DoF kinematics and dynamics, with other studies focusing solely on pattern recognition. Moreover, except for the studies [26], [27], the remaining studies have only been tested on the high-density sEMG (HD-sEMG), thereby overlooking the performance of GNNs on commonly used sparse sEMG electrodes.

Our previous systematic review [6] and recent research [29] have showcased that Convolutional-Bidirectional Long Short-Term Memory hybrid networks (CNN-BiLSTMs) integrated with Multi-Head Self-Attention mechanisms (MSA) exhibit outstanding performance in predicting motion intentions. The

CNN autonomously extracts advanced features, BiLSTM establishes long-term contextual dependencies among extracted features, and MSA finally enhances the weighting of global key features. Therefore, to comprehensively extract the MS features by including both the spatial and temporal components of sEMG signals, this study constructs four models based on the CNN-BiLSTM-MSA framework. Each model incorporates the same BiLSTM-MSA (BM) module, while modifying only the CNN module, enabling the extraction of MS features from four distinct perspectives.

These four models are as follows: multi-scale CNN with additional spatiotemporal convolutional kernels (MCNN-BM), convolutional AE based on sEMG signal compression and reconstruction loss (CAE-BM), GraphSAGE based on the graph topology (GraphSAGE-BM), and 3D-CNN based on the anatomical structure of target muscles and sEMG electrode placement (3DCNN-BM). Unlike traditional 1D-CNN and 2D-CNN algorithms, the 3DCNN-BM adopts a video analysis approach, instead of treating sEMG signals as 1D time series or 2D grid-like images composed of time and channels. It reconstructs 1D sEMG samples collected at individual time point into a 2D sEMG time frame according to the anatomy of target muscles and sEMG electrode placement. Each 2D sEMG frame represents the muscle activation intensity and the corresponding co-contraction/relaxation ratio of the target muscles at the anatomical cross-section related to the electrode placement at each time point. Subsequently, all 2D frames from each sliding time window are stacked into video segments for MS feature extraction using the 3DCNN-BM model.

These four models were validated on our wrist motion dataset and the Ninapro DB2 dataset [30], respectively, and were compared to the baseline models (CNN-BiLSTM, CNN, BiLSTM, LSTM) by predicting the wrist flexion/extension angles. Results indicate that the novel 3DCNN-BM model outperforms other models in terms of prediction accuracy, robustness, training efficiency, and MS feature extraction.

Overall, the four DL-based models proposed in this study collectively establish a multi-perspective research framework for MS feature extraction. Among these, the 3DCNN-BM serves as the core model, representing the major contribution of this paper. MCNN-BM, CAE-BM, and GraphSAGE-BM offer valuable explorations and supplements, and their comparison with the 3DCNN-BM further emphasizes the effectiveness of extracting MS features from the perspective of muscle anatomy. To our knowledge, this is the first study to extract MS features from sEMG for motion intention prediction from the muscle anatomy perspective. Given that the sEMG can be decomposed into the spatial and temporal components, the 3DCNN-BM represents sEMG frames as spatial components and then stacks multiple frames as videos to concurrently represent both spatial and temporal components, thereby providing BSS with prior knowledge based on the muscle anatomy to overcome the limitations of MFAs. Additionally, the 3DCNN-BM's training time on both datasets was less than one minute, simplifying the retraining process for both intra-subject and inter-subject scenarios and reducing the reliance on transfer learning.

II. MATERIALS

A. IRASS Dataset

The collection of this dataset was approved by the MaPS and Engineering Joint Faculty Research Ethics Committee of the University of Leeds (MEEC 18-002), involving 12 subjects (8 males and 4 females, age 28.9 ± 2.8 years) who signed informed consent forms. As for the experimental protocol, as depicted in Figure 1, all subjects were instructed to sit in an armchair with their trunks upright, shoulders abducted to 90 degrees, elbows flexed to 90 degrees, and hands positioned neutrally. Before the experiment, the maximum voluntary contraction (MVC) of each subject was recorded. Subsequently, each subject was required to perform continuous wrist flexion and extension movements while keeping the metacarpophalangeal (MCP) joints fully extended and avoiding the ulnar/radial deviation, with the protocol repeated five times. To prevent muscle fatigue, subjects rested for three minutes between each trial.



Fig. 1. Experimental protocol of IRASS dataset [31].

Regarding the experimental equipment, the joint angles were recorded by the VICON motion capture system (VICON Motion Systems Ltd., UK) at a sampling rate of 250 Hz. Sixteen reflective markers were placed on the right arm of subjects to correspond with VICON's upper limb model, and the wrist joint angles were calculated using VICON Nexus software based on the upper limb model. sEMG signals were collected using Avanti sensors (Delsys Trigno™) at a sampling frequency of 2000 Hz. Prior to signal acquisition, hair on the subject's arm was shaved, and the skin was cleaned with alcohol to reduce motion artifacts and impedance, thereby ensuring signal quality. In addition, the placement of sEMG electrodes was determined by palpation and quantitative signal evaluation, following the SENIAM recommendation. The five primary wrist muscles sampled consist of two wrist flexors and three wrist extensors: flexor carpi radialis (FCR), flexor carpi ulnaris (FCU), extensor carpi ulnaris (ECU), extensor carpi radialis longus (ECRL), and extensor carpi radialis brevis (ECRB). Finally, the wrist angle data and sEMG data were synchronized and resampled at 1000 Hz through the VICON Nexus software's trigger module.

B. Open-Access Ninapro DB2 Dataset

Specifically, the primary purpose of the Ninapro DB2 dataset [30] is to improve the accuracy of gesture recognition for myoelectric prostheses. The majority of studies utilizing this dataset focus on gesture classification rather than continuous regression. Nevertheless, this dataset still recorded joint angles during different gestures, making it also applicable to continuous regression studies of joint kinematics. Therefore, this dataset was selected for this study, and 12 subjects (8 males

and 4 females, age 29.9 ± 3.9 years) were chosen to match the number of subjects in the IRASS dataset.

Regarding the experimental protocol, as shown in Figure 2(a), subjects were asked to sit in an optimally adjusted chair with relaxed arms, and then mimicked the movements displayed on a laptop screen in front of them with right hand for each task. Each motion was repeated six times and lasted five seconds, with three-second rest intervals between motions to prevent muscle fatigue. To maintain consistency with the IRASS dataset, as shown in Figure 2(b), only wrist flexion/extension movements from Exercises B were selected for prediction.

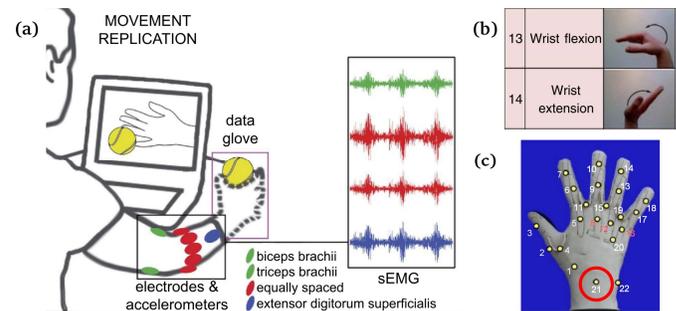


Fig. 2. (a) Experimental protocol of Ninapro DB2 dataset. (b) Wrist flexion/extension in Exercises B. (c) The CyberGlove II dataglove and the glove's 21st sensor.

As for the experimental equipment, as depicted in Figure 2(c), the joint angle was measured using the CyberGlove II dataglove equipped with 22 sensors, and the sEMG signals were captured using the same 12 Delsys Trigno wireless electrodes as in the IRASS dataset. As shown in Figure 2(a), the placement of the sEMG electrodes was determined by palpation, with eight electrodes evenly distributed around the forearm and the remaining four placed on the superficial flexor digitorum, superficial extensor digitorum, biceps brachii, and triceps brachii. Additionally, since the focus of this study was solely on wrist flexion/extension, only angle data from the glove's 21st sensor was utilized. Lastly, to maintain consistency with the sampling rate of the IRASS dataset, both angle data and sEMG data were resampled at a frequency of 1000 Hz.

It is important to note that the CyberGlove II dataglove of Ninapro DB2 dataset was not calibrated, leading to angle data variations due to differences in hand size among subjects and the susceptibility to other joint movements. Therefore, the angle data utilized in this study was obtained from the Ninapro DB9 dataset [32], which underwent 13 calibration trials to determine specific gain values of the 21st sensor, subsequently calculating the calibrated wrist flexion/extension angles. Additionally, since wrist flexion/extension movements are minimally related to the biceps and triceps brachii, only the data from the other 10 sEMG electrodes were considered.

C. Signal Preprocessing

Preprocessing of the collected data encompasses three stages: filtering and denoising, normalization, and sliding window segmentation. To minimize variability, the same preprocessing methods were applied to both the IRASS dataset and the Ninapro DB2 dataset. According to [33], the power of sEMG signals is mostly between 10-400Hz, with components outside this range mainly consisting of noise and motion artifacts, and

the highest frequency component of sEMG signals is between 450-500 Hz. Therefore, initial preprocessing of raw sEMG data involved using a 5th-order Butterworth band-pass filter from 15-450 Hz, followed by a 50 Hz notch filter to eliminate power line interference. Subsequently, the full-wave rectification and a 4Hz 5th-order Butterworth low-pass filter were employed to obtain a smoother sEMG signal envelope, ensuring the stability of model training. Finally, the envelope sEMG signals were normalized based on the MVC values measured before the experiment, while the Ninapro dataset's envelope sEMG signals

underwent Z-score normalization. As for the joint angle signals, as recommended in [30], they were initially smoothed with a 1Hz 5th-order Butterworth low-pass filter, and then linearly scaled to the [-1, 1] range using max-min normalization. As for sliding window segmentation, considering that time delays in real-time control should not exceed 300 milliseconds (ms), and shorter sliding windows could degrade prediction performance [34], this study used a 200 ms window size with an 80% overlap rate to balance the model latency and predictive performance.

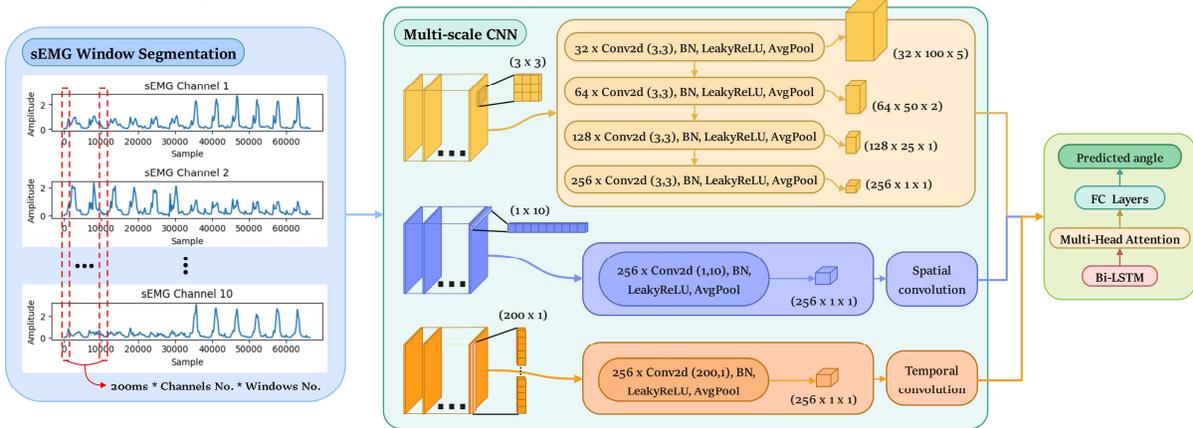


Fig. 3. The proposed MCNN-BM model demonstrated using the Ninapro DB2 dataset.

III. METHODS

In this section, the details of the four proposed models for extracting MS features from four distinct perspectives, along with baseline models, are introduced. These details include the main framework of models, network architecture, and designed loss functions. Furthermore, the hyperparameter settings and the training configurations applied during the model training are discussed, as well as the model performance metrics.

A. MCNN-BM Model

Figure 3 illustrates the MCNN-BM model. To address the limitations imposed by the limited sizes of conventional CNN kernels, which constrain the extraction of spatial and temporal components from grid-structured sEMG images, the MCNN-BM incorporates additional temporal and spatial convolutional kernels alongside the standard 3×3 kernels to enhance the spatiotemporal feature extraction.

The temporal kernel size is defined as (sliding window size \times 1), and the spatial kernel size as ($1 \times$ number of electrode channels). Consequently, the temporal kernel extracts the temporal components of each target muscle corresponding to the sEMG electrode within each sliding window. Spatial kernel, on the other hand, captures the spatial components by extracting synergistic relationships among all target muscles at each time point within each sliding window. Additionally, to balance the weights of features extracted by the standard, temporal, and spatial kernels, the average pooling layers are employed to maintain consistent output feature shapes from these kernels.

These output features are then concatenated and fed into the subsequent BM module and fully connected (FC) layers. The single-layer BiLSTM establishes the long-term contextual dependencies among features, while the four-headed MSA refines these dependencies and increases the weight of key

dependencies from four distinct global perspectives, with the FC layers ultimately establishing nonlinear mappings between extracted MS features and joint angles. Notably, to ensure a fair comparison of MS feature extraction capabilities across various models, the subsequent three models adopt the same BM module and FC layers configuration as the MCNN-BM model.

B. CAE-BM Model

Figure 4 illustrates the CAE-BM model, which integrates CNN with AE to extract MS features based on the sEMG signal reconstruction loss. Centered around the latent vector, the CAE module of the model can be divided into two parts: the encoder on the left and the decoder on the right.

The encoder consists of four down-sampling modules based on CNNs with 3×3 kernels and the FC encoder layer, while the decoder comprises a FC reconstruction layer and four CNNs up-sampling modules corresponding to the down-sampling modules. The down-sampling modules progressively compress the input sEMG images by gradually doubling the number of CNN channels and using average pooling layers to extract advanced features. The FC encoder layer further compresses and reduces the dimensionality of advanced features to reduce redundant information, ultimately obtaining the latent vector with key MS information. Following preliminary upscaling and reconstruction of the latent vector by the FC decoder layer, the up-sampling modules based on transposed CNNs progressively reconstruct the latent vector into the high-dimensional representations that match the shape of the corresponding down-sampling modules. The final and most crucial step involves introducing the Mean Squared Error (MSE)-based reconstruction loss function between each down-sampling and corresponding up-sampling module, thereby minimizing the difference between the extracted and reconstructed features to ensure the integrity of MS feature extraction.

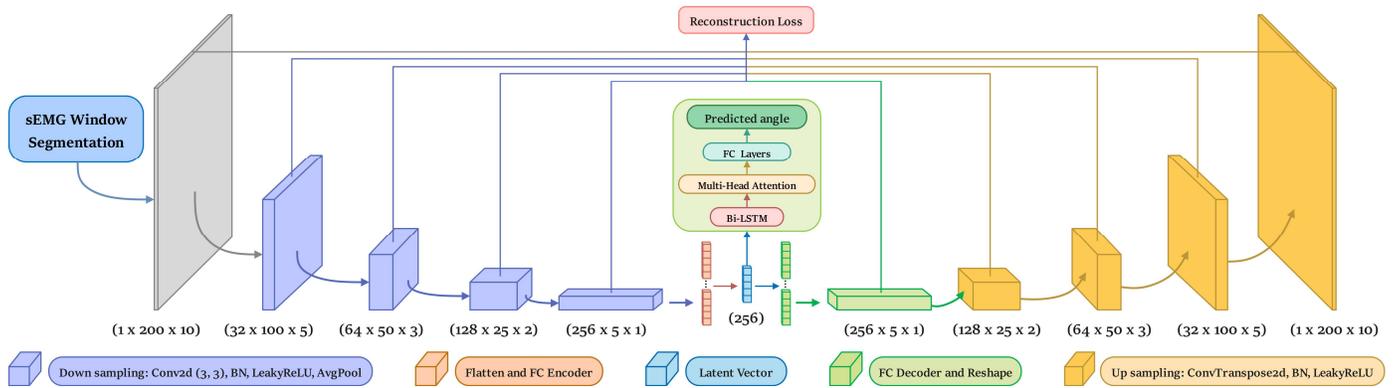


Fig. 4. The proposed CAE-BM model demonstrated using the Ninapro DB2 dataset.

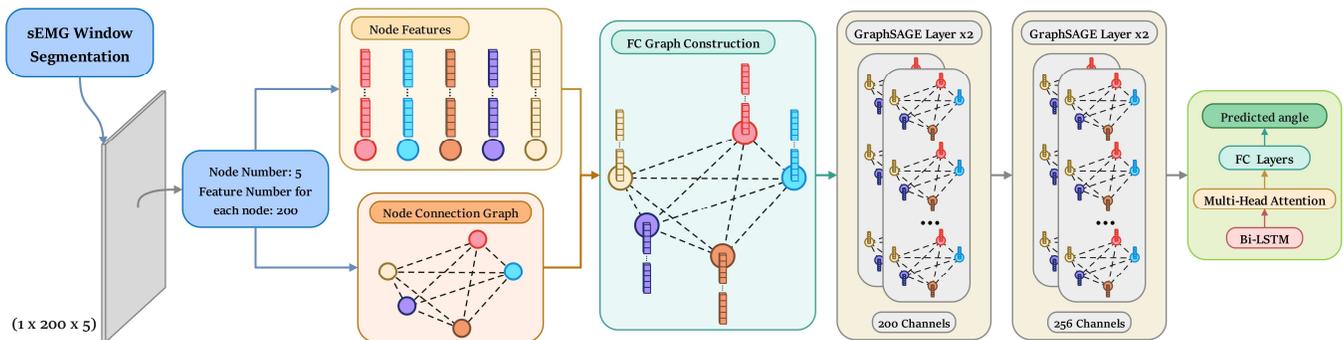


Fig. 5. The proposed GraphSAGE-BM model demonstrated using the IRASS dataset.

C. GraphSAGE-BM Model

Figure 5 illustrates the process of transforming sEMG signal images into graph-structured data, followed by the MS feature extraction through the GraphSAGE-BM model. In this model, each sEMG electrode is considered an independent node, with the connections between nodes defined as edges [35]. Since the length of sEMG sliding window is set to 200 ms, each node possesses 200 features.

Notably, the form of edge connections significantly impacts the performance of GNNs [23]. Due to the inherent sparsity of MSS, not all sEMG electrode channels are equally significant at any given moment. For instance, during the wrist flexion movements, signals from the flexor muscles are more crucial than those from the extensor muscles. Consequently, not all edges convey useful information for prediction at the same time and the redundant edges can lead to overfitting and performance degrade, while overlooking edges with critical MS relationships can also reduce the model's predictive performance [24].

Unlike traditional GCNs that rely on the overall structure of the graph and process all nodes simultaneously at each layer, the inductive learning based GraphSAGE randomly samples a fixed number of neighbors during each aggregation step instead of relying on all neighbors. By experimenting with different sampling combinations, GraphSAGE focuses on nodes and edges rich in critical information after several iterations, not only improving prediction accuracy but also serving as a regularization strategy to mitigate the risk of overfitting. Hence, this study initially established FC edges among nodes, then employed GraphSAGE to identify the key nodes and edges that precisely represent MS relationships within this FC graph.

D. 3DCNN-BM Model

Figure 6 illustrates the construction process of the core model proposed in this paper. Based on the biomechanical principles of joint movements, the flexion/extension of the wrist joint are primarily driven by the coordinated contraction/relaxation (C/R) of the flexors and extensors, respectively. This antagonistic effect transmits the muscle-generated force to the bones through the tendons, thereby driving the corresponding movement of the wrist joint.

Therefore, the core idea of the 3DCNN-BM model is that if this process is observed on the anatomical cross-section of the forearm, it can be observed that the individual muscles on the cross-section have varying degrees of C/R at different moments during the wrist flexion and extension. Notably, at any given moment, the relative C/R ratio among target muscles is the MS representing the spatial component in the sEMG signal, while the C/R level of each individual muscle is the muscle activation coefficient representing the temporal component in the sEMG signal. Therefore, the 3DCNN-BM model firstly converts the 1D sEMG samples collected at each time point into the 2D anatomical slices, and then stacks these 2D slices from multiple time points within the sliding window to synthesize the 3D C/R process of the target muscles on the anatomical cross-section. Finally, this 3D C/R segment is input into the 3DCNN-BM model in video format for extracting MS features.

Figures 6(a) and 6(b) depict the process of reconstructing 1D sEMG samples from the IRASS datasets and the Ninapro DB2 dataset into 2D sEMG frames based on the anatomical structure of forearm cross-section. Taking IRASS dataset as an example,

the anatomical distribution of the five target muscles can be obtained from the Complete Anatomy software (3D4Medical Ltd.). However, a certain distance still exists between the center of the target muscles and the sEMG electrodes placed on the skin surface. Therefore, to minimize the prediction errors caused by the imprecise matching between the localization of target muscles and sEMG samples in 2D frames, the midpoint of the line connecting the sEMG electrodes and the center of target muscles is utilized as the new positioning of target muscles in the 2D frame matrix. Ultimately, five electrodes from the IRASS dataset are reconstructed into a 4×6 sEMG

frame, and ten electrodes from the Ninapro DB2 dataset are reconstructed into a 7×7 sEMG frame. Figure 6(c) illustrates the specific framework of the 3DCNN-BM model. Taking the IRASS dataset as an example, after stacking the sEMG frames as the sEMG video segments, multi-scale 3DCNN convolution kernels similar to those used in the MCNN-BM model are employed to extract MS features from these video segments. Specifically, the 1×6 and 4×1 3DCNN kernels are employed to extract MS features from other two different perspectives, thereby complementing the 4×6 full-coverage kernels.

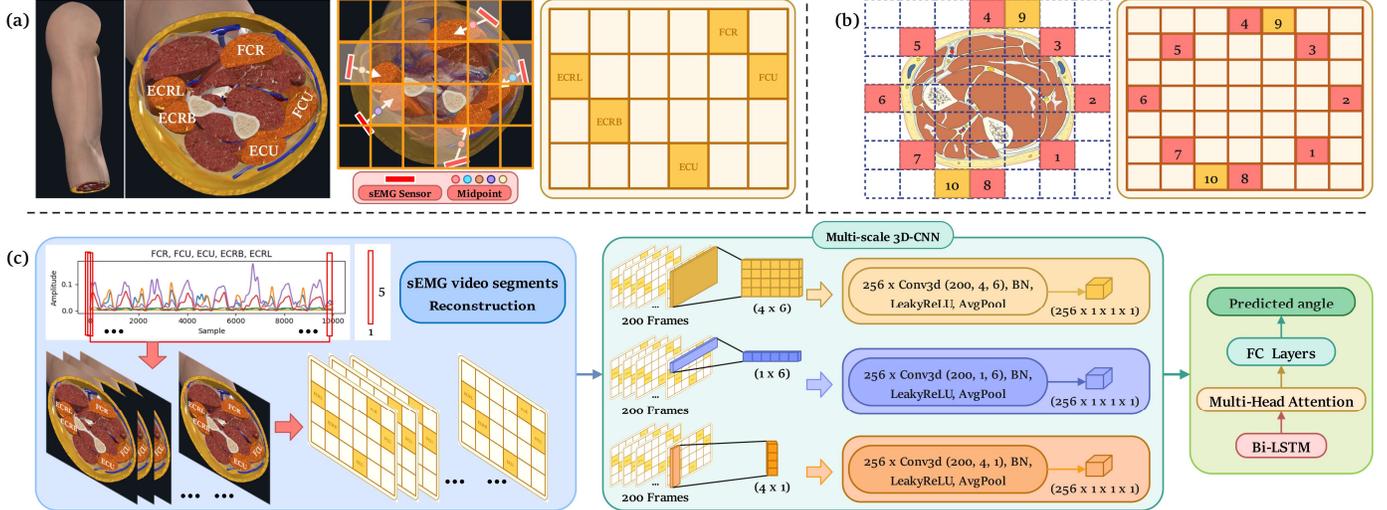


Fig. 6. (a) The sEMG frame reconstruction process for the IRASS dataset. (b) The sEMG frame reconstruction process for the Ninapro DB2 dataset. (c) The proposed 3DCNN-BM model demonstrated using the IRASS dataset.

E. Baseline Models

The four selected baseline models (CNN-BiLSTM, CNN, BiLSTM, and LSTM) have been widely used in recent years for predicting motion intention based on sEMG signals. To fairly verify and compare the predictive performance of the proposed MCNN-BM, CAE-BM, GraphSAGE-BM, and 3DCNN-BM models against these baseline models, these four baseline models were modified based on the MCNN-BM model.

Specifically, the CNN-BiLSTM model omitted the temporal and spatial convolutional kernels from the MCNN module, retaining only the conventional 3×3 kernels shown in Figure 3. Additionally, this model removed the MSA from the BM module, retaining only the single-layer BiLSTM and FC layers. The CNN model further removed the BiLSTM from the CNN-BiLSTM model, retaining only the CNN with 3×3 convolution kernels and the FC layers. As for the BiLSTM model, it consists of two layers of BiLSTM and FC layers, with each BiLSTM layer's input dimension matching the size of the sEMG sliding window. Lastly, the LSTM model shares the same framework as the BiLSTM model, except that the BiLSTM is replaced with the LSTM.

F. Hyperparameter Settings and Model Training

All models were constructed and trained on Google's Colab platform using the Pytorch DL framework, with identical hyperparameter settings employed. Specifically, all models were trained using the adaptive moment estimation with weight decay (AdamW) optimizer, with the dynamic learning rate set

at 0.001 and the mini-batch size of 256. Moreover, the Dropout rate was set at 20% and the slope for the LeakyReLU activation function was set at 0.1, where the learning rate, mini-batch size, and dropout rate were manually adjusted through multiple experimental tests. Additionally, to ensure the complete convergence of all models and to facilitate the comparison of training time across models, the number of training epochs was set to 1000. Finally, all models utilized the loss function based on the MSE between predicted and actual angles. Notably, the CAE-BM model's loss function incorporated not only the angle-based MSE loss but also included reconstruction losses across five different levels of down-sampling and corresponding up-sampling modules. Therefore, the final loss functions of the CAE-BM model are as follows:

$$MSE_{\text{reconstruct}} = \sum_{i=1}^5 MSE_{\text{Level } i} \quad (1)$$

$$L_{\text{CAE-BM}} = MSE_{\text{angle}} + \alpha MSE_{\text{reconstruct}} \quad (2)$$

Where α is the hyperparameter used to balance the weight between the angle loss and reconstruction loss, which was set at 0.5 through multiple experimental tests.

Regarding the training and testing data, data collected from each subject was divided into 80% for training and 20% for testing. Moreover, considering the limited computational resources available on rehabilitation devices in practical applications, training was performed using only the entry-level NVIDIA Tesla T4 GPU on the Colab platform, which offers

similar computational performance to the NVIDIA GeForce RTX 2070 GPU.

G. Performance Metrics

The normalized Root Mean Square Error (nRMSE) and the Coefficient of Determination (R^2) were employed to assess the magnitude of errors and the discrepancy in correlations between the actual and predicted wrist joint angles, thereby quantifying the predictive performance of all models. Compared to RMSE, nRMSE is less sensitive to the size of the data volume and the range of outputs, making it more suitable for comparing model performance across different datasets. Moreover, R^2 provides a more comprehensive evaluation of model's overall predictive accuracy compared to nRMSE. Specifically, the lower nRMSE and the R^2 closer to 1 indicate better predictive performance of the model. The definitions of nRMSE and R^2 are as follows:

$$\text{nRMSE} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\theta_i - \hat{\theta}_i)^2}}{\theta_{\max} - \theta_{\min}} \times 100\% \quad (3)$$

$$R^2 = \left(1 - \frac{\text{Var}(\theta - \hat{\theta})}{\text{Var}(\theta)}\right) \times 100\% \quad (4)$$

Where θ and $\hat{\theta}$ represent the actual and predicted wrist joint angles, respectively, and N denotes the sample size.

Additionally, to comprehensively assess the performance of models in real-world applications, the training time of models was also considered as one of the evaluation criteria.

IV. RESULTS

A. Results of All Models

1) IRASS Dataset

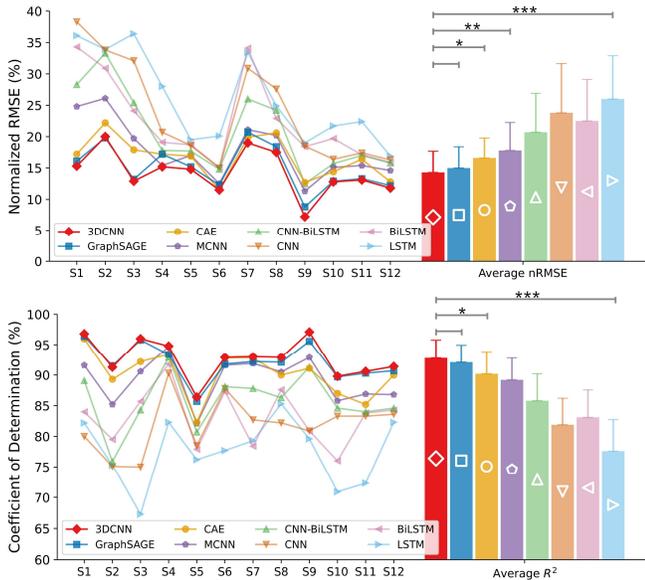


Fig. 7. nRMSE/ R^2 results for all models on the IRASS dataset, as well as the average nRMSE/ R^2 values and the corresponding standard deviations for each model. The significance level is set as 0.05 ($***p < 0.001$, $**p < 0.01$, and $*p < 0.05$).

Figure 7 illustrates the nRMSE and R^2 results of the four proposed models and the four baseline models on the IRASS

dataset. It is evident that the 3DCNN-BM model consistently outperforms all other models for each subject. Specifically, this model achieved the lowest average nRMSE (14.3%) and the highest average R^2 (92.8%). Additionally, the error bars on the histograms indicate that 3DCNN-BM offers the best stability, with the least variation in prediction results. The performance of other models, ranked from highest to lowest, along with their corresponding average nRMSE and average R^2 are as follows: GraphSAGE-BM (15%/92.1%), CAE-BM (16.6%/90.2%), MCNN-BM (17.8%/89.2%), CNN-BiLSTM (20.7%/85.8%), BiLSTM (22.5%/83.1%), CNN (23.8%/81.9%), and LSTM (26%/77.6%).

To evaluate the performance advantage of 3DCNN-BM over other comparative models, Repeated Measures ANOVA was employed for paired analysis between the 3DCNN-BM and other models, followed by the post-hoc analysis using Tukey's Honest Significant Difference (HSD) test, with the significance level set at $p < 0.05$. Statistical analysis revealed significant differences between 3DCNN-BM and all other models except GraphSAGE-BM ($p < 0.05$). Although 3DCNN-BM exhibited better predictive performance than GraphSAGE-BM, the difference was not statistically significant ($p > 0.05$).

Overall, the 3DCNN-BM and GraphSAGE-BM significantly outperform other models on the IRASS dataset, with 3DCNN-BM demonstrating the greatest accuracy and stability in extracting MS features.

2) Ninapro DB2 Dataset

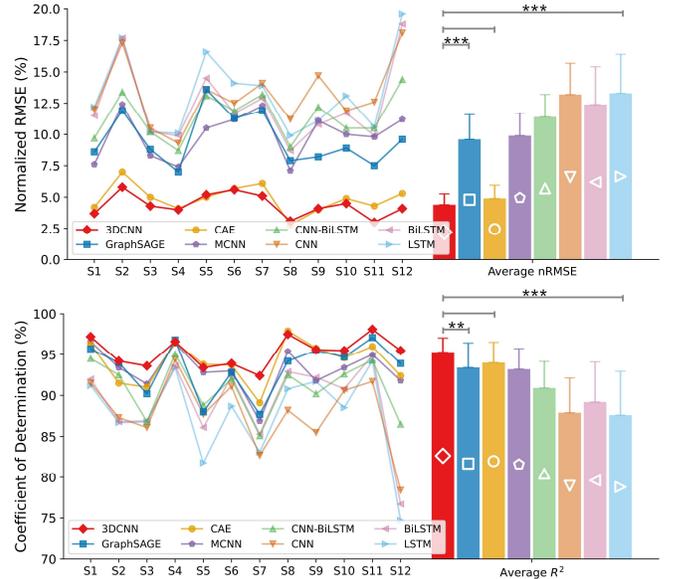


Fig. 8. nRMSE/ R^2 results for all models on the Ninapro DB2 dataset, as well as the average nRMSE/ R^2 values and the corresponding standard deviations for each model. The significance level is set as 0.05 ($***p < 0.001$ and $**p < 0.01$).

Figure 8 presents the nRMSE and R^2 results for all models on the Ninapro DB2 dataset. Similarly, 3DCNN-BM maintains the best predictive performance with the lowest variation. Specifically, it achieved the lowest average nRMSE (4.4%) and the highest average R^2 (95.3%). The performance of other models in descending order along with their corresponding average nRMSE and R^2 are as follows: CAE-BM (4.9%/94.1%), GraphSAGE-BM (9.6%/93.4%), MCNN-BM (9.9%/93.2%),

CNN-BiLSTM (11.4%/90.9%), BiLSTM (12.4%/89.2%), CNN (13.2%/87.9%), and LSTM (13.3%/87.6%). Statistical analysis showed that 3DCNN-BM's predictive performance was significantly better than that of all models except CAE-BM ($p < 0.05$). Although 3DCNN-BM slightly outperformed CAE-BM, the difference was not statistically significant ($p > 0.05$).

Notably, compared to the IRASS dataset, the CAE-BM outperforms GraphSAGE-BM on the Ninapro DB2 dataset. Additionally, due to fewer repetitions of movements by subjects in this dataset, nRMSE provides a more accurate reflection of model performance than R^2 . As evident from Figure 8, both 3DCNN-BM and CAE-BM have significantly lower nRMSE compared to other models, with the gap exceeding half.

Overall, the 3DCNN-BM and CAE-BM exhibit significantly superior predictive performance on the Ninapro DB2 dataset, with the 3DCNN-BM model continuing to demonstrate the highest accuracy and stability in MS feature extraction.

3) Training Times Comparison

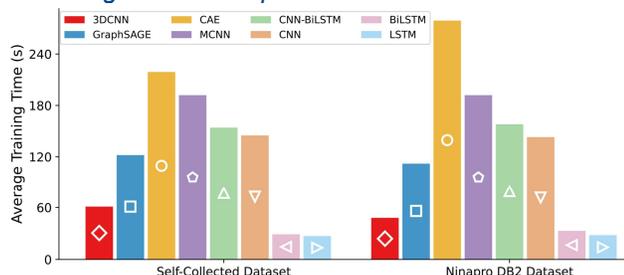


Fig. 9. The training times of all models on the IRASS dataset and the Ninapro DB2 dataset.

Figure 9 illustrates the training times of all models on both the IRASS and Ninapro DB2 datasets, thereby evaluating the training efficiency of each model and combining this with the predictive performance (nRMSE/ R^2) for more comprehensive assessment. The order of training time from shortest to longest is as follows: LSTM, BiLSTM, 3DCNN-BM, GraphSAGE-BM, CNN, CNN-BiLSTM, MCNN-BM, CAE-BM. Moreover, the ranking of training times is consistent across both datasets. Notably, although the LSTM and BiLSTM models have shorter training times than the 3DCNN-BM model, their predictive performance is significantly inferior. Overall, the 3DCNN-BM model not only excels in prediction accuracy and stability but also has a relatively shorter training time.

B. Examples of Predicted Wrist Angles and Errors

Figures 10(a) and 10(b) present representative examples of wrist joint angle predictions by the four proposed models on the IRASS dataset and the Ninapro DB2 dataset, respectively. Three representative subjects from each dataset were selected for an intuitive comparison of actual versus predicted wrist joint angles, along with corresponding error values for each model. The prediction error curves for these six subjects reveal that, despite differences among individuals resulting in varied error ranges, the 3DCNN-BM consistently shows smaller prediction errors compared to GraphSAGE-BM, CAE-BM, and MCNN-BM, and closely matches the actual angles. However, slight oscillations are still observed at the peak values of predicted wrist angles for the 3DCNN-BM.

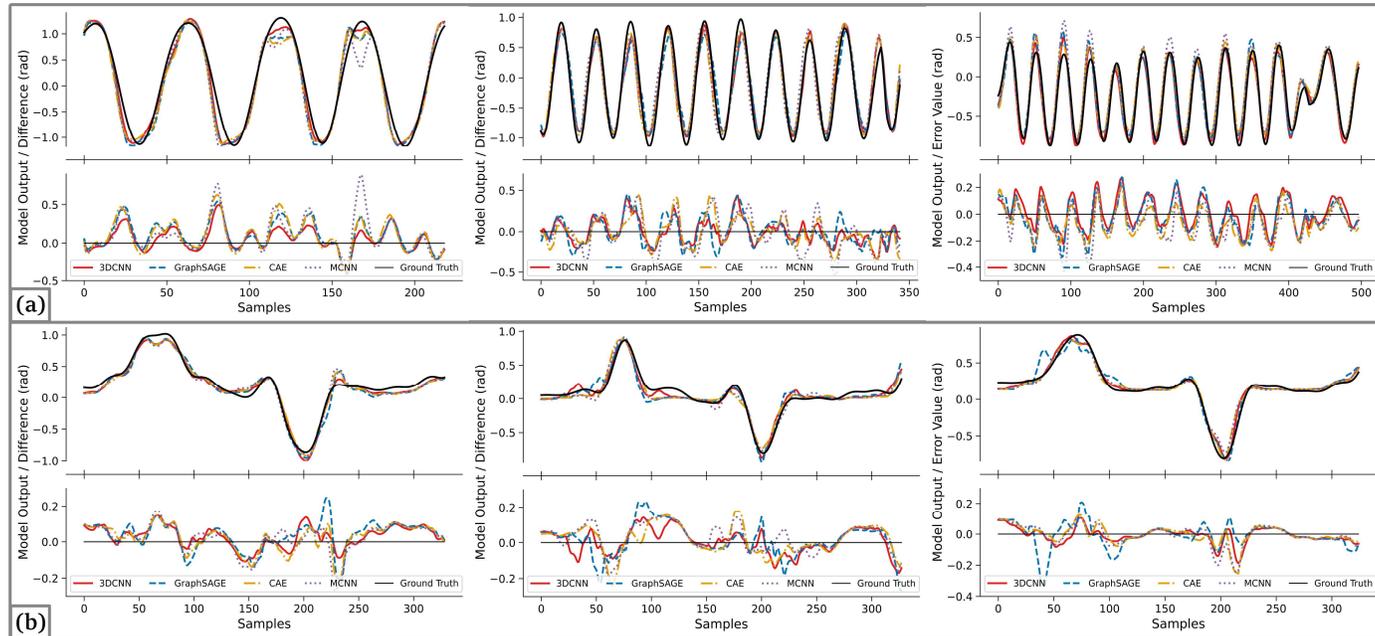


Fig. 10. (a) and (b) are representative examples of the intuitive comparison between actual and predicted wrist joint angles by the four proposed models on the IRASS dataset and the Ninapro DB2 dataset, respectively.

C. Visualization of MS Features Based on t-SNE

Figure 11(a), 11(b), 11(c), and 11(d) respectively illustrate the visualization of MS features (i.e., final features obtained from the BM module) extracted by the 3DCNN-BM, GraphSAGE-BM, CAE-BM, and MCNN-BM models after the

dimensionality reduction by using the t-distributed stochastic neighbor embedding (t-SNE). These features are projected into 3D space while preserving local similarities among the high-dimensional MS features, enabling a visual comparison of MS feature structures across the four proposed models. The three coordinate axes represent three primary features, with the color

gradient from dark blue to dark red indicating the transition from the normalized maximum negative angle to the maximum positive angle during the wrist flexion/extension.

The t-SNE results of the four models were evaluated based on five criteria: cluster separation, intra-cluster compactness, smoothness of the color gradient, global structure morphology, and distribution of outliers. It was observed that, although the extreme values of 3DCNN-BM were more dispersed along the third coordinate axis compared to GraphSAGE-BM, 3DCNN-BM demonstrated better cluster separation and smoother color transitions than GraphSAGE-BM. In contrast, the distribution of extreme values in GraphSAGE-BM was more scattered overall compared to 3DCNN-BM, resulting in more outliers in the transitional regions. CAE-BM and MCNN-BM performed worse in terms of intra-cluster compactness and global structure morphology, with MCNN-BM exhibiting the worst inter-cluster coherence.

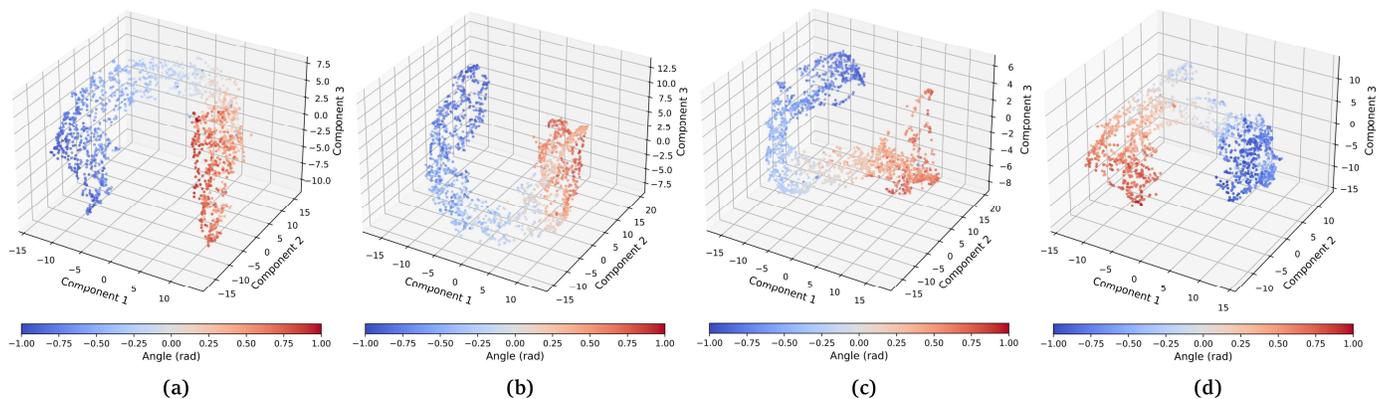


Fig. 11. Using t-SNE for the dimensionality reduction and 3D projection of the extracted MS features for visual comparison among the (a) 3DCNN-BM model, (b) GraphSAGE-BM model, (c) CAE-BM model, and (d) MCNN-BM model.

V. DISCUSSION

A. Analysis of Prediction Results for All Models

For the analysis of predictive results across all models, the 3DCNN-BM demonstrated superior accuracy and stability on both datasets, according to the evaluation criteria of nRMSE, R^2 , and training time, with an average training time of less than one minute. Consequently, 3DCNN-BM achieved an optimal balance between predictive performance and training efficiency, demonstrating its exceptional ability in extracting MS features for continuous joint kinematics prediction and highlighting its significant potential for practical applications. Notably, the variations in wrist flexion/extension speed, amplitude, and duration among the six subjects depicted in Figure 10(a) posed additional challenges to motion estimation. Nonetheless, the 3DCNN-BM still consistently performed long-term accurate joint kinematic estimations in these varying scenarios, further validating its exceptional robustness and generalizability. Moreover, the t-SNE-based comparative visualization results further corroborated that the 3DCNN-BM can more accurately and comprehensively extract MS features from sEMG.

To further quantitatively evaluate these five criteria, cluster separation and global structural morphology were measured using the Trustworthiness and Continuity metrics, respectively, to assess the local and global preservation of high-dimensional MS features in low-dimensional space (i.e., Figure 11). For intra-cluster compactness, the mean and standard deviation of the Average Local Neighbor Distance metric were applied to assess the local density of data points in the low-dimensional space. The Local Gradient Smoothness metric quantified the smoothness of color gradient by calculating the rate of color change between adjacent data points. Finally, the Local Outlier Factor metric was used to quantify the outlier degree of each data point, thereby evaluating the distribution of outliers. Based on the quantitative evaluation results of these five metrics, the comprehensive ranking of t-SNE results for the four models from best to worst was as follows: 3DCNN-BM, GraphSAGE-BM, CAE-BM, and MCNN-BM.

B. Analysis of Four Perspectives for MS Feature Extraction

Generally, models with superior predictive performance are capable of more comprehensively extracting both spatial and temporal features from sEMG signals, and these spatiotemporal features together constitute the ultimate MS features. Notably, BiLSTM outperforms CNN in both datasets, indicating that conventional CNNs are limited by the size of convolutional kernels, leading to a limited extraction of spatial and temporal features. BiLSTM, on the other hand, leverages the strong long-term contextual relationship between sEMG signals and joint angles to extract ample temporal features, thereby establishing more accurate sEMG-joint angle mappings. However, using the BiLSTM alone neglects the spatial feature extraction. Furthermore, although CNN-BiLSTM incorporates BiLSTM, the contextual relationships within BiLSTM are based on the features extracted by conventional CNNs, rather than directly from the sEMG signals. Therefore, MCNN-BM, through the additional temporal and spatial convolutional kernels, can more comprehensively identify and extract spatiotemporal features related to MSs from sEMG, thereby overcoming the limitations of regular CNN kernel sizes and addressing the shortcomings of CNN, LSTM, BiLSTM, and CNN-BiLSTM.

Regarding CAE-BM, previous study [36] has demonstrated the superior performance of CAE in the compression and reconstruction of ECG and sEMG signals. Consequently, CAE-BM exhibits significant potential in extracting MS features. By employing the unsupervised learning-based CAE structure, CAE-BM can adaptively refine low-dimensional latent vector representations from sEMG signals, which can be considered as spatiotemporal features representing MSs. Additionally, CAE-BM establishes reconstruction losses between five different levels of down-sampling and up-sampling modules, thereby implementing soft constraints on the features extracted by the down-sampling modules. This compels the model to focus on MS features rather than irrelevant noise or redundant features, thereby achieving predictive performance superior to that of MCNN-BM. However, the multi-level soft constraints of CAE-BM are established based on doubling the number of CNN modules, resulting in prolonged training time and reducing its potential for practical applications.

Unlike CNNs, the input to the GraphSAGE-BM is no longer grid-structured sEMG images composed of time and electrode channel axes, but rather data based on graph topology composed of nodes and edges. Although CNNs are widely used, the uniform spacing between pixels in sEMG images, coupled with the limitations imposed by the size of CNN kernels, impedes the capture of relationships between distant pixels. Consequently, CNNs based on sEMG images tend to overlook the coherence relationships between electrode nodes, namely the synergistic interactions among different target muscles [25], [37]. The results section indicates that although MCNN-BM employs additional spatiotemporal convolutional kernels to overcome the limitations of pixel distance and kernel size, its predictive performance remains inferior to that of GraphSAGE-BM. This underscores that more comprehensive MS features can be extracted from graph topology perspective. Additionally, previous GNN-based studies primarily focused on HD-EMG, overlooking the prediction performance of GNNs on commonly used sparse sEMG electrodes. The sole study [28] employing GCN-LSTM based on HD-sEMG for continuous estimation of wrist joint angles reports the prediction performance for wrist flexion/extension (nRMSE/R²) at (13%/90.6%), which is comparable to that of GraphSAGE-BM, thereby validating the feasibility of GNNs on sparse sEMG electrodes. However, due to the lower data quality and fewer motion repetitions in the Ninapro DB2 dataset compared to the IRASS dataset, GraphSAGE-BM's predictive performance on the Ninapro DB2 is inferior to that of CAE-BM, reflecting its lower robustness and generalizability compared to the 3DCNN-BM.

Lastly, 3DCNN-BM reconstructs 1D sEMG samples into 2D sEMG frames based on the anatomical distribution of the target muscles and the placement of sEMG electrodes, which are further stacked into 3D sEMG video segments to effectively extract MS features. Essentially, this approach represents 2D sEMG frames as the spatial components of sEMG signals at various time points based on the muscle anatomy, and by stacking multiple sEMG frames, it can concurrently represent the spatial and temporal components of sEMG signal within the sliding window. Overall, this method achieves the non-negative decomposition of sEMG based on the spatial and temporal components, similar to NMF, allowing the 3DCNN to simultaneously account for the synergistic relationships and

activation coefficients of target muscles within sEMG video segments, thereby fully extracting the spatiotemporal features of the sEMG. Moreover, 3DCNN-BM integrates additional anatomically based prior knowledge into BSS, effectively overcoming the limitations of NMF, such as the minimum sEMG channel number requirement, dependency on complete motion cycle data, and the need to manually set the number of synergies and the VAF threshold for reconstruction errors. In addition, since the pixel count in sEMG frames is significantly lower than in sEMG images, the 3DCNN kernel can cover the entire frame in a single operation, eliminating the need for multiple 2DCNNs to compress features, thereby significantly reducing the training time for 3DCNN-BM compared to MCNN-BM and CAE-BM. This characteristic simplifies the model retraining for intra-subject and inter-subject scenarios, thereby reducing the reliance on transfer learning.

Additionally, three previous studies, which are also based on our IRASS dataset, can be compared with this study. Study [38] proposed a personalized wrist musculoskeletal (MSK) model based on direct collocation optimization of physiological parameters. Study [39] introduced the GAN based on inverse dynamics muscle modeling and Lagrangian equations of motion, utilizing the physics-informed policy gradient to improve the efficiency of model training. Study [40] proposed an architecture called the Deep Kalman Filter Network (DKFN), which first employs the CNN to extract sEMG features and then uses the LSTM-KF module for sequence regression, where the Kalman filter's parameters and gains are learned through the LSTM. Compared to the predictive performance of 3DCNN-BM (average R² = 0.93), the average R² values of these three studies were 0.88, 0.92, and 0.87, respectively. Therefore, it is evident that the predictive performance of 3DCNN-BM surpasses that of these previous studies, particularly the MSK model [38] and DKFN [40].

C. Limitations and Future Work

This study has several limitations. Firstly, the two datasets employed were restricted to 1-DoF wrist flexion/extension movements. The predictive performance of the four proposed models on the other 2-DoFs of the wrist (i.e., ulnar/radial deviation and pronation/supination movements) and on other joints remain unexplored. Moreover, in addition to further testing the model's robustness and generalizability in cross-day and cross-subject predictions, it is also crucial to evaluate its performance in predicting random and complex motions, as well as in long-duration and real-time predictions. This would ensure that the proposed model maintains good predictive performance in practical applications, even in non-standardized scenarios such as missing data and irregular motion cycles. Secondly, this study solely included healthy subjects and did not encompass stroke patients, which limits the evaluation of the model's effectiveness in real rehabilitation treatment scenarios. Therefore, future work should include samples from disabled populations to further assess the applicability of the proposed model to the intended target groups. Lastly, the predictive performance of the 3DCNN-BM model largely depends on the accurate construction of sEMG frames, which requires precise localization of both the target muscles and sEMG electrodes. Consequently, this model may be more suitable for cases where the target muscles are close to the skin

surface, as sEMG signals from deeper muscles are susceptible to crosstalk from adjacent muscles, leading to inaccuracies in matching the muscle activations of the target muscles with their corresponding anatomical localization in the sEMG frames.

Regarding the extension of this study, the 3DCNN-BM model currently processes sEMG frames limited to the 2D anatomical slice of target muscles, thus requiring the placement of sEMG electrodes on the same anatomical cross-section. Future studies could consider employing additional sEMG or HD-EMG electrodes oriented along the direction of the muscle fibers of the target muscles, thereby enabling the extension of 2D sEMG frames into higher-dimensional multi-channel sEMG frames, akin to RGB images. Subsequently, the construction of such multi-channel frames could enable the model to extract more comprehensive MS features, thereby further enhancing the model's predictive performance. Moreover, since all DL models proposed in this study are data-driven and inherently "black-box" tools, they may overlook the neurophysiological mechanisms between sEMG and joint kinematics, leading to a lack of interpretability. Inspired by study [41], future work could incorporate loss functions based on MSK model motion equations or integrate Hill models to establish GAN, thereby utilizing soft constraints based on the physics-informed prior knowledge to enhance the interpretability of the DL models.

VI. CONCLUSION

Considering that sEMG signals can be regarded as a linear combination of temporal (activation coefficients) and spatial (synergistic relationships) components, and that the MS feature extraction is critical for accurately predicting motion intentions, thereby this paper proposes four DL models that extract MS features from four distinct perspectives. These models include: the MCNN-BM based on additional multi-scale spatiotemporal convolutional kernels; the CAE-BM based on the compression and reconstruction losses of sEMG signals; the GraphSAGE-BM based on graph topological structure; and the 3DCNN-BM based on the target muscles' anatomy and sEMG electrode placement. The core concept of the 3DCNN-BM is that if the contraction/relaxation process of the target muscle in the anatomical cross-section during joint movement is viewed as a video, then each frame of this video represents the spatial component of the target muscle's sEMG at that moment, while the stacked frames form a video segment that simultaneously represents both the temporal and spatial components of the sEMG signal during that period.

Subsequently, utilizing both the IRASS dataset and the Ninapro DB2 dataset, the performance of four proposed models and four baseline models (CNN-BiLSTM, CNN, BiLSTM, LSTM) in continuously predicting the wrist flexion/extension angles was evaluated and compared. Experimental results indicate that GraphSAGE-BM validates the feasibility of GNNs in continuous motion intention prediction with sparse sEMG electrodes, while 3DCNN-BM model surpasses other models in terms of prediction accuracy, robustness, training efficiency, and MS feature extraction. Overall, the 3DCNN-BM offers a novel and unique perspective for future research on continuous motion intention prediction using sEMG signals.

VII. REFERENCES

- [1] B. Sheng, J. Zhao, Y. Zhang, S. Xie, and J. Tao, "Commercial device-based hand rehabilitation systems for stroke patients: State of the art and future prospects," *Heliyon*, vol. 9, no. 3, Mar. 2023, doi: 10.1016/J.HELIYON.2023.E13588.
- [2] A. A. Blank, J. A. French, A. U. Pehlivan, and M. K. O'Malley, "Current Trends in Robot-Assisted Upper-Limb Stroke Rehabilitation: Promoting Patient Engagement in Therapy," *Curr Phys Med Rehabil Rep*, vol. 2, no. 3, p. 184, Sep. 2014, doi: 10.1007/S40141-014-0056-Z.
- [3] L. Bi, A. Feleke, and C. Guan, "A review on EMG-based motor intention prediction of continuous human upper limb motion for human-robot collaboration," *Biomed Signal Process Control*, vol. 51, pp. 113–127, May 2019, doi: 10.1016/J.BSPC.2019.02.011.
- [4] D. Xiong, D. Zhang, X. Zhao, and Y. Zhao, "Deep Learning for EMG-based Human-Machine Interaction: A Review," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 3, pp. 512–533, Mar. 2021, doi: 10.1109/JAS.2021.1003865.
- [5] S. P. Sitole and F. C. S. IV, "Continuous Prediction of Human Joint Mechanics using EMG Signals: A Review of Model-based and Model-free Approaches," *IEEE Trans Med Robot Bionics*, 2023, doi: 10.1109/TMRB.2023.3292451.
- [6] Z. Wei, Z. Q. Zhang, and S. Q. Xie, "Continuous Motion Intention Prediction Using sEMG for Upper-Limb Rehabilitation: A Systematic Review of Model-Based and Model-Free Approaches," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 1466–1483, 2024, doi: 10.1109/TNSRE.2024.3383857.
- [7] I. Campanini, C. Disselhorst-Klug, W. Z. Rymer, and R. Merletti, "Surface EMG in Clinical Assessment and Neurorehabilitation: Barriers Limiting Its Use," *Front Neurol*, vol. 11, p. 556522, Sep. 2020, doi: 10.3389/FNEUR.2020.00934/BIBTEX.
- [8] A. Ebied, E. Kinney-Lang, L. Spyrou, and J. Escudero, "Evaluation of matrix factorisation approaches for muscle synergy extraction," *Med Eng Phys*, vol. 57, pp. 51–60, Jul. 2018, doi: 10.1016/J.MEDENGGPHY.2018.04.003.
- [9] K. Zhao et al., "Evaluation of Methods for the Extraction of Spatial Muscle Synergies," *Front Neurosci*, vol. 16, p. 732156, Jun. 2022, doi: 10.3389/FNINS.2022.732156/BIBTEX.
- [10] Y. Ma, C. Shi, J. Xu, S. Ye, H. Zhou, and G. Zuo, "A Novel Muscle Synergy Extraction Method Used for Motor Function Evaluation of Stroke Patients: A Pilot Study," *Sensors* 2021, Vol. 21, Page 3833, vol. 21, no. 11, p. 3833, Jun. 2021, doi: 10.3390/S21113833.
- [11] D. Zhao et al., "MCR-ALS-based muscle synergy extraction method combined with LSTM neural network for motion intention detection," *Front Neurobot*, vol. 17, p. 1174710, Jun. 2023, doi: 10.3389/FNBOT.2023.1174710/BIBTEX.
- [12] Z. Teng, G. Xu, X. Zhang, X. Chen, S. Zhang, and H. Y. Huang, "Concurrent and continuous estimation of multi-finger forces by synergy mapping and reconstruction: a pilot study," *J Neural Eng*, vol. 20, no. 6, p. 066024, Dec. 2023, doi: 10.1088/1741-2552/AD10D1.
- [13] L. T. Hu, C. Xu, L. Chen, X. Y. Wu, and W. S. Hou, "Study on the Establishment Process of Muscle Synergy Based on Cosine Similarity," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 6590–6593, 2021, doi: 10.1109/EMBC46164.2021.9630701.
- [14] K. Zhao et al., "Muscle synergies for evaluating upper limb in clinical applications: A systematic review," *Heliyon*, vol. 9, no. 5, p. e16202, May 2023, doi: 10.1016/J.HELIYON.2023.E16202.
- [15] C. Camardella, M. Junata, K. C. Tse, A. Frisoli, and R. K. Y. Tong, "How Many Muscles? Optimal Muscles Set Search for Optimizing Myocontrol Performance," *Front Comput Neurosci*, vol. 15, p. 668579, Oct. 2021, doi: 10.3389/FNCOM.2021.668579/BIBTEX.
- [16] Z. He, Z. Qin, and Y. Koike, "Continuous Estimation of Finger and Wrist Joint Angles Using a Muscle Synergy Based Musculoskeletal Model," *Applied Sciences (Switzerland)*, vol. 12, no. 8, p. 3772, Apr. 2022, doi: 10.3390/AP12083772/S1.
- [17] D. Xu, H. Zhou, W. Quan, F. Gusztav, J. S. Baker, and Y. Gu, "Adaptive neuro-fuzzy inference system model driven by the non-negative matrix factorization-extracted muscle synergy patterns to estimate lower limb joint movements," *Comput Methods Programs Biomed*, vol. 242, p. 107848, Dec. 2023, doi: 10.1016/J.CMPB.2023.107848.
- [18] Y. Cho, P. Kim, and K. S. Kim, "Estimating Simultaneous and Proportional Finger Force Intention Based on sEMG Using a Constrained Autoencoder," *IEEE Access*, vol. 8, pp. 138264–138276, 2020, doi: 10.1109/ACCESS.2020.3012741.

- [19] D. Buongiorno et al., "An undercomplete autoencoder to extract muscle synergies for motor intention detection," *Proceedings of the International Joint Conference on Neural Networks*, vol. 2019-July, Jul. 2019, doi: 10.1109/IJCNN.2019.8851975.
- [20] D. Buongiorno, G. D. Cascarano, C. Camardella, I. De Feudis, A. Frisoli, and V. Bevilacqua, "Task-Oriented Muscle Synergy Extraction Using An Autoencoder-Based Neural Model," *Information 2020*, Vol. 11, Page 219, vol. 11, no. 4, p. 219, Apr. 2020, doi: 10.3390/INFO11040219.
- [21] H. Lee et al., "Stretchable array electromyography sensor with graph neural network for static and dynamic gestures recognition system," *npj Flexible Electronics* 2023 7:1, vol. 7, no. 1, pp. 1–13, Apr. 2023, doi: 10.1038/s41528-023-00246-3.
- [22] M. Xu, X. Chen, Y. Ruan, and X. Zhang, "Cross-User Electromyography Pattern Recognition Based on a Novel Spatial-Temporal Graph Convolutional Network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 72–82, 2024, doi: 10.1109/TNSRE.2023.3342050.
- [23] S. M. Massa, D. Riboni, and K. Nazarpour, "Explainable AI-Powered Graph Neural Networks for HD EMG-Based Gesture Intention Recognition," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 4499–4506, Feb. 2024, doi: 10.1109/TCE.2023.3333421.
- [24] S. M. Massa, D. Riboni, and K. Nazarpour, "Graph Neural Networks for HD EMG-based Movement Intention Recognition: An Initial Investigation," *RASSE 2022 - IEEE International Conference on Recent Advances in Systems Science and Engineering, Symposium Proceedings, 2022*, doi: 10.1109/RASSE54974.2022.9989657.
- [25] W. Zhong, Y. Zhang, P. Fu, W. Xiong, and M. Zhang, "A Spatio-Temporal Graph Convolutional Network for Gesture Recognition from High-Density Electromyography," *2023 29th International Conference on Mechatronics and Machine Vision in Practice, M2VIP 2023*, 2023, doi: 10.1109/M2VIP58386.2023.10413402.
- [26] X. Jia et al., "GTGR-Net: Graph Attentional-Temporal Network for Surface-Electromyography-Based Gesture Recognition," *Proceedings - 2022 3rd International Conference on Computing, Networks and Internet of Things, CNIOT 2022*, pp. 182–185, 2022, doi: 10.1109/CNIOT55862.2022.00039.
- [27] Z. Lai et al., "STCN-GR: Spatial-Temporal Convolutional Networks for Surface-Electromyography-Based Gesture Recognition," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13110 LNCS, pp. 27–39, 2021, doi: 10.1007/978-3-030-92238-2_3/FIGURES/5.
- [28] D. Li, P. Kang, Y. Yu, and P. B. Shull, "Graph-Driven Simultaneous and Proportional Estimation of Wrist Angle and Grasp Force via High-Density EMG," *IEEE J Biomed Health Inform*, vol. 28, no. 5, pp. 2723–2732, May 2024, doi: 10.1109/JBHI.2024.3373432.
- [29] L. Liu, J. Feng, J. Li, W. Chen, Z. Mao, and X. Tan, "Multi-layer CNN-LSTM network with self-attention mechanism for robust estimation of nonlinear uncertain systems," *Front Neurosci*, vol. 18, p. 1379495, Apr. 2024, doi: 10.3389/FNINS.2024.1379495/BIBTEX.
- [30] M. Atzori et al., "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Scientific Data* 2014 1:1, vol. 1, no. 1, pp. 1–13, Dec. 2014, doi: 10.1038/sdata.2014.53.
- [31] Y. Zhao, Z. Zhang, Z. Li, Z. Yang, A. A. Dehghani-Sanij, and S. Xie, "An EMG-Driven Musculoskeletal Model for Estimating Continuous Wrist Motion," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 12, pp. 3113–3120, Dec. 2020, doi: 10.1109/TNSRE.2020.3038051.
- [32] N. J. Jarque-Bou, M. Atzori, and H. Müller, "A large calibrated database of hand movements and grasps kinematics," *Scientific Data* 2020 7:1, vol. 7, no. 1, pp. 1–10, Jan. 2020, doi: 10.1038/s41597-019-0349-2.
- [33] L. McManus, G. De Vito, and M. M. Lowery, "Analysis and Biophysics of Surface EMG for Physiotherapists and Kinesiologists: Toward a Common Language With Rehabilitation Engineers," *Front Neurol*, vol. 11, p. 576729, Oct. 2020, doi: 10.3389/FNEUR.2020.576729/BIBTEX.
- [34] Y. Niu, W. Chen, H. Zeng, Z. Gan, and B. Xiong, "Optimizing sEMG Gesture Recognition: Leveraging Channel Selection and Feature Compression for Improved Accuracy and Computational Efficiency," *Applied Sciences* 2024, Vol. 14, Page 3389, vol. 14, no. 8, p. 3389, Apr. 2024, doi: 10.3390/APP14083389.
- [35] A. Demir, T. Koike-Akino, Y. Wang, M. Haruna, and D. Erdogmus, "EEG-GNN: Graph Neural Networks for Classification of Electroencephalogram (EEG) Signals," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 1061–1067, 2021, doi: 10.1109/EMBC46164.2021.9630194.
- [36] K. Dinashi, A. Ameri, M. A. Akhaee, K. Englehart, and E. Scheme, "Compression of EMG Signals Using Deep Convolutional Autoencoders," *IEEE J Biomed Health Inform*, vol. 26, no. 7, pp. 2888–2897, Jul. 2022, doi: 10.1109/JBHI.2022.3142034.
- [37] L. Zhao, R. Liu, S. Li, X. Wang, and D. Bao, "Spatio-Temporal Variable Structure Graph Neural Network for EEG Data Classification," *2023 6th International Symposium on Autonomous Systems, ISAS 2023*, 2023, doi: 10.1109/ISAS59543.2023.10164356.
- [38] Y. Zhao et al., "Computationally Efficient Personalized EMG-Driven Musculoskeletal Model of Wrist Joint," *IEEE Trans Instrum Meas*, vol. 72, 2023, doi: 10.1109/TIM.2022.3225023.
- [39] Y. Shi, S. Ma, Y. Zhao, C. Shi, and Z. Zhang, "A Physics-Informed Low-Shot Adversarial Learning for sEMG-Based Estimation of Muscle Force and Joint Kinematics," *IEEE J Biomed Health Inform*, vol. 28, no. 3, pp. 1309–1320, Mar. 2024, doi: 10.1109/JBHI.2023.3347672.
- [40] T. Bao, Y. Zhao, S. A. R. Zaidi, S. Xie, P. Yang, and Z. Zhang, "A deep Kalman filter network for hand kinematics estimation using sEMG," *Pattern Recognit Lett*, vol. 143, pp. 88–94, Mar. 2021, doi: 10.1016/J.PATREC.2021.01.001.
- [41] J. Zhang et al., "Physics-Informed Deep Learning for Musculoskeletal Modeling: Predicting Muscle Forces and Joint Kinematics From Surface EMG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 484–493, 2023, doi: 10.1109/TNSRE.2022.3226860.