A Transformer-Based Multi-Task Learning Framework for Myoelectric Pattern Recognition Supporting Muscle Force Estimation

Xinhui Li, Xu Zhang, *Member, IEEE*, Liwei Zhang, Xiang Chen, *Member, IEEE*, and Ping Zhou, *Senior Member, IEEE*

***Abstract*— Simultaneous implementation of myoelectric pattern recognition and muscle force estimation is highly demanded in building natural gestural interfaces but a challenging task due to the gesture classification accuracy degradation under varying muscle strengths. To address this problem, a novel method using transformer-based multi-task learning (MTL-Transformer) for the prediction of both myoelectric patterns and corresponding muscle strengths was proposed to describe the inherent characteristics of an individual gesture pattern under different force conditions, thereby improving the accuracy of myoelectric pattern recognition. In addition, the transformer model enabled the characterization of long-term temporal correlations to ensure precise and smooth estimation of the muscle force. The performance of the proposed MTL-Transformer framework was evaluated via experiments of classifying eleven hand gestures and estimating the corresponding muscle force simultaneously, using high-density surface electromyogram (HD-sEMG) recordings from forearm flexor muscles of eleven intact-limbed subjects. The MTL-Transformer framework yielded high classification accuracy (98.70±1.21%) and low root mean square deviation (12.59±2.76%), and outperformed other two common temporally modelling methods significantly ($p < 0.05$) in terms of both improved gesture recognition accuracies and reduced muscle force estimation errors. The MTL-Transformer framework is demonstrated as an effective solution for simultaneous implementation of myoelectric pattern recognition and muscle force estimation. This study promotes the development of robust and smooth myoelectric control systems, with wide applications in gestural interfaces, prosthetic and orthotic control.**

***Index Terms*— Myoelectric pattern recognition, muscle force estimation, varying muscle strengths, transformer model, multi-task learning.**

Manuscript received 10 February 2023; revised 17 June 2023; accepted 12 July 2023. Date of publication 25 July 2023; date of current version 18 August 2023. This work was supported by the National Natural Science Foundation of China under Grant 62271464. *(Corresponding author: Xu Zhang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Review Board of the University of Science and Technology of China (USTC), Hefei, Anhui, China, under Application No. 2022-N(H)-163, in February 2022.

Xinhui Li, Xu Zhang, and Xiang Chen are with the School of Microelectronics, University of Science and Technology of China, Hefei, Anhui 230027, China (e-mail: xuzhang90@ustc.edu.cn).

Liwei Zhang is with the First Affiliated Hospital, University of Science and Technology of China, Hefei, Anhui 230001, China.

Ping Zhou is with the Department of Biomedical and Rehabilitation Engineering, University of Health and Rehabilitation Sciences, Qingdao, Shandong 266024, China (e-mail: dr.ping.zhou@outlook.com).

Digital Object Identifier 10.1109/TNSRE.2023.3298797

I. Introduction

MYOELECTRIC control is a technology that converts human movement intentions into machine commands by sensing and processing electromyographic (EMG) signals to control peripheral devices. It has been widely used as gestural interfaces in prosthetic and orthotic robots [1], [2], [3]. Due to its favorable non-invasive property, the surface EMG (sEMG) is usually used as the command source in the myoelectric control systems [4], [5], [6]. In recent years, a number of studies in myoelectric control have been devoted to the interpretation of movement patterns from the sEMG signals [7], [8], [9]. Many machine learning methods such as linear discriminant classifier [10], Gaussian mixture model [11], support vector machine [12], have been adopted to process the sEMG signals and improve the number of recognizable patterns and recognition accuracy, with significant progresses [13], [14], [15]. In particular, the rapid development of deep learning algorithms in recent years has significantly advanced the techniques for myoelectric control [16], [17], [18]. To reduce the adverse interference when exploring the feasibility of the recognition methods, these studies are usually carried out with different movement patterns under constant medium force levels, without considering potential variations of the muscle force. Intuitively, both movement pattern recognition and muscle force estimation are not separate tasks. For instance, when griping on an object by a prosthetic control system, both the movement pattern and muscle force are generated in a well-coordinated manner so as to achieve natural and smooth control. Consequently, it's necessary to validate the gesture pattern recognition algorithm for the sEMG signals under the condition of varying forces, and this further motivates the research on simultaneous implementation of both gesture recognition and muscle force estimation.

Towards advanced myoelectric pattern recognition supporting muscle force estimation, several studies have been

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

conducted. For example, Baldacchino et al. [19] proposed a multivariate Bayesian hybrid model based on the gate function, which can achieve the pattern recognition of nine finger movements and force estimation of the fingertip. Fang et al. [20] proposed an attribute-driven granular (AGrM) model for recognizing eight finger pinch patterns and estimating fingertip forces. Despite the achievements of these works, their pattern recognition performance degraded significantly under variable forces. One main reason was that both the gesture recognition and the muscle force estimation were treated as independent tasks, ignoring their complementary properties underlying complex muscle coordination.

Since the gesture pattern and muscle force can both be predicted from the sEMG signal, the multi-task learning (MTL) framework is naturally considered. The MTL approaches aim to learn multiple related tasks simultaneously, by sharing the feature representations of different tasks, it can achieve better generalization ability than learning individual task independently [21]. Hua et al. [22] used a MTL framework based on the multi-stream temporal convolutional neural network (TCNN) to simultaneously make decisions within eight movement patterns and three corresponding force levels. This method just considered three fixed force levels for each pattern, which had certain limitations in real world applications. Hu et al. [23] proposed a MTL framework based on the long-short term memory (LSTM) network and the multi-layer perceptron (MLP), incorporating a post-processing approach. It enabled the recognition of eleven gestures while supporting instantaneous estimation of the muscle force of the activated gesture. However, the post-processing algorithm can lead to a large time delay, which was not conducive to the real-time requirement of the myoelectric control system. Besides, these methods have just achieved unsatisfying and limited performance (an average accuracy of just around 90%).

In the simultaneous control task, gesture recognition is usually a more important issue, and the prediction of muscle force is meaningful only when the gesture patterns are recognized correctly. Meanwhile, the main difficulty of the simultaneous control task also lies in overcoming the degradation of gesture recognition accuracy under the influence of variable forces. Movement patterns have been frequently characterized by sEMG features. Most of them are associated with sEMG amplitudes [6], such as time domain (TD) features [24], [25], and they may change obviously with varying forces, leading to decreased pattern recognition performance. This places a higher demand on the user's operational normality in the application of myoelectric control systems, resulting in poor user experience [22]. To deal with this problem, some methods have been proposed to improve the generalization ability of the classification algorithm by extracting sEMG features that are insensitive to force changes, thus reducing the variation in feature space caused by variable contraction forces [26], [27], [28]. For example, Al-Timmy et al. [27] used the time-dependent power spectral descriptors (TD-PSD) of sEMG signals on a six-class classification task under three force levels, reducing the classification error significantly when compared to the conventional characteristics such as autoregressive model coefficient, discrete Fourier transform coefficient, and wavelet transform coefficient. Pancholi and Joshi [28] proposed an energy kernel-based feature extraction method, with an average classification accuracy of 92% for six gestures under three force levels, which achieved a 2%-9% improvement over TD-PSD and wavelet transform coefficients. Although some progresses have been made, the decoding of movement patterns under varying forces is still unsatisfactory.

Due to the sequential properties of sEMG data, it is essential to mine the temporal relevance along the data sequence. It is hypothesized that temporal modeling of signal sequences helps to learn robust features of one gesture by aggregating information from sEMG data over varying force levels, thus improving the accuracy of gesture pattern recognition. In recent years, the transformer model has attracted wide attention due to its powerful temporal modeling capability and has been successfully applied in speech recognition, machine translation and many other computer vision tasks [29], [30], [31]. The key of the transformer model lies in the self-attention mechanism, it allows the data point in the input sequence to interact with each other by computing the similarity score (attention weight) among them [32]. The self-attention mechanism can help to capture long-term dependencies in the time sequence and aggregate global information of the data, instead of only focusing on the local context information as in convolutional neural networks. Besides, compared with recurrent neural networks (RNNs) such as LSTM [33], [34], [35] with the similar capability of aggregating global context information, the transformer has a property of parallel computation [36], which can reduce the training time cost and improve the execution efficiency. Although the transformer model has been utilized in the myoelectric pattern recognition tasks with promising performance [37], [38], [39], its effectiveness has not been investigated in simultaneous implementation of both the myoelectric pattern recognition task and the muscle force estimation task.

To reduce the negative impacts of varying muscle strengths and simultaneously predict both the gestural pattern and the force, we proposed a novel transformer-based multi-task learning (MTL-Transformer) method for myoelectric pattern recognition supporting muscle force estimation. In this method, the sEMG data samples were characterized as features and fed into the transformer model, and then went through the classification module and the regression module simultaneously to obtain decisions of both gesture pattern and instantaneous muscle force. Our proposed method can achieve efficient and robust myoelectric control, which is of great significance to gestural interfaces, prosthetic and orthotic control.

II. Methodology

Figure 1 demonstrates the flowchart of the proposed method. First, high-density sEMG (HD-sEMG) from the forearm flexor and the corresponding grasping force are collected simultaneously when the gestures are executed. The HD-sEMG data are used to extract sEMG envelopes in channel-wise manner. The multi-channel sEMG signals and the corresponding multi-channel envelope signals are stitched together, which are segmented into a series of multi-channel time windows and then fed into the transformer model. For

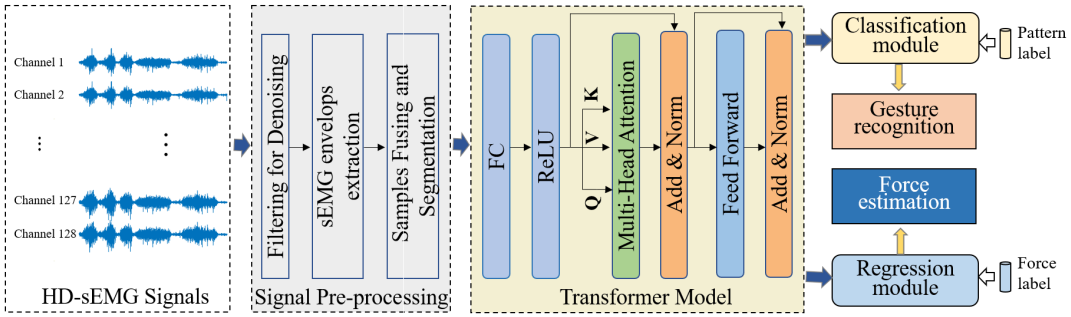

Fig. 1. The flowchart of the proposed MTL-Transformer method.

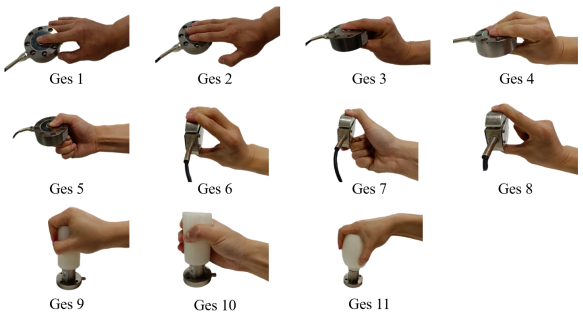

Fig. 2. Illustration of 11 different gesture patterns.

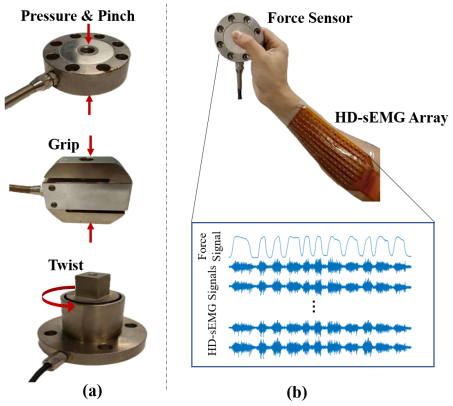

Fig. 3. The diagram of the experiment set-up. (a) Three force sensors including two pressure sensors and a torque sensor. (b) Schematic diagram of force sensor and HD-sEMG array placement in a gesture task.

each sEMG sample in one window, features obtained from the transformer model are fed into the classification module and the regression module simultaneously to obtain the gesture pattern and instantaneous muscle force.

A. Subjects and Experiments

Eleven male subjects aged from 23 to 27 years old were recruited in this study. All subjects did not have any neuromuscular diseases and were informed of the experimental procedures and signed the informed consent. The study was approved by the Ethics Review Board of the University of Science and Technology of China (Hefei, China).

Eleven gestures involving pressure, pinch, grip and twist were selected from commonly used daily gestures to form the target gesture set in this study, as shown in Fig. 2. Several hand molds from 3-D printing were adopted to assist the data collection of twist gestures. The diagram of the experiment set-up was shown in Fig. 3. As shown in Fig. 3(a), two pressure sensors (LOADING SEN, LDCZL-FA & LDCZL-SC, China) and a torque sensor (LOADING SEN, LDN-08A, China) were used to record the grasping force. A HD-sEMG electrode array consisting of 128 electrodes arranged in a 16×8 grid form was used to collect HD-sEMG signals. Each electrode had a circular recording probe of 3-mm diameter, and the center-by-center inter-electrode distance between two neighboring electrodes was 8 mm. Each electrode in the array worked in a monopolar manner concerning the common reference electrode that was placed on the back of the other hand, constituting 128 recording channels.

At the beginning of the experiment, the skin of the subjects' forearms was cleaned with medical-grade alcohol to reduce the skin-electrode impedance. As shown in Fig. 3(b), the HD-sEMG array was placed to the skin surface of the forearm flexor muscle (containing ulnar carpal flexor, radial carpal flexor, and intra-hand flexor), whose primary function corresponds to the eleven gestures, and an extra inelastic bandage was used to secure the HD-sEMG array and reduce the motion artifacts. Then the subjects were asked to perform three maximal voluntary contractions (MVCs) while their corresponding forces were recorded, and the largest force was used as the final MVC for every subject. Then, subjects were instructed to perform target gestures in a variable force generation pattern that rises smoothly from the initial baseline (resting state) to 60% MVC force level and then falls smoothly to baseline. The duration of each force generation pattern was maintained between 2s and 3s, and the force generation pattern was repeated 20 times for each user performing each gesture. The target force and the actual force generation curve were displayed on a human-computer interaction interface to help the subjects to better complete the force generation task, and the actual measured force was used in the subsequent signal processing process.

A homemade recording system was used for both force and sEMG data collection. There are a two-stage amplifier with a total gain of 60dB and a band-pass filter at 20–500 Hz per

channel. The HD-sEMG and force signals were sampled at 2 kHz and digitized via a 16-bit A/D converter (ADS1299, Texas Instruments, TX). All collected sEMG and force data were saved to the hard disk of a computer through a high-speed USB cable. All subsequent data processing and analyses were conducted on a desktop PC equipped with an Intel i7 CPU and an NVIDIA GeForce GTX 1080Ti.

B. Signal Pre-Processing

A 20-500 Hz band-pass filter was applied to HD-sEMG signals first to eliminate low-frequency noise artifacts. Then a set of second-order notch filters were adopted to remove the 50-Hz power line interference and its harmonics for each sEMG channel. For the three twist gestures (Ges 9, Ges 10, and Ges 11), the torque recorded by the force sensor was converted into the corresponding twist force:

$$\hat{D} = I/d, \tag{1}$$

where $\hat{D}$ represents the calculated twist force, I is the measured torque, and d is the arm of the hand mold's force. For each gesture and each subject, both the HD-sEMG data and force data were normalized separately. We first calculated a global maximum absolute value of the sEMG amplitude over all channels. Once the maximum absolute amplitude value was obtained, it was used to normalize each channel of sEMG signal between -1 and 1. In addition, the force signal was normalized between 0 and 1 using the previously recorded MVC value. Furthermore, we used the full-wave rectification and low-pass filtering (cutoff frequency 3 Hz, finite impulse response filter, Hanning window, 80th order) to process each channel of the normalized HD-sEMG signals in the temporal dimension, obtaining a corresponding signal envelope. These multi-channel envelopes were also normalized in the same way according to the maximum absolute value across all channels. Considering following supervised learning analyses, all of the above parameters for normalization were just derived from the training data and they were stored and applied to any test data. Thus a channel-augmented data stream was obtained by concatenating the normalized multi-channel envelopes and the normalized multi-channel sEMG data. These data also carried their original labels including both the gesture pattern and the corresponding measured force. The sEMG data along with the force signal were segmented into several overlapping analysis windows with a window length of 64 ms and a window increment of 32 ms. These multi-channel analysis windows were considered as the basic samples for both myoelectric pattern recognition and muscle force estimation tasks.

C. Model Structure and Model Training

In this section, we introduce the detailed design of our MTL-Transformer model. Given the input data $X \in R^{B\times T\times M}$ with pattern labels $Y^{cls} \in R^{B}$ and force labels $Y^{reg} \in R^{B\times T}$ in a mini-batch, where B is the batch size, T is the length of data samples in a window and M is the channel number of fusion samples. To fuse the feature information of different electrodes, we first adopted a fully connected layer and a ReLU activation layer to map X into P channels (i.e., empirically 256 channels in this study) as follows:

$$U = Max(XW + b, 0), \tag{2}$$

where $W \in R^{M\times P}, b \in R^{P}$. Note that the T data samples in a window correspond to the same gesture pattern label but have different muscle force values, and the muscle force is not inconsecutive and it's important to consider the smoothness of the muscle force prediction results. To correctly recognize the gesture pattern of these data samples and estimate the muscle force, temporal modeling is essential for improving the feature robustness and smoothness. Consequently, the output U was then fed into the multi-head attention module to aggregate temporal information. Specifically, the multi-head attention module consists of N heads, and each head processes the input independently. For the n-th head, we first map U into query Q_n, key K_n and value V_n:

$$Q_n = UW_n^Q, K_n = UW_n^K, V_n = UW_n^V \tag{3}$$

where $W_n \in R^{P\times \frac{P}{N}}$ represents the learnable weight. Then query Q_n and key K_n were used to calculate the similarity matrix among data samples:

$$S_n = softmax\left(\frac{Q_n K_n'}{\tau}\right) \tag{4}$$

where $\tau = \sqrt{P/N}$ is the scale factor, K_n' is the transpose of the K_n. With the similarity matrix S_n, the output H_n of the n-th head can be obtained:

$$H_n = S_n V_n. \tag{5}$$

By concatenating the output of N heads together, we can obtain the aggregated temporal feature for each data sample according to Equation (6). In this paper, the heads number N was set to be 4, so the dimension of H is $R^{B\times T\times 256}$. To keep the original feature so as to prevent over smoothing for muscle force estimation, U and H were fused together and then fed into a layer-normalization layer to obtain U' according to Equation (7). At last, U' was fed into a feed forward module FFN and a layer-normalization layer to obtain the output of the transformer model according to Equation (8).

$$H = Concat\,(H_1, \ldots, H_N) \tag{6}$$

$$U' = LayerNormalization(U + H) \tag{7}$$

$$U'' = LayerNormalization\left(FFN\left(U'\right) + U'\right) \tag{8}$$

The feed forward module FFN consists of two fully connected layers, with a ReLU activation layer between them and a normalization layer after them. For the FFN module, each FC layer had 256 input channels and 256 output channels.

With the output features of each sample $U'' \in R^{256\times T}$, the classification scores $\hat{Y}^{cls} \in R^{B\times C}$ and muscle force estimation results $\hat{Y}^{reg} \in R^{B\times T}$ were obtained by the classification module N^{cls} and regression module N^{reg} according to Equation (9). In the classification module, we used the temporal average pooling operation on U'' in the second dimension to obtain the output $U''' \in R^{256}$, which was then fed into the fully connected layer. For the classification module, the

number of the input channels was 256, the number of output channels was the pattern number of the gestures (*i.e.*, 11 in this study). In addition, the regression module was composed of a fully connected layer and a sigmoid activation layer. For the regression module, there were 256 input channels, and just 1 output channel.

$$\hat{Y}^{cls} = N^{cls}\left(U''\right), \hat{Y}^{reg} = N^{reg}\left(U''\right) \tag{9}$$

In the training stage, the real gesture label Y^{cls} and muscle force value Y^{reg} were available. The network was trained with the stochastic gradient descent (SGD) algorithm [40], with the batch size of B (i.e., 10 in this study). Consequently, the classification loss L^{cls} and muscle force estimation loss L^{reg} can be calculated as the cross entropy loss and mean square error loss according to Equation (10) and Equation (11), respectively, where C represents the pattern number of the gestures, $Y^{cls}_{j,c}$ means the gesture label value of the j-th sample labeled as pattern c in a mini-batch, $\hat{Y}^{cls}_{j,c}$ means the predicted probability that the j-th sample belongs to pattern c, $Y^{reg}_{j,t}$ means the real force value of the t-th sampling points of the j-th sample, $\hat{Y}^{reg}_{j,t}$ means the corresponding predicted force value. The final loss used for model training was the weighted sum of two losses defined as Equation (12), where the weight α was used to balance the contribution of the muscle force estimation loss. Since the regression loss was found to be about ten times smaller than the classification loss, α was set from 0 to 10 by every increment step of 1 to find an appropriate value leading to optimal performance. The learning rate was set to be 1×10^{-3} in this paper, and the number of training epochs was set to be 20.

$$L^{cls} = -\frac{1}{B}\sum_{j=1}^{B}\sum_{c=1}^{C} Y^{cls}_{j,c}\log\left(\hat{Y}^{cls}_{j,c}\right) \tag{10}$$

$$L^{reg} = \frac{1}{B}\sum_{j=1}^{B}\sum_{t=1}^{T}\left(Y^{reg}_{j,t} - \hat{Y}^{reg}_{j,t}\right)^2 \tag{11}$$

$$L = L^{\text{cls}} + \alpha L^{\text{reg}} \tag{12}$$

D. Model Testing and Decision Making

In the testing stage, given the testing data $X \in R^{L\times M}$, we feed it into the MTL-Transformer model to obtain the gesture classification result $\hat{Y}_{cls} \in R^{C}$ and muscle force estimation result $\hat{Y}_{reg} \in R^{T}$. Since $\hat{Y}_{cls}$ was a pattern distribution, the predicted pattern label c can be obtained:

$$c = \underset{i}{argmax}\,\hat{Y}_{cls}(i) \tag{13}$$

E. Performance Evaluation

To evaluate the effectiveness of the proposed method, the training set and testing set were divided in the proportion of 3:7 for each subject's data. The classification accuracy (CA) described in Equation (14) and the root mean square deviation (RMSD) defined in Equation (15) were used to evaluate the performance of gesture recognition and force estimation respectively, where F and $\hat{F}$ are the predicted force and the measured force, respectively.

$$CA = \frac{Correct\ Instances}{Total\ Instances} \times 100\% \tag{14}$$

$$RMSD = \sqrt{\frac{\sum_{t=1}^{T}[\hat{F}(t) - F(t)]^2}{L}} \times 100\% \tag{15}$$

To validate the advantage of the proposed method, two common temporal modeling approaches were applied to construct MTL framework, which can realize simultaneous gesture recognition and force estimation. One was based on the LSTM model (termed MTL-LSTM), where a FC layer, a ReLU layer and a LSTM layer were used to obtain the features of each sample that then went through the classification and regression modules, thus the predicted gesture pattern and muscle force can be obtained. The other one was based on the multi-stream temporal convolutional neural network (termed MTL-TCNN) according to the previous study [10]. In this work, data from each channel of the fusion samples were used as the input of each stream. Three *conv* blocks containing a BN layer, a *conv* layer and a maxpooling were utilized to extract features of each sample, which were then fed into the classification and regression modules. In both methods, the structure of the classification and regression modules was the same as the proposed method. All the experiments were conducted under a single GTX 2080 GPU and an Intel(R) Xeon(R) CPU E5-2695 v4 @ 2.10GHz. The details of the implemented neural network layers and parameters were shown in Table I.

F. Statistical Analysis

Two one-way repeated-measure ANOVAs were performed on the CA and RMSD respectively, to examine the effect of gesture recognition and muscle force estimation using different methods. The LSD post hoc test was employed for multiple pairwise comparisons. The significance level was set as $p < 0.05$. All statistical analysis were implemented by SPSS software (version 24.0, SPSS Inc. Chicago, IL, USA).

III. Results

Table II shows the performance of gesture recognition and muscle force estimation respectively, when the weight coefficient α was set from 0 to 10. Please note that $\alpha = 0$ means that only the gesture recognition task was performed without force estimation, thus there was no result of force estimation. In this case, the CA of the single gesture recognition task was 95.00±5.15%. When the α was increased from 1 to 10, the CA was all improved obviously with statistical significance ($p < 0.05$), and the maximal value was 98.70±1.21% when α was set to be 6. In this case, the muscle force estimation performance was also competitive. Thus, $\alpha = 6$ was selected and it was consistently applied in subsequent analyses.

Fig. 4 reports the CA when varying the model depth from 1 to 4 layers, both the MTL-Transformer method and the MTL-LSTM method achieved their best performance with just 1 layer, and the MTL-TCNN method had the optimal performance with 3 layers. Notably, the MTL-Transformer

TABLE I
THE DETAILS OF THE DESIGNED NEURAL NETWORK. *B* REPRESENTS THE BATCH SIZE

Layer index	Layer type	Output shape	Parameter number
1	Linear	[*B*, 256, 256]	65792
2	ReLU	[*B*, 256, 256]	0
3	Linear	[*B*, 256, 256]	196608
4	Attention	[*B*, 256, 256]	0
5	LayerNorm	[*B*, 256, 256]	512
6	Linear	[*B*, 256, 256]	65792
7	ReLU	[*B*, 256, 256]	0
8	Linear	[*B*, 256, 256]	65792
9	LayerNorm	[*B*, 256, 256]	512
10	Pooling	[*B*, 256]	512
11	Linear	[*B*, 11]	2827
12	Linear	[*B*, 256, 1]	256

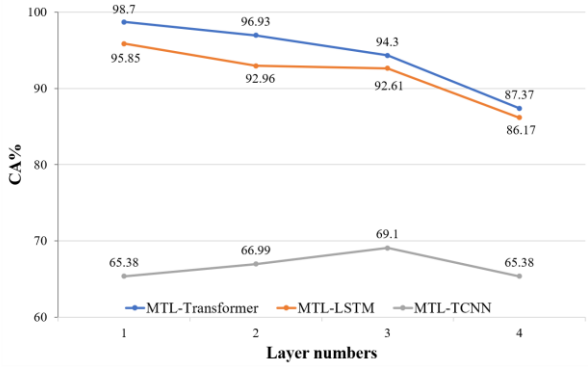

Fig. 4. The CA achieved by three methods as the number of layers varies from 1 to 4.

method outperformed any other method under any model depth setting.

Fig. 5 and Fig. 6 exhibit the CA and RMSD values of all subjects using the proposed method and two comparison methods, respectively, and the mean value averaged over all subjects was shown in the right side of each figure. It can be seen that the CA of the MTL-TCNN, MTL-LSTM, and the proposed method were 69.10±20.816%, 95.85±4.63% and 98.70±1.21%, respectively. The proposed method had the highest average CA, which outperformed two contrast methods significantly ($p < 0.05$). At the same time, the average RMSD of the proposed method was 12.59±2.76%, achieving a reduction in estimation error compared to 13.79±3.20% by the MTL-LSTM method and 21.69±1.92% by MTL-TCNN method, with statistical significance ($p < 0.05$).

Besides, the computational time cost in the testing stage was also calculated as the average time cost over all windows from the testing dataset for three methods. The mean time cost of the proposed method was 0.20 ms, and it was a bit longer than 0.13 ms achieved by the MTL-TCNN method, but much shorter than 0.36 ms resulted from the MTL-LSTM method. The resultant time cost of any method was much less than the window increment (32 ms). In addition, the sum of individual window length (64 ms) and the time cost per window, i.e., the total time delay for a testing window was less than 300 ms (the tolerance for real-time myoelectric control).

Fig. 7 displays representative examples of the estimated force curve with respect to actually measured force curve, using three methods (the MTL-Transformer, MTL-LSTM and MTL-TCNN methods), respectively. It can be found that the predicted muscle force curve of MTL-TCNN was very noisy and therefore failed to fit the measured force well. The MTL-LSTM method had a predicted muscle force curve much smoother than the MTL-TCNN method. However, it still had many fluctuations not in accordance with the true force curve. By contrast, the estimated force curve derived from the proposed method fits better with the measured force curve, by capturing the fluctuations of the actual force precisely. In this case, the proposed method achieved the lowest average RMSD value (9.27%), outperforming the MTL-LSTM method (9.72%) and the MTL-TCNN method (26.48%).

To evaluate the real-time classification performance of the proposed MTL-Transformer method, we visualized the real-time classification results [6], [41] derived from a representative subject S4, as shown in Figure 8. In this test, S4 was asked to cycle through every gesture. It can be found that most of the samples were classified correctly, and misclassifications usually occur during gesture transitions.

IV. DISCUSSION

This paper presents a novel method for simultaneous implementation of gesture recognition and muscle force estimation using the MTL-Transformer method. The transformer model was embedded in the MTL framework to mine context information of sEMG sequences. The innovations and major contributions are as follows: (1) A novel Transformer-based multi-task learning method is proposed for simultaneous gesture recognition and muscle force estimation. (2) The

TABLE II
THE CA OF THE GESTURE RECOGNITION AND THE RMSD OF THE FORCE ESTIMATION RESPECTIVELY, WHEN THE WEIGHT COEFFICIENT α RANGED FROM 0 TO 10

Subject	Indicators (%)	a=0	a=1	a=2	a=3	a=4	a=5	a=6	a=7	a=8	a=9	a=10
S1	CA	85.51	99.78	99.76	99.8	**99.88**	99.73	99.8	99.8	99.8	99.66	99.73
	RMSD	-	11.87	11.55	11.37	11.41	10.88	10.95	10.76	10.72	10.93	**10.66**
S2	CA	96.3	99.43	98.95	99.28	**99.92**	99.68	99.72	98.44	99.28	99.46	99.9
	RMSD	-	11.49	**11.23**	11.5	11.38	11.56	11.79	11.7	11.75	12.06	11.92
S3	CA	92.44	95.71	**97.29**	94.29	92.72	96.13	96.74	96.31	95.69	94.43	96.04
	RMSD	-	13.91	13.52	12.39	11.96	11.67	11.12	11.96	**10.52**	11.23	11.12
S4	CA	98.97	99.16	99.09	99.12	99.12	**99.16**	99.11	99.12	99.12	99.11	99.24
	RMSD	-	11.46	11.25	10.97	10.81	10.71	10.66	10.63	10.61	10.59	**10.12**
S5	CA	98.07	98.52	98.76	**98.79**	98.7	98.61	98.67	98.43	98.46	98.49	98.79
	RMSD	-	14.10	14.29	14.12	13.84	13.54	13.30	13.17	13.00	12.87	**12.66**
S6	CA	98.85	99.77	99.82	99.77	99.82	**99.82**	99.85	99.77	99.75	99.7	99.67
	RMSD	-	14.25	13.9	13.48	12.91	12.29	11.75	11.56	11.5	**11.49**	12.53
S7	CA	97.98	**100**	**100**	**100**	**100**	**100**	**100**	99.98	99.98	99.98	**100**
	RMSD	-	20.02	18.67	18.26	17.82	17.16	17.02	16.88	16.79	16.68	**15.76**
S8	CA	84.96	97	96.8	96.77	**97.62**	97.11	97.07	96.7	96.39	95.92	95.23
	RMSD	-	10.81	10.36	10.09	9.86	9.7	9.54	9.43	9.38	**9.25**	9.53
S9	CA	97.17	98.35	97.88	98.08	97.96	98.06	97.57	98.59	**98.68**	98.68	98.41
	RMSD	-	14.13	13.66	12.93	12.7	12.35	12.83	12.9	12.99	13.04	**12.11**
S10	CA	98.28	99.1	99.08	99	99.13	98.95	**99.41**	99.03	98.98	98.95	99.03
	RMSD	-	18.52	18.55	18.68	18.61	18.41	18.48	18.29	18.11	18.15	**17.96**
S11	CA	96.44	96.89	**98.04**	98	97.89	97.37	97.8	97.55	96.26	96.94	96.84
	RMSD	-	11.68	11.41	11.29	11.16	11.14	**11.06**	11.36	11.66	11.27	11.47
Avg	CA	95.00 ±5.15	98.52 ±1.41	98.68 ±1.06	98.45 ±1.67	98.43 ±2.09	98.60 ±1.28	**98.70 ±1.21**	98.52 ±1.23	98.40 ±1.55	98.30 ±1.78	98.44 ±1.66
	RMSD	-	13.84 ±2.99	13.49 ±2.85	13.19 ±2.86	12.95 ±2.82	12.67 ±2.73	**12.59 ±2.76**	12.60 ±2.69	12.46 ±2.70	12.51 ±2.66	12.35 ±2.48

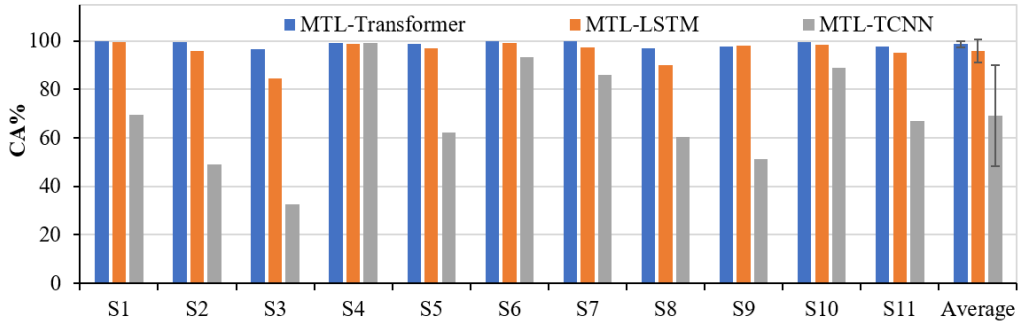

Fig. 5. The CA of gesture recognition for all subjects using the proposed method and two contrast methods, respectively.

invariance of the sEMG characteristics inherent to patterns under variable forces is explored by temporal modeling using the transformer model, and the smoothness of both gesture recognition and muscle force estimation is ensured simultaneously. (3) Better recognition performance is achieved by sharing feature representations between both the muscle force estimation task and the gesture recognition task through the MTL design, as compared with the method implementing just one individual task.

In the proposed method, MTL framework was used to improve the generalization ability of the model by learning a shared feature space of gesture recognition and force

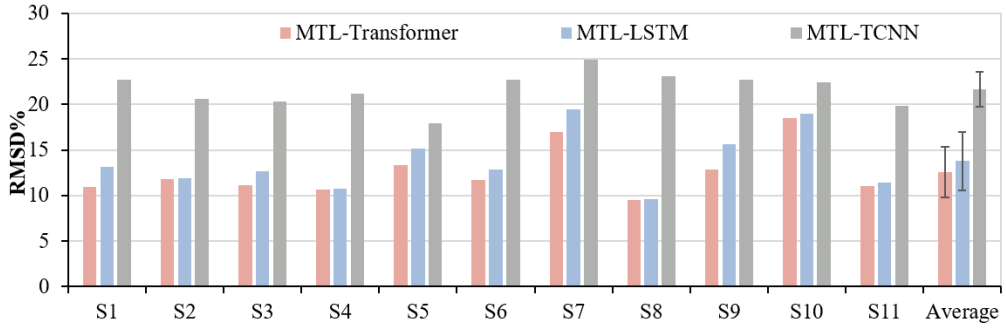

Fig. 6. The RMSD of muscle force estimation for all subjects using the proposed method and two contrast methods, respectively.

estimation, compared to the single-task independent learning way. Besides, the end-to-end implementation improves the convenience of the myoelectric control systems. In the MTL framework, it is essential to balance the contribution of each task. Since gesture recognition is often a more important task than muscle force estimation, muscle force estimation can be considered as an auxiliary task, and the model can work better by dynamically adjusting the importance of the force estimation. In this paper, this was achieved by adding a weighting factor α to the regression loss corresponding to the muscle force estimation. When α is too small, the model training may be dominated by the classification loss, neglecting the contribution of the regression task. Conversely, when α is too large, the model training may emphasize too much on muscle force estimation loss and lead to degradation on gesture classification. As shown in Table I, without the involvement of muscle force estimation (i.e., $\alpha = 0$), the lowest performance of gesture recognition was obtained. This demonstrated the effectiveness of MTL, i.e., the addition of the muscle force estimation task has a positive effect on the performance of gesture recognition. In contrast, when the auxiliary task was added (with a non-zero balance factor), the model performance was found to be improved and the best average classification accuracy was achieved when α equals to 6. Since gesture recognition is more important than muscle force estimation, we chose the value of α when the highest accuracy of gesture recognition and a competitive RMSD of muscle force estimation was obtained, thus α was determined as 6. Notably, there was no significant difference in recognition results when α ranged from 1 to 10 ($p > 0.05$), suggesting that the model was not sensitive to the value of the balance factor for this task. This is consistent with a previous finding [22], and can provide guidance for similar tasks based on the MTL framework in myoelectric control.

As clarified in the Introduction section, the changes in sEMG feature space under variable forces inevitably may degrade gesture recognition accuracy. In this paper, we adopted transformer model to implement simultaneously gesture recognition and muscle force estimation. As shown in Fig. 5 and Fig. 6, the proposed MTL-Transformer method had the CA of 98.70±1.21% and the RMSD of 12.59±2.76%, demonstrating the best performance by both the gesture recognition and the muscle force estimation. This verified the previous scientific hypothesis that the negative effect of force variation on the sEMG features can be mitigated by the temporal modeling

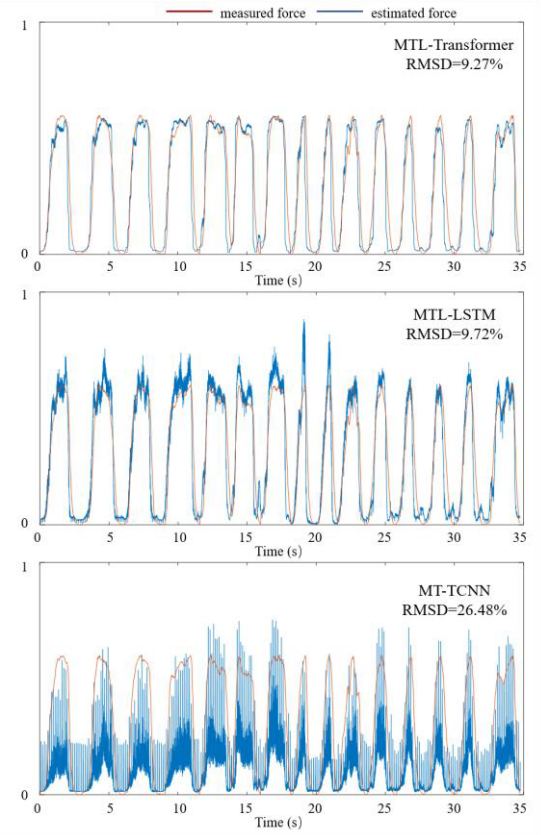

Fig. 7. Representative examples of the estimated force curve (blue) with respect to actually measured force curve (red), selected from a gesture performed by the subject S4, using the proposed MTL-Transformer method (a), the MTL-LSTM method (b) and the MTL-TCNN method (c), respectively.

ability of the transformer model through aggregating features of each sample with different force values. Meanwhile, the proposed MTL-Transformer method ensured the smoothness of the muscle force estimation curve, as visualized in Fig. 7.

As a commonly used powerful model for temporal modeling, the LSTM model was designed as a MTL structure and used for comparison in this paper. Not surprisingly, the proposed MTL-Transformer outperformed the MTL-LSTM method in terms of both gesture recognition and muscle force estimation. This is due to the fact that the LSTM relies on historical memory, thus the initial performance of the model is limited. When compared with recurrent neural networks

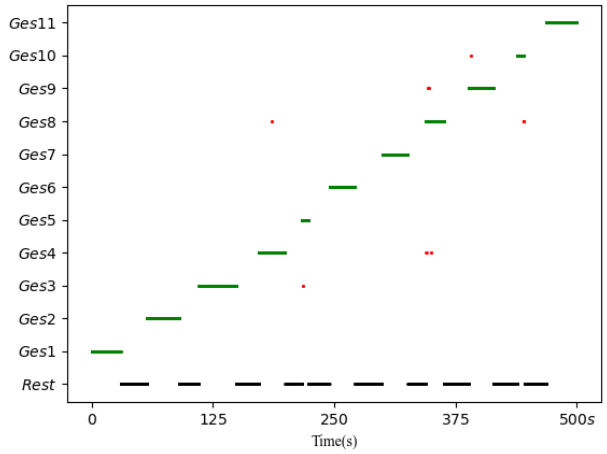

Fig. 8. The classification visualization results during a real-time test performed on S4, in which S4 was asked to cycle through every gesture. Green blocks indicate correct gesture decisions, black blocks indicate correct predictions of the resting state, and just some sporadic red blocks indicate incidental errors of the gesture classification.

(including LSTM) that process temporal data in a sequential manner, the self-attention operation can be conducted in a parallel way for all data samples, which makes it very time efficient [32], as verified by less time delay derived from the proposed method.

Besides, temporal CNN (TCNN) is also a typical method used to characterize temporal relationships, which has been designed and validated as a multi-stream structure by previous work [22]. However, the performance of the method in terms of both gesture recognition and muscle force estimation was unsatisfactory. There are two possible reasons. The first is the simple structure of TCNN with a small number of parameters. Although this property can reduce the inference time, the simple model cannot be adapted to the complex gesture recognition and muscle force estimation tasks. Previous work only carried out gesture recognition and force level estimation for sEMG signals at three fixed force levels, whereas the force varied continuously in a great range from resting (almost 0% MVC) to 60% MVC when performing different gestures in this study. Greater force changes make the distribution of sEMG features more variable, and more powerful models are needed for achieving satisfactory results. Secondly, the receptive field of TCNN is limited, and it can only focus on the information within a short period of time in the long time series, so the stationarity of gesture recognition and muscle force estimation cannot be guaranteed. Compared with the TCNN model, the transformer model is able to well characterize global context information through self-attention mechanism and can improve the model performance. All of these advantageous aspects of the proposed MTL-Transformer method can be used to explain its significant performance gains (over 30%) at the cost of slightly longer time delay, with respect to the MTL-TCNN method. Meanwhile, the prolonged testing time delay is too small to affect the common usability of the myoelectric control system. Therefore, the proposed method is regarded to achieve superior gesture recognition and muscle force estimation performance along with sufficient computational efficiency.

Although the results are promising, there are still some limitations in this study. First, the target gestures in the experimental scheme in this paper only include several comprehensive gestures involving press, pinch, grasp and twist. More complex and dexterous gestures from daily life can be considered to expand the gesture set. Second, the proposed method in this paper is based on the user-specific condition where each new user needs to provide certain training data, which may be inconvenient in practical use. Thus, this method can be explored in the future in conjunction with strategies such as unsupervised domain adaptation for cross-user simultaneous gesture recognition and muscle force estimation.

V. Conclusion

In this paper, a novel MTL-Transformer method is presented using the transformer model embedded in the MTL framework for predicting both gesture pattern and muscle force, which can mitigate the negative effect of force variation on the sEMG features through temporal modeling. The proposed MTL-Transformer method outperformed common temporal modelling methods-based MTL framework in terms of both gesture recognition and muscle force estimation ($p < 0.05$). The experimental results demonstrated the effectiveness of the transformer model in mining the context information of sEMG sequences. This study offers a promising method for robust and smooth myoelectric control systems, with wide applications in gestural interfaces, prosthetic and orthotic control.

References

[1] M. Asghari Oskoei and H. Hu, "Myoelectric control systems—A survey," *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275–294, Oct. 2007.

[2] C. Cipriani, F. Zaccone, S. Micera, and M. C. Carrozza, "On the shared control of an EMG-controlled prosthetic hand: Analysis of user–prosthesis interaction," *IEEE Trans. Robot.*, vol. 24, no. 1, pp. 170–184, Feb. 2008.

[3] A. A. Adewuyi, L. J. Hargrove, and T. A. Kuiken, "An analysis of intrinsic and extrinsic hand muscle EMG for improved pattern recognition control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 4, pp. 485–494, Apr. 2016.

[4] X. Li et al., "Decoding muscle force from individual motor unit activities using a twitch force model and hybrid neural networks," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103297.

[5] S. Lee, M.-O. Kim, T. Kang, J. Park, and Y. Choi, "Knit band sensor for myoelectric control of surface EMG-based prosthetic hand," *IEEE Sensors J.*, vol. 18, no. 20, pp. 8578–8586, Oct. 2018.

[6] V. E. Kosmidou and L. J. Hadjileontiadis, "Sign language recognition using intrinsic-mode sample entropy on sEMG and accelerometer data," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 12, pp. 2879–2890, Dec. 2009.

[7] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Trans. Syst. Man, Cybern. A, Syst. Humans*, vol. 41, no. 6, pp. 1064–1076, Nov. 2011.

[8] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng, "Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation," *Sensors*, vol. 17, no. 3, p. 458, Feb. 2017.

[9] X. Chen, Y. Li, R. Hu, X. Zhang, and X. Chen, "Hand gesture recognition based on surface electromyography using convolutional neural network with transfer learning method," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1292–1304, Apr. 2021.

[10] 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, Jul. 2003.

[11] Y. Huang, K. B. Englehart, B. Hudgins, and A. D. C. Chan, "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1801–1811, Nov. 2005.

[12] M. A. Oskoei and H. Hu, "Support vector machine-based classification scheme for myoelectric control applied to upper limb," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1956–1965, Aug. 2008.

[13] X. Wu, B. Zhou, Z. Lv, and C. Zhang, "To explore the potentials of independent component analysis in brain-computer interface of motor imagery," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 3, pp. 775–787, Mar. 2020.

[14] G. R. Naik, D. K. Kumar, and Jayadeva, "Twin SVM for gesture classification using the surface electromyogram," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 301–308, Mar. 2010.

[15] L. J. Hargrove, K. Englehart, and B. Hudgins, "A comparison of surface and intramuscular myoelectric signal classification," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 847–853, May 2007.

[16] A. Kashizadeh, K. Peñan, A. Belford, A. Razmjou, and M. Asadnia, "Myoelectric control of a biomimetic robotic hand using deep learning artificial neural network for gesture classification," *IEEE Sensors J.*, vol. 22, no. 19, pp. 18914–18921, Oct. 2022.

[17] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 370–379, Feb. 2020.

[18] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "A fully embedded adaptive real-time hand gesture classifier leveraging HD-sEMG and deep learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 232–243, Apr. 2020.

[19] T. Baldacchino, W. R. Jacobs, S. R. Anderson, K. Worden, and J. Rowson, "Simultaneous force regression and movement classification of fingers via surface EMG within a unified Bayesian framework," *Frontiers Bioeng. Biotechnol.*, vol. 6, p. 13, Feb. 2018.

[20] Y. Fang, D. Zhou, K. Li, Z. Ju, and H. Liu, "Attribute-driven granular model for EMG-based pinch and fingertip force grand recognition," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 789–800, Feb. 2021.

[21] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*.

[22] S. Hua, C. Wang, Z. Xie, and X. Wu, "A force levels and gestures integrated multi-task strategy for neural decoding," *Complex Intell. Syst.*, vol. 6, no. 3, pp. 469–478, Oct. 2020.

[23] R. Hu, X. Chen, H. Zhang, X. Zhang, and X. Chen, "A novel myoelectric control scheme supporting synchronous gesture recognition and muscle force estimation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1127–1137, 2022.

[24] S.-H. Park and S.-P. Lee, "EMG pattern recognition based on artificial intelligence techniques," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 4, pp. 400–405, Dec. 1998.

[25] G. N. Saridis and T. P. Gootee, "EMG pattern analysis and classification for a prosthetic arm," *IEEE Trans. Biomed. Eng.*, vol. BME-29, no. 6, pp. 403–412, Jun. 1982.

[26] J. He, D. Zhang, X. Sheng, S. Li, and X. Zhu, "Invariant surface EMG feature against varying contraction level for myoelectric control based on muscle coordination," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 3, pp. 874–882, May 2015.

[27] A. H. Al-Timemy, R. N. Khushaba, G. Bugmann, and J. Escudero, "Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 6, pp. 650–661, Jun. 2016.

[28] S. Pancholi and A. M. Joshi, "Advanced energy kernel-based feature extraction scheme for improved EMG-PR-based prosthesis control against force variation," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3819–3828, May 2022.

[29] A. T. Liu, S.-W. Li, and H.-y. Lee, "TERA: Self-supervised learning of transformer encoder representation for speech," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 2351–2366, 2021.

[30] Y. Kawara, C. Chu, and Y. Arase, "Preordering encoding on transformer for translation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 644–655, 2021.

[31] Y. Shi et al., "Emformer: Efficient memory transformer based acoustic model for low latency streaming speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 6783–6787.

[32] C. Yang et al., "Lite vision transformer with enhanced self-attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11988–11998.

[33] A. Shrestha, H. Li, J. L. Kernec, and F. Fioranelli, "Continuous human activity classification from FMCW radar with bi-LSTM networks," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13607–13619, Nov. 2020.

[34] X. Yuan, L. Li, Y. A. W. Shardt, Y. Wang, and C. Yang, "Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development," *IEEE Trans. Ind. Electron.*, vol. 68, no. 5, pp. 4404–4414, May 2021.

[35] A. Zollanvari, K. Kunanbayev, S. Akhavan Bitaghsir, and M. Bagheri, "Transformer fault prognosis using deep recurrent neural network over vibration signals," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[36] X. Dong et al., "CSWin transformer: A general vision transformer backbone with cross-shaped windows," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 12114–12124.

[37] S. Shen, X. Wang, F. Mao, L. Sun, and M. Gu, "Movements classification through sEMG with convolutional vision transformer and stacking ensemble learning," *IEEE Sensors J.*, vol. 22, no. 13, pp. 13318–13325, Jul. 2022.

[38] Z. Chen, H. Wang, H. Chen, and T. Wei, "Continuous motion finger joint angle estimation utilizing hybrid sEMG-FMG modality driven transformer-based deep learning model," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 105030.

[39] R. Song et al., "Decoding silent speech from high-density surface electromyographic data using transformer," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104298.

[40] W. E. L. Ilboudo, T. Kobayashi, and K. Sugimoto, "Robust stochastic gradient descent with student-t distribution based first-order momentum," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1324–1337, Mar. 2022.

[41] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "Intuitive real-time control strategy for high-density myoelectric hand prosthesis using deep and transfer learning," *Sci. Rep.*, vol. 11, no. 1, May 2021, Art. no. 11275.