SurfMyoAiR: A Surface Electromyography-Based Framework for Airwriting Recognition

Ayush Tripathi, Prathosh A. P., Suriya Prakash Muthukrishnan, and Lalan Kumar, Member, IEEE

Abstract—Airwriting recognition is the task of identifying letters written in free space with finger movement. It is a dynamic gesture recognition with the vocabulary of gestures corresponding to letters in a given language. Electromyography (EMG) is a technique used to record electrical activity during muscle contraction and relaxation as a result of movement and is widely used for gesture recognition. Most of the current research in gesture recognition is focused on identifying static gestures. However, dynamic gestures are natural and user-friendly for being used as alternate input methods in human–computer interaction (HCI) applications. Airwriting recognition using EMG signals recorded from forearm muscles is, therefore, a viable solution. Since the user does not need to learn any new gestures and a large range of words can be formed by concatenating these letters, it is generalizable to a wider population. There has been limited work in recognition of airwriting using EMG signals and forms the core idea of the current work. The SurfMyoAir dataset comprising of EMG signals recorded during writing English uppercase alphabets is constructed. Several different time-domain features to construct EMG envelope and two different time–frequency image representations: short-time Fourier transform and continuous wavelet transform were explored to form the input to a deep learning model for airwriting recognition. Several different deep learning architectures were exploited for this task. In addition, the effect of various parameters, such as signal length, window length, and interpolation techniques on the recognition performance, is comprehensively explored. The best-achieved accuracy was 78.50% and 62.19% in user-dependent and user-independent scenarios, respectively, by using short-time Fourier transform in conjunction with a 2-D convolutional neural network (CNN)-based classifier. Airwriting has great potential as a user-friendly modality to be used as an alternate input method in HCI applications.

Index Terms—Airwriting, deep learning, electromyography (EMG), gesture recognition, human–computer interaction (HCI), muscle computer interface, wearables.

I. INTRODUCTION

A. Background

The ability to communicate is one of the most important of all life skills that humans possess. With the rapid emergence and evolution of digital devices, interaction of humans with these devices has also increased. However, the medium of user input still remains traditional that includes keyboard, mouse, and touchscreen. Hence, alternate input methods need to be explored for human–computer interaction (HCI) that can reduce the burden of carrying additional devices. This is to be achieved by modalities that can act as natural extensions of human cognition, thereby leading to seamless connection with the digital world [1], [2]. The example includes virtual and augmented reality-based devices where output of the system is directly fed to the user’s eyes. This leaves no scope of using external peripherals for providing input to the system. Various ways have been explored for giving input to such devices. Providing speech-based input is one of the most commonly used solutions [3]. However, in the presence of noise and reverberations, or in the case when the user suffers from a speech disorder, the performance of speech recognition systems degrades significantly [4]. In addition, this method is not feasible when user is required to maintain privacy, such as at a public place. An alternate approach that is suitable for the task is gesture recognition [5], [6]. Although privacy and silent transmission of information are taken care of by using gesture recognition, it suffers from additional limitations. A fixed dictionary of gestures is utilized for communication, thus limiting the range of interaction. In addition, the user is required to learn the set of gestures and memorize them. Airwriting recognition, a special case of gesture recognition, overcomes these shortcomings and can be captured using various physiological signals.

In this work, electromyography (EMG)-based airwriting recognition is explored. Airwriting is the process of writing letters in free space using unrestricted finger movements [7], [8], [9], [10], [11]. Airwriting recognition has been tackled using different modalities as detailed in Table I. EMG is a physiological signal generated due to muscle contraction
and relaxation during the movement. EMG signal can be acquired in two ways: 1) intramuscular EMG, where electrodes are inserted into the skeletal muscle and 2) surface EMG, where the electrodes are placed on the skin above the muscle. Surface EMG (sEMG) has been utilized widely because of its noninvasive and user-friendly nature. sEMG has been used for applications, such as prosthetic control [12], rehabilitation [13], and human–machine interaction [14] and gesture recognition [15]. Airwriting is a dynamic gesture recognition in which the vocabulary of gestures corresponds to letters in a particular language. Since the user is not required to learn any new set of gestures and a wide range of words can be formed by concatenating letters, airwriting can provide the user with a wide range of interaction capabilities. This makes such a system easy to use and generalizable to a larger section of population.

### B. Related Work

The recognition of hand gestures using sEMG signals has garnered wide attention for various HCI applications, including robot control [18], rehabilitation [19], sign language recognition [20], and user authentication [21]. The field of hand gesture recognition can be further subdivided into two categories: static gesture recognition and dynamic gesture recognition [22]. In static gesture recognition, the prime focus is on identification of gestures formed by specific hand shapes without any temporal dimension. There have been various attempts at tackling this problem by using either handcrafted features along with traditional machine learning-based approaches [14] or deep learning-based approaches [15]. Duan et al. [23] proposed a nine-class gesture classification using time-domain features and linear discriminant analysis classifier from three-channel sEMG signals. In [24], a hand-gesture classification system using time-domain features and a neural network architecture using sEMG signals from a Myo Armband was proposed. In addition to handcrafted features, deep learning methods have been applied for the task with reasonably high performance. A convolutional neural network (CNN) has been widely used to extract the spatial relationship present in the multidimensional sEMG signals. Côté Allard et al. [25] proposed a CNN-based gesture classification scheme using frequency-domain features for application in robotic arm. An attention-based CNN and a recursive neural network (RNN) architecture with six different types of images as an input for sEMG-based gesture recognition were proposed by Hu et al. [26]. Rahimian et al. [27] proposed a few-shot training strategy in order to minimize the need for recalibration and allow the user to retrain the model for additional gestures. Botros et al. [28] investigated the use of sEMG signals recorded from the wrist for classification of different single-finger, multifinger, and wrist gestures for HCI applications. Various studies also propose gesture classification schemes using different image representations of the sEMG signals, such as short-time Fourier transform (STFT) [29], continuous wavelet transform (CWT) [30], empirical mode decomposition [32], raw sEMG images [33], sEMG muscle activation maps [34], and gray-scale sEMG images [35].

Dynamic gesture recognition deals with recognition of hand motion trajectory in space. It becomes crucial to comprehensively consider position, shape, and trajectory of the movement simultaneously, making the task of dynamic gesture recognition particularly challenging. In [36], a recognition model based on CNN trained with time–frequency images was proposed for identifying five dynamic hand movements.

### B. Related Work

The recognition of hand gestures using sEMG signals has garnered wide attention for various HCI applications, including robot control [18], rehabilitation [19], sign language recognition [20], and user authentication [21]. The field of hand gesture recognition can be further subdivided into two categories: static gesture recognition and dynamic gesture recognition [22]. In static gesture recognition, the prime focus is on identification of gestures formed by specific hand shapes without any temporal dimension. There have been various attempts at tackling this problem by using either handcrafted features along with traditional machine learning-based approaches [14] or deep learning-based approaches [15]. Duan et al. [23] proposed a nine-class gesture classification using time-domain features and linear discriminant analysis classifier from three-channel sEMG signals. In [24], a hand-gesture classification system using time-domain features and a neural network architecture using sEMG signals from a Myo Armband was proposed. In addition to handcrafted features, deep learning methods have been applied for the task with reasonably high performance. A convolutional neural network (CNN) has been widely used to extract the spatial relationship present in the multidimensional sEMG signals. Côté Allard et al. [25] proposed a CNN-based gesture classification scheme using frequency-domain features for application in robotic arm. An attention-based CNN and a recursive neural network (RNN) architecture with six different types of images as an input for sEMG-based gesture recognition were proposed by Hu et al. [26]. Rahimian et al. [27] proposed a few-shot training strategy in order to minimize the need for recalibration and allow the user to retrain the model for additional gestures. Botros et al. [28] investigated the use of sEMG signals recorded from the wrist for classification of different single-finger, multifinger, and wrist gestures for HCI applications. Various studies also propose gesture classification schemes using different image representations of the sEMG signals, such as short-time Fourier transform (STFT) [29], continuous wavelet transform (CWT) [30], empirical mode decomposition [32], raw sEMG images [33], sEMG muscle activation maps [34], and gray-scale sEMG images [35].

Dynamic gesture recognition deals with recognition of hand motion trajectory in space. It becomes crucial to comprehensively consider position, shape, and trajectory of the movement simultaneously, making the task of dynamic gesture recognition particularly challenging. In [36], a recognition model based on CNN trained with time–frequency images was proposed for identifying five dynamic hand movements. Yang et al. [37] proposed a multistream residual network (MResLSTM) for sEMG-based classification of six dynamic gestures involving the entire hand movement. The problem of recognition of handwritten alphabets may be considered as a specific case of dynamic gesture recognition. A dynamic time warping-based method was proposed in [38] for handwriting recognition, which was further improved in [39]. Linderman et al. [40] proposed a template matching framework for recognition of handwriting from sEMG signals. Beltran-Hernandez et al. [41] proposed a methodology based on sEMG signals to recognize multiuser free-style handwriting characters. A CNN-LSTMBased framework for classification of 36 handwritten characters (A–Z, 0–9) using filtered EMG signals was proposed in their study. A brief description of gesture recognition based on sEMG is presented in Table II.

### B. Related Work

The recognition of hand gestures using sEMG signals has garnered wide attention for various HCI applications, including robot control [18], rehabilitation [19], sign language recognition [20], and user authentication [21]. The field of hand gesture recognition can be further subdivided into two categories: static gesture recognition and dynamic gesture recognition [22]. In static gesture recognition, the prime focus is on identification of gestures formed by specific hand shapes without any temporal dimension. There have been various attempts at tackling this problem by using either handcrafted features along with traditional machine learning-based approaches [14] or deep learning-based approaches [15]. Duan et al. [23] proposed a nine-class gesture classification using time-domain features and linear discriminant analysis classifier from three-channel sEMG signals. In [24], a hand-gesture classification system using time-domain features and a neural network architecture using sEMG signals from a Myo Armband was proposed. In addition to handcrafted features, deep learning methods have been applied for the task with reasonably high performance. A convolutional neural network (CNN) has been widely used to extract the spatial relationship present in the multidimensional sEMG signals. Côté Allard et al. [25] proposed a CNN-based gesture classification scheme using frequency-domain features for application in robotic arm. An attention-based CNN and a recursive neural network (RNN) architecture with six different types of images as an input for sEMG-based gesture recognition were proposed by Hu et al. [26]. Rahimian et al. [27] proposed a few-shot training strategy in order to minimize the need for recalibration and allow the user to retrain the model for additional gestures. Botros et al. [28] investigated the use of sEMG signals recorded from the wrist for classification of different single-finger, multifinger, and wrist gestures for HCI applications. Various studies also propose gesture classification schemes using different image representations of the sEMG signals, such as short-time Fourier transform (STFT) [29], continuous wavelet transform (CWT) [30], empirical mode decomposition [32], raw sEMG images [33], sEMG muscle activation maps [34], and gray-scale sEMG images [35].

Dynamic gesture recognition deals with recognition of hand motion trajectory in space. It becomes crucial to comprehensively consider position, shape, and trajectory of the movement simultaneously, making the task of dynamic gesture recognition particularly challenging. In [36], a recognition model based on CNN trained with time–frequency images was proposed for identifying five dynamic hand movements. Yang et al. [37] proposed a multistream residual network (MResLSTM) for sEMG-based classification of six dynamic gestures involving the entire hand movement. The problem of recognition of handwritten alphabets may be considered as a specific case of dynamic gesture recognition. A dynamic time warping-based method was proposed in [38] for handwriting recognition, which was further improved in [39]. Linderman et al. [40] proposed a template matching framework for recognition of handwriting from sEMG signals. Beltran-Hernandez et al. [41] proposed a methodology based on sEMG signals to recognize multiuser free-style handwriting characters. A CNN-LSTMBased framework for classification of 36 handwritten characters (A–Z, 0–9) using filtered EMG signals was proposed in their study. A brief description of gesture recognition based on sEMG is presented in Table II.

Unlike the traditional setting of writing, airwriting recognition aims to identify characters written in free space with wrist/finger movements. The absence of a support for the finger during the writing process and lack of visual and haptic feedback adds a level of challenge to this problem. There have been various attempts by using inertial sensors [9], [10], [11] and computer vision-based approaches [16], [17] for tackling the airwriting recognition problem. However, to the best of the authors’ knowledge, sEMG-based airwriting recognition has not been explored hitherto. Inspired by the success of the deep learning-based approaches in sEMG-based gesture recognition, different types of sEMG envelopes and time–frequency representations to serve as input to a deep learning model are explored for the purpose of airwriting recognition.

### Table I

| Method | Reference | Approach | Description |
|--------|-----------|----------|-------------|
| Gesture gesture with tongue sensors | Aurrekoeta et al. [8] | CNN-based language model | sEMG-based 2D and 3D models |
| Wiki remote | Chon et al. [8] | CNN-based model | sEMG-based 2D and 3D models |
| Wrist-worn inertial sensor | Yoon et al. [9] | CNN-based classification | sEMG-based 2D and 3D models |
| Wrist-worn inertial sensor | Yoon et al. [9] | CNN-based classification | sEMG-based 2D and 3D models |
| Wrist-worn inertial sensor | Trigida et al. [10] | Image segmentation | sEMG-based 2D and 3D models |
| Camera motion | Kim et al. [10] | Spatio-temporal residual architecture | sEMG-based 2D and 3D models |
| Computer Vision | MultiSource et al. [11] | Region-based CNN | sEMG-based 2D and 3D models |

### Table II

| Approach | Gesture type | Reference |
|----------|--------------|-----------|
| Time domain features and frequency-domain analysis | Static | Duan et al. [15] |
| Frequency domain features and CNN | Static | Allard et al. [25] |
| CNN and RNN | Static | Hu et al. [26] |
| Few-Shot Learning | Static | Rahman et al. [27] |
| Time-Frequency Images and CNN | Dynamic | Song et al. [34] |
| Multi-stream residual network (MResLSTM) | Dynamic | Yang et al. [37] |
| Dynamic Time Warping | Dynamic | Huang et al. [39] |
| Template Matching | Dynamic | Linderman et al. [40] |
| CNN-LSTM | Dynamic | Hernandez et al. [41] |
C. Objectives and Contributions

Airwriting is a dynamic gesture characterized by continuous motion of the wrist to write a specific letter. Unlike static gesture recognition, there has been limited focus on the recognition of dynamic gestures. The focus in this work is on the identification of the dynamic airwriting gestures from sEMG signals. First, the SurfMyoAiR dataset is established, which comprises of sEMG signals recorded from 50 subjects during the task of airwriting. To the best of the authors' knowledge, this is the first instance of a large-scale dataset for the sEMG-based airwriting recognition task. Subsequently, different processing strategies and deep learning architectures for recognition of airwritten letters are explored. The specific contributions of this article are listed below.

1) A surface EMG-based dataset (SurfMyoAiR) recorded from the forearm muscles of 50 subjects while writing the English uppercase alphabets (ten repetitions) in air is created.

2) The performance of different time-domain and time–frequency domain-based approaches in conjunction with deep learning-based classification schemes for the task of airwriting recognition is analyzed.

3) The effects of varying different parameters, such as interpolation techniques, signal duration, and window size, are comprehensively explored, and the optimum choice of parameters suited for the airwriting recognition task is reported.

4) The experiments have been performed in both user-dependent and user-independent manner to ensure generalizability of the proposed airwriting recognition framework.

5) All the source codes and the collected data used in this article will be made available for usage in the HCI community.

The remainder of this article is organized as follows. Section II details the EMG data collection protocol (Section II-A), preprocessing steps (Section II-B), time-domain feature extraction (Section II-C), time–frequency feature extraction (Section II-D), and deep learning models (Section II-E). Experimental details and results are presented in Section III, and Section IV concludes this article.

II. MATERIALS AND METHODS

In this section, the sEMG data collection process, preprocessing of the raw sEMG signals, time-domain and time–frequency-based feature extraction, and the deep learning models for classification are presented.

A. EMG Data Collection

Surface EMG data for this work were collected in accordance with the guidelines of the Helsinki Declaration and were approved by the Institute Ethics Committee, All India Institute of Medical Sciences, New Delhi, India. Fifty healthy subjects (40 male and ten female, all right handed) with an average age of 23.12 years participated in the experiment. Before starting the recording session, the contents of the experiment were explained in detail to the participants, and written consent was obtained. The sEMG data were recorded from each participant’s dominant hand. The signals were recorded using the Noraxon Ultium wireless sEMG sensor [42] at a sampling frequency of 2000 Hz. Disposable, wet-gel-based, self-adhesive Ag/AgCl dual electrodes (having an interelectrode spacing of 20 mm) were placed on the skin over the target muscle according to the orientation of the muscle fiber. In order to keep the contact impedance low, the electrode placement location was cleaned with an alcohol solution. Based on the anatomical bony landmarks of the arm, which is the standard method for surface EMG electrode placement, Pronator teres, Flexor Carpi Radialis, Flexor Digitorum, Extensor Digitorum, and Brachio Radialis were selected as the target muscles for our experiment. The electrode placement location was kept consistent for all the participants that took part in the study.

The participant sat comfortably on a chair while firmly placing the elbow on the table. A user interface designed using the Tkinter module was used to give the participant visual cue regarding the character to be written and also to store annotations during the experiment. The user interface was operated by an experimenter overseeing the data recording session. On pressing the spacebar, a character randomly appeared on the screen, and the participant was asked to write the character in a manner that they were writing on a whiteboard with their finger as the marker. In this manner, each of the participant wrote ten sets of 26 English uppercase alphabets, thus resulting in $26 \times 10 = 260$ samples per subject and a total of $260 \times 50 = 13,000$ samples. The alphabets within a set were randomly shuffled and did not appear in the usual alphabetical order. Furthermore, each alphabet was individually recorded, and the subject was provided rest at the end of two sets of recording. The duration of the rest duration was not fixed, and the participant could take as much rest as required during this period. This was done to increase the variance between different repetitions of the same letter. The entire data collection setup and the sEMG electrode placement locations are presented in Figs. 1 and 2, respectively.

B. EMG Preprocessing

Given that the EMG signals are recorded from different users writing at different speeds and sizes, there exists large variation in the length of signals. Moreover, since different alphabets require different motion to be written, this adds
Fig. 2. Depiction of the sEMG electrode placement locations. The electrode orientation is in accordance with the muscle fiber orientation.

Fig. 3. Distribution of writing times of all samples across all subjects and alphabets. The red dashed line, black solid line, and green dashed-dotted lines depict the mean, median, and 99.9 percentile of the lengths, respectively.

to the variance in the signal lengths. A histogram depicting the duration of writing across all subjects and alphabets is presented in Fig. 3. The mean, median, and 99.9 percentile of duration taken to write an alphabet are 1.96, 1.9, and 3.86 s, respectively. However, feeding the data to any deep learning architecture requires the input dimensions to be fixed. In order to achieve this goal, interpolation is employed if the signal is of a length less than \( L \), while discarding the extra samples otherwise. The effect of variation in the signal length, different values of the parameter \( L \in \{2, 2.5, 3, 3.5, 4\} \) s, is utilized in the experiments. Several standard 1-D interpolation techniques that include linear, quadratic, cubic, and nearest neighbor interpolation are utilized for this task. Suppose the length of a recorded sEMG signal is \( l \ (<L) \); then, the first step is to define a list of \( L \) time stamps that are linearly spaced: \([0, (1 \cdot L - 1), (2 \cdot L - 1), \ldots, l]\). Subsequently, for each time point, two values recorded just before and after this instant from the original signal are taken. Interpolation is then performed using one of the aforementioned techniques to obtain the value at the required time instant.

### C. Time-Domain Analysis Approaches

To the fix length sEMG signals, a sliding window method is applied to extract different sEMG envelopes from the raw signals. Features extracted from sliding rectangular windows of length \( W \) with an overlap of 50% were used to construct envelopes from the five-channel EMG signals. The window length is varied with \( W \in \{25, 50, 75, 100, 125, 150\} \) ms to comprehensively analyze the effect of variation of window length on the sEMG-based airwriting recognition task. Several standard time-domain features were used for constructing the envelope, which are depicted in Fig. 4. Given a windowed segment of signal from a single sEMG electrode \( x \) of length \( W \), the following methods have been applied to obtain the corresponding sEMG envelopes.

1) **Mean Absolute Envelope:** The mean absolute value (MAV) is an average of rectified sEMG amplitude within a segment and is used widely for EMG onset detection tasks. Mathematically, it is expressed as follows:

\[
\text{MAV} = \frac{1}{W} \sum_{i=1}^{W} |x_i|.
\]  

MAV provides an indication of the level of muscle contraction with the value being proportional to the amount of contraction. In addition, it has also been used as a tool for detecting muscle movement onset.

2) **Energy Envelope:** Energy envelope is constructed by using the sum of square of the segmented sEMG signals as follows:

\[
\text{Energy} = \frac{1}{W} \sum_{i=1}^{W} x_i^2.
\]

3) **Variance Envelope:** The variance of a signal segment is defined as the averaged square of deviation from the mean of the segment (denoted by \( \mu_x \)). It is a power index of the sEMG signal and computed as follows:

\[
\text{Variance} = \frac{1}{(W - 1)} \sum_{i=1}^{W-1} (x_i - \mu_x)^2.
\]

Since sEMG can be assumed as a zero mean process, the value \( \mu_x \) can be taken to be 0 for computing the variance feature.

4) **Root-Mean-Square Envelope:** The root-mean-square value is a measure of the strength of the segment and is mathematically computed as follows:

\[
\text{rms} = \sqrt{\frac{1}{W} \sum_{i=1}^{W} x_i^2}.
\]

It is an optimal method for estimating the standard deviation of a signal segment under the assumption of normal distribution.

5) **Absolute Temporal Moment Envelope:** The first and second temporal moments are the same as MAV and energy, respectively, while the three different higher order temporal
Fig. 4. Example of a five-channel EMG signal for the alphabet “A,” interpolated to a length of 3 s (6000 samples) using linear interpolation [I(a)–I(e)], and the different envelopes extracted from the signal: mean absolute envelope [II(a)–II(e)], energy envelope [III(a)–III(e)], variance envelope [IV(a)–IV(e)], root-mean-squared energy envelope [V(a)–V(e)], temporal moment 3 envelope [VI(a)–VI(e)], temporal moment 4 envelope [VII(a)–VII(e)], temporal moment 5 envelope [VIII(a)–VIII(e)], and log detector envelope [IX(a)–IX(e)].

Temporal moment of an sEMG signal has been widely used for applications in gesture recognition and prosthetic arm control in prior literature [43], [44]. The absolute value taken during computation of a temporal moment is to ensure that the within class separation of the different movements is reduced.

6) Log Detector Envelope: It is a nonlinear detector based on exponential-logarithm computation and is mathematically expressed as follows:

$$\log D = \exp \left( \frac{1}{W} \sum_{i=1}^{W} \log(|x_i|) \right).$$  (8)

Moments are computed as follows:

- **TM3**
  $$TM3 = \frac{1}{W} \sum_{i=1}^{W} |x_i|^3.$$  (5)

- **TM4**
  $$TM4 = \frac{1}{W} \sum_{i=1}^{W} x_i^4.$$  (6)

- **TM5**
  $$TM5 = \frac{1}{W} \sum_{i=1}^{W} |x_i|^5.$$  (7)
The log detector feature provides an implicit estimation of the exerted muscle force and has been used in movement control of upper prostheses [45].

### D. Time–Frequency Analysis Approaches

A time–frequency analysis is used to analyze the energy distribution of a signal jointly over both time and frequency domains. It is an effective way to convert the 1-D time series to a corresponding 2-D image representation. Several previous studies have utilized TF images for sEMG-based classification tasks [46], [47]. In particular, two different approaches: short-time Fourier transform and continuous wavelet transform are investigated for the task of EMG-based airwriting recognition (depicted in Fig. 5), which are briefly introduced in the following subsections.

1) Short-Time Fourier Transform: STFT is a conventional approach used for representation and analysis of nonstationary signals. A sliding window is used to segment the signal, and subsequently, Fourier transform is computed for each segment, which results in a joint time–frequency representation of the signal. Mathematically, for a signal \( y(t) \), the STFT is computed as follows:

\[
Y[t, f] = \int_{-\infty}^{\infty} y(\tau) e^{-j2\pi f \tau} d\tau.
\]

In this equation, \( \omega(t) \) is a window function, which is taken to be a Hanning window. In practice, the fast Fourier transform (FFT) is computed from a windowed sEMG signal while maintaining an overlap of 50\% between successive windows. The number of points for computing the FFT is taken to be the same as the window length, and one-sided spectrum is used for subsequent analysis. The STFT magnitude is further computed by taking the absolute value of the complex 2-D signal \( Y[t, f] \), which is fed to the deep learning model for the task of airwriting recognition.

2) Continuous Wavelet Transform: CWT is a multiresolution analysis technique that provides a representation of a signal by varying the translation and scale parameter of mother wavelets [48]. The basis functions for CWT [denoted by \( \Psi_{\sigma, \tau}(t) \)] are described by scaling and translating a single mother wavelet function \( \Psi(t) \). Mathematically, these basis functions can be represented as follows:

\[
\Psi_{\sigma, \tau}(t) = \frac{1}{\sqrt{\sigma}} \Psi \left( \frac{t - \tau}{\sigma} \right).
\]

In the aforementioned equation, \( \tau \) is the translation parameter that shifts the mother wavelet over time, and \( \sigma \) is the scale factor. The normalizing factor \( \frac{1}{\sqrt{\sigma}} \) ensures that the basis function has unit energy. In order to compute the CWT for a signal \( y(t) \), the inner product of the signal with basis functions at different \( \tau \) and \( \sigma \) parameters is computed. Mathematically, it may be represented as follows:

\[
W_{y}[\sigma, \tau] = y(t) \cdot \Psi_{\sigma, \tau}(t) = \frac{1}{\sqrt{\sigma}} \int_{-\infty}^{\infty} y(\tau) \Psi \left( \frac{t - \tau}{\sigma} \right) d\tau.
\]

The wavelet coefficients are obtained by considering all possible shifts and scales of the Morlet mother wavelet. Subsequently, the absolute value of the obtained CWT is taken, which is used for feeding to the deep learning model for classification.

### E. Deep Learning Frameworks

Inspired by the success of deep learning models for solving classification tasks, several different architectures are used to curb and harness the task of airwriting recognition. The model architectures used for the sEMG envelope-based classification are detailed below. Extensive hyperparameter tuning was performed to identify the model architecture parameters. A detailed tabular description of all the models used in the study is provided in Tables IX–XIV (Supplementary Material).
1) **1-DCNN:** The 1-D-CNN architecture is made up of four convolutional layers, with a pooling layer sandwiched between pairs of convolutional layers on either side. This is followed by a global average pooling layer, a dropout layer, and a 26-neuron dense layer activated by the softmax activation function. In the first two convolution layers, the number of filters is set to be 100, while it is set to 160 in the last two convolution layers. The kernel size is set to 10 for all the convolution layers and 3 for the pooling layer.

2) **LSTM/BiLSTM:** In this architecture, two LSTM/BiLSTM layers each with 512 number of units are stacked. This is followed by dropout and dense layers for classification.

3) **1-DCNN-LSTM/BiLSTM:** Here, the extracted sEMG envelope is first passed through two 1-D convolution layers each having 200 filters with a size of 10. This is followed by a pooling layer with a pool size of 3. The features extracted from the CNN are subsequently fed to stacked LSTM/BiLSTM layers having 512 units each. Subsequently, dropout and 26-neuron dense layer activated by the softmax activation function are added. The effect of removing the pooling layer on the classification performance has also been explored. The models with and without the maxpooling layer are denoted by 1-DCNN-Pool-LSTM/BiLSTM and 1-DCNN-LSTM/BiLSTM, respectively.

In case of the models with LSTM/BiLSTM layers, the effect of adding the attention mechanism \([49, 50]\) has also been explored. The attention mechanism aims at relating the different positions of a sequence to compute a weighted representation of the given sequence.

For the time–frequency image-based airwriting recognition, two different schemes are adopted. In the first scenario, the time–frequency images for each of the five channels are separated and fed to parallel models (with shared weights) having same configuration as that in the case of time-domain-based classification. The feature vectors obtained from the five parallel models are then concatenated, which is followed by dropout and dense layers. In addition, a 2-DCNN-based model is also used for recognizing airwriting gestures from the time–frequency images. The 2-DCNN architecture comprises of four convolutional layers with the number of filters on each layer, which are double the number of filters on the previous layer, starting with 32 filters. Each convolutional layer is followed by a pooling layer, and the output of the last layer is flattened and succeeded with a dropout and a dense layer. The kernel size of all convolution and pooling layers is set to (3, 3).

### III. Experimental Setup and Results

#### A. Experimental Details

The recorded sEMG signals from different users correspond to letters written at different speeds and sizes. This leads to a large variation in amplitude of the sEMG signals and, subsequently, the extracted envelopes and time–frequency representations. Therefore, z-normalization is applied to each channel of the obtained sEMG envelopes and time–frequency images. In order to check for the robustness of the proposed airwriting recognition system, two different validation schemes are adopted.

1) **User-Independent Validation:** In this setting, fivefold validation is performed while ensuring that there is no subject overlap in the training and test sets. For this purpose, data corresponding to 40 subjects (40 × 26 × 10 = 10 400 samples) are used for training the model, and data from remaining ten subjects (10 × 26 × 10 = 2600 samples) are used for testing the performance of the trained model. This process is then repeated five times to ensure that accuracy on each subject is evaluated.

2) **User-Dependent Validation:** In this setting, the train-test split is done by dividing the entire data based on the repetition number during the airwriting data collection process. Training set is made by combining data from eight (out of the ten) repetitions from all subjects and the remaining two repetitions from the test set. Similar to the user-independent setting, the process is repeated five times to ensure that each repetition is a part of test set in one of the folds.

The training data in each fold are further split into a training and a validation set in a 80:20 ratio. The parameters of the model are learned by minimizing the categorical cross-entropy loss. A mini-batch training process with a batch size of 256 is employed, and the parameters of the model are updated using the Adam optimizer \([51]\). To prevent overfitting of the model, in addition to the 50% dropout, early stopping with a patience of ten epochs while monitoring the accuracy on the validation set is applied.

#### B. Metrics

Let \(C\) denote the confusion matrix obtained by using the predictions of the model and the true labels. The following metrics have been reported for validating the robustness of the proposed airwriting recognition system.

1) **Accuracy:** It is the total number of correctly classified alphabets. Mathematically, it is given as follows:

\[
\text{Accuracy} = \frac{\sum_{i=1}^{26} C_{ii}}{\sum_{i=1}^{26} \sum_{j=1}^{26} C_{ij}}. \tag{12}
\]

2) **Error Contribution per Letter Pair:** The metric gives an idea of the most confusing letter pairs. For a letter pair \(i\) and \(j\), the measure (denoted by \(EC\)) is computed as follows:

\[
EC_{ij} = \frac{C_{ij} + C_{ji}}{\sum_{k=1}^{26} \sum_{l=0}^{26} C_{kl}}. \tag{13}
\]

3) **Precision:** For an alphabet \(i\), it is defined as the number of correctly predicted alphabet \(i\) out of all predicted alphabet \(i\)

\[
\text{Precision}_i = \frac{C_{ii}}{\sum_{j=1}^{26} C_{ij}}. \tag{14}
\]
TABLE III
RECOGNITION ACCURACIES FOR THE AIRWRITING RECOGNITION TASK BY USING DIFFERENT TIME-DOMAIN sEMG ENVELOPES AND MODEL ARCHITECTURE COMBINATION IN USER-DEPENDENT AND USER-INDEPENDENT SETTINGS. VALUES ARE INDICATED AS (MEAN ± STANDARD DEVIATION) OF ACCURACIES ACROSS THE FIVEFOLDS

| Approach | MA | Energy | RMS | Var | T3M | T3M4 | logD | Raw-data |
|----------|----|--------|-----|-----|-----|------|------|----------|
| 1DCNN    | 75.95 ± 1.66 | 71.22 ± 2.31 | 75.70 ± 2.47 | 71.95 ± 1.46 | 65.93 ± 1.65 | 57.00 ± 1.28 | 50.83 ± 1.69 | 75.55 ± 2.12 | 70.02 ± 2.51 |
| LSTM     | 74.79 ± 1.85 | 69.01 ± 1.86 | 73.77 ± 1.50 | 69.93 ± 1.51 | 61.57 ± 1.05 | 54.18 ± 2.10 | 46.92 ± 1.85 | 73.02 ± 2.13 | 71.40 ± 2.49 |
| LSTM-Att | 75.34 ± 2.98 | 70.48 ± 2.33 | 73.84 ± 2.46 | 70.02 ± 3.06 | 63.02 ± 1.55 | 54.84 ± 3.95 | 47.62 ± 0.95 | 73.61 ± 2.05 | 73.95 ± 2.29 |
| BiLSTM   | 74.34 ± 2.85 | 68.12 ± 2.53 | 72.72 ± 1.50 | 69.19 ± 1.58 | 60.96 ± 0.96 | 53.44 ± 3.56 | 46.54 ± 1.57 | 71.66 ± 2.44 | 71.65 ± 2.65 |
| BiLSTM-Att | 75.59 ± 2.34 | 70.18 ± 3.22 | 74.57 ± 2.83 | 70.69 ± 2.66 | 61.60 ± 2.36 | 55.42 ± 2.71 | 48.04 ± 2.04 | 72.91 ± 2.76 | 73.29 ± 2.65 |
| User     | 1DCNN-Pool-LSTM | 74.39 ± 3.04 | 70.02 ± 2.49 | 73.08 ± 2.43 | 68.65 ± 2.22 | 63.17 ± 1.65 | 54.12 ± 0.90 | 47.63 ± 1.17 | 72.69 ± 2.36 | 73.88 ± 2.54 |
| Dependent | 1DCNN-Pool-LSTM-Att | 73.88 ± 3.59 | 68.05 ± 2.42 | 73.24 ± 2.23 | 68.51 ± 2.37 | 61.18 ± 1.55 | 52.72 ± 0.00 | 46.05 ± 2.78 | 71.99 ± 3.43 | 69.32 ± 3.53 |
| 1DCNN     | 74.54 ± 2.61 | 68.53 ± 2.17 | 72.84 ± 2.39 | 69.52 ± 2.02 | 61.78 ± 2.00 | 54.97 ± 3.00 | 49.44 ± 2.35 | 71.94 ± 2.54 | 72.16 ± 2.58 |
| LSTM      | 73.78 ± 2.79 | 68.03 ± 1.40 | 72.35 ± 2.68 | 68.74 ± 1.82 | 60.86 ± 1.15 | 53.22 ± 1.94 | 46.73 ± 1.94 | 72.08 ± 2.38 | 73.79 ± 0.25 |
| 1DCNN-Pool-BiLSTM | 74.30 ± 1.85 | 66.62 ± 1.29 | 72.35 ± 1.66 | 67.99 ± 2.20 | 61.14 ± 1.74 | 52.60 ± 3.02 | 46.69 ± 2.73 | 72.62 ± 0.79 | 69.56 ± 2.48 |
| Dependent | 1DCNN-Pool-BiLSTM-Att | 72.89 ± 2.69 | 67.71 ± 3.54 | 71.56 ± 1.88 | 67.83 ± 2.49 | 60.21 ± 2.58 | 51.73 ± 1.30 | 45.54 ± 1.43 | 73.13 ± 2.23 | 64.46 ± 2.79 |
| 1DCNN     | 73.29 ± 2.67 | 68.70 ± 2.08 | 73.18 ± 3.32 | 68.75 ± 2.49 | 61.79 ± 1.47 | 54.06 ± 2.14 | 47.98 ± 1.02 | 72.28 ± 2.43 | 70.39 ± 2.00 |
| LSTM      | 74.67 ± 2.71 | 68.89 ± 2.71 | 74.98 ± 3.02 | 69.80 ± 2.27 | 62.93 ± 2.28 | 54.37 ± 1.70 | 48.08 ± 1.76 | 73.38 ± 1.28 | 67.32 ± 2.86 |

Fig. 6. Effect of variation of different parameters on mean recognition accuracies for the airwriting recognition system using MA envelope and 1-DCNN model. The left plot depicts the variation of accuracy with window length while keeping the signal length fix to 8000 samples (4 s), while the plot on the right depicts the variation of accuracy with signal length while keeping the window length fix to 250 samples (125 ms). Effect of different interpolation techniques is marked by the legend in the subplot.

4) **Recall:** For an alphabet $i$, it is defined as the number of correctly predicted alphabet $i$ out of all alphabets with true label $i$

$$
Recall_i = \frac{C_{ii}}{\sum_j C_{ij}}.
$$

(15)

5) **F1-Score:** It is computed as the harmonic mean of precision and recall. Mathematically, it is represented as follows:

$$
F1 - \text{Score}_i = \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}.
$$

(16)

C. Results

Table III lists the recognition accuracies by using different time-domain sEMG envelopes and deep learning models. For all the results presented in the table, the sEMG signals are interpolated to 4-s duration (8000 samples) by using the cubic interpolation technique. Subsequently, sEMG envelope is constructed by using a window size of 125 ms (250 samples). In Table IV, the mean recognition accuracy for the airwriting recognition task by using the time–frequency images and model architecture combination in user-dependent and user-independent settings, values are indicated as (mean ± standard deviation) of accuracies across the fivefolds.

TABLE IV
RECOGNITION ACCURACIES FOR THE AIRWRITING RECOGNITION TASK BY USING DIFFERENT TIME–FREQUENCY IMAGES AND MODEL ARCHITECTURE COMBINATION IN USER-DEPENDENT AND USER-INDEPENDENT SETTINGS. VALUES ARE INDICATED AS (MEAN ± STANDARD DEVIATION) OF ACCURACIES ACROSS THE FIVEFOLDS

| Approach | STFT | CTW |
|----------|------|-----|
| 1DCNN    | 75.11 ± 1.45 | 68.97 ± 2.70 |
| LSTM     | 69.90 ± 3.17 | 66.73 ± 3.05 |
| LSTM-Att | 73.23 ± 3.54 | 67.94 ± 3.03 |
| BiLSTM   | 70.40 ± 2.47 | 65.82 ± 3.14 |
| BiLSTM-Att | 73.87 ± 2.24 | 68.76 ± 2.64 |
| User     | 1DCNN-Pool-LSTM | 71.65 ± 3.69 | 65.68 ± 2.56 |
| Dependent | 1DCNN-Pool-LSTM-Att | 68.43 ± 4.77 | 66.38 ± 2.32 |
| 1DCNN     | 71.67 ± 1.67 | 65.77 ± 1.57 |
| LSTM      | 71.01 ± 2.23 | 64.53 ± 3.12 |
| 1DCNN     | 71.44 ± 2.79 | 64.96 ± 2.41 |
| LSTM      | 70.34 ± 3.50 | 64.92 ± 2.61 |
| 1DCNN     | 72.35 ± 3.26 | 67.31 ± 2.12 |
| LSTM      | 75.80 ± 2.33 | 72.98 ± 2.16 |

Fig. 6. Effect of variation of different parameters on mean recognition accuracies for the airwriting recognition system using MA envelope and 1-DCNN model. The left plot depicts the variation of accuracy with window length while keeping the signal length fix to 8000 samples (4 s), while the plot on the right depicts the variation of accuracy with signal length while keeping the window length fix to 250 samples (125 ms). Effect of different interpolation techniques is marked by the legend in the subplot.
other envelope construction strategies. This is also evident as individual clusters may be attributed to its capability of capturing the dynamic nature of the gestures. The superior performance of the mean absolute envelope coupled with 1-DCNN-based model architecture yields the best recognition accuracy of 75% in user-independent and user-dependent settings by performing one-tailed t-test. This implies that there is a statistically significant difference between the results obtained by using different sEMG envelopes with the 1-DCNN model. Further, an analysis of alphabetwise precision, recall, and F1-Scores is also performed and reported in Table XV (Supplementary Material).

| Rank | Letter Pair | KC  |
|------|-------------|-----|
| 1    | D,P         | 4.75% |
| 2    | N,W         | 3.36% |
| 3    | O,U         | 2.55% |
| 4    | U,V         | 2.49% |
| 5    | B,R         | 2.33% |

| Rank | Letter Pair | KC  |
|------|-------------|-----|
| 1    | D,P         | 6.48% |
| 2    | N,W         | 4.78% |
| 3    | U,V         | 3.47% |
| 4    | O,U         | 2.68% |
| 5    | K,R         | 2.52% |

From Table III, it may be noted that using the mean absolute sEMG envelope coupled with 1-DCNN-based model architecture yields the best recognition accuracy of 75.75% and 61.13% in user-dependent and user-independent scenarios. The superior performance of the mean absolute envelope may be attributed to its capability of capturing the fine details contained in the sEMG signal compared with other envelope construction strategies. This is also evident from the sample envelopes presented in Fig. 4. Likewise, the envelopes constructed by using root mean square and log detector features yield a comparable performance to the mean absolute envelope due to the aforementioned reason. It may also be noted that, on using higher order features, such as energy and temporal moments, the fine-level details of the sEMG signal are lost, which leads to a significant reduction in the classification performance. The use of mean absolute envelope is also superior to the use of raw sEMG signals for airwriting recognition. This may be attributed to the fact that extracting the sEMG envelopes makes smoothen the sEMG signal, which, in turn, helps in boosting the airwriting recognition performance. Among the different classifiers, the 1-DCNN model outperforms all the other architectures due to its ability to exploit the spatial correlation in the sEMG envelopes and, hence, suiting well for this task. The other deep learning architectures used in the study also perform fairly well, yielding comparable recognition accuracies. The performance of LSTM/BilSTM and 1-DCNN-LSTM/BilSTM models is boosted by using the attention mechanism. However, in case of the 1-DCNN-Pool-LSTM/BilSTM model, employing attention does not improve the model performance, since the signal dimension gets shortened due to the pooling layer. On using the time–frequency images, the best accuracy is achieved on using the STFT and 2-DCNN model architecture combination. An accuracy of 78.50% and 62.19% is achieved in user-independent and user-dependent settings, respectively.

The representation of the sEMG signal in joint time and frequency space helps the 2-DCNN model to extract diverse characteristics and learn the complex patterns, which is helpful for the airwriting recognition task. It is to be noted that airwriting is a dynamic gesture, and the gesture vocabulary comprises of 26 classes (chance accuracy of 3.84%). This implies that the accuracies achieved by the models are well within acceptable limits for practical usage. A single-factor analysis of variance (ANOVA) performed on the accuracies obtained from the fivefolds by using different sEMG envelopes with the 1-DCNN-based classification model yielded the p-values of 0.00047 and 8.21 × 10⁻²¹ in user-independent and user-dependent settings, respectively. Similarly, the p-values of 0.2054 and 0.0082 were obtained by using the 2-DCNN model with the time–frequency-based approach in user-independent and user-dependent settings by performing one-tailed t-test. This implies that there is a statistically significant difference between the results obtained by using different features for the task of airwriting recognition using sEMG signals. A detailed t-test between all possible pairs of sEMG envelopes with the 1-DCNN model was also performed, and the results are presented in Tables VII and VIII. From Figs. 6 and 7, it may be observed that on increasing the length of the signal, the performance of the airwriting recognition system improves. This increasing trend is due to the fact that discarding samples to constraint the signal length leads to a loss of information, which is vital for the model to learn nuances of the airwritten letter. The choice of window length parameter is influenced by a trade-off between smoothness of the sEMG signal and capturing the fine details of the sEMG signal. Hence, the accuracy at first increases with

![Fig. 7. Effect of variation of different parameters on mean recognition accuracies for the airwriting recognition system using the STFT and 2-DCNN model. The left plot depicts the variation of accuracy with window length while keeping the signal length fix to 8000 samples (4 s), while the plot on the right depicts the variation of accuracy with signal length while keeping the window length fix to 250 samples (125 ms). Effect of different interpolation techniques is marked by the legend in the subplot.](image-url)
increasing window size but then converges (and drops slightly) upon further increment. Therefore, a window size of 125 ms (250 samples) was chosen as optimum for all the analysis. Similar trade-off between time and frequency resolution of the resulting STFT image yields a window size of 100 ms (200 samples) as optimum for obtaining the time–frequency image.

The classical ML algorithms, when used along with handcrafted features, led to poorer results due to lack of incorporation of temporal information. For instance, an SVM model trained with a set of 12 features per sEMG channel (MAV, energy, variance, temporal moment 3, temporal moment 4, temporal moment 5, root mean square, log detector, kurtosis, skewness, mode, and median) yielded an accuracy of 7.34% and 8.73% in user-independent and user-dependent scenarios, respectively. A similar trend was observed with other classical algorithms, such as logistic regression and decision trees, as well. Since airwriting recognition falls under the category of dynamic gesture recognition, it is essential to capture the temporal dynamics for making accurate predictions. The poor performance of handcrafted features with classical ML algorithms may be attributed to the fact that the summary statistics features fail to capture the dynamics of the time-series data, which is critical for recognition accuracies.

A list of top-5 pairs with highest contribution to the error in user-independent and user-dependent settings by using the mean absolute envelope and 1-DCNN model is presented in Tables V and VI, respectively. It may be noted that letters pairs with visible similarity contribute the highest to the misclassification. For instance, “D” and “P” are the most widely confusing letter pairs, which is attributed to the fact that both the alphabets involve similar movement required to write them in the air (vertical line and semicircle). A user can easily lose spatial orientation while writing the letters, thereby leading to high misclassification rate.

TABLE VII

| Feature | MAV | MAE | RMS | Vari | TMA1 | TMA2 | TMA3 | Log5 |
|---------|-----|-----|-----|------|------|------|------|------|
| MAV     | 0.325 | 0.209 | 0.209 | 0.029 | 0.004 | 0.028 | 0.379 |
| MAE     | 0.209 | 0.115 | 0.115 | 0.029 | 0.004 | 0.028 | 0.379 |
| RMS     | 0.029 | 0.014 | 0.014 | 0.004 | 0.028 | 0.379 |

TABLE VIII

| Feature | MAV | MAE | RMS | Vari | TMA1 | TMA2 | TMA3 | Log5 |
|---------|-----|-----|-----|------|------|------|------|------|
| MAV     | 0.325 | 0.209 | 0.209 | 0.029 | 0.004 | 0.028 | 0.379 |
| MAE     | 0.209 | 0.115 | 0.115 | 0.029 | 0.004 | 0.028 | 0.379 |
| RMS     | 0.029 | 0.014 | 0.014 | 0.004 | 0.028 | 0.379 |

IV. Conclusion

In this article, an sEMG-based airwriting recognition framework was proposed. The SurfMyoAiR dataset comprising of sEMG signals measured from five forearm muscles of 50 subjects while writing the 26 English uppercase alphabets was constructed. To the best of our knowledge, this is the first large-scale dataset for the airwriting recognition task. Several sEMG envelope extraction methods using time-domain features, such as MAV, energy, variance, root mean square, different temporal moments, and log detector, and time–frequency images, such as short-time Fourier transform and continuous wavelet transform, were explored to form input to deep learning models for airwriting recognition. The effect of different parameters, such as signal length, window length, and different interpolation techniques, on the system performance in both user-dependent and user-independent settings was also comprehensively examined. Among the different EMG envelopes, the mean absolute envelope yielded the best recognition accuracy of 61.13% and 75.75% in the user-independent and user-dependent settings, respectively. This is attributed to the fact that the mean absolute envelope captures finer details of the EMG signal, which leads a superior representation compared with other envelope construction approaches. The STFT image outperforms all the techniques explored in the article giving a recognition accuracy of 62.19% and 78.50% in the user-independent and user-dependent settings, respectively. This is expected as complex attributes corresponding to different airwritten letters are better observable in a joint time–frequency space. Future work may be focused on improving the performance of the system by exploring other sophisticated feature extraction and/or deep learning techniques. In conclusion, the fairly high recognition accuracies form a great baseline for future work in the domain of sEMG-based airwriting recognition. Such an approach has a potential to be used as an alternate input method for HCI applications.

ACKNOWLEDGMENT

The authors would like to thank Shivanii Ranjan, Anant Jain, and Mirdoddi Kaushik for their valuable insights and assistance during data collection and all the subjects for their participation in the experiment.

REFERENCES

[1] A. Quito. (2019). An MIT Researcher Demonstrated How we Can Now Search the Internet Through a Wearable Sticker. [Online]. Available: https://lgz.com/1602401/mits-arxiv-kapur-performs-a-live-demo-of-alterego-at-ted

[2] Tech@Facebook. (2021). Inside Facebook Reality Labs: The Next Era of Human-Computer Interaction. [Online]. Available: https://tech.fb.com/inside-facebook-reality-labs-the-next-era-of-human-computer-interaction

[3] D. Rothmann. (2018). Speech as Input in Virtual Reality. [Online]. Available: https://towardsdatascience.com/speech-as-input-in-virtual-reality-b68929f9b41

[4] M. Moore, H. Venkateswara, and S. Panchanathan, “Whistle-blowing ASRs: Evaluating the need for more inclusive speech recognition systems,” in Proc. Interspeech, 2018, pp. 466–470, doi: 10.21437/Interspeech.2018-2391.
[5] M. Yasen and S. Jusoh, “A systematic review on hand gesture recognition techniques, challenges and applications,” PeerJ Comput. Sci., vol. 5, p. e218, Sep. 2019.

[6] M. J. Cheok, Z. Omar, and M. H. Jaward, “A review of hand gesture and sign language recognition techniques,” Int. J. Mach. Learn. Cybern., vol. 10, pp. 131–153, Aug. 2019, doi: 10.1007/s13041-017-0705-5.

[7] C. Amma and T. Schultz, “Airwriting: Bringing text entry to wearable computers,” XRDS: Crossroads, ACM Mag. for Students, vol. 20, no. 2, pp. 50–55, 2013.

[8] M. Chen, G. AlRegib, and B.-H. Juang, “Air-writing recognition—Part I: Modeling and recognition of characters, words, and connecting motions,” IEEE Trans. Human-Mach. Syst., vol. 46, no. 3, pp. 403–413, Jun. 2016.

[9] T. Yanay and E. Shmueili, “Air-writing recognition using smart-bands,” Sensors, vol. 16, no. 10, 2016, Art. no. 101183.

[10] A. Tripathi, A. K. Mondal, L. Kumar, and P. Ap, “SClAIR: Supervised contrastive learning for user and device independent airwriting recognition,” IEEE Sensors Lett., vol. 6, no. 2, pp. 1–4, Feb. 2022.

[11] A. Tripathi, A. K. Mondal, L. Kumar, and P. A. Prathosh, “ImAIR: Airwriting recognition framework using image representation of IMU signals,” IEEE Sensors Lett., vol. 6, no. 10, pp. 1–4, Oct. 2022.

[12] O. W. Samuel et al., “Pattern recognition of electromyography signals based on novel time–domain features for amputees’ limb motion classification,” Comput. Electr. Eng., vol. 67, pp. 646–655, Apr. 2018.

[13] I. Campanini, C. Disselhorst-Klug, W. Z. Rymer, and R. Merletti, “Surface EMG in clinical assessment and neurorehabilitation: Barriers limiting its use,” Frontiers Neurosci., vol. 11, p. 934, Sep. 2020.

[14] M. Simão, N. Mendes, O. Gibaru, and P. Neto, “A review on electromyography decoding and pattern recognition for human-machine interaction,” IEEE Access, vol. 7, pp. 39564–39582, 2019.

[15] W. Li, P. Shi, and H. Yu, “Gesture recognition using surface electromyography and deep learning for prostheses hand: State-of-the-art, challenges, and future,” Frontiers Neurosci., vol. 15, Apr. 2021, Art. no. 621885.

[16] U.-H. Kim, Y. Hwang, S.-K. Lee, and J.-H. Kim, “Writing in the air: Unconstrained text recognition from finger movement using spatio-temporal convolution in 2021,” arXiv:2104.09021.

[17] S. Mukherjee, S. A. Ahmed, D. P. Dogra, S. Kar, and P. P. Roy, “Fingertip detection and tracking for recognition of air-writing in videos,” Expert Syst. Appl., vol. 136, pp. 217–229, Dec. 2019.

[18] B. Boru and K. Erin, “Novel technique for control of industrial robots with wearable and contactless technologies,” Measurement, vol. 192, Mar. 2022, Art. no. 110850.

[19] H. Zhou et al., “Toward hand pattern recognition in assistive and rehabilitation robotics using EMG and kinematics,” Frontiers Neurorobot., vol. 15, p. 50, Jan. 2021.

[20] S. Tateno, H. Lii, and J. Ou, “Development of sign language motion detection and tracking for recognition of air-writing in videos,” IEEE Access, vol. 7, pp. 39564–39582, 2019.

[21] Y. Xu and Y. Dai, “Review of hand gesture recognition study and application,” Contemp. Eng. Sci., vol. 10, pp. 375–384, 2017.

[22] F. Duan, X. Ren, and Y. Yang, “A gesture recognition system based on time domain features and linear discriminant analysis,” IEEE Trans. Cognit. Devop. Syst., vol. 13, no. 1, pp. 200–208, Mar. 2021.

[23] A. A. Neacsu, G. Cioroiu, A. Radoi, and C. Burileanu, “Automatic EMG-based hand gesture recognition system using time-domain descriptors and fully-connected neural networks,” in Proc. 42nd Int. Conf. Telecommun. Signal Process. (TSP), Jul. 2019, pp. 232–235.

[24] U. C. Allard et al., “A convolutional neural network for robotic arm guidance using SEMG based frequency-features,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Oct. 2016, pp. 2464–2470.

[25] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, “A novel attention-based hybrid CNN-BNN architecture for sEMG-based gesture recognition,” PLoS ONE, vol. 13, no. 10, Oct. 2018, Art. no. e0200649.

[26] E. Rahimian, S. Zabihi, A. Asif, D. Farina, S. F. Atashzar, and A.Mohammadi, “FS-HGR: Few-shot learning for hand gesture recognition via electromyography,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, pp. 1004–1015, 2021.

[27] 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.
Ayush Tripathi received the B.Tech. degree in electrical engineering from the Visvesvaraya National Institute of Technology, Nagpur, India, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, IIT Delhi, New Delhi, India. During the B.Tech. degree, he had worked as a Research Aide with Arizona State University, Tempe, USA, and as a Research Intern with IIT Guwahati, Guwahati, India. He was a Systems Engineer with the TCS Research and Innovation Labs Mumbai, Mumbai, India, until September 2020. His research interests include biomedical signal processing, human–computer interface, and machine learning. Mr. Tripathi is a recipient of Prime Minister’s Research Fellows (PMRF) Scheme, Ministry of Education, Government of India.

Prathosh A. P. received the Ph.D. degree in temporal data analysis from the Indian Institute of Science (IISc), Bengaluru, India, in 2015. He submitted the Ph.D. degree thesis three years after the B.Tech. degree in 2011, with many top-tier journal publications. He worked with the corporate research labs including Xerox Research, Philips research, Bangalore, India, and a startup in CA, USA. His work in the industry, focusing on healthcare analytics, led to the generation of several IP. He joined IIT Delhi in 2017 as an Assistant Professor with the Computer Technology Group. He is currently a Faculty Member with the Department of ECE, IISc, Bengaluru. His current research interests include deep-representational learning, cross-domain generalization, signal processing, and their applications in computer vision and speech analytics.

Suriya Prakash Muthukrishnan received the M.D. degree in physiology from the All India Institute of Medical Sciences (AIIMS), New Delhi, India, in 2014. He is currently a Faculty of Physiology, AIIMS. His research interests include investigating the neurophysiological mechanisms underlying visuospatial working memory and object recognition using functional and effective EEG connectivity of brain networks, identifying cost-effective biomarkers for diagnosis and prognosis of brain disorders, and devising novel treatment strategies using neuromodulation and neurofeedback in neurodegenerative and neuropsychiatric disorders.

Lalan Kumar (Member, IEEE) received the B.Tech. degree in electronics engineering from IIT BHU, Varanasi, India, in 2008, and the Ph.D. degree from IIT Kanpur, Kanpur, India, in 2015. He has worked with IIT Bhubaneshwar, Bhubaneshwar, India, as an Assistant Professor and a Research Fellow with Nanyang Technological University, Singapore. He has also worked with Motorola, Bengaluru, India, as a Software Engineer with a multimedia team from 2008 to 2009. He is currently an Assistant Professor with the Department of Electrical Engineering, IIT Delhi, New Delhi, India. His current research interests include array signal processing, brain–computer interface (BCI) for soft-exosuit, and (sub)cortical source localization.