

# A Layered sEMG–FMG Hybrid Sensor for Hand Motion Recognition From Forearm Muscle Activities

Peiji Chen , Ziyi Li, Shunta Togo , *Member, IEEE*, Hiroshi Yokoi, *Member, IEEE*,  
and Yinlai Jiang , *Member, IEEE*

**Abstract**—The activities of muscles in the forearm have been widely investigated to develop human interfaces involving hand motions, especially in the fields of prosthetic hands and teleoperation. Although surface electromyography (sEMG) is considered as an effective biological signal from which hand motions can be recognized, the availability and quality of sEMG data can limit the usability and intuitiveness of human interfaces. This article introduces force myography (FMG) as a supplementary signal and proposes a layered sEMG–FMG hybrid sensor that can measure both sEMG and FMG at the same skin surface location. Meanwhile, a layer fusion convolution neural network (LFC) is designed to extract multiscale features from sEMG and FMG. To evaluate the effectiveness of the hybrid sEMG–FMG sensor and LFC, a 22-hand motion classification experiment was conducted on nine able-bodied subjects. The recognition results indicated a significantly improved classification accuracy ( $p < 0.001$ ) of the hybrid sEMG–FMG modality with respect to single sEMG or FMG modality. The classification accuracies (CAs) of LFC were compared with conventional machine learning methods, including support vector machine, random forest classifier, xgboost, and k-nearest neighbor. Compared with the single-modality sEMG, the CAs of the dual-modality sEMG–FMG using conventional methods, and LFC were improved by 21.31% and 16.71%, respectively. These results suggest that the layered sEMG–FMG sensing approach can effectively enhance the performance of human interfaces, which offers great potential in the clinical applications of sophisticated prosthetic hands and teleoperation.

**Index Terms**—Force myography (FMG), hybrid sensor, layer fusion convolutional neural network (LFC), surface electromyography (sEMG).

Manuscript received 28 April 2023; accepted 13 June 2023. Date of publication 10 July 2023; date of current version 18 September 2023. This work was supported in part by JSPS KAKENHI under Grant JP23H00166, Grant JP23H03785, and Grant JP22K04025, in part by a project commissioned by JSPS and NSFC under Japan-China Scientific Cooperation Program, and in part by JKA through its promotion funds from KEIRIN RACE. This article was recommended by Associate Editor H. Yu. (*Corresponding author: Yinlai Jiang.*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by University of Electro-Communications under Application No. 10006(5), and performed in line with the Helsinki.

Peiji Chen, Ziyi Li, Shunta Togo, and Hiroshi Yokoi are with the Department of Mechanical and Intelligent Systems Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan (e-mail: p.chen@hi.mce.uec.ac.jp; lzy@hi.mce.uec.ac.jp; s.togo@hi.mce.uec.ac.jp; yokoi@hi.mce.uec.ac.jp).

Yinlai Jiang is with the Center for Neuroscience and Biomedical Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan (e-mail: jiang.yinlai@uec.ac.jp).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/THMS.2023.3287594>.

Digital Object Identifier 10.1109/THMS.2023.3287594

## I. INTRODUCTION

THE human hand plays an indispensable role in daily life as a tool to interact with the environment and communicate with each other. The anatomical fact that the muscles related to the movement of the fingers are mostly located in the forearm enables the possibility of estimating hand motions by measuring muscle activities on the forearm. This has been widely used to develop human interfaces, especially for prosthetic hands and teleoperation.

To date, muscle activity has mainly been investigated by measuring surface electromyography (sEMG) with electrodes attached to the skin surface. Many studies have confirmed the possibility of controlling myoelectric prosthetic hands [1] and teleoperation [2] by identifying hand motions from sEMG signals. The use of sEMG signals has long been studied using many effective signal processing methods proposed in the literature. However, in daily life, the human–machine interface based on sEMG is unstable and unreliable because of the characteristics of sEMG signals, which are extremely susceptible to interference, such as sweat and muscle fatigue. Hence, the development of methods for measuring effective control signals has become a focal issue that requires urgent attention.

Previous studies have proposed solutions to improve the acquisition quality of sEMG signals, including targeted muscle reinnervation (TMR) surgery, high-density sEMG measurement, and the introduction of supplementary signals. Kuiken et al. [3] showed that amputees treated with TMR surgery achieved encouraging results in ten different forearm motion tasks. However, this method involves some clinical inconveniences, including high invasiveness and high cost. The preliminary experimental results reported by Zhang and Zhou [4] also demonstrated the feasibility of applying pattern-recognition technology to high-density sEMG signals to recognize the hand motions of stroke survivors. However, the complicated measurement system induces both physical and mental stress in subjects, which can lead to classification degradation.

In contrast, the introduction of supplementary signals has been gradually attracting attention [5] in the field of human–machine interfacing. Force myography (FMG), a structural signal of muscle activity, is generated by muscles during contraction and relaxation. It is relatively stable and immune to sweat conditions. Some studies on prosthesis control indicated that FMG was more stable than sEMG when a pattern was held [6]. Therefore, the introduction of FMG could compensate for the limitations of sEMG. Previous studies [7], [8], [9], [10], [11] have demonstrated that the sEMG–FMG dual-modal approach can effectively improve the performance of hand gesture recognition tasks.

The aim of this study is to develop a hybrid sensor that can measure sEMG and FMG signals simultaneously, without causing any extra burden to the user. Previous studies [7], [8], [9] measured sEMG and FMG separately at different locations, which required extra space to accommodate the sensors. It is desirable to develop a hybrid sensor system that can record the sEMG and FMG signals at the same location. The measurement of sEMG and FMG at the same location has been investigated in previous studies [10], [11]. Both studies acquired FMG by connecting a deformable material directly to force-sensitive sensors, so that the sensors can detect the deformation due to force related to muscle contraction. When subjected to position drift due to movement, the force sensor based FMG unit developed in previous studies [10], [11], [12] can easily lead to errors owing to position-drift-induced horizontal forces. In contrast to these studies, we design an optical FMG measurement unit using reflected infrared light to detect deformation due to pressure changes, which has not yet been reported in the literature. In the FMG unit structure, the phototransistor embedded in the reflectance sensor is placed vertically on top of hollow silicone. Therefore, the output of the reflectance sensor in the proposed optical FMG unit is only affected by the vertical deformation of the silicone, which makes it less susceptible to horizontal deformation owing to position drift.

Furthermore, we designed a deep layer fusion convolutional neural network (LFC) to perform an end-to-end hand motion recognition task. The layer fusion structure was first proposed by Yang and Ramanan [13] in computer vision field [14] transferred the raw sEMG into signal image and proposed a 2D-CNN-based multiview network to perform a late fusion and early fusion for sEMG data. Considering the parameters and the extra preprocessing process for using 2D-CNN, we develop a 1D-CNN-based multi-input layer fusion network, which reduces preprocessing and can perform an independent convolution for every signal channel. This network structure is developed by considering the characteristics of the multichannel one-dimensional physiological time-series signals, and also the feasibility for clinical applications. To the best of our knowledge, this is the first study to introduce a layer-fusion 1D-CNN-based network to perform end-to-end classification of sEMG or FMG signals.

The rest of this article is organized as follows. Section II describes the development of the layered sensor system, the verification experiments we conducted are described in Section III, and the detailed structure of the LFC is introduced in Section IV. Sections V and VI report and discuss the experimental results, respectively. Finally, Section VII concludes this article.

## II. sEMG–FMG HYBRID SENSOR

Muscle contraction is associated with bioelectrical activity and changes in muscle geometry. Bioelectrical activity can be captured using sEMG [15]. The change in muscle geometry can be captured via FMG [16] as the pressure distribution changes. In contrast to the sEMG signal, FMG is not susceptible to electromagnetic interference, muscle fatigue, or unreliable electrode–skin impedance caused by sweat. Therefore, considering the advantages of both sEMG and FMG, cost, comfort, and the available measurement area of the user, we combine the two measurement units into a layered sensor system and design a sensor band to measure three-channel sEMG and FMG signals simultaneously at the same location on the skin surface.

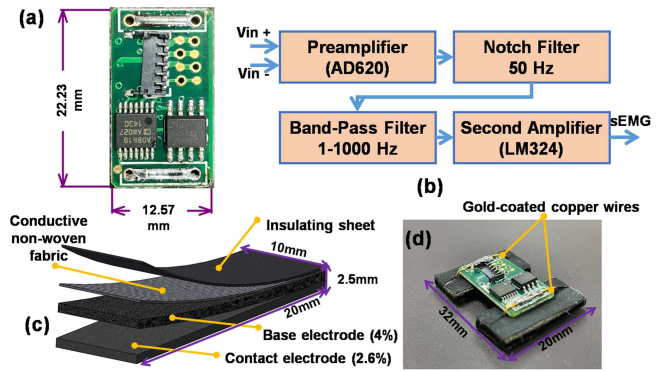


Fig. 1. sEMG measurement unit. (a) sEMG sensor board. (b) Amplifiers (CMR is 120 dB, gain is 5000 in total) and filter circuit. (c) Configuration of double-layer electrode. (d) sEMG sensor with conductive silicone electrodes.

### A. sEMG Measurement Unit

The sEMG measurement unit consists of a circuit board [Fig. 1(a)] and two conductive electrodes. The circuit board includes amplifier and filter function modules, as shown in Fig. 1(b). Differential input signals are preamplified by an instrument amplifier AD620 (Analog Devices, USA), with a gain of 40 in the first stage amplification. The power-line interference is removed with a notch filter (50 Hz), and a bandpass filter (bandwidth 1–1000 Hz) is used to limit the frequency range. Finally, the processed sEMG signals are converted into digital signals by the second amplifier, LM324 (Texas Instruments, USA), with a gain of 125 in the second-stage amplification. The output voltage ranges from 0 to 5 V.

The sEMG sensor is used to measure biopotentials by placing electrodes on the skin. The conductive silicone electrode consisted of silicone (TSG-E30, Tanac Co., Ltd., Japan) and carbon black (EC600JD, Lion Specialty Chemicals Co., Ltd., Japan). This dry electrode was hypoallergenic and sufficiently flexible for long-term daily use. The electrode size was adjustable and set to 10 mm × 20 mm × 2.5 mm [Fig. 1(c)] for the experiment. The composite electrode was divided into two layers: the base electrode and contact electrode with carbon-to-silicone weight ratios of 4% and 2.6%, respectively. A previous study has shown that double-layer electrodes under such concentration combinations exhibit lower skin–electrode contact impedances and higher signal-to-noise ratios [17]. A layer of conductive nonwoven fabric was placed on the base electrode to prevent damage. The sEMG circuit board was connected to the electrodes via gold-coated copper wires and was isolated using an insulating sheet. The total size of the sEMG measurement unit was 32 mm × 20 mm × 6.5 mm [Fig. 1(d)].

### B. FMG Measurement Unit

An optical FMG measurement unit that uses reflected infrared light to detect pressure changes is designed, as shown in Fig. 2. The unit is composed of a base layer, an isolation layer, an elastic silicone structure between the base and isolation layers, and a reflectance sensor to measure the interlayer distances. The reflectance sensor (QTR-1A, Pololu Co., Ltd., USA) carried a single infrared light-emitting diode and phototransistor pair in a 12.7 mm × 7.62 mm × 2.54 mm module. The phototransistor was connected to a pull-up resistor, which functioned as a voltage divider, that generated an analog voltage output between 0 and

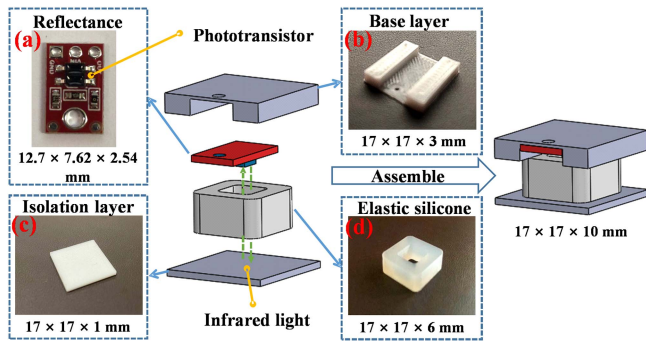


Fig. 2. Implementation of FMG measurement unit. (a) Reflectance sensor. (b) Base layer. (c) Isolation layer. (d) Elastic silicone. The base and isolation layers were made by a 3D printer using PLA material.

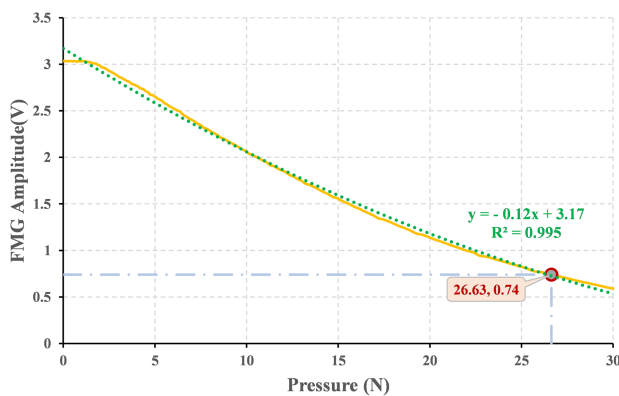


Fig. 3. FMG sensor calibration curve. The yellow line represents the measured data, the green line represents the approximate result and the red dot represents the max pressure 26.63 N at the maximum displacement of 3 mm, whereas FMG is 0.74 V.

5 V as a function of the reflected infrared light. The ideal perception distance was 3–6 mm. The reflectance sensor was fixed to the base layer, and a 6-mm-thick square of hollow silicone was placed between the base and the isolation layers. The infrared light emitted by the reflectance sensor reached the isolation layer via the hollow silicone and was reflected by the isolation layer. Lower output voltages indicate greater reflection and shorter distances. The change in distance detected by the reflectance sensor can be expressed as the pressure deformation of the elastic silicone under the action of an external force.

The optical FMG measurement is based on the relationship between the reflectance sensor's output and the displacement caused by muscle contractions. As the muscle contracts, the elastic silicone deforms vertically due to the pressure change, resulting in a displacement between the reflectance sensor and the isolation layer surface. This displacement leads to a corresponding change in the intensity of the reflected infrared light. The phototransistor converts the light intensity into an electrical signal that reflects the pressure caused by muscle contraction.

To investigate the relationship between the FMG signal and pressure, we used a tension/compression testing machine (IMADA SL-6002) to perform a compression test with a load range from 0 to 30 N. Fig. 3 shows the results.

The trend line that fits the FMG amplitude pressure is a linear approximation (approximation degree 99.5%) which can

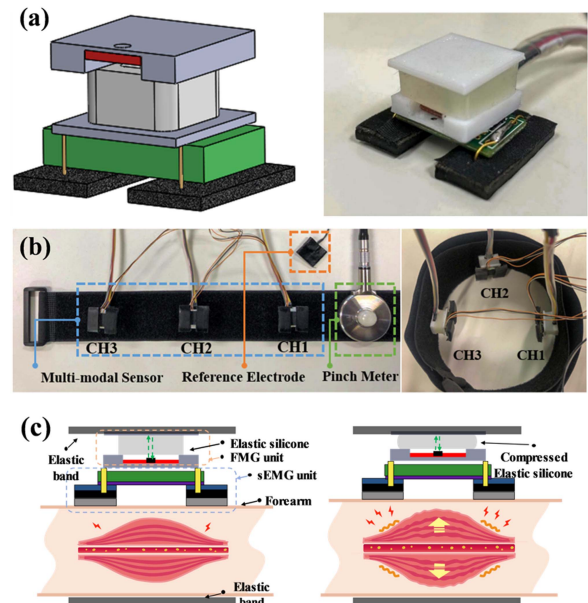


Fig. 4. Dual-modal sensor implementation. (a) Single combined sensor. (b) Dual-modal sensor band with three channels. (c) Diagram of sEMG and FMG measurement.

be expressed as follows:

$$y = -0.12x + 3.17. \quad (1)$$

### C. Layered Hybrid Sensor and Sensor Band

As shown in Fig. 4(a), the sEMG and FMG measurement units were vertically combined. The height of the combined sensor was  $\sim 16.5$  mm and the weight was  $\sim 4.66$  g (without wires). This vertical combination facilitates the simultaneous detection of sEMG and FMG signals on the same skin area. Three hybrid sensors were attached to a strap to realize the three-channel sEMG-FMG sensor band [Fig. 4(b)] used for verification. Fig. 4(c) illustrates the mechanism of simultaneous detection of sEMG and FMG. During muscle contraction, sEMG signals are measured as the amplified potential changes on both electrodes. FMG signals are measured via the deformation of the silicone caused by the pressure change between the elastic band and the skin.

## III. EXPERIMENT

To investigate the effects of the FMG signals, we conducted a 22-class hand gesture recognition experiment using a three-channel sensor band. The experimental results were used to compare the classification performances under three modality settings (sEMG, FMG, and sEMG-FMG).

### A. Subjects

Nine able-bodied, right-handed male subjects (age:  $27.1 \pm 1.9$ , height:  $175.2 \pm 4.2$  cm, weight:  $72.2 \pm 10.5$  kg, numbered S1–S9) participated in the experiment. None of the participants had previously undergone similar sEMG recognition experiments. All experimental procedures were conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of The University of Electro-Communications, Tokyo, Japan. Each participant received an



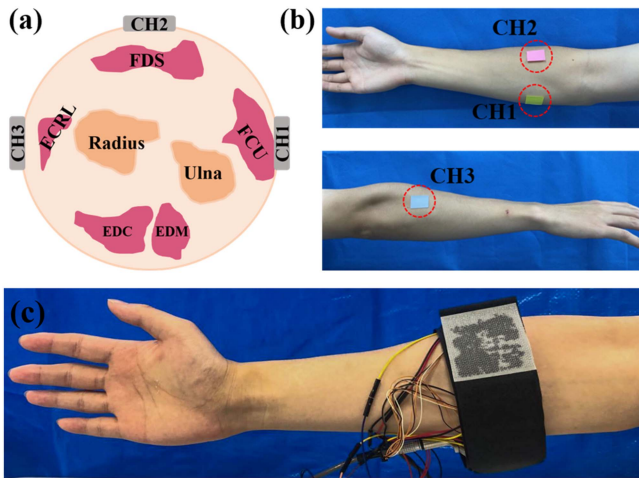


Fig. 5. Sensor placement on forearm. (a) Sectional view of the forearm and sensor positions. (b) CH1 and CH2 are attached near the flexor carpi ulnaris and flexor digitorum superficialis muscles, respectively, CH3 is placed near the extensor carpi radialis longus muscle, a reference electrode is set on the elbow. (c) Photograph of a subject wearing the sensor band.

explanation of the experimental procedure and provided their informed consent.

## B. Experimental Protocol

1) *Sensor Attachment*: The wearing positions of the sensors are shown in Fig. 5. All subjects used their right hands to perform gestures. The forearm skin was cleaned with alcohol before each experiment. Each participant was fitted with a sensor band on the right forearm and tightened to a comfortable but secure level. During the experiment, we used a pinch meter (K800, Biometrics Co., Ltd., U.K.) to monitor the skin–electrode contact characteristics. The initial pressure between the band and the forearm was set to 0.2 kg.

2) *Hand Gestures*: To meet the daily human–computer interaction needs, we focused on gestures frequently used in daily life. Bullock et al. [18] investigated the occurrence frequencies of different human grasps in several daily activities and extracted ten most-frequent grasp types that account for  $\sim 80\%$  of all grasping tasks. Furthermore, inspired by a previous study [19], we included eight hand gestures and four finger gestures. Detailed descriptions of the 22 hand gestures are shown in Fig. 6.

3) *Experimental Paradigm*: The participants were seated with their elbows resting on the table and their right forearm perpendicular to the table. To compensate for the difference in the initial states of each participant, ten seconds of relaxed-state signals were recorded as a baseline. Each measurement period lasted for ten seconds, including two, 3 s of action and two, 2 s of relaxation. Each subject repeated the trial five times. The participants were allowed to rest for several minutes between trials to prevent muscle fatigue. The sEMG sampling rate was set to 1000 Hz using a 16-bit-precision professional analog-to-digital converter (AIO-160802AY-USB, Contec Co., Ltd., Japan) The hybrid sEMG–FMG signal waveform during wrist flexion was demonstrated as in Fig. 7.

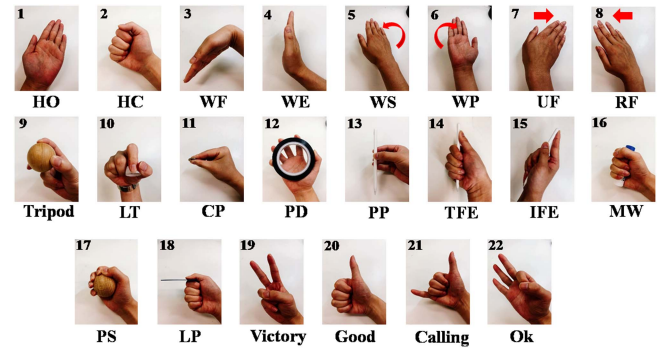


Fig. 6. In total, 22 hand motions used in the experiment. (1) Hand open. (2) Hand close. (3) Wrist flexion. (4) Wrist extension. (5) Wrist supination. (6) Wrist pronation. (7) Ulnar flexion. (8) Radial flexion. (9) Tripod. (10) Lateral tripod. (11) Coin pinch. (12) Precision disk. (13) Pen pinch. (14) Thumb finger extension. (15) Index finger extension. (16) Medium wrap. (17) Power sphere. (18) Lateral pinch.

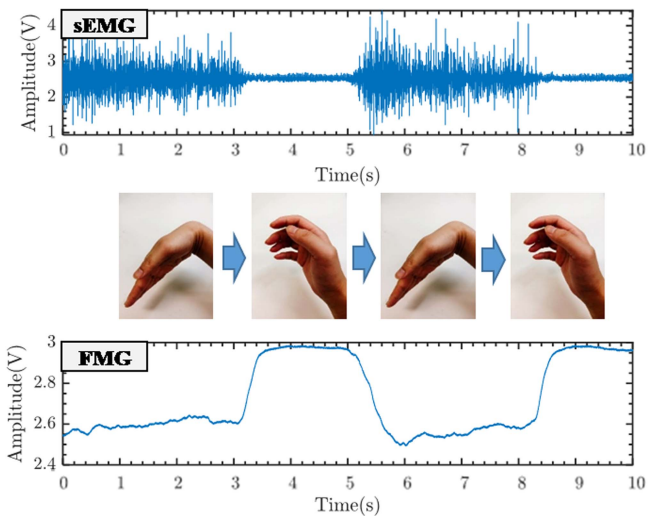


Fig. 7. Time-domain waveform of sEMG and FMG signals during wrist flexion. The signals are from CH1 in Fig. 5.

## IV. LAYER FUSION CONVOLUTION NEURAL NETWORK

### A. Multiscale Representation

The output of each layer in the CNN can be regarded as a scale representation of the input data. As the layers deepen, the model can extract higher-level features. The shallow layers of the CNN are good at extracting low-level features, and the deep layers of the CNN are good at extracting high-level features. The general CNN structures used to process physiological signals only consider high-level features, which can be competent for some simple tasks. However, when the category difference of the input data is small and contains noise signals, high-level features cannot fully represent the full information of the input data [12].

### B. Layer Fusion CNN

Recent studies [20], [21] have demonstrated that deep learning methods outperform conventional machine learning (ML) methods in gesture recognition tasks; however, their low generalization ability makes them rarely used in real life. Compared with picture and video data, the acquisition of physiological signals is difficult, costly, and limited by the collection equipment and

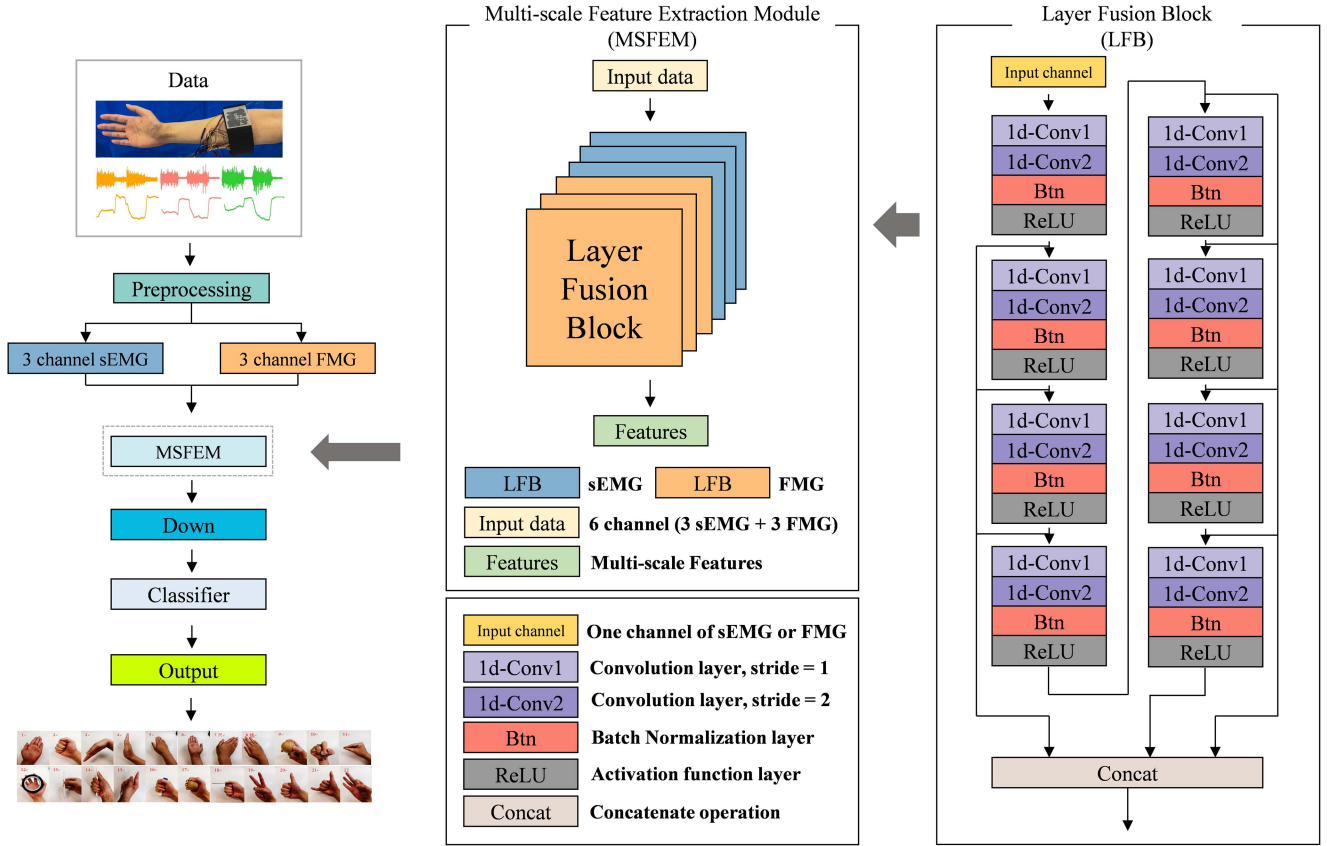


Fig. 8. Proposed LFC for 22-class gesture recognition.

environment. The formats of physiological signals are usually different, making it difficult to collect enough data to train a model with a strong generalization ability. Most deep learning models have been proposed in the fields of computer vision and natural language processing, and are optimized to process image data and text data. When processing time-series signals with noise, some models cannot converge easily [22]. To improve the generalization ability of the CNN model, we took advantage of the multiscale input data and adopted LFC (Fig. 8) to process the sEMG and FMG signals.

The entire gesture-recognition network can be represented by the following decision function:

$$y_{pred} = F_w(x) \quad (2)$$

where  $x$  represents the data input to the deep learning model,  $F_w(x)$  is the decision function,  $w$  is the parameter to be learned for the entire network, and  $y_{pred}$  represents the output gesture category. The network consists of feature extraction and classification layers.

The feature extraction layers contained multiple convolution blocks and a layer fusion structure. The convolutional block primarily consists of two 1d convolution layers with strides 1 and 2, respectively. The convolutional block can be defined as follows:

$$\delta_{n+1} = H_{\lambda_n}(O_n) \quad (3)$$

where  $O_n$  is the output result of the  $n$ th convolution block,  $\delta_{n+1}$  is the input of the subsequent convolution block,  $H_{\lambda_n}(O_n)$  is the calculation performed in the  $n$ th convolution block, and  $\lambda_n$  are the parameters to be learned.

The fusion layer is defined as follows:

$$F = Concat(f_1 + f_2 + \dots + f_n) \quad (4)$$

where  $f_n$  denotes the results of the  $n$ th convolution block,  $F$  denotes the fusion results of all layers, and  $Concat(x)$  denotes the concatenation operation. This structure offers two advantages. First, during forward propagation, information from each layer is directly input to the decision layer, which makes the data propagation more efficient than structures only using the final layer's results. Second, during backpropagation, gradients are directly propagated to each layer through the structure of layer fusion, which enables the parameters of each layer to be updated effectively. Both advantages contribute to the improvement of the generalization ability [13], [23].

### C. Model Complexity

A good trade-off between improved performance and increased model complexity is required to make a model of practical use. Our designed LFC (sEMG/FMG) and LFC (sEMG–FMG) require almost the same FLOPs in a single forward for a  $3 \times 1000$  and  $6 \times 1000$  input signal, respectively. For params, the sEMG–FMG LFC model only increases the number of parameters by 0.3% compared with the single-modality LFC. And about the training time and testing time, we will discuss in Section VI. The detailed structural codes are available in our Github.<sup>1</sup>

<sup>1</sup>[Online]. Available: <https://github.com/peijichen0324/a-layered-sensor-unit>

TABLE I  
CLASSIFICATION RESULTS OF EVERY SUBJECTS

Classifiers	Modalities	Subject									Mean $\pm$ SD
		S1	S2	S3	S4	S5	S6	S7	S8	S9	
RFC	sEMG	71.82%	69.27%	60.27%	45.96%	69.7%	63.72%	60.33%	57.32%	56.81%	61.69% $\pm$ 8.08%
	FMG	59.67%	52.21%	52.5%	64.04%	63.33%	62.08%	50.89%	64.78%	56.99%	58.5% $\pm$ 5.51%
	Hybrid	83.98%	80.72%	79.91%	74.89%	87.37%	83.63%	82.61%	76.99%	84.51%	81.62% $\pm$ 3.91%
XGBoost	sEMG	65.75%	65.06%	64.73%	53.19%	62.12%	63.72%	53.26%	49.79%	55.61%	59.25% $\pm$ 6.22%
	FMG	55.25%	56.1%	54.02%	62.34%	61.75%	62.12%	50.11%	62.62%	53.47%	57.53% $\pm$ 4.73%
	Hybrid	76.8%	77.71%	77.23%	67.23%	79.29%	77.1%	76.63%	74.06%	76.06%	75.79% $\pm$ 3.5%
KNN	sEMG	61.88%	66.87%	62.5%	49.79%	59.6%	63.72%	64.13%	51.88%	52.58%	59.22% $\pm$ 6.2%
	FMG	41.44%	50.84%	45.98%	62.13%	60.6%	63.72%	51.26%	56.48%	49.3%	53.53% $\pm$ 7.66%
	Hybrid	87.29%	77.11%	75.89%	68.94%	90.9%	86.29%	88.59%	76.57%	82.63%	81.58% $\pm$ 7.32%
SVM	sEMG	65.75%	69.28%	69.64%	59.15%	69.7%	66.81%	69.02%	58.58%	60.09%	65.34% $\pm$ 4.75%
	FMG	51.93%	55.42%	56.7%	64.68%	57.07%	58.85%	52.72%	64.85%	55.87%	57.57% $\pm$ 4.6%
	Hybrid	87.29%	84.94%	87.5%	75.32%	90.4%	89.38%	87.5%	83.68%	91.55%	86.4% $\pm$ 4.83%
LFC	sEMG	81.22%	87.35%	79.91%	70.64%	82.32%	80.53%	76.09%	71.13%	76.53%	78.41% $\pm$ 5.39%
	FMG	70.17%	78.31%	60.71%	77.02%	81.81%	78.32%	77.72%	75.73%	73.71%	74.83% $\pm$ 6.22%
	Hybrid	91.71%	96.39%	92.86%	86.81%	97.98%	92.92%	92.39%	91.21%	95.31%	93.06% $\pm$ 3.26%

#### D. Comparison With Conventional ML

Several conventional ML methods (support vector machine (SVM), random forest classifier, K-nearest neighbors (KNN), and XGBoost) were used as baseline for comparison. To perform data preprocessing for these methods, the sEMG and FMG signals were divided into a series of nonoverlapping 500 ms shift-analysis windows, and features were extracted from each window. Five sEMG time-domain (TD) features were extracted, including the root mean square, mean absolute value, waveform length, zero crossings, and slope sign changes, all of which have been confirmed to be effective in previous studies [15], [24]. There are also many effective sEMG features [25] that have been proven to be capable of representing complex self-regulatory systems. However, they are not used in this study, only five common sEMG features were used. In addition, 12 FMG TD features were extracted, including the maximum (Max), minimum (Min), average value (Avg), standard deviation (SD), and degree of deformation in various states (1, 5, 10, 25, 50, 75, 90, and 99%) [26].

Because LFC does not rely on prior knowledge, the raw signals are fed directly to the model after being extracted by a fixed-length window and normalized. The dataset was divided into training and testing sets of 80% and 20%, respectively. A five-fold cross-validation was conducted. We implemented our model based on the PyTorch framework. The proposed model was trained by an Adam optimizer with learning rate of 0.001. The learning rate was halved every 20% of the total epochs.

## V. RESULTS

#### A. Classification Results of Individual Subject

A comparison of the classification accuracies (CAs) of the nine subjects is presented in Table I. The highest CAs among the ML methods are shown in red, and the highest CAs among all methods are in blue. Compared with the sEMG modality, the CAs in every subject of the hybrid sEMG-FMG modality using conventional ML and LFC both showed a significant improvement (t-test, RFC:  $p = 2.8E-05 < 0.01$ , XGBoost:  $p = 1.2E-05 < 0.01$ , KNN:  $p = 3.5E-05 < 0.01$ , SVM:  $p =$

$7.2E-08 < 0.01$ , LFC:  $p = 8.9E-06 < 0.01$ ). Specifically, SVM was the best-performing conventional ML, whose average CAs of the sEMG and hybrid sEMG-FMG modality settings across all subjects were  $65.34\% \pm 4.75\%$  and  $86.4\% \pm 4.83\%$ , respectively. The average CAs of LFC were  $78.41\% \pm 5.39\%$  and  $93.06\% \pm 3.26\%$ . Notably, no significant difference (two-way analysis of variance (ANOVA), SVM:  $p = 0.93 > 0.05$ , LFC:  $p = 0.13 > 0.05$ ) was observed in the CAs of the nine subjects with respect to the classifier, which suggests that the proposed sensor does not cause significant differences between individuals and offers high generalization ability. Meanwhile, the difference between the three modality settings was significant (two-way ANOVA, SVM:  $p = 1.17E-08 < 0.001$ , LFC:  $p = 4.58E-07 < 0.001$ ).

#### B. Classification Results of All Subjects

To verify the generalization ability of the methods, we used different subjects for training and testing (e.g., S1-S7 for training, and S8 and S9 for testing). Considering the domain difference from the intra and interindividual, the following baseline subtraction and normalization were performed before training. 1) Baseline subtraction: For each subject, the baseline of sEMG and FMG signals for each channel was calculated by averaging the values during a period of relaxation without muscle contraction. These baseline values were then subtracted from the corresponding channel signals to focus on the relative in signal changes resulting from muscle contractions. 2) Normalization: To account for the domain difference from intra and interindividual variations, a normalization step was performed before training. The mean and standard deviation of the sEMG and FMG signals for each channel on the training datasets (i.e., S1-S7) were calculated. During the training and testing phases, the sEMG and FMG data for each channel were standardized based on the mean and standard deviation calculated from the training datasets. Results are shown in Fig. 9, where the hybrid sEMG-FMG outperformed the sEMG and FMG. Same as the results for the individual subject, SVM performed the best among the conventional ML methods. The CAs of the hybrid sEMG-FMG, sEMG, and FMG using SVM were  $82.32\% \pm$



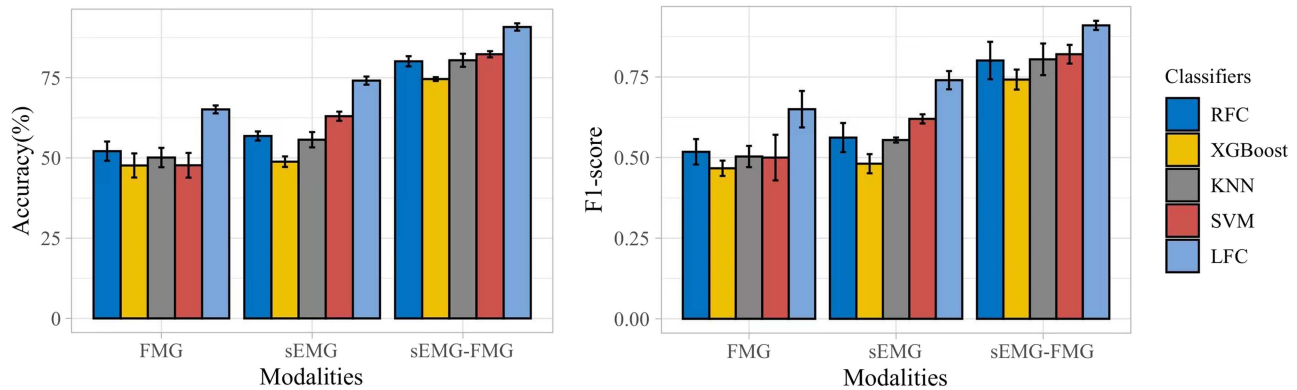


Fig. 9. CAs (left) and F1 scores (right) of conventional ML and our proposed LFC under three modality settings.

0.75%,  $62.01\% \pm 1.14\%$ ,  $47.72\% \pm 4.25\%$ , and the f1-score were  $0.82 \pm 0.05$ ,  $0.61 \pm 0.02$ ,  $0.44 \pm 0.14$ . Meanwhile, the CAs and f1-score of the hybrid sEMG-FMG, sEMG, and FMG using the proposed LFC were  $90.81\% \pm 0.8\%$ ,  $74.1\% \pm 0.95\%$ ,  $65.13\% \pm 0.86\%$  and  $0.91 \pm 0.02$ ,  $0.73 \pm 0.03$ ,  $0.64 \pm 0.08$ . Compared with the sEMG modality, the CAs of the hybrid sEMG-FMG modality improved by 21.31% with SVM and 16.71% with LFC. There was no significant difference ( $p = 0.27 > 0.05$ ) between the random training method and the group training method, implying that the trained LFC has a strong generalization ability and potential to be used for new subject.

The confusion matrix for the LFC is illustrated in Fig. 10. The average recall was  $90.93\% \pm 5.3\%$ , and the average precision of the 22 hand motions was  $90.74\% \pm 4.54\%$ , which suggests that the LFC achieved a strong classification performance for each gesture. Combining precision and recall, the classification performance for wrist flexion was the highest (F1 score = 0.97), whereas that for pen pinches was the lowest (F1 score = 0.82). The ten grasp gestures (from tripod to lateral pinch) suffered more misclassifications than the other gestures, as shown by the blue dotted region in Fig. 10. The correlation between these gestures is relatively high, because they require a strong force to hold the object, which may lead to similar FMG.

## VI. DISCUSSION

The advantages of combining sEMG and FMG signals at the same location on the skin surface are reflected not only in the significant improvement in CAs, but also in the classification stability. The SD of the average LFC-obtained dual-modality CAs for all subjects was smaller than that of either single signal (hybrid:3.26%; sEMG:5.39%; FMG:6.22%). Because the CAs of the LFC exceeded those of conventional ML methods, we further discuss the performance and characteristics of the dual-modality approach based on the LFC.

The loss and classification accuracy curves of the three modality settings using LFC are shown in Fig. 11. The hybrid sEMG-FMG modality achieved the highest performance in terms of loss and CA. In the hybrid sEMG-FMG modality, the loss converged to the lowest value of the three modality settings (training:  $0.22 \pm 0.01$  and test:  $0.4 \pm 0.01$ ), with a generalization gap at 0.18. Furthermore, it achieved a faster convergence speed during the first 100 epochs for both training and testing sets. The performance of FMG was poorest, and its final loss converged to  $0.98 \pm 0.01$  (training set) and  $1.27 \pm$

0.02 (test set), with a generalization gap at 0.29. Compared with sEMG, FMG had a smaller generalization gap between the final convergence stabilities for the training and testing sets, which indicates that, although sEMG achieved a higher CA, FMG might offer superior robustness in the dual-modal approach. Thus, the dual-modal approach can further improve hand gesture classification performance by combining sEMG and FMG.

To elucidate the effect of layer fusion, we trained multiple classifiers using multiscale representations of the input signals in the conventional CNN. The results of the different scales of the CNN are shown in Fig. 12 and compared with the LFC that uses all scales. The heatmap shows that not all gestures obtain the best classification performance in the scale representation of the last layer scale (scale8), which means that the optimal scale representation of the signals of different gestures is different. When all the layer results from the CNN with layer fusion are used, almost all gestures can achieve the best classification performance. The traditional CNN structure can only extract the same-scale representation of the last layer for the input signals. And the addition of the layer-fusion structure allows the extraction of different-scale representation from the input signals.

A comparison of our research with previous studies that measured sEMG and FMG at the same location is shown in Table II. Jiang et al. [10], who used an armband consisting of eight colocated EMG-FMG sensing units, achieved a CA of 91.6% on a ten-hand gesture recognition task. Ke et al. [11] presented a modular EMG-FMG sensor unit, with four of them on an elastic belt to obtain a CA of up to 99.42% on a six-hand gesture recognition task. In our research, we developed a layered sEMG-FMG hybrid sensor and a layer-fusion convolution neural network. An equivalent level of CA (93.06%) was obtained with fewer sensors for a more challenging hand gesture recognition task. The effectiveness of both the hybrid sensor and layer-fusion convolution neural network was indicated by the experiment and comparison.

The computation time of the proposed LFCNN (sEMG-FMG) was investigated on a GPU (GeForce RTX 3090, NVIDIA, USA). Table III displays the recognition accuracy and computation time using different window lengths on different subjects. In the forward propagation stage, the time of a short window sample was significantly reduced compared with a long window sample. However, due to the loss of some low-frequency information, the recognition accuracy was also reduced. The longest time required for the training stage using datasets of different window lengths was about 20 min for one subject,

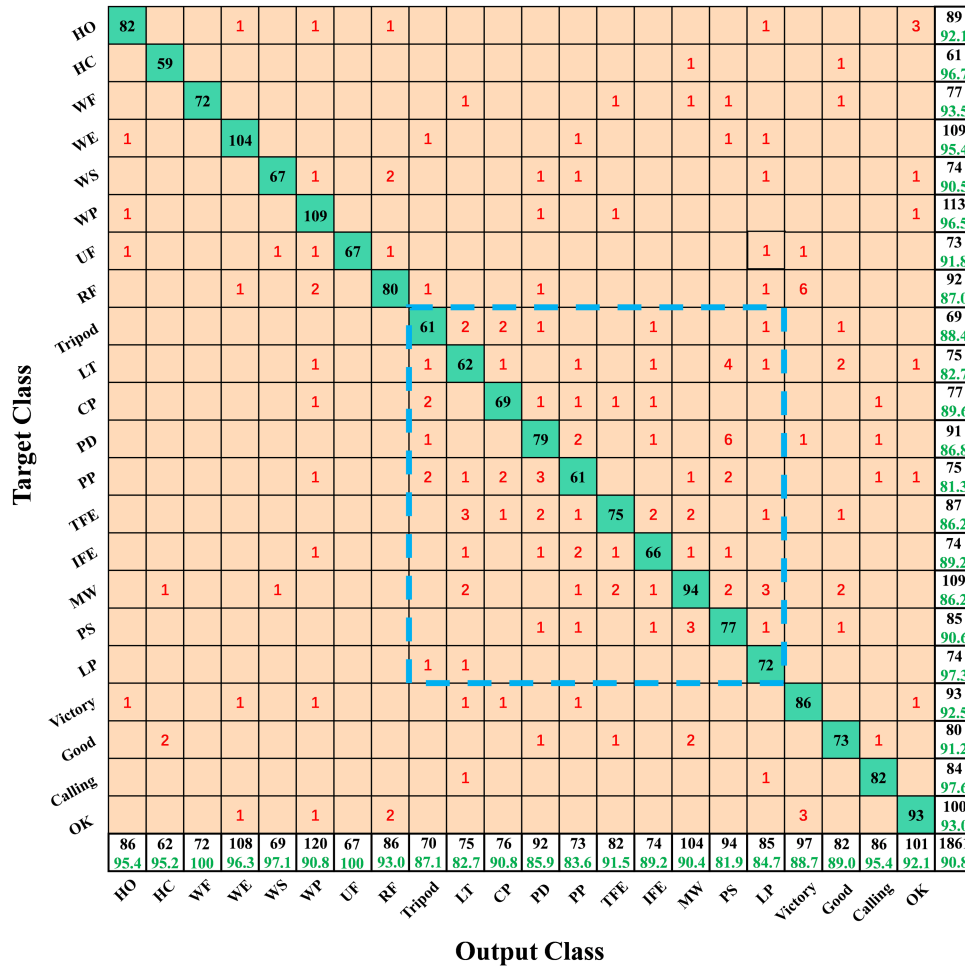


Fig. 10. Confusion matrix of proposed LFC using sEMG-FMG signals. The rightmost column and the bottom row of the matrix represent the recall and precision of each gesture, respectively. The diagonal of the matrix represents the number of correct classifications under both conditions of precision and recall. The black numbers on the edges denote the total number of samples for each gesture, the red numbers denote the number of misclassifications, and the green numbers denote the percentage of correct classification.

TABLE II  
COMPARISON WITH PREVIOUS STUDIES

Research	Modality	Number of motions	Subjects	Number of channels	Method	Results
[10]	FMG			8		80.6%
	sEMG	10	5	8	LDA	81.5%
	FMG&sEMG			16 (8FMG + 8 sEMG)		91.6%
[11]	FMG			4		95.83% ± 3.5%
	sEMG	6	8	4	SVM	98.76% ± 0.75%
	FMG&sEMG			8 (4 FMG + 4 sEMG)		99.42% ± 0.41%
Our	FMG			3		74.83% ± 6.2%
	sEMG	22	9	3	LFC	78.41% ± 5.4%
	FMG&sEMG			6 (3 FMG + 3 sEMG)		93.06% ± 3.3%

which is acceptable in clinical applications to train a specified model for a new subject.

## VII. CONCLUSION

In this article, we present a layered hybrid sEMG-FMG sensor that can measure sEMG and FMG at the same location of the

skin surface. The unique layered structure of the sensor enables the measurement of an additional signal without requiring additional space. An LFC was designed to process sEMG and FMG signals. The results of a 22-hand motion classification experiment on nine subjects showed that the hybrid sEMG-FMG approach can achieve high performance with good generalization ability. The introduction of FMG improved the robustness



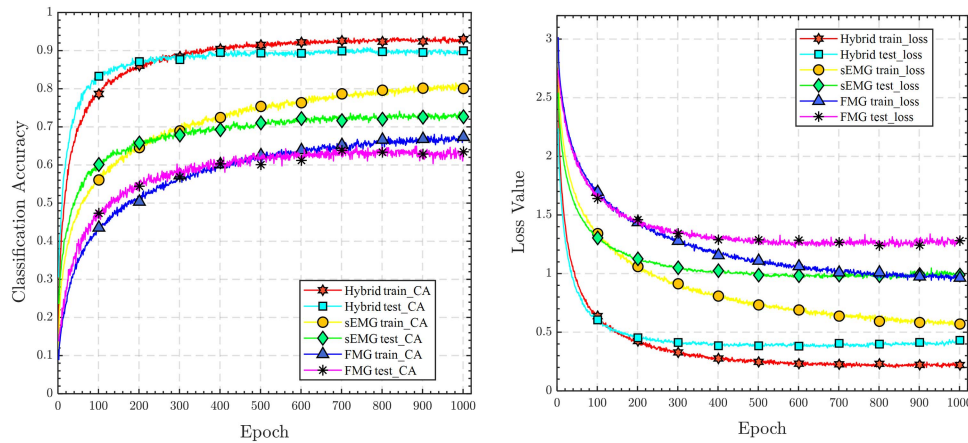


Fig. 11. CAs (left) and loss (right) curves for the three modality settings.



Fig. 12. Heatmap of results (f1-score) for gesture recognition using different scale representation from CNN and LFC. The vertical axis represents different scale representation (scale1 ~ scale8) from different layers' results of CNN. Results of all scales using LFC are shown in the bottom for comparison.

TABLE III  
RECOGNITION ACCURACY AND COMPUTATION TIME WITH DIFFERENT WINDOW LENGTH

Window length (ms)	Accuracy (%)	Flops	Training time		Testing time	
			All samples	One sample (only FP)	All sample	One sample (only FP)
500	93.06% ± 3.26%	32.8 M	19 min 56 s ± 1 min 30 s	506 μs ± 10 μs	7 min 48 s ± 8 s	364 μs ± 7 μs
200	89.77% ± 1.96%	16.2 M	19 min 48 s ± 1 min 15 s	474 μs ± 8 μs	7 min 46 s ± 16 s	343 μs ± 10 μs
100	82.03% ± 2.75%	10.2 M	20 min 10 s ± 1 min 50 s	424 μs ± 5 μs	7 min 16 s ± 9 s	311 μs ± 3 μs

of the classification performance. The hybrid sensor and the LFC have promising potential applications in fields such as teleoperation and human-robot interfaces.

The limitations of this study include the small sample size and potential variations in muscle activation depending on demographics. Future studies will focus on increasing the sample size and involving a wider range of participant demographics. Additionally, the design of the layered sensor system will be improved to reduce its thickness by shortening the effective range of the reflectance sensor and to develop deep learning models better suited for processing multimodal physiological signals.

## REFERENCES

- [1] A. Dwivedi, Y. Kwon, A. J. McDaid, and M. Liarokapis, "A learning scheme for EMG based decoding of dexterous, in-hand manipulation motions," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2205–2215, Oct. 2019.
- [2] A. Hagengruber, U. Leipscher, B. M. Eskofier, and J. Vogel, "Electromyography for teleoperated tasks in weightlessness," *IEEE Trans. Human-Mach. Syst.*, vol. 51, no. 2, pp. 130–140, Apr. 2021.
- [3] T. A. Kuiken et al., "Targeted muscle reinnervation for real-time myoelectric control of multifunction artificial arms," *JAMA*, vol. 301, no. 6, pp. 619–628, Feb. 2009.
- [4] X. Zhang and P. Zhou, "Classification of upper limb motions in stroke using high density surface EMG," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2011, pp. 3367–3370.
- [5] W. Guo, X. Sheng, H. Liu, and X. Zhu, "Development of a multi-channel compact-size wireless hybrid sEMG/NIRS sensor system for prosthetic manipulation," *IEEE Sensors J.*, vol. 16, no. 2, pp. 447–456, Jan. 2016.
- [6] M. Connan, E. R. Ramirez, B. Vodermayr, and C. Castellini, "Assessment of a wearable force- and electromyography device and comparison of the related signals for myocontrol," *Front. Neurobot.*, vol. 10, no. 17, pp. 1–13, 2016.
- [7] C. Ahmadizadeh, L.-K. Merhi, B. Pousett, S. Sangha, and C. Menon, "Toward intuitive prosthetic control: Solving common issues using force myography, surface electromyography, and pattern recognition in a pilot case study," *IEEE Robot. Automat. Mag.*, vol. 24, no. 4, pp. 102–111, Dec. 2017.
- [8] J. McIntosh, C. McNeill, M. Fraser, F. Kerber, M. Löchtefeld, and A. Krüger, "EMPress: Practical hand gesture classification with wrist-mounted EMG and pressure sensing," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2016, pp. 2332–2342.

- [9] M. Nowak, T. Eiband, and C. Castellini, "Multi-modal myocontrol: Testing combined force- and electromyography," in *Proc. Int. Conf. Rehabil. Robot.*, 2017, pp. 1364–1368.
- [10] S. Jiang, Q. Gao, H. Liu, and P. B. Shull, "A novel, co-located EMG/FMG-sensing wearable armband for hand gesture recognition," *Sensors Actuators A, Phys.*, vol. 301, 2020, Art. no. 111738.
- [11] A. Ke, J. Huang, L. Chen, Z. Gao, and J. He, "An ultra-sensitive modular hybrid EMG-FMG sensor with floating electrodes," *Sensors*, vol. 20, no. 17, Jan. 2020, Art. no. 4775.
- [12] P. Fang et al., "Fabrication, structure characterization, and performance testing of piezoelectret-film sensors for recording body motion," *IEEE Sensors J.*, vol. 18, no. 1, pp. 401–412, Jan. 2018, doi: [10.1109/JSEN.2017.2766663](https://doi.org/10.1109/JSEN.2017.2766663).
- [13] S. Yang and D. Ramanan, "Multi-scale recognition with DAG-CNNs," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1215–1223.
- [14] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli, and W. Geng, "Surface-electromyography-based gesture recognition by multi-view deep learning," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2964–2973, Oct. 2019, doi: [10.1109/TBME.2019.2899222](https://doi.org/10.1109/TBME.2019.2899222).
- [15] N. Nazmi, M. A. A. Rahman, S. I. Yamamoto, S. A. Ahmad, H. Zamzuri, and S. A. Mazlan, "A review of classification techniques of EMG signals during isotonic and isometric contractions," *Sensors*, vol. 16, no. 8, Aug. 2016, Art. no. 1304.
- [16] E. Cho, R. Chen, L. K. Merhi, Z. Xiao, B. Pousett, and C. Menon, "Force myography to control robotic upper extremity prostheses: A feasibility study," *Front. Bioeng. Biotechnol.*, vol. 4, pp. 1–12, 2016.
- [17] S. Togo, Y. Murai, Y. Jiang, and H. Yokoi, "Development of an sEMG sensor composed of two-layered conductive silicone with different carbon concentrations," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 13996.
- [18] I. M. Bullock, J. Z. Zheng, S. D. L. Rosa, C. Guertler, and A. M. Dollar, "Grasp frequency and usage in daily household and machine shop tasks," *IEEE Trans. Haptics*, vol. 6, no. 3, pp. 296–308, Jul.–Sep. 2013.
- [19] S. Jiang et al., "Feasibility of wrist-worn, real-time hand, and surface gesture recognition via sEMG and IMU Sensing," *IEEE Trans. Ind. Inform.*, vol. 14, no. 8, pp. 3376–3385, Aug. 2018.
- [20] W. Yang, D. Yang, Y. Liu, and H. Liu, "Decoding simultaneous multi-DOF wrist movements from raw EMG signals using a convolutional neural network," *IEEE Trans. Hum. Mach. Syst.*, vol. 49, no. 5, pp. 411–420, Oct. 2019.
- [21] T. Triwiyanto, I. P. A. Pawana, and M. H. Purnomo, "An improved performance of deep learning based on convolution neural network to classify the hand motion by evaluating hyper parameter," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 7, pp. 1678–1688, Jul. 2020.
- [22] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Inform.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [24] S. M. Khan, A. A. Khan, and O. Farooq, "Selection of features and classifiers for EMG-EEG-based upper limb assistive devices—A review," *IEEE Rev. Biomed. Eng.*, vol. 13, pp. 248–260, 2020.
- [25] Y. Xue, X. Ji, D. Zhou, J. Li, and Z. Ju, "SEMG-based human in-hand motion recognition using nonlinear time series analysis and random forest," *IEEE Access*, vol. 7, pp. 176448–176457, 2019.
- [26] Z. G. Xiao and C. Menon, "Does force myography recorded at the wrist correlate to resistance load levels during bicep curls?," *J. Biomech.*, vol. 83, pp. 310–314, Jan. 2019.



**Peiji Chen** received the B.S. and M.S. degrees in electrical and electronic engineering from Tianjin University of Technology, Tianjin, China, in 2018 and 2021, respectively. He is currently working toward the Ph.D. degree in engineering with The University of Electro-Communications, Tokyo, Japan.

His research interests include affective computing, brain-computer interface, deep learning, and computer vision.



**Ziyi Li** received the B.S. degrees from Xuhai College, China University of Mining and Technology, Xuzhou, China and The University of Electro-Communications, Tokyo, Japan, in 2015, and the M.S. degree from The University of Electro-Communications, in 2021.

His research interests include multimodal sensor fusion and myoelectric prosthetic hand.



**Shunta Togo** (Member, IEEE) received the B.E., M.E., and Ph.D. degrees from Nagoya University, Nagoya, Japan, in 2009, 2011, and 2014, respectively.

From 2012 to 2014, he was a Japan Society for the Promotion of Science (JSPS) Research Fellow (DC2) with the Graduate School of Engineering, Nagoya University. From 2012 to 2016, he was a JSPS Research Fellow (PD) with Cognitive Mechanisms Laboratories, Advanced Telecommunications Research Institute International (ATR). Since 2016, he has been an Assistant Professor with the Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan. His current interests are human motor control, motor coordination, and myoelectric control device. His research interests include control mechanisms of human multijoints and multimuscles.

His research interests are human motor control, motor coordination, and myoelectric control device. His research interests include control mechanisms of human multijoints and multimuscles.



**Hiroshi Yokoi** (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in precision engineering from Hokkaido University, Sapporo, Japan, in 1986, 1990, and 1993, respectively.

From 1986 to 1987, he was an Engineer with Toyota Motor Cooperation, and from 1993 to 1995, he was a Researcher with the Institute of Bioscience and Human Technology, AIST. From 1995 to 2004, he was an Associate Professor with the Department of Complex System Engineering, Faculty of Engineering, Graduate School of Hokkaido University,

and from 2004 to 2009, he was with the Department of Precision Engineering, Faculty of Engineering, University of Tokyo. He is currently a Professor with the Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan. His current research interests include computational intelligence in robotics, artificial life, and medical engineering.



**Yinlai Jiang** (Member, IEEE) received the B.S. and M.S. degrees in computer science and technology from Northeastern University, Shenyang, China, in 2002 and 2005, respectively, and the Ph.D. degree in engineering from Kochi University of Technology, Kami, Japan, in 2008.

From 2008 to 2012, he was a Research Associate, and from 2013 to 2014, he was an Assistant Professor with Kochi University of Technology. He is currently an Associate Professor with the Center for Neuroscience and Biomedical Engineering, The University of Electro-Communications, Tokyo, Japan. His current research interests include biological engineering, robotics, and human-robot interface.

His current research interests include biological engineering, robotics, and human-robot interface.