

From Forearm to Wrist: Deep Learning for Surface Electromyography-Based Gesture Recognition

Jiayuan He¹, Member, IEEE, Xinyue Niu, Penghui Zhao, Chuang Lin², Member, IEEE, and Ning Jiang³, Senior Member, IEEE

Abstract—Though the forearm is the focus of the prostheses, myoelectric control with the electrodes on the wrist is more comfortable for general consumers because of its unobtrusiveness and incorporation with the existing wrist-based wearables. Recently, deep learning methods have gained attention for myoelectric control but their performance is unclear on wrist myoelectric signals. This study compared the gesture recognition performance of myoelectric signals from the wrist and forearm between a state-of-the-art method, TDLDA, and four deep learning models, including convolutional neural network (CNN), temporal convolutional network (TCN), gate recurrent unit (GRU) and Transformer. It was shown that with forearm myoelectric signals, the performance between deep learning models and TDLDA was comparable, but with wrist myoelectric signals, the deep learning models outperformed TDLDA significantly with a difference of at least 9%, while the performance of TDLDA was close between the two signal modalities. This work demonstrated the potential of deep learning for wrist-based myoelectric control and would facilitate its application into more sections.

Index Terms—Myoelectric control, surface electromyography (sEMG), gesture recognition, deep learning.

Manuscript received 19 April 2023; revised 8 November 2023; accepted 6 December 2023. Date of publication 8 December 2023; date of current version 16 January 2024. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant YJ202373, in part by the Science and Technology Major Project of Tibetan Autonomous Region of China under Grant XZ202201ZD0001G, in part by the 1.3.5 Project for Disciplines of 1435 Excellence Grant from the West China Hospital under Grant ZYYC22001, and in part by the Key Project from the Med-X Center for Manufacturing under Grant 0040206107007. (Corresponding author: Jiayuan He.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Office of Research Ethics of the University of Waterloo under Application 31346, and performed in line with the Declaration of Helsinki.

Jiayuan He and Ning Jiang were with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada. They are now with the National Clinical Research Center for Geriatric, West China Hospital, Sichuan University, Chengdu, Sichuan 610017, China, and also with the Med-X Center for Manufacturing, Sichuan University, Chengdu, Sichuan 610017, China (e-mail: jiayuan.he@wchscu.cn).

Xinyue Niu, Penghui Zhao, and Chuang Lin are with the School of Information Science and Technology, Dalian Maritime University, Dalian, Liaoning 116026, China.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TNSRE.2023.3341220>, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3341220

I. INTRODUCTION

SURFACE electromyography (sEMG) is the electrical manifestation of the human skeletal muscle contractions, recorded noninvasively from the skin surface [1]. As related to the action potentials from the motor neurons in the spinal cord, it contains neural information of human movements and can be used to decode hand gestures and control the electronics and robots [2], which is termed myoelectric control. Its conventional application is to help the amputees intuitively operate the advanced five-fingered prosthetic hand with the signals from the residual muscles [3]. With the development of human-machine interface (HMI) technology, the application of myoelectric control has been extended to more areas, including sign language interpretation for the deaf and mute [4], powered exoskeleton control for the patients with motor dysfunction [5], consumer electronics play for the entertainment [6], human-robot collaboration and teleoperation for the industrial production [7], [8], and so on.

Current research mainly focused on the performance of myoelectric control with the electrodes attached on the forearm for the application in upper-limb prostheses was clear and dominant. Forearm sEMG signals were available from the transradial amputees, and the signal quality was high enough to detect the intentions of the gestures. Wearable devices were one main embodiment of HMI [9]. As the requirement of long-term wearing, wearable devices are expected to be unobtrusive, subtle, and socially acceptable [10], [11], [12]. Though forearm sEMG signals achieved success in prosthetic hand control, few commercial wearable devices were designed to be positioned on the forearm. Besides, the forearm is normally clothes-covered, and the electrodes need to be skin-touched. As such, it would be cumbersome for forearm sEMG to be paired with wearable devices. Currently, many wearable devices adopt the wrist-based form, such as smartwatches, and fitness bands. Compared to the forearm-based, myoelectric control with the electrodes on the wrist could be incorporated into these existing wearables, and empower them with recognizing hand gestures to achieve control of other devices and machines, benefiting life and production. The successful incorporation would expand the application and

bring new opportunities to both wearables and myoelectric control.

As the potential for the wearables, researchers have turned to investigating the performance of gesture recognition from wrist sEMG. Jiang et al. examined the signal-to-noise ratio (SNR) of wrist sEMG and found that its difference from forearm sEMG was not significant [11]. They then presented a wristband design fusing sEMG and inertial measurement unit (IMU) data for hand gesture recognition. S. Botros et al. systematically compared the quality of wrist sEMG to that of forearm sEMG and found that wrist sEMG was better in finger movements [10]. Additionally, the classification accuracy of wrist sEMG was also better than that of forearm sEMG in recognizing single and multi-finger gestures. Though the number of relevant studies was still limited, they demonstrated the feasibility of decoding gestures from wrist sEMG with the traditional pattern recognition-based methods.

In the last decade, deep learning methods have gained increased attention and become an effective tool for processing and decoding sEMG signals for gesture recognition [13], [14]. In the traditional pattern recognition-based methods, feature extraction or feature engineering was an important step for the final classification performance. Deep learning methods combined the steps of feature extraction and the subsequent classification, achieving feature-free or end-to-end learning [8]. Geng et al. employed a convolutional neural network (CNN), translating sEMG signals as an image form to exploit spatial information [15]. It attained better classification performance than the state-of-the-art methods. Asif et al. investigated the effect of the hyper-parameters for CNN on gesture recognition from sEMG [16]. To reduce the number of parameters and training data, Tsinganos et al proposed a temporal convolutional network structure and showed higher classification accuracy than that of CNN on a public sEMG dataset [17]. As sEMG signals were time-series data, deep learning methods designed for sequential data processing, such as recurrent neural network (RNN), long short-term memory (LSTM), gate recurrent unit (GRU), and transformer, were also investigated [18]. Rahimian et al. proposed a transformer structure that outperformed LSTM in classification accuracy and algorithm complexity [19]. Chen et al. proposed a gate recurrent unit (GRU) structure and improved the gesture recognition performance with different forearm postures compared to the traditional pattern recognition methods [20]. These studies showed that deep learning methods could achieve superior performance to traditional pattern recognition methods with feature engineering. It offered a new strategy to explore and learn the information from sEMG for gesture recognition.

Though wrist sEMG was a potential modality for future wearables, current studies of deep learning all focused on forearm sEMG, and its performance on wrist sEMG is yet to be investigated. To explore the power of deep learning methods on wrist sEMG, this study selected four effective deep learning structures: CNN, GRU, TCN, and Transformer. They were all representative structures over the development of deep learning

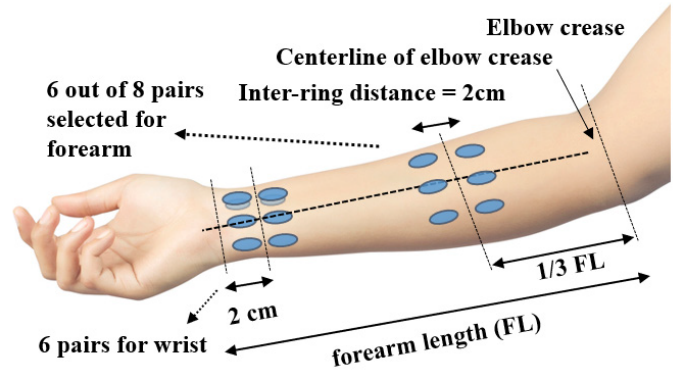


Fig. 1. Electrode placement in this study. Eight and six bipolar electrodes were equally spaced on the forearm and wrist, respectively. The inter electrode distance is 2 cm. Six pairs with the lowest error rate of a traditional method (TDLDA) are selected from eight pairs of the forearm to keep the number of the electrodes equal between forearm and wrist.

and employed for forearm sEMG-based gesture recognition in previous studies [21]. CNN was widely used in biosignal processing for extracting spatial information [14]. GRU, TCN, and Transformer were all proposed in the last decade, introducing different mechanisms to improve the accuracy and efficiency of sequential data [21]. The performance of the four deep learning methods was compared between forearm and wrist sEMG with a large dataset consisting of forty-three participants and seventeenth gesture classes. A comparison was also made between the deep learning methods and the state-of-the-art methods for sEMG-based gesture recognition.

This study focused on the effect of deep learning models on myoelectric control, and investigated the performance of gesture recognition from forearm and wrist sEMG with one traditional and four common deep learning methods, *i.e.*, CNN, GRU, TCN, and Transformer. The difference of decoding wrist and forearm sEMG would be explored, as well as their interactions with the methods. The results would provide insights into the characteristics of wrist sEMG, and benefit the development of wrist-based HMI and wearables.

II. METHOD

A. Dataset

Forty-three healthy participants (23 Males, and 20 Females with average age of 26.35 ± 2.89) were recruited for the study. The informed consent was obtained before the experiment, and the protocol was in accordance with the Declaration of Helsinki, and approved by the Office of Research Ethics of the University of Waterloo (ORE# 31346).

Myoelectric signals were collected from both the forearm and the wrist. Prior to the electrode placement, the forearm length between the olecranon process and the ulnar styloid process was measured. The electrodes for the forearm and wrist were positioned one-third of the forearm length from the olecranon process and 2 cm away from the ulnar styloid process, respectively (Fig. 1). For the forearm placement, sixteen monopolar electrodes were positioned in the form of

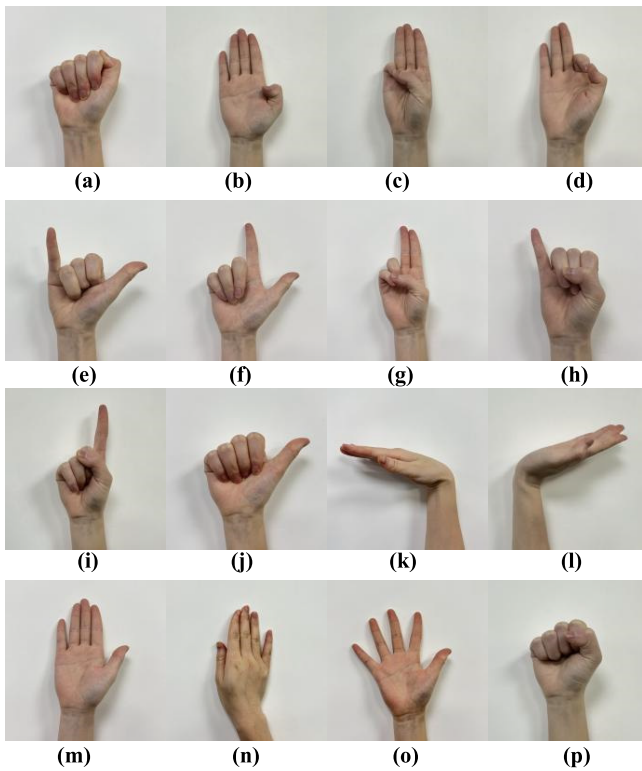


Fig. 2. Active gesture classes investigated in this study. (a) lateral prehension (LP), (b) thumb adduction (TA), (c) thumb and little finger opposition (TLFO), (d) thumb and index finger opposition (TIFO), (e) thumb and little finger extension (TLFE), (f) thumb and index finger extension (TIFE), (g) index and middle finger extension (IMFE), (h) little finger extension (LFE), (i) index finger extension (IFE), (j) thumb extension (TE), (k) wrist flexion (WF), (l) wrist extension (WE), (m) forearm supination (FS), (n) forearm pronation (FP), (o) hand open (HO), (p) hand close (HC).

two rings, of which each had eight electrodes equally spaced, making up eight bipolar pairs. The distance between the two centers of the rings was 2 cm. For the wrist placement, there were twelve monopolar electrodes positioned in a similar way to the forearm setup, where two rings and six bipolar pairs were formed.

There were seventeen gesture classes investigated in this study (Fig. 2), which were lateral prehension (LP), thumb adduction (TA), thumb and little finger opposition (TLFO), thumb and index finger opposition (TIFO), thumb and little finger extension (TLFE), thumb and index finger extension (TIFE), index and middle finger extension (IMFE), little finger extension (LFE), index finger extension (IFE), thumb extension (TE), wrist flexion (WF), wrist extension (WE), forearm supination (FS), forearm pronation (FP), hand open (HO), hand close (HC), and rest (R). The subject was instructed to perform each gesture following the cues on the screen. There were seven trials for each subject. In one trial, each gesture was held for five seconds, and a ten-second rest was provided between the contractions to avoid muscle fatigue. The data were recorded using a commercial device (EMGUSB2+, OT Bioelettronica, Italy) at a sampling rate of 2048 Hz. The dataset was uploaded and available in PhysioNet, *i.e.* the data of the first day in [22].

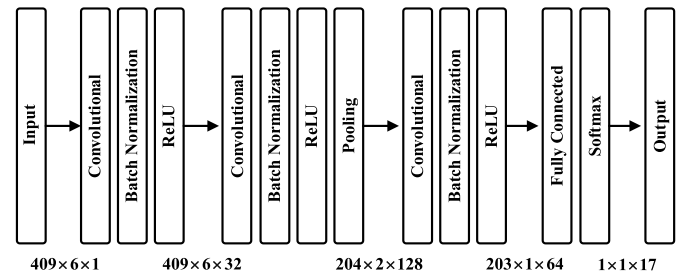


Fig. 3. Structure of the convolutional neural network (CNN) used in this study. It has three convolutional blocks and one full connection block. The input and output dimension of each block is annotated under each block.

B. Gesture Recognition Models

Four representative deep learning models, *i.e.*, CNN, GRU, TCN, and transformer, and one traditional method were chosen to compare the performance of gesture recognition between the wrist and forearm. The four networks employed different structures and were all reported to have a high accuracy of decoding gestures from forearm myoelectric signals [16], [17], [19], [20] with end-to-end learning, *i.e.*, bypassing the feature engineering step¹. The traditional method was the combination of time domain features and linear discriminant analysis, termed TDLDA, which was the state-of-the-art method for recognizing gestures from forearm myoelectric signals [23]. The details are as follows:

1) **CNN**: CNN is the most widely adopted deep-learning model in the field of myoelectric gesture recognition. It has an excellent ability to exploit spatial information from the input data by performing convolutional operations sliding over the input. Asif et al. investigated how various hyper-parameters affected CNN performance optimization [16]. The learning rate was set as 0.0001 for its better convergence compared to 0.1, 0.01 and 0.001 tested in a pilot study. Adapted from the results of [16], the structure of CNN used in this study is depicted in Fig. 3. There were three convolutional blocks after the input layer. For the first and third block, they both consisted of a convolutional layer and a batch normalization layer. For the second block, there was a pooling layer after the convolutional and batch normalization layer to discard useless features and reduce computation. The size of the input layer was one sEMG image matrix of $C \times L$, where C was the channel number and L was the window length. The convolutional layers in three blocks had 32 filters of $1 \times 3 \times 3$ with the stride 1-by-1, 128 filters of $32 \times 3 \times 3$ with the stride 1-by-1, and 64 filters of $128 \times 2 \times 2$ with the stride 1-by-1, respectively. After each batch normalization layer in three blocks, a ReLU layer was adopted as an activation function. A fully connected layer with 17 units was applied at last for classification.

2) **GRU**: GRU is a type of RNN, which is well suited to processing time series data. It employs a gating mechanism to control the information flow and has a relatively simple architecture and comparable performance compared to other similar neural networks, such as LSTM. The structure of GRU was

¹The python codes of four deep learning models would be available from https://github.com/Easyeasyworld/HGR_wrist.

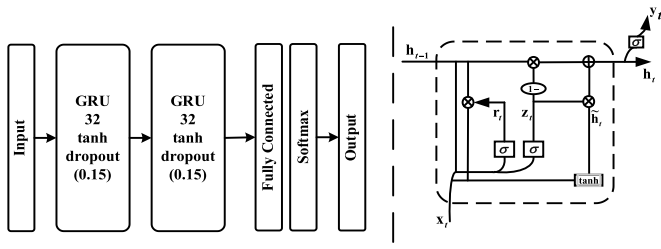


Fig. 4. Structure of gated recurrent unit (GRU) and its cell in this study. It has two GRU block and one full connection block.

adopted from [20] for its robustness against forearm postures in recognizing gestures from forearm myoelectric signals. The whole model consisted of two GRU layers and a fully linked layer. There were 32 hidden units in both GRU layers and 17 hidden units in the fully connected layer. A dropout layer was set after every three layers with a probability of 0.15 to avoid overfitting. The structure of GRU and the particulars of one cell is displayed in Fig. 4. The calculation of the reset gate r_t and the update gate z_t depend on the current input x_t and the hidden state from the last step h_{t-1} as shown in Equation (1) and (2). The reset gate r_t decides how to combine the current input with the before memory, and the update gate z_t decides how much of the before memory can affect in the current step. Then the candidate activation \tilde{h} and actual activation h_t are computed based on the r_t and z_t as Equation (3) and (4). The output y_t is finally calculated based on \tilde{h} as described in Equation (5).

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (1)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (2)$$

$$\tilde{h} = \tanh(W_h \cdot [r_t * h_{t-1}, x_t]) \quad (3)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h} \quad (4)$$

$$y_t = \sigma(W_o \cdot h_t) \quad (5)$$

3) *TCN*: TCN is a variant of CNN, which employs a one-dimensional (1-D) fully convolutional network architecture to capture temporal patterns from the sequential data. The length of each hidden layer in TCN is the same as that of the input layer. Compared to GRU, its requirement for memory is much lower in training. As displayed in Fig. 5, the TCN model of this study consisted of two parts, a TCN structure to extract features and a full connection block for classification. The structure of TCN was adapted from [17] where its performance was demonstrated better than that of CNN on a 53-class task of gesture recognition from forearm myoelectric signals. The TCN structure consisted of three 1-D convolutional layers with dilation of 1, 2, and 4, respectively. They had 32, 64, and 17 filters of 3×3 with a stride of 1×1 , respectively. The generated features were then fed into a full connection block with 17 hidden units for classification.

4) *Transformer*: A transformer is a deep learning model designed to process sequential data. Different from RNN, it employs a self-attention mechanism, and processes the entire input data all at once, not sequentially, allowing for more parallelization to reduce the training time [24]. As displayed

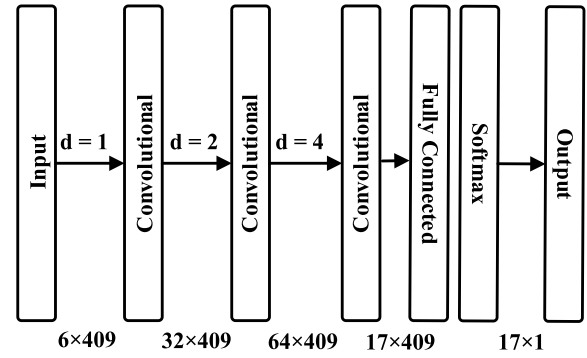


Fig. 5. Structure of temporal convolutional network (TCN) in this study. It has three convolutional layers and one full connection block. The dilation factors are 1, 2, and 4, respectively. The input and output dimension of each part is annotated under each layer.

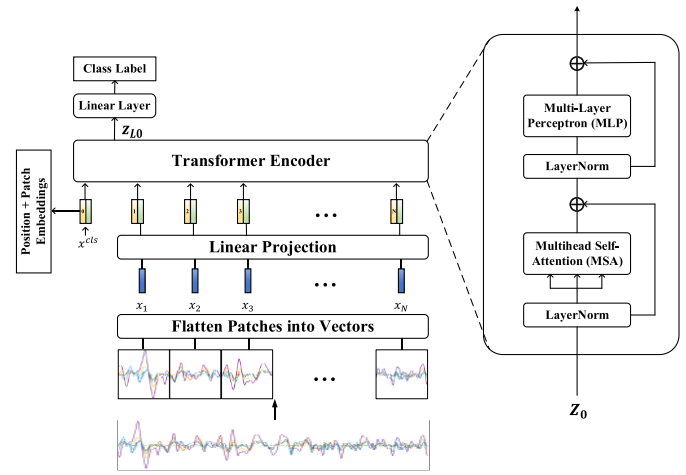


Fig. 6. Structure of Transformer in this study. The myoelectric signals are divided, flattened into vectors, and then linearly projected for transformer encoder. The encoder consists of L ($L=5$) layers, and each layer has two LayerNorm, one MLP and one MSA.

in Fig. 6, this study adopted a transformer structure from [19], of which the performance surpassed the state-of-the-art methods in terms of overall accuracy and algorithm complexity. Each input of $C \times L$ was first divided into N ($N = L/C$) patches of $C \times C$ without overlapping. These patches were then flattened into a vector $x_j^p \in \mathbb{R}^{C^2}$ and linearly projected with a matrix $E \in \mathbb{R}^{C^2 \times d}$ shared between patches. The output of this step was called patch embedding. Then the patch embeddings were added with standard 1-D trainable position embeddings. The output of this step is defined as

$$Z_0 = [x^{cls}; x_1^p E; x_2^p E; \dots; x_N^p E] + E^{pos} \quad (6)$$

where x^{cls} is a trainable $[cls]$ token appended similarly as in the BERT (Bidirectional Encoder Representation from Transformers) framework [8], and $E^{pos} \in \mathbb{R}^{(N+1) \times d}$ denotes the position embeddings. The output Z_0 was then fed into the Transformer encoder, a L -layers ($L=5$ in this study) module each of which contained two LayerNorm, a multi-layer perceptron (MLP), and a multi-head self-attention (MSA). The

output of the Transformer encoder Z_L can be described as

$$Z_L = [z_{L0}; z_{L1}; \dots; z_{LN}] \quad (7)$$

where z_{L0} is used for classification and finally once again linear projected to the final values.

5) **TDLDA**: TDLDA was a traditional machine learning model. It extracted four time domain features from the myoelectric signals [25], including mean absolute value (MAV), zero crossing (ZC), sign slope changes (SSC), and wavelet length (WL). The extracted features were concatenated as a vector and sent to LDA for classification. As TDLDA was not only widely used in gesture recognition from forearm myoelectric signals, but also proved effective in processing wrist myoelectric signals [10], it was regarded as a benchmark to evaluate the performance of the deep learning models in this study.

C. Data Analysis

Both forearm and wrist myoelectric signals were collected from the bipolar electrodes. As the number of bipolar electrodes was eight and six for the forearm and wrist, respectively, only six pairs were used in this study for the forearm to make the comparison fair. The selection was based on the results of a pre-analysis, where the classification accuracy of TDLDA with each combination of six pairs was calculated. The combination with the lowest error rate was selected, of which the electrode ID was 2, 3, 5, 6, 7, and 8. Before the classification, the raw signals were filtered with a bandpass filter between 10 and 500 Hz, and a notch filter at 60 Hz for powerline noise. The data was then segmented, and each segment was fed into the models. The segment lengths were 200 ms, and the increment was 10 ms, resulting in the overlap between the two consecutive segments being 190 ms. Seven-fold cross-validation was adopted for classification. There were seven rounds. In each round, six out of seven trials were chosen for training and the remaining trial was used for testing.

D. Feature Space Analysis

For comparing the separability of the deep learning and TD features between forearm and wrist, t-distributed stochastic neighbor embedding (t-SNE) and Davies-Bouldin index (DBI) were employed to visualize the feature distribution and quantitatively measure the separation of the seventeen classes, respectively. t-SNE is a nonlinear and unsupervised dimension reduction method preserving neighborhood relationships by embedding the original data into low dimensions [26]. It was often employed in deep learning studies for data distribution visualization [27].

DBI evaluated the data clustering performance by calculating the ratio of within-class similarity to between-class similarity. Suppose S_i , S_j and $D_{i,j}$ represent the within-class and between-class similarity for and between the i th and j th class, respectively. DBI measures the worst separability of each class with its neighborhood and calculates the average,

$$DBI = \frac{1}{K} \sum_{i=1}^K \left(\frac{S_i + S_j}{D_{i,j}} \right) \quad (8)$$

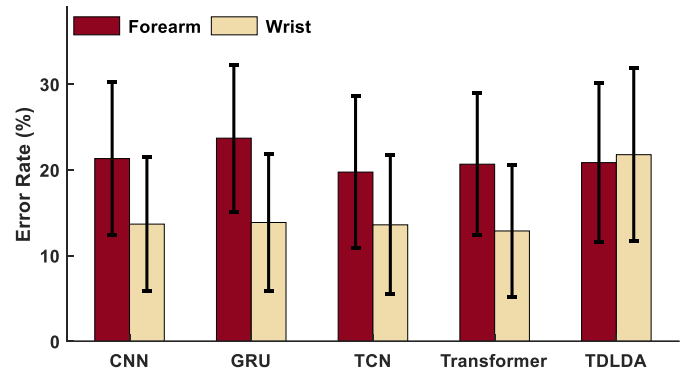


Fig. 7. Classification error rates of five models for forearm and wrist myoelectric signals. The performance of four deep learning models, i.e., CNN, GRU, TCN, and Transformer is similar, which achieves lower error rates from the wrist compared to the forearm. For the traditional method TDLDA, the classification performance between wrist and forearm myoelectric signals is similar.

where K is the number of classes. Normally the Euclidean distance is used to measure the magnitude of the similarity, and the computation of S_i and $D_{i,j}$ are as follows:

$$S_i = \sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} (x_{i,j} - \mu_i)^T (x_{i,j} - \mu_i)} \quad (9)$$

$$D_{i,j} = \sqrt{(\mu_i - \mu_j)^T (\mu_i - \mu_j)} \quad (10)$$

where $x_{i,j}$, μ_i and N_i are the j th feature vector, the mean feature vector, and the number of the feature vectors of the i th class, respectively. A small DBI value indicates a large class separation.

III. RESULTS

A. Classification Performance

The classification performance of the gestures was different between the myoelectric signals of the wrist and forearm for all the five models (Fig. 7). For the four deep learning models, i.e., CNN, GRU, TCN and Transformer, the averaged error rates of wrist-based gesture recognition were close, between 12.9% and 13.7%, and they were all lower than the corresponding error rates of forearm-based gesture recognition. With the same deep learning model, the difference between the error rates of the wrist-based and forearm-based gesture recognition was between 6.2% and 9.8%. TCN obtained the smallest error rate of recognizing gestures from the forearm, i.e., 19.7%, and the error rate of GRU was the biggest, which was 23.7%. For TDLDA, its recognition performance between wrist and forearm myoelectric signals was similar, which was different from the scenarios of the four deep learning models. The average error rate of the wrist-based gesture recognition of TDLDA was 21.8%, around 9% larger than the performance of the deep learning models. The average error rate of the forearm-based gesture recognition of TDLDA was 20.8%, larger than the error rates of TCN and Transformer but smaller than the error rates of the other two deep learning models.

The confusion matrices of all five models are displayed in Fig. 8. For TDLDA, the matrices of the wrist and the forearm were similar. For the four deep learning models, the

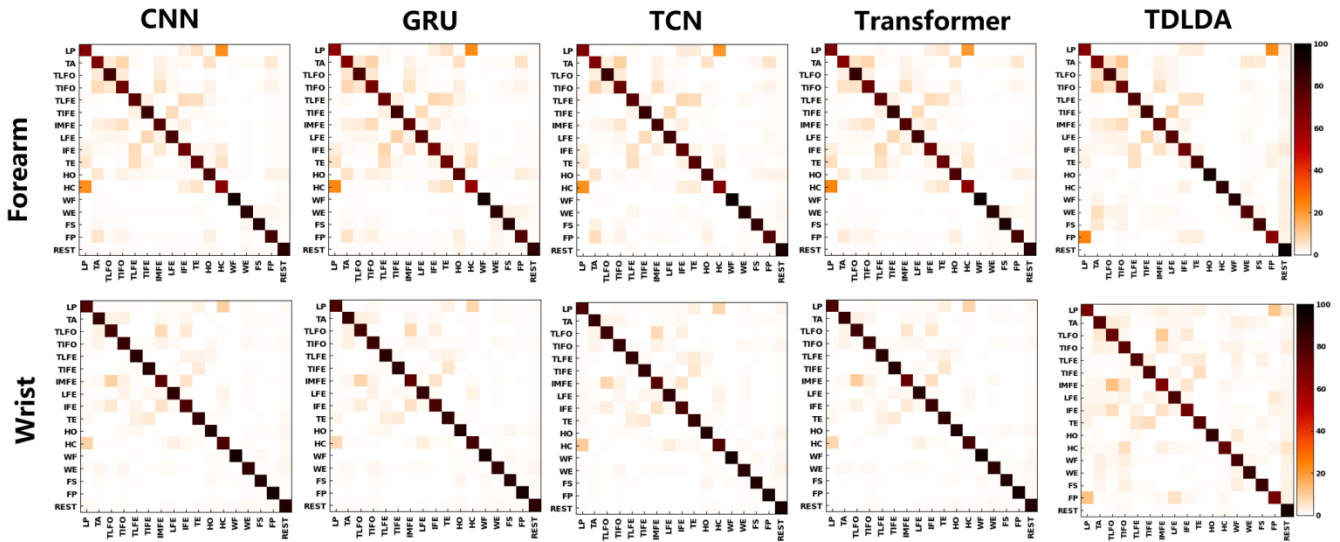


Fig. 8. Confusion Matrices of five models for forearm and wrist myoelectric signals. For the four deep learning models from CNN to Transformer, the misclassification in the upper left part of the forearm matrices are decreased in the wrist matrices, indicating the classification of the finger gestures are improved with the wrist myoelectric signals. For TDLDA, the matrices are similar between the forearm and the wrist.

misclassification of the forearm myoelectric signals mainly collected in the upper left part of the matrices. And the misclassification of this part was decreased in the matrices of the wrist. As the gesture sequence of the matrix was from finger to gross hand and wrist movements, it indicated that the recognition of the finger movements was confused by each other with the forearm myoelectric signals, and the confusion was improved with the wrist myoelectric signals. The other evaluation metrics, including the true positive rate (sensitivity), true negative rate (specificity), positive predictive value (Precision) and negative predictive value, and F1 Score of each movement, are in Supplementary Materials.

Statistical analysis confirmed the above observations. A three-way ANOVA was conducted to evaluate the effect of the three factors, *i.e.*, gesture, model, and electrode position, on the classification performance. It was revealed that there was a significant interaction among the three factors ($p = 0.000$, $F = 1.72$). To compare the recognition performance between the forearm and wrist myoelectric signals, the level of the model was fixed and the two-way ANOVAs were performed. There was a significant interaction between the factors of gesture and electrode position, and the one-way ANOVAs were subsequently performed with the fixed levels of gesture (See Supplement for p and F values). As displayed in Table I, for the four deep learning models, the recognition performance of most gestures (≥ 11 classes) was significantly improved with wrist myoelectric signals compared to that with forearm myoelectric signals. For TDLDA, there was no such trend, and forearm and wrist myoelectric signals had their respective advantages in recognizing different gestures. Besides, similar analyses were performed to compare the performance of the five models. The two-way ANOVAs were first performed with the fixed level of electrode position, and after the significant interaction was detected between the factor of gesture and model, one-way ANOVAs were performed with the fixed level of gesture. As displayed in Table II, with

TABLE I
STATISTICAL ANALYSIS OF CLASSIFICATION ERRORS BETWEEN WRIST AND FOREARM MYOELECTRIC SIGNALS

Gesture	CNN	GRU	TCN	Transformer	TDLDA
LP	*	*	*	*	~
TA	*	*	*	*	*
TLFO	~	*	~	~	X
TIFO	*	*	*	*	*
TIFE	*	*	*	*	~
TLFE	*	*	*	*	~
IMFE	~	~	*	*	X
LFE	*	*	*	*	~
IFE	*	*	*	*	~
TFE	*	*	*	*	~
WF	~	~	*	~	X
WE	~	~	~	~	X
FS	~	~	*	~	X
FP	*	*	*	*	*
HO	*	*	*	*	~
HC	*	*	*	*	*
R	~	~	~	~	X

Note: * represents the classification errors of wrist myoelectric signals is significantly lower than the corresponding values of the forearm ($p < 0.05$), and X represents it is significantly higher than that of the forearm. ~ represents the difference is not significant.

forearm myoelectric signals, a majority of the comparisons were not significant. With wrist myoelectric signals displayed in Table III, there was no significant difference among the four deep learning models, and they were all significantly better than TDLDA in recognizing gestures except for the Rest class, where the performance of TCN, TDLDA, and Transformer was comparable, and TCN and TDLDA significantly outperformed GRU and CNN.

B. Visual and Quantitative Comparison of Features

Deep learning features were the inputs of the full connection layer for CNN, TCN, and GRU, as well as the outputs of

TABLE II
STATISTICAL ANALYSIS OF CLASSIFICATION ERRORS AMONG FIVE MODELS FOR FOREARM MYOELECTRIC SIGNALS

Gesture	TCN	TCN	TCN	Transformer	Transformer	CNN	TDLDA	TDLDA	TDLDA	TDLDA
	vs. Transformer	vs. CNN	vs. GRU	vs. CNN	vs. GRU	vs. GRU	vs. TCN	vs. CNN	vs. Transformer	vs. GRU
LP	~	~	~	~	~	~	~	~	~	~
TA	~	~	~	~	~	~	~	~	~	~
TLFO	~	~	*	~	*	~	x	~	x	~
TIFO	~	~	~	~	~	~	~	~	~	~
TIFE	~	~	~	~	~	~	~	~	~	*
TLFE	~	~	~	~	~	~	~	~	~	~
IMFE	~	*	*	~	~	~	x	~	~	~
LFE	~	~	~	~	*	~	x	~	x	~
IFE	~	~	*	~	~	~	~	~	~	~
TFE	*	~	*	~	~	~	~	*	*	*
WF	~	~	~	~	*	~	x	~	x	~
WE	~	~	~	~	~	~	~	~	~	~
FS	x	x	~	~	~	~	~	~	~	~
FP	~	~	~	~	~	~	~	~	~	~
HO	~	~	*	~	~	~	~	~	~	*
HC	~	~	~	~	~	~	~	~	~	~
R	~	~	*	~	~	~	*	*	*	*

Note: * represents the classification errors of the first model is significantly lower than the corresponding values of the counterpart ($p < 0.05$), and x represents it is significantly higher than that of the counterpart. ~ represents the difference is not significant.

TABLE III
STATISTICAL ANALYSIS OF CLASSIFICATION ERRORS AMONG FIVE MODELS FOR WRIST MYOELECTRIC SIGNALS

Gesture	TCN	TCN	TCN	Transformer	Transformer	CNN	TDLDA	TDLDA	TDLDA	TDLDA
	vs. Transformer	vs. CNN	vs. GRU	vs. CNN	vs. GRU	vs. GRU	vs. TCN	vs. CNN	vs. Transformer	vs. GRU
LP										
TA										
TLFO										
TIFO										
TIFE										
TLFE										
IMFE										
LFE										
IFE			~							
TFE										
WF										
WE										
FS										
FP										
HO										
HC										
R	~	*	*	~	~	~	~	*	~	*

Note: * represents the classification errors of the first model is significantly lower than the corresponding values of the counterpart ($p < 0.05$), and x represents it is significantly higher than that of the counterpart. ~ represents the difference is not significant.

the encoder layer for Transformer. As the feature dimensions of CNN and TCN were too large, which were 12992 and 6953, respectively, the DBI and t-SNE were only performed for GRU, Transformer, and TDLDA, of which the feature dimensions were 32, 32, and 24, respectively.

The data distribution of one representative subject after t-SNE is displayed in Fig. 9. For the deep learning models, GRU and Transformer, the clusters of wrist myoelectric signals were more concentrated compared to those of forearm myoelectric signals. For TDLDA, the cluster size was similar between wrist and forearm myoelectric signals. For quantitative comparison, the DBI values are displayed in Fig. 10. For GRU and Transformer, the DBI of wrist myoelectric signals was smaller than that of forearm myoelectric signals,

indicating the separability was improved. As for TDLDA, the DBI of wrist myoelectric signals were larger compared to forearm myoelectric signals.

Two-way ANOVA revealed that there was a significant interaction between the factors of the model and electrode position ($p = 0.000$, $F = 109.86$). With the fixed level of the model, the subsequent one-way ANOVA showed that the DBI differences between the forearm and the wrist were all significant for the three models (GRU: $p = 0.000$, $F = 427.32$; Transformer: $p = 0.000$, $F = 408.91$; TDLDA: $p = 0.000$, $F = 38.04$). Both the visual and quantitative results indicated that the data separability was significantly improved with wrist myoelectric signals compared to forearm myoelectric signals for GRU and Transformer, which was different from TDLDA.

C. Computational Cost

The training cost of each method was evaluated by a server with 40-thread Intel(R) Xeon(R) Silver 4316 CPU and an NVIDIA Geforce RTX 3090 GPU. The amount of training data was 6 trials per class, resulting in 49062 samples in total. The average training time of each model is listed in Table V. The time of TDLDA was much smaller than those of the deep learning models, which was expected. For the deep learning models, TCN and GRU had the least and most training cost, respectively. Further, we also compared the time each model took to generate decisions from the inputs. Two scenarios were tested, *i.e.*, 128 samples and one sample, and the results are listed in Table V. For the five models, the sequence of testing time cost was the same as that of training time cost, and the average time cost of a sample from 128 samples was faster than that from one single sample due to parallel computation of the machine. All the models were able to generate an output in less than 10 ms, which was the interval of the two input samples. Additionally, the time cost of TCN was also tested with a Raspberry Pi 4B, a 4-thread, 1.5GHz-main frequency microcontroller. The inference time was 986.89 ms and 8.36 ms for 128 samples and 1 sample, respectively, which were both less than 10 ms for one decision. It indicated that the deep learning methods had the potential to be employed in embedded systems for the wearables.

IV. DISCUSSION

This study investigated the performance of four deep learning models for a thirteen-class task with myoelectric signals from the wrist and forearm, respectively, and they were compared to that of a traditional method TDLDA. The four deep learning models were good at extracting either spatial or temporal information, and all were reported to have a good classification performance of hand movements with forearm myoelectric signals [16], [17], [19], [20]. The structures and parameters of the deep learning models were initially proposed based on forearm myoelectric signals. The results showed that with forearm myoelectric signals, the difference between deep learning models and TDLDA was insignificant for most gestures, and the average classification errors were close as well. With wrist myoelectric signals, the classification errors of deep learning models were significantly reduced compared to the errors with forearm myoelectric signals. The difference was between 6.2% and 9.8%. However, for TDLDA, the error difference between wrist and forearm myoelectric signals was around 1.4%, *i.e.*, the gesture recognition performance was little affected by the signal modality. As the performance between deep learning models and TDLDA was similar with forearm myoelectric signals, with wrist myoelectric signals, the gesture recognition performance was significantly improved by deep learning models compared to TDLDA. Previous studies showed the possibility of recognizing hand movements from wrist myoelectric signals using traditional methods specifically designed for processing forearm myoelectric signals [10]. However, as the power of deep learning on myoelectric control from forearm muscles was demonstrated in more and more studies [8], [14], its performance with wrist myoelectric signals was not reported yet. The wrist was a

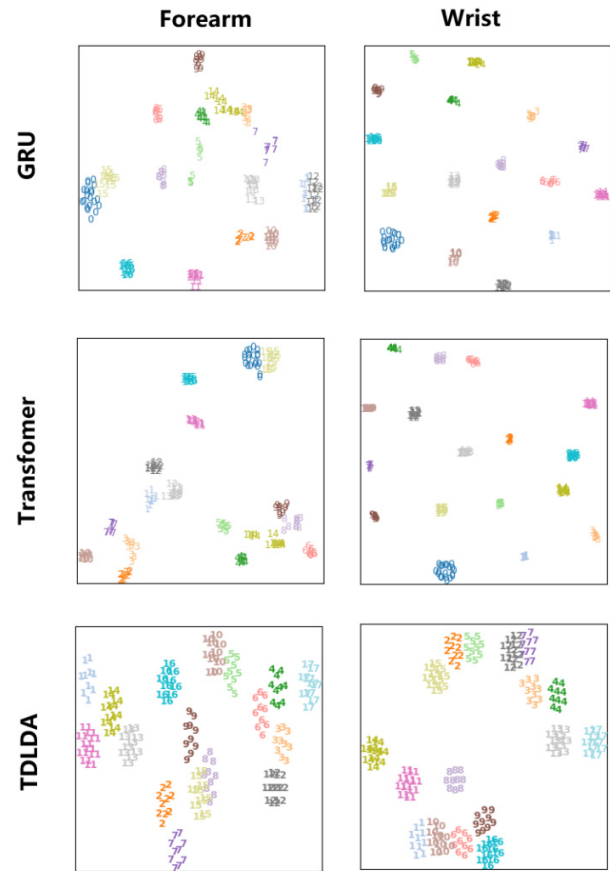


Fig. 9. Visual distribution of the features after t-distributed stochastic neighbor embedding (t-SNE) for GRU, Transformer and TDLDA. The class separability of the wrist is better than that of the forearm for GRU and Transformer.

more favorable part for the wearables compared to the forearm. The results would benefit the development of wrist-based HMI and the extension of myoelectric control technology into more applications.

It was shown that deep learning models were better than the traditional method in recognizing gestures with myoelectric signals from the wrist, but not the forearm. It might be related to the difference of the physiological structure between the forearm and wrist (See Supplement). For the forearm, there were mainly muscle bellies, which were the widest parts of the muscles. As such, an electrode with limited space could only pick up information from a few muscles. However, for the wrist, there were mainly muscle tendons. They were the ends of the muscles and distributed narrowly around the wrist. As the size was small, the pickup area of an electrode covered multiple tendons, and the related information was collected. As such, the information richness of one channel would be higher from the wrist compared to the forearm. With the same number of channels, the signals from the wrist might provide more discriminative information than the forearm. With the help of deep learning power, the discriminative information was captured and learned, increasing the accuracy of gesture recognition. From Table I, the advantage of deep learning models was mainly on finger gestures with wrist myoelectric signals. The muscles associated with finger movements, such

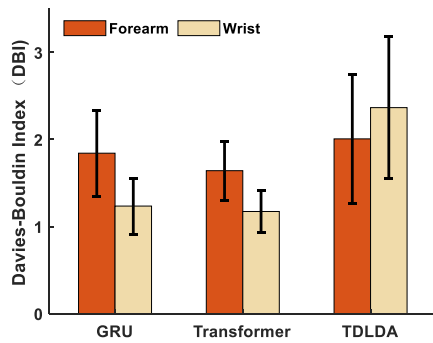


Fig. 10. Davies-Bouldin index (DBI) for quantitative comparison of the features among the three models. For GRU and Transformer, the values of the wrist are smaller than the values of the forearm, corresponding to a better class separation.

as the superficial flexor and the deep flexor, were small or deeply buried in the forearm. The electrodes on the forearm could not capture much information about these muscles, while the electrodes on the wrist could have their activities from the tendons. The information was well decoded by deep learning models, resulting in an improvement in finger gesture recognition performance.

Except for the classification error rates, visual and quantitative measurements were also employed to illustrate the feature separability of each method. There was consistency among these measurements with the recognition performance comparison between wrist and forearm myoelectric signals. In Fig. 7, the error rates of GRU and Transformer were much lower with wrist myoelectric signals. And in Fig. 10, for GRU and Transformer, the DBI values of wrist myoelectric signals were smaller compared to the values of forearm myoelectric signals (GRU: 1.24 vs. 1.84, Transformer: 1.17 vs. 1.64), indicating a better class separation. The clusters after t-SNE from wrist myoelectric signals were also smaller in Fig. 9. It should be noted that for TDLDA, the DBI difference between wrist and forearm myoelectric signals was significant, while the difference of the error rates was not. This indicated that DBI might be more sensitive than error rates. The results of DBI and t-SNE both confirmed the advantage of deep learning models on wrist myoelectric signal-based gesture recognition.

The results of TDLDA were different from a previous study [10], which employed four channels and showed lower classification errors of finger gestures with myoelectric signals from the wrist compared to the forearm. As this study employed six channels, a reduced number of electrodes would lead to a decline in classification performance. The degradation of classification performance would be different between the wrist and forearm myoelectric signal. The decrease might be larger in forearm myoelectric signals for less information was contained. Further, the locations of the employed electrodes in this study were optimized from eight pairs of electrodes equidistantly distributed around the forearm. Its classification errors of forearm myoelectric signals were lower than that of the other locations.

Four deep learning models were investigated in this study. Considering factors including model complexity,

TABLE IV

AVERAGE TIME OF MODEL TRAINING PER SUBJECT IN SECONDS

	CNN	GRU	TCN	Transformer	TDLDA
Time (s)	154.67	345.05	44.77	107.75	0.13

Note: The size of training data is 49062 samples.

TABLE V

AVERAGE TIME OF MODEL TESTING PER SUBJECT IN MILLISECONDS

	CNN	GRU	TCN	Transformer	TDLDA
128 samples	3.09	14.66	1.33	2.42	0.22
1 sample	0.58	6.76	0.71	1.22	0.0068

computational cost, recognition performance, *etc.*, TCN was considered the best option for gesture recognition from wrist myoelectric signals. The structure of TCN was simple with 1d convolutional operation as the major operation. As shown in Table IV and V, its computational cost was relatively low compared to the other three deep learning models. Besides, structural simplicity did not lead to a decline in recognition performance. The classification error rates of TCN were the lowest on both wrist and forearm myoelectric signals in Fig. 7. Only offline results were presented. Based on the computational cost in Table V, all the models had the possibilities to be used in real-time for the time cost of one sample was smaller than the increment (10 ms). Besides, compared to discrete gesture recognition, continuous estimation of hand kinematics would provide more detailed information about the movements, which might be more useful in practical applications. Future studies will focus on these directions.

V. CONCLUSION

This study compared the performance of gesture recognition between TDLDA and four deep learning models with wrist and forearm myoelectric signals, respectively. Though the deep learning models were designed for forearm myoelectric signals, they showed much lower error rates and higher class separability than TDLDA for gesture recognition from wrist myoelectric signals. With deep learning methods, the error rates of gesture recognition from wrist myoelectric signals were also lower than those from forearm myoelectric signals. This study highlighted the potential of deep learning methods for decoding movements from wrist myoelectric signals, benefiting the development of wrist-based sEMG wearables and facilitating the extension of myoelectric control technology into more sections.

REFERENCES

- [1] K. R. Mills, "The basics of electromyography," *J. Neurol., Neurosurg. Psychiatry*, vol. 76, no. 2, pp. 32–35, Jun. 2005, doi: 10.1136/jnnp.2005.069211.
- [2] D. Farina et al., "Man/machine interface based on the discharge timings of spinal motor neurons after targeted muscle reinnervation," *Nature Biomed. Eng.*, vol. 1, no. 2, p. 25, Feb. 2017, doi: 10.1038/s41551-016-0025.
- [3] N. Jiang, S. Dosen, K.-R. Müller, and D. Farina, "Myoelectric control of artificial limbs—Is there a need to change focus? [In the spotlight]," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 150–152, Sep. 2012, doi: 10.1109/MSP.2012.2203480.

- [4] X. Yang, X. Chen, X. Cao, S. Wei, and X. Zhang, "Chinese sign language recognition based on an optimized tree-structure framework," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 4, pp. 994–1004, Jul. 2017, doi: [10.1109/JBHI.2016.2560907](https://doi.org/10.1109/JBHI.2016.2560907).
- [5] C. G. McDonald, J. L. Sullivan, T. A. Dennis, and M. K. O'Malley, "A myoelectric control interface for upper-limb robotic rehabilitation following spinal cord injury," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 978–987, Apr. 2020, doi: [10.1109/TNSRE.2020.2979743](https://doi.org/10.1109/TNSRE.2020.2979743).
- [6] H. Zeng, W. Yu, D. Chen, X. Hu, D. Zhang, and A. Song, "Exploring biomimetic stiffness modulation and wearable finger haptics for improving myoelectric control of virtual hand," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1601–1611, 2022, doi: [10.1109/TNSRE.2022.3181284](https://doi.org/10.1109/TNSRE.2022.3181284).
- [7] L. M. Doornebosch, D. A. Abbink, and L. Peternel, "Analysis of coupling effect in human-commanded stiffness during bilateral tele-impedance," *IEEE Trans. Robot.*, vol. 37, no. 4, pp. 1282–1297, Aug. 2021, doi: [10.1109/TRO.2020.3047064](https://doi.org/10.1109/TRO.2020.3047064).
- [8] D. Yang and H. Liu, "An EMG-based deep learning approach for multi-DOF wrist movement decoding," *IEEE Trans. Ind. Electron.*, vol. 69, no. 7, pp. 7099–7108, Jul. 2022, doi: [10.1109/TIE.2021.3097666](https://doi.org/10.1109/TIE.2021.3097666).
- [9] A. Kamišalić, I. Fister, M. Turkanović, and S. Karakatić, "Sensors and functionalities of non-invasive wrist-wearable devices: A review," *Sensors*, vol. 18, no. 6, p. 1714, May 2018, doi: [10.3390/s18061714](https://doi.org/10.3390/s18061714).
- [10] F. S. Botros, A. Phinyomark, and E. J. Scheme, "Electromyography-based gesture recognition: Is it time to change focus from the forearm to the wrist?" *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 174–184, Jan. 2022, doi: [10.1109/TII.2020.3041618](https://doi.org/10.1109/TII.2020.3041618).
- [11] S. Jiang et al., "Feasibility of wrist-worn, real-time hand, and surface gesture recognition via sEMG and IMU sensing," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3376–3385, Aug. 2018, doi: [10.1109/TII.2017.2779814](https://doi.org/10.1109/TII.2017.2779814).
- [12] J. He and N. Jiang, "Fast multi-level mental stress identification from bispectrum-based heart rate variability feature," *IEEE Trans. Ind. Informat.*, pp. 1–10, 2023, doi: [10.1109/TII.2023.3262248](https://doi.org/10.1109/TII.2023.3262248).
- [13] D. Xiong, D. Zhang, X. Zhao, and Y. Zhao, "Deep learning for EMG-based human-machine interaction: A review," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 3, pp. 512–533, Mar. 2021, doi: [10.1109/JAS.2021.1003865](https://doi.org/10.1109/JAS.2021.1003865).
- [14] N. Jiang et al., "Bio-robotics research for non-invasive myoelectric neural interfaces for upper-limb prosthetic control: A 10-year perspective review," *Nat. Sci. Rev.*, vol. 10, no. 5, pp. 1–12, Feb. 2023, doi: [10.1093/NSR/NWAD048](https://doi.org/10.1093/NSR/NWAD048).
- [15] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, p. 36571, Nov. 2016, doi: [10.1038/srep36571](https://doi.org/10.1038/srep36571).
- [16] A. R. Asif et al., "Performance evaluation of convolutional neural network for hand gesture recognition using EMG," *Sensors*, vol. 20, no. 6, p. 1642, Mar. 2020, doi: [10.3390/s20061642](https://doi.org/10.3390/s20061642).
- [17] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, and A. Skodras, "Improved gesture recognition based on sEMG signals and TCN," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1169–1173, doi: [10.1109/ICASSP.2019.8683239](https://doi.org/10.1109/ICASSP.2019.8683239).
- [18] M. Simão, P. Neto, and O. Gibaru, "EMG-based online classification of gestures with recurrent neural networks," *Pattern Recognit. Lett.*, vol. 128, pp. 45–51, Dec. 2019, doi: [10.1016/j.patrec.2019.07.021](https://doi.org/10.1016/j.patrec.2019.07.021).
- [19] E. Rahimian, S. Zabihi, A. Asif, D. Farina, S. F. Atashzar, and A. Mohammadi, "TEMGNet: Deep transformer-based decoding of upperlimb sEMG for hand gestures recognition," 2021, *arXiv:2109.12379*.
- [20] R. Chen, Y. Chen, W. Guo, C. Chen, Z. Wang, and Y. Yang, "sEMG-based gesture recognition using GRU with strong robustness against forearm posture," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Jul. 2021, pp. 275–280, doi: [10.1109/RCAR52367.2021.9517639](https://doi.org/10.1109/RCAR52367.2021.9517639).
- [21] J.-S. Ang, K.-W. Ng, and F.-F. Chua, "Modeling time series data with deep learning: A review, analysis, evaluation and future trend," in *Proc. 8th Int. Conf. Inf. Technol. Multimedia (ICIMU)*, Aug. 2020, pp. 32–37, doi: [10.1109/ICIMU49871.2020.9243546](https://doi.org/10.1109/ICIMU49871.2020.9243546).
- [22] N. Jiang, A. Pradhan, and J. He, "Gesture recognition and biometrics electroMyogram (GRABMyo)," *PhysioNet*, 2022. Accessed: Jan. 10, 2023. [Online]. Available: <https://doi.org/10.13026/rtfc-np50>
- [23] J. He, X. Sheng, X. Zhu, C. Jiang, and N. Jiang, "Spatial information enhances myoelectric control performance with only two channels," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 1226–1233, Feb. 2019, doi: [10.1109/TII.2018.2869394](https://doi.org/10.1109/TII.2018.2869394).
- [24] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Jun. 2017, pp. 6000–6010.
- [25] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, Jun. 2003, doi: [10.1109/TBME.2003.813539](https://doi.org/10.1109/TBME.2003.813539).
- [26] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
- [27] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).