Real-Time Multiple Gesture Recognition: Application of a Lightweight Individualized 1D CNN Model to an Edge Computing System

Zengjie Yu, Yiqian Lu, Qi An, Cong Chen, Ye Li, *Member, IEEE*, and Yishan Wang

Abstract— The human–machine interface (HMI) detects electrophysiological signals from the subject and controls the machine based on the signal information. However, most applications are still only in the testing stage and are generally unavailable to the public. In recent years, researchers have been devoted to making wearable HMI devices smarter and more comfortable. In this study, a wearable, intelligent eight-channel electromyography (EMG) signal–based system was designed to recognize 21 types of gestures. An analog front end (AFE) integrated chip (IC) was developed to detect the EMG signals, and an integrated EMG signal acquisition device integrating an elastic armband was fabricated. An SIAT database of 21 gestures was established by collecting EMG gesture signals from 10 volunteers. A lightweight 1D CNN model was constructed and subjected to individualized training by using the SIAT database. The maximum signal recognition accuracy was 89.96%, and the average model training time was 14 min 13 s. Given its small size, the model can be applied on lower-performance edge computing devices and is expected to be applied to smartphone terminals in the future. The source code is available at https://github.com/Siat-F9/EMG-Tools.

Index Terms— Convolutional neural network, electromyography, gesture recognition, lightweight individualized model, wearable device.

I. Introduction

HUMAN–MACHINE interfaces (HMIs) can detect and understand electrophysiological signals transmitted from

Zengjie Yu and Yiqian Lu are with the Center for Biomedical Information Technology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China.

Qi An, Cong Chen, Ye Li, and Yishan Wang are with the Center for Biomedical Information Technology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (e-mail: ys.wang@siat.ac.cn).

individuals. According to the information read, machines are commanded to execute orders. Although HMIs have been under development for many years, HMI devices generally remain confined to laboratory settings. To promote their acceptability to the public, researchers have been devoted to making devices smarter and more comfortable in recent years.

Electromyography (EMG) signals are a type of physiological spatiotemporal electrical signal measured from the skin, which contains a large amount of motion information due to muscle stretching activity. EMG signals are rich in signal characteristics, thus enabling improvements in the accuracy of signal recognition and control. These signals are widely used in HMI research [1]–[3]. EMG signal-based personal identification methods and personal verification methods are proposed to match identity information and protect the security of personal information [4].

Hand gesture recognition based on EMG signals is a topic of interest, and the diversity of arm and wrist movements allows for the capture of various hand gestures for training and application to real-world scenarios. Kim *et al.* achieved more favorable results in using four different gestures than in using a remote to control an RC car [5]. Zhang *et al.* designed a virtual Rubik's cube game. Gesture control was employed to evaluate the performance of their gesture classification system [6]. Although most studies were based on hand gesture recognition studies with arm EMG signals, one study verified the feasibility of hand gesture recognition studies based on wrist EMG signals [7].

Multiple studies on EMG-based gesture recognition use signal acquisition devices that are uncomfortable to wear [8]–[11]. The overall power consumption of such acquisition devices is high because of the high number of sensors. This characteristic lowers the acceptability to and popularity of HMI devices among the general public.

To improve the accuracy of gesture recognition based on EMG signals, a considerable number of signal collection studies has involved the design of multichannel EMG signal acquisition systems, which contain more spatiotemporal information than do single-channel systems. Thus, the corresponding accuracy of gesture recognition is also enhanced. However, gesture recognition accuracy improves only to a certain extent when the number of channels is increased. Stango *et al.* reported that the improvement in recognition accuracy attributable to raising the number of channels was not

significant [12]. Multichannel devices introduce a relatively high amount of noise because a greater number of electrodes are in contact with the skin [13]. Furthermore, the use of multichannel EMG acquisition systems inevitably increases device costs and the spatiotemporal computational overhead. This is ascribable to the latency in gesture recognition and the large amount of storage space used. Hargrove *et al.* observed that signals from three critical channels were able to achieve 97% classification accuracy for individual gestures [14]. To balance performance and recognition accuracy, an eight-channel EMG signal acquisition system was constructed in the present study.

HMI devices should not only be comfortable to wear but also smart to use. Deep learning methods have high potential for the extraction of features from signals, and with the active development of deep learning in recent years, growing attention has been placed on the combination of deep learning and gesture recognition based on EMG signals.

Coôteé-Allard *et al.* used an armband called Myo for data collection and employed transfer learning to facilitate the classification of gestures by convolutional neural networks (CNN). Seven gestures were recognized with 98.31% accuracy. However, due to the limited range of gestures involved, the researchers could not determine whether the model could achieve favorable recognition accuracy under multigesture conditions [15]. Zia ur Rehman *et al.* input original EMG signals into a CNN network model for gesture classification, comparing the classification performance with those achieved using linear discriminant analysis, stacked sparse autoencoders with features, and raw samples. The CNN model outperformed the linear discriminant analysis approach. The performance of the CNN model and the stacked sparse autoencoders with features did not differ significantly [16].

Geng *et al.* introduced a CNN for the classification of high-density EMG signals and tested it comparatively on three large, publicly available data sets [17]: Ninapro [18], CapgMyo, and CSL-HDEMG [19]. The CNN model outperformed classical classifiers such as *k*-NN, SVM, LDA, and RF. The self-developed CapgMyo is a static gesture database with 128 EMG signal channels. The researchers tested 8 gestures corresponding to 18 participants 10 times. The data in this database have been preprocessed, and the validity of the gesture data was confirmed through 1-s trials. By setting various time windows and overlapping windows, data segmentation can be performed to generate a sufficient amount of training data. Because the CapgMyo database has high-density channels, subtle changes in hand gestures can be captured with favorable accuracy. However, this database contains only eight gestures. Moreover, the original study used a two-dimensional (2D) CNN model for gesture signal recognition [17]. Such models are not suitable for applications involving wearable devices because of their high performance overhead, although they can achieve high accuracy in gesture signal recognition.

The evidence indicates that CNN models have a greater advantage in the EMG-based gesture recognition system. However, when the network model is overly complex, the recognition speed decreases as the data volume increases.

A moderately complex network model can achieve high recognition accuracy, and increasing its complexity does not significantly improve recognition accuracy. To balance the recognition accuracy and recognition speed, this paper mainly focuses on a one-dimensional (1D) CNN model.

A considerable number of studies have achieved high accuracy in gesture recognition through testing on public data sets. However, when a new individual gesture signal is input, it is often not well differentiated and does not have favorable generalizability. An ideal classification model should be lightweight, accurate, and generically useful. The realization of all three features in practical applications is challenging. Therefore, the design idea of this paper is to balance the three metrics to achieve better results in all aspects. Regarding generalizability we replace it with individualization, through the design of a lightweight 1D CNN model—which can achieve high recognition accuracy for individual gesture signals—an individualized model was trained on new participant data in a relatively short period of time.

In this study, a highly integrated, low-power wearable EMG signal acquisition device was designed, and a lightweight CNN model was designed using publicly available data sets. Due to the lightweight of the model, we innovatively put forward the concept of individual model. It compensates the deficiency that the generic model does not recognize new individual data well. The three model metrics of lightness, accuracy, and generalizability achieved a favorable performance balance. Moreover, an integrated system of gesture acquisition and recognition based on EMG signals that can be applied to edge computing devices is constructed. This provides a good solution for portable gesture recognition system. In addition, we constructed a 21-class static multigesture database based on ethical specifications.

II. EMG Signal Acquisition System

To enhance the comfort of the EMG acquisition system, an EMG signal acquisition device with high integration was fabricated. An integrated chip was designed to acquire eight channels of EMG signals to reduce power consumption, and the electrodes of these eight channels were fixed on an elastic band to ensure ease of wear.

A. AFE IC

The AFE has eight channels, as shown in Fig. 1. Each channel consists of four stages: preamplifier (PreAmp), fixed gain amplifier (FGA), band-pass filter (BPF), and programmable gain amplifier (PGA). The PreAmp and FGA each contributes 20 dB gain. To reduce the noise caused by the PreAmp and FGA, a capacitor-coupled structure is employed for both stages. The gain of the PGA is programmed by two control pins SEL_Gain with four options, which controls the total feedback resistance of the PGA. The total gain of the AFE reaches 500–2000 dB. The bandwidth of the band-pass filter is adjustable by the off-chip capacitor and the control pin SEL BW for low and high pass, respectively. The control pins for all eight channels are connected, which means the gains and bandwidths of the channels can only be set to the
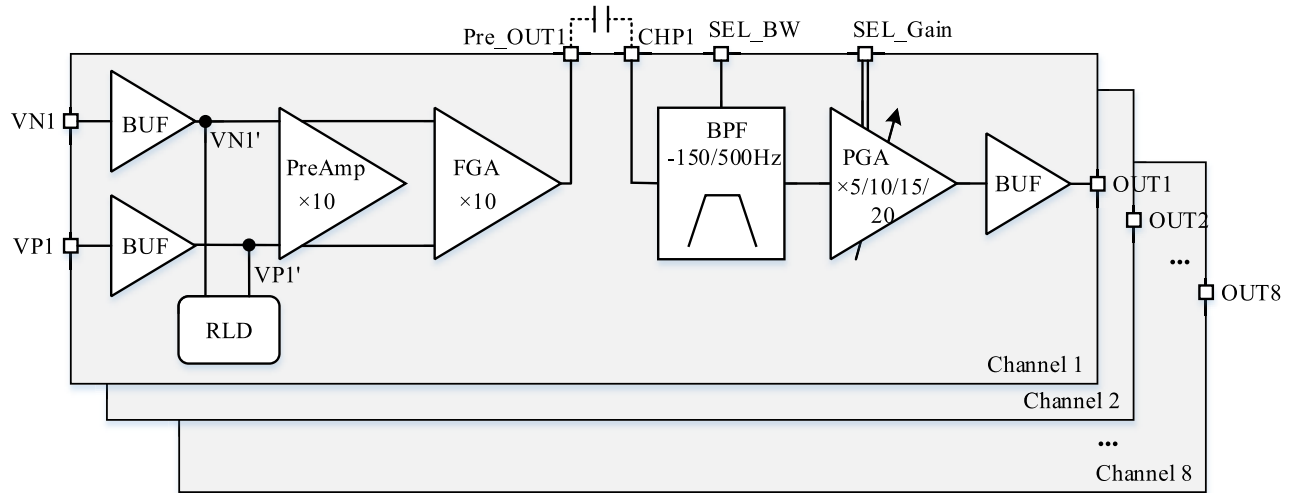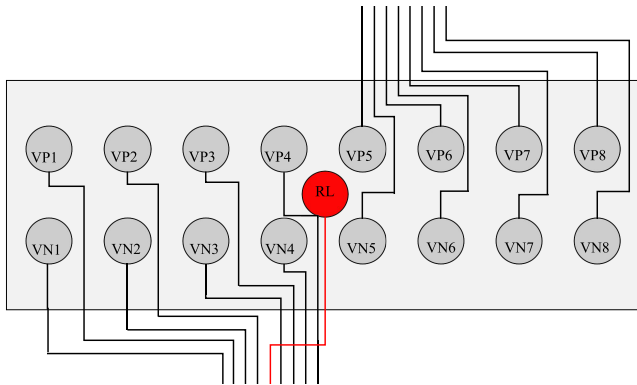
Fig. 1. Architecture of the AFE.

Fig. 2. Model diagram of the signal acquisition armband.

Fig. 3. Signal acquisition process of the proposed system.

same value. The buffer at the output supplies the mA-level drive ability. The right leg driven (RLD) circuit is applied in the AFE to reduce the 50 Hz/60Hz inference.

B. Acquisition Equipment

As the number of channels increases to a certain level, the accuracy of gesture recognition does not improve substantially, but an excessive number of channels leads to computational complexity, increasing the device cost and spatiotemporal computational overhead and causing the system to become bloated. If the number of channels is insufficient, detailed information cannot be collected, reducing the accuracy of gesture recognition to a certain extent. On balance, we used an eight-channel EMG signal acquisition device. To ensure the ease of wear of the acquisition device, we designed an integrated elastic armband in which all electrodes were embedded (Fig. 2).

To enable repeatable EMG signal acquisition, copper-nickel-plated dry electrodes were used, and the distance between positive and negative electrodes was set to 3 cm to ensure a reasonable voltage amplitude. Eight pairs of positive and negative electrodes are spaced evenly on the armband, and the driven-right-leg electrode is placed in the middle of the armband. Elastic was selected as the armband material to
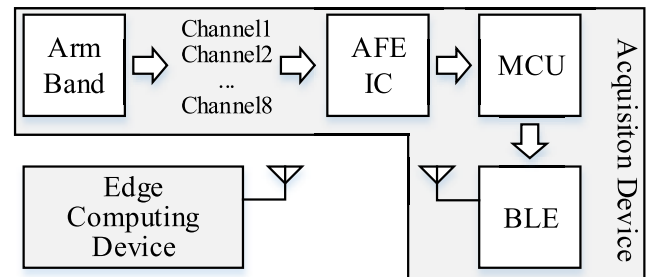
ensure the electrodes were in contact with the skin when the device was worn. To ensure the consistency of the collected EMG data, we considered the location of the electrodes in relation to the skin carefully. We decided to anchor the first pair of dry electrodes at the ulnar carpal extensors 3 cm from the elbow of the left arm. The other seven pairs of electrodes were evenly wrapped around the forearm by using the elastic armband to prevent measurement inconsistencies caused by differences in individual arm thickness.

The signal acquisition process of the system is shown in Fig. 3. The signals captured by the electrode armband are amplified and filtered by the AFE IC. They are then converted from analog to digital signals by the ADC of the microcontroller (MCU) before being sent to the edge computing device through the Bluetooth low-energy module.

Jetson Nano, Nvidia's development platform for edge computing, has a graphics processing unit (GPU) with a 0.5-TFLOPS configuration, a four-core 1.43-G Arm A57, 4 GB memory, BT/WIFI and Ethernet resources and expansion pins, and screen connectivity through an HDMI interface. Because this is an embedded platform, we performed cross-compilation and examined the version dependency of development tools on Ubuntu 1804, adapting the deep learning framework PyTorch 1.4.0, which is highly compatible with Cuda Toolkit 10.2 in Nvidia Developer.

Model training is carried out on high-performance computers. The pretrained model can be downloaded into the edge platform. In the application phase, the real-time identification

of the subject gesture can be performed by the edge computing device. The configuration of this edge computing device is comparable to that of a smartphone; therefore, the model can potentially be imported to the smartphones in the future, making the system truly highly portable.

III. LIGHTWEIGHT MODEL

A. Signal Preprocessing

Signal preprocessing contributes crucially to the final gesture recognition results. The signal data may contain a large amount of redundancy and may also have several outlier interferences that reduce the gesture recognition accuracy. Therefore, signal filtering must be performed. The signal is first smoothed using a Savitzky–Golay filter, and the Savitzky–Golaysmoothing equation is expressed in (1),

$$x_{k,s\ mooth} = \overline{x_k} = \frac{1}{H} \sum_{i=-w}^{+w} x_{k+i} h_i \qquad (1)$$

where k is the index value of the data point, w is the value of the smoothing window which is 50 in this work, $\frac{h_i}{H}$ is a smooth coefficient which is obtained by the least squares method [20]. Then the filtered function is detected for mutations by using the slope method. Absolute value changes in the data are detected over a 50-ms interval, and the segment is discarded if the change exceeds a certain threshold value. In experiments, the typical value of the experimental process is that the variation cannot exceed 0.12 V within 50 ms. This is the result of actual measurements made by on a large number of observations.

B. 1D CNN Model

Using the CapgMyo database [17], we pruned the VCG16 network model and constructed a 1D CNN network model (Fig. 4). The input data are the preprocessed EMG signal data, which is convolved and computed by the convolutional layers. The operation is expressed in (2), where input signal vector $\boldsymbol{x}$ of length N is convolved with a filter vector $\boldsymbol{w}$ of length k, k is 3 in this work, and b is the bias term, f is a non-linear function, which is Rectified Linear Unit (ReLU) [21] in this work. The output layer $\boldsymbol{y}$ of length N with zero-padding.

$$\boldsymbol{y}(j) = f\left(\sum_{i=0}^{k-1} \boldsymbol{\omega}(i)\boldsymbol{x}(j-i) + b\right), \quad j = 0, 1, \ldots, N-1 \qquad (2)$$

The feature map generated by the convolution of the input data is reduced by the pooling layer, where the maximum value in a kernel window function u, with size $m \times 1$ and stride s, is taken over an input vector $\boldsymbol{y}$ resulting in an output vector $\boldsymbol{z}$ defined as

$$z = \max(u(m \times 1, s)\boldsymbol{y}) \qquad (3)$$

where m is 3 and s is 3 in this work.

After the average pooling layer the size of data is reduced, and data are connected to the dense layer through the flatten layer. The dropout layer is used to reduce the complexity and
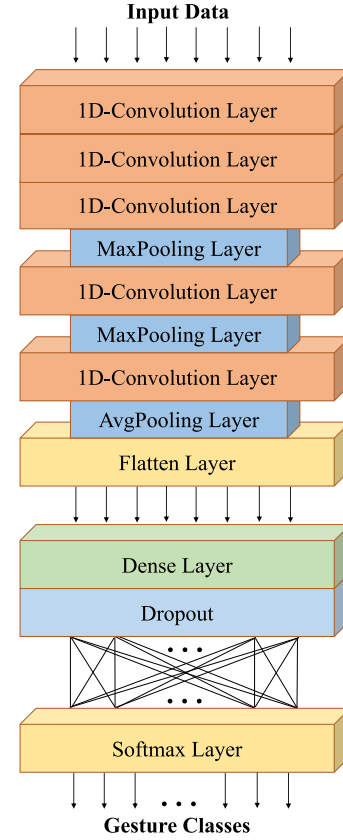

Fig. 4. Block diagram of the 1D CNN model.

computation of the model and enhance model generalizability. Finally, the softmax classifier is used for gesture category output. The probability of each gesture is calculated by softmax function, the gesture with the highest probability is selected as the recognition result,

$$z = \arg max_{z_i} \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}, \quad i = 1, 2, \ldots, n \qquad (4)$$

where z_i is the output value of each gesture through the fully connected layer, n is the number of gesture types which is 21 in this work.

The reason for this optimization is determined by the data; we wanted to ensure that the capacity of the network matched the size of the data. The input EMG data have a 1D signal with a extremely low number of features compared to a 2D picture, and a shallow layer of the network would be able to parse the features well.

IV. INDIVIDUALIZED MODEL

A. SIAT Database Establishment

Using the acquisition device, we designed a data collection visualization software. This software allows for the real-time observation of signals transmitted from individuals, as well as for the determination of signal quality. Through the spatiotemporal frequency domain display of the eight-channel signals by the acquisition device, we observed industrial frequency interference and special signal changes, such as baseline drift and data mutation. We marked these special signal conditions for subsequent signal processing.

TABLE I
BASIC CHARACTERISTICS OF THE VOLUNTEERS

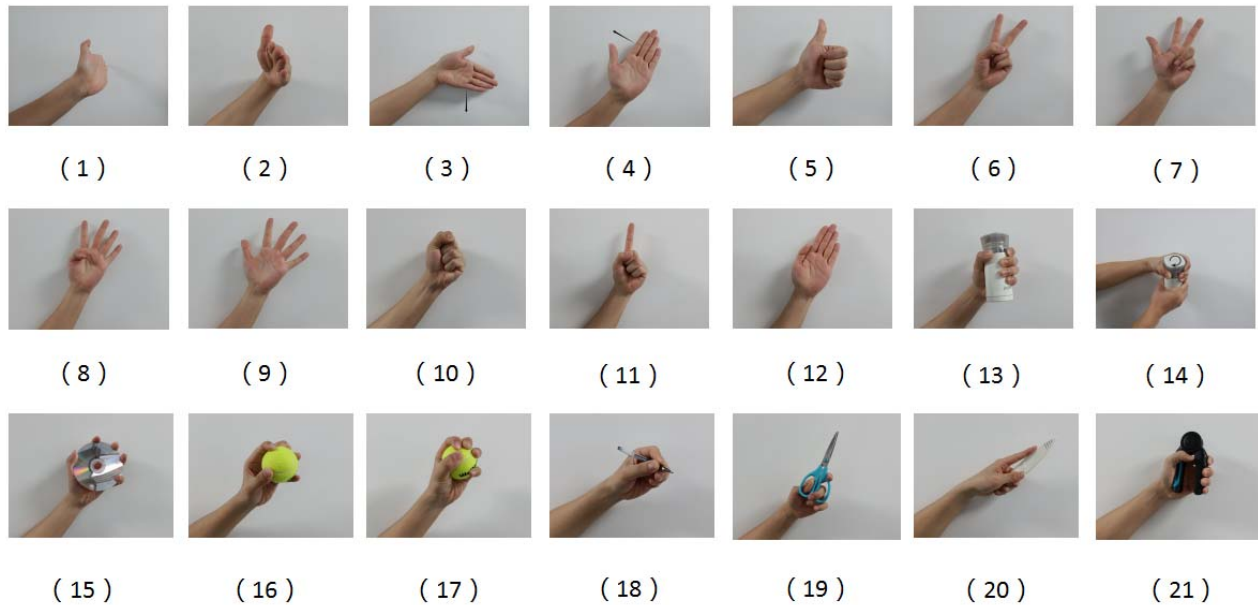| ID	Gender	Age	Height(cm)	Weight(kg)	Arm Length(cm)	Elbow Width(cm)	Wrist Width(cm)
1	Male	23	190	100	26	27	17.5
2	Female	24	156	55	26	21.5	15.5
3	Female	24	158	47	21	20.5	14.5
4	Male	24	172	78	28	26.5	17.5
5	Male	24	180	75	26	26.5	17
6	Male	25	166	55	21	24	16
7	Male	25	171	67	24	24	16
8	Male	26	172	63	25	26.5	16
9	Female	29	159	49	23	20.5	14.5
10	Male	29	171	65	24	25	15

Fig. 5. SIAT database of 21 gestures.

In accordance with ethical norms, the 10 volunteers signed an informed consent form prior to signal acquisition (ethical approval number: SIAT-IRB-201115-H0540). The personal information of the 10 volunteers is presented in Table I. During signal acquisition, the volunteer's elbow was placed on the test bench for support to ensure that only the signals from the forearm were firing. The arm was kept parallel to the wrist, and each movement was performed at 70% of the volunteer's maximum force level until it was felt that the corresponding muscle could be mobilized.

Each volunteer repeated each gesture, recorded over 5-s increments, eight times. The consistency of the endurance of individual muscles varies greatly, and some gestures are easily fatigued by excessive force. Thus, the volunteers were given sufficient rest periods between signal acquisitions. The 21 gestures were performed sequentially, and the force exerted was determined by the researchers' gesture guidelines to ensure the consistency of the collected data. Finally, we constructed a database of 21 types of gestures in 8 channels for 10 volunteers, as displayed in Fig. 5.

Gestures 1 – 4 are wrist operations up, down, left, and right. Gestures 5 – 12 are derived from CapgMyo database and these gestures are isometric and equal-tension gestures, which require the same force position and force angle for the fingers. Gestures 13 – 21 are functional gestures that will be used in daily life.

B. Individualization Adjustment and Real-Time Recognition

As mentioned, the 1D CNN model was pretrained by the CapgMyo database to balance accuracy and performance overhead. Next, the model was trained separately for each individual's data such that recognition accuracy could be improved. The trained model was downloaded to the edge computing device.

When training the model, we focus on the hyperparameters that exert substantial impacts on the neural network and use a grid search approach to find the most appropriate hyperparameter configuration. In general, hyperparameters affect

TABLE II
COMPARISON OF THE PRESENT SYSTEM WITH THOSE IN OTHER STUDIES

	[23]	[24]	[25]	This work
Channel	1	8	4/6/8	8
Detected signals	EEG/ECG/EMG	EMG	EEG/ECG	EEG/ECG/EMG
Technology	$0.18\,\mu$m CMOS	$0.13\,\mu$m CMOS	-	$0.18\,\mu$m CMOS
Supply (V)	3.3	3.3	5	3.3
Current (μA)	3.8	4500	4060/5570/7140	379
Gain (dB)	47.3-71.9	18.2-66.6	0-27.6	54-66.0
Bandwidth (Hz)	0.5-100 (programmable)	3.2k	27k-662k (programmable)	adjusted-150/500 (programmable)
Input referred noise	$0.94\,\mu$V (100 Hz BW)	$0.82\,\mu$V (1-100 Hz)	$1\,\mu$V (0.01-70 Hz)	$0.21\,\mu$V/$\sqrt{\text{Hz}}$@100 Hz
Area (mm^2)	$4.9 * 4.9$	2.19	$8 * 8$ (with package)	$1.6 * 1.6$ (die)
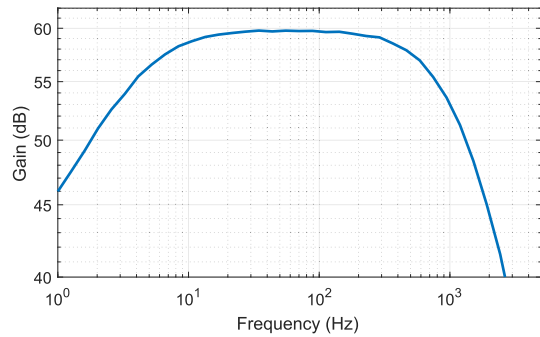
(a) Total gain and bandwidth of the AFE IC when the gain of the AFE is set to 60 dB and the bandwidth of the band-pass filter is set to 5–500 Hz.
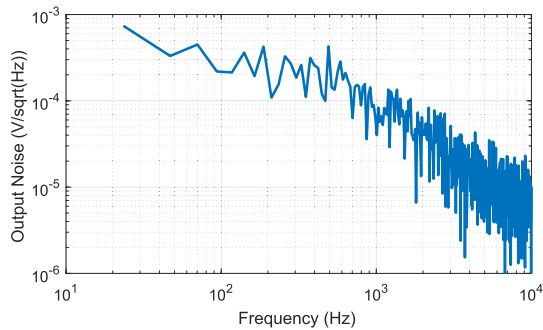
(b) Noise output by the AFE IC when the total gain is set to 60 dB.

Fig. 6. Gain and noise performance of the AFE IC.

the accuracy, stability, lightness, and training time of the model. We divide them into two categories. One is network hyperparameters that require debugging, including the learning rate [18], batch size [22], network depth and network width, and regularized decay factor. The other category is data hyperparameters relevant to this experiment. These comprise the time window length and overlapping window length, which determine the accuracy and data size, respectively; the channel amplitude normalization process; and the filter configuration. The hyperparameters were optimized to produce a sufficient amount of qualitative data.

V. RESULTS

A. Portable EMG Signal Acquisition System

1) Implementation and Measurement Results of the AFE IC: The chip is fabricated in a CMOS (0.18μm) with a 1.6mm × 1.6 mm die area. The current is 379μA.
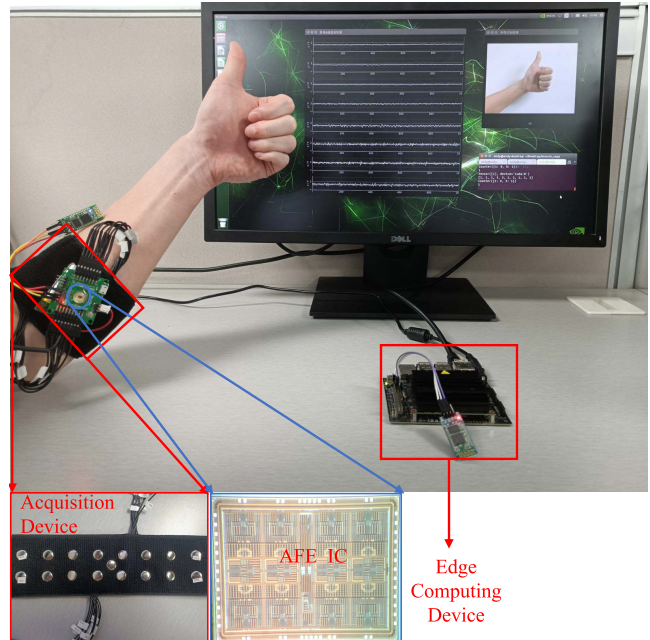

Fig. 7. Setup of the signal acquisition system.

The supply voltage is 3.3V. In the measurement, the control pins of the chip are connected to the MCU. Therefore, the input mode, gain, and bandwidth can have varying settings. Fig. 6(a) presents the frequency response measured when the total gain is set to 60dB with the control pins SEL_Gain and when the bandwidth is set to 5–500 Hz by the off-chip capacitor and the control pin SEL_BW. Fig. 6(b) displays the noise spectrum output under a 60-dB gain setting, when the input pins of the channel were shorted and connected to a reference voltage. The input referred noise was calculated to be 213nV/$\sqrt{\text{Hz}}$ at 100Hz.

The performance of the AFE was compared with those of systems in recent studies (Table II). The AFE not only supports multiple electrophysiological signals and input modes for HMIs but also has low noise and low power consumption.

2) Acquisition System: Fig. 7 presents the construction of the acquisition system, which consists of the AFE IC, MCU, Bluetooth low-energy module, and elastic armband. The edge computing device is Jetson Nano. The eight-channel EMG signals are first detected by the AFE IC and then digitalized and transmitted by the MCU with the Bluetooth low-energy module. The edge computing device processes the received

TABLE III
GESTURE RECOGNITION ACCURACY AND TRAINING TIME
UNDER TIME WINDOWS OF VARYING LENGTH

Time Window Length(ms)	Acc(%)	Time(s)
100	85.9	152
50	79.3	115
20	63.2	103

TABLE IV
RECOGNITION ACCURACY AND TRAINING
TIME FOR EACH PARTICIPANT

Subject ID	Accuracy	Training Time
1	84.72%	13min 33s
2	83.32%	13min 46s
3	77.81%	14min 41s
4	80.04%	16min 18s
5	89.96%	17min 07s
6	83.10%	13min 00s
7	80.79%	13min 34s
8	85.20%	14min 18s
9	86.62%	12min 36s
10	77.70%	13min 14s
Average	82.93%	14min 13s

signals and recognizes the current gesture with the pretrained model. Finally, the monitor is used to display the results output by the acquisition system in real time.

B. Lightweight Model Analysis

1) Data Preprocessing: We performed time slicing on EMG signals over 1-s durations for data expansion. The smaller the time slice is set, the more data can be obtained. The effective frequency band of EMG signals ranges from 10 to 500 Hz, which corresponds to a period of 2 to 100 ms. Thus, a 100-ms EMG signal contains complete frequency domain data.

Table III displays the gesture recognition accuracy and training time under time windows of varying length. When the time window is set to 50 ms (corresponding to the frequency band covered by 20–500 Hz), the loss of low-frequency information from 10 to 20 Hz results in an accuracy 6.6% lower than that under a 100-ms time window. However, the training time is significantly reduced. The accuracy can be improved by using the voting mechanism, which compensates for the loss of low-frequency features attributable to the small amount of time data.

The recognition result of each voting window is R_i $(i = 1, 2, \ldots, N)$, N is the number of time windows. The recognition accuracy of gesture G is P, P can be calculated as,

$$P = \frac{M}{N} \times 100\% \qquad (5)$$

where M is the number of recognition results that satisfy $R_i = G$. The result of the Nth window is the gesture G with the largest P value, which is the classification result with the most occurrences in a period of time, and the accuracy of gesture recognition will be improved. The specific implementation is to set N to 11.Furthermore, the window overlap method can be employed to expand the data during signal segmentation. We set the length of the final overlapping window to 45 ms.

2) Model Performance Analysis: Given that high-density EMG signals can be regarded as three-dimensional (3D) spatiotemporal signals, 1D, 2D, and 3D CNNs alike can be used as the main framework of the gesture classification network. In general, 1D CNN models are structurally simpler than are 2D or 3D CNN model but sacrifice accuracy to a certain extent.In our previous work [26], we compared the recognition accuracy and the performance overhead of 1D CNN and 2D CNN models on CapgMyo database. To balance the model overhead and recognition accuracy, we use a 1D CNN model. The present 1D CNN model is lightweight, with a floating point operations(FLOPs) of only 5 708 288. We compensate for the loss in accuracy of this 1D model through individualized modeling.

Because eight-channel EMG signals were used, we selected a model with a smaller number of parameters and a model file size of no more than 50 kB. It has a negligible volume overhead. With guaranteed gesture recognition accuracy, the training time of the model is shortened and the performance requirements of the edge computing device become less stringent, thereby promoting device portability.

C. Individualized Model and Gesture Recognition

The model was trained individually for each participant, and reasonable hyperparameters were set through automatic optimization. The number of epochs was 50, the weight decay is 0.1, the learning rate is 5×10^{-5} under the cosine descent method, the time window length was 50 ms, and the overlapping window length was 45 ms. The training set was composed of 1, 3, 5, and 7 trials, and the validation set consisted of 2, 4, 6, and 8 trials. Because of the window setting, 991 pieces of data were segmented from each trial, for a total of 3964 over 4 trials. Thus, the training and validation sets comprised a total of 83 244 data entries (3964 × 21 classes of gestures). The actual data volume was slightly lower than that because data preprocessing removed abnormal data entries. By comparing the experimental results of time domain data with frequency domain data as input, we found that the model accuracy is higher when frequency domain data of EMG signal is used as input. To enhance the discriminative power of the neural network model, frequency domain data was used as input.

Regarding the hardware, the GPU was the GeForce TITAN X (12 GB; Nvidia). Moreover, PyTorch version 1.4.0 and CUDA Toolkit 11.0 were employed. The maximum gesture recognition accuracy, mean gesture recognition accuracy, and mean training time was 89.96%, 82.93%, and 14 min 13 s, respectively. Thanks to the lightweight 1D CNN model, the training time for each individual is very short, and the new individual data can be trained quickly to derive an individualized model. The information for each subject is shown in TABLE IV.
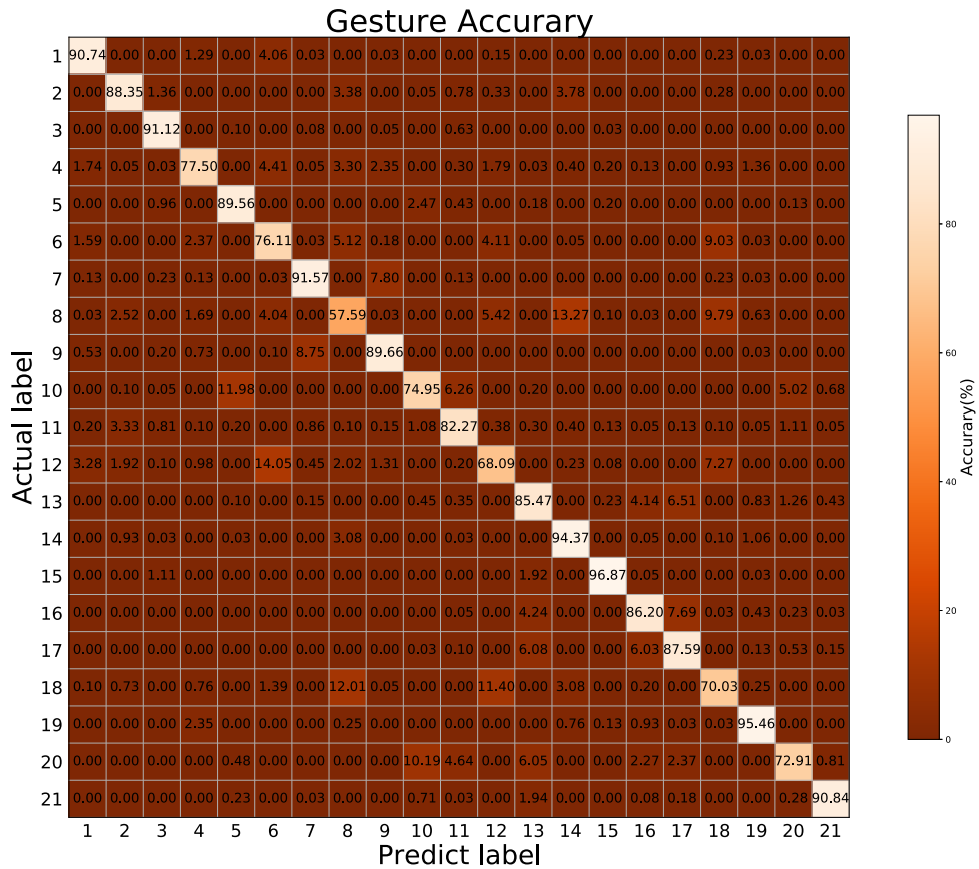
Fig. 8. Confusion matrix of the 21 gestures for participant 1. The horizontal and vertical axes display the predicted and true values, respectively, and the meaning of each element is the ratio of the predicted value to the total value under the true value for each row.
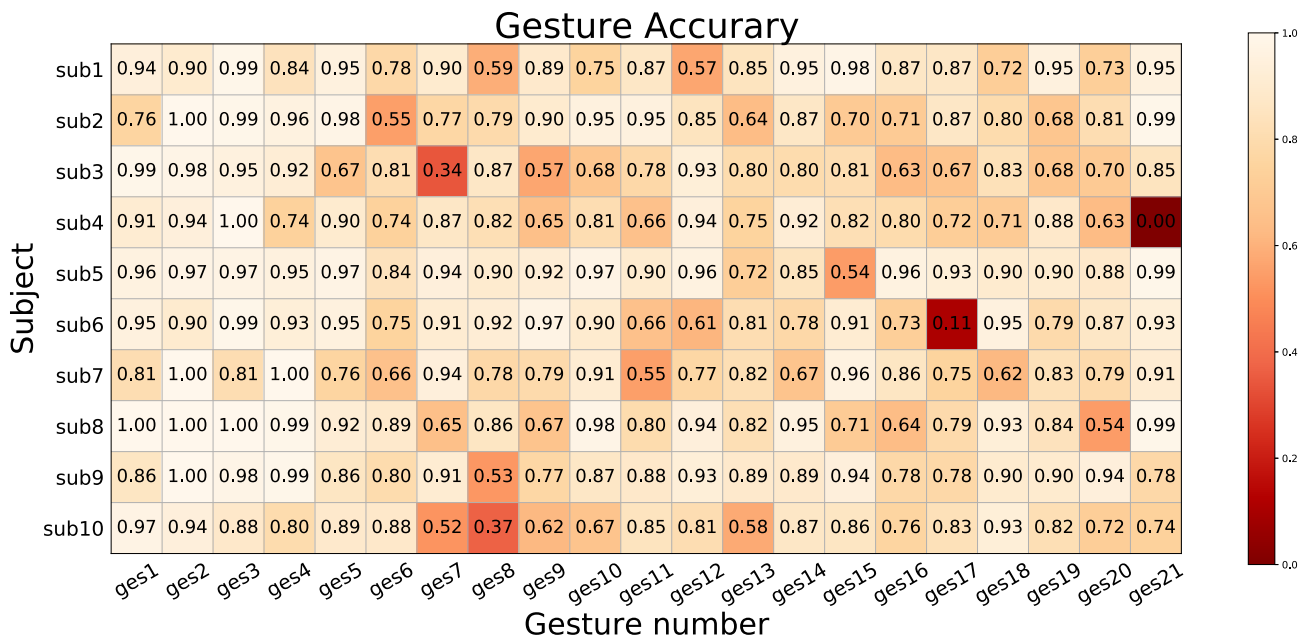

Fig. 9. Matrix of individual 21-class gesture recognition accuracy. The horizontal axis represents the 21 gesture categories, the vertical axis represents the participant number, and each value in the matrix represents the accuracy with which a given participant's gesture was recognized.

High model accuracy for each class of gestures is the aim when faced with a multiple classification task. A confusion matrix can be employed in differentiating classification results.

In a confusion matrix, the horizontal axis presents the predicted values, the vertical axis displays the true values, and the meaning of each element is the ratio of the predicted

value to the total value under the true value for each row. In a perfect confusion matrix, the most prominent elements are the diagonal elements. These represent the elements for which the predicted and true values accord. The confusion matrix of participant 1 after one training session demonstrates that most of the true gestures were consistent with the predictions (Fig. 8).

Fig. 9 displays the matrix of gesture recognition accuracy. The horizontal axis of the matrix shows the participant number, the vertical axis shows the gesture number, and the value of each element is the accuracy with which that gesture was recognized. Observation of the matrix columns revealed no significant convergence in the accuracy of gesture type recognition across participants. This provides evidence to support the premise that the data are inconsistently distributed across individuals. The accuracy of the first four columns of gestures is higher, according to analysis because gestures $1-4$ are wrist movements, and the signal response of the corresponding channel is strong when force is applied, which is easier to identify. The accuracy of gesture 21 in participant 4 was 0.0 because the participant was too fatigued to perform the gesture. The accuracy of gesture 17 in participant 6 was 0.11, because its force pattern was easily confused with gesture 13 and gesture 16.

VI. Conclusion

In this study, a portable gesture acquisition and recognition system based on EMG signals was established. First, an integrated EMG chip was designed to detect EMG signals on eight channels simultaneously. The 0.18-μm CMOS chip has a die area of 1.6 mm $\times$ 1.6 mm and consumes only 379 μA of power under a 3.3-V supply voltage. Using this device, we constructed an SIAT database of 21 gestures performed by 10 volunteers. A lightweight individualized 1D CNN model was designed and trained separately on the data of each participant. The gesture recognition accuracy for the 10 participants was 82.93% on average and peaked at 89.96%. The average training time of the model was only 14 min 13 s. The model can be deployed on edge computing devices such as smartphones, freeing them from the reliance on laboratory environments. Thus, it can be embedded into EMG signal–based wearable gesture recognition systems in the future.

References

[1] F. Palermo, M. Cognolato, A. Gijsberts, H. Müller, B. Caputo, and M. Atzori, "Repeatability of grasp recognition for robotic hand prosthesis control based on sEMG data," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 1154–1159.

[2] M. Ison, I. Vujaklija, B. Whitsell, D. Farina, and P. Artemiadis, "High-density electromyography and motor skill learning for robust long-term control of a 7-DoF robot arm," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 4, pp. 424–433, Apr. 2016.

[3] A. Moin *et al.*, "An EMG gesture recognition system with flexible high-density sensors and brain-inspired high-dimensional classifier," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–5.

[4] L. Lu, J. Mao, W. Wang, G. Ding, and Z. Zhang, "A study of personal recognition method based on EMG signal," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 4, pp. 681–691, Aug. 2020.

[5] J. Kim, S. Mastnik, and E. André, "EMG-based hand gesture recognition for realtime biosignal interfacing," in *Proc. Int. Conf. Intell. User Interfaces (IUI)*, New York, NY, USA, Jan. 2008, pp. 30–39.

[6] X. Zhang, X. Chen, W.-H. Wang, J.-H. Yang, V. Lantz, and K.-Q. Wang, "Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors," in *Proc. 14th Int. Conf. Intell. User Interfaces (IUI)*, New York, NY, USA, Feb. 2009, pp. 401–406.

[7] F. S. Botros, A. Phinyomark, and E. J. Scheme, "Electromyography-based gesture recognition: Is it time to change focus from the forearm to the wrist?" *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 174–184, Jan. 2022.

[8] X. Chen, X. Zhang, Z.-Y. Zhao, J.-H. Yang, V. Lantz, and K.-Q. Wang, "Hand gesture recognition research based on surface EMG sensors and 2D-accelerometers," in *Proc. Int. Symp. Wearable Comput. (ISWC)*, Oct. 2007, pp. 11–14.

[9] 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.

[10] K. Lin, C. Wu, X. Huang, Q. Ding, and X. Gao, "A robust gesture recognition algorithm based on surface EMG," in *Proc. 7th Int. Conf. Adv. Comput. Intell. (ICACI)*, Mar. 2015, pp. 131–136.

[11] Z. Chen, N. Zhang, Z. Wang, Z. Zhou, and D. Hu, "Hand gestures recognition from multi-channel forearm EMG signals," in *Cognitive Systems and Signal Processing* (Communications in Computer and Information Science), F. Sun, H. Liu, and D. Hu, Eds. Singapore: Springer, 2017, pp. 119–125.

[12] A. Stango, F. Negro, and D. Farina, "Spatial correlation of high density EMG signals provides features robust to electrode number and shift in pattern recognition for myocontrol," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 2, pp. 189–198, Mar. 2015.

[13] M. Stachaczyk, S. F. Atashzar, and D. Farina, "Adaptive spatial filtering of high-density EMG for reducing the influence of noise and artefacts in myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 7, pp. 1511–1517, Jul. 2020.

[14] 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.

[15] U. Côté-Allard *et al.*, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Oct. 2019.

[16] M. Z. U. Rehman *et al.*, "Multiday EMG-based classification of hand motions with deep learning techniques," *Sensors*, vol. 18, no. 8, p. 2497, 2018.

[17] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, pp. 1–8, Dec. 2016.

[18] M. Atzori *et al.*, "Building the ninapro database: A resource for the biorobotics community," in *Proc. 4th IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Jun. 2012, pp. 1258–1265.

[19] C. Amma, T. Krings, J. Böer, and T. Schultz, "Advancing muscle-computer interfaces with high-density electromyography," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, Apr. 2015, pp. 929–938.

[20] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964.

[21] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 1–8.

[22] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 464–472.

[23] H. Song, Y. Park, H. Kim, and H. Ko, "Fully integrated biopotential acquisition analog front-end IC," *Sensors*, vol. 15, pp. 25139–25156, Oct. 2015.

[24] S. Benatti *et al.*, "A versatile embedded platform for EMG acquisition and gesture recognition," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 5, pp. 620–630, Oct. 2015.

[25] *ADS1299-xLow-Power,4-,6-,8-Channel,24-Bit, Analog Front-End for Biopotential Measurements*, Texas Instrum., Dallas, TX, USA, 2017.

[26] S. Hao, R. Wang, Y. Wang, and Y. Li, "A spatial attention based convolutional neural network for gesture recognition with HD-sEMG signals," in *Proc. IEEE Int. Conf. E-Health Netw., Appl. Services (HEALTHCOM)*, Mar. 2021, pp. 1–6.