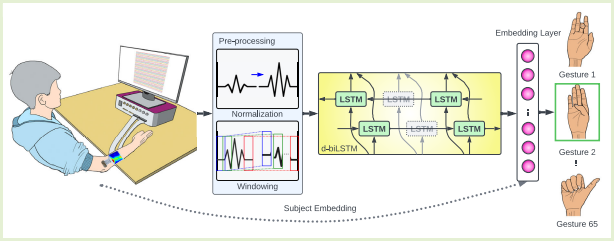
A Deep Learning Sequential Decoder for Transient High-Density Electromyography in Hand Gesture Recognition Using Subject-Embedded Transfer Learning

Golara Ahmadi Azar, Qin Hu, *Graduate Student Member, IEEE*, Melika Emami, Alyson Fletcher, Sundeep Rangan, *Fellow, IEEE*, and S. Farokh Atashzar, *Senior Member, IEEE*

***Abstract*—Hand gesture recognition (HGR) has gained significant attention due to the increasing use of AI-powered human–computer interfaces (HCIs) that can interpret the deep spatiotemporal dynamics of biosignals from the peripheral nervous system, such as surface electromyography (sEMG). These interfaces have a range of applications, including the control of extended reality, agile prosthetics, and exoskeletons. However, the natural variability of sEMG among individuals has led researchers to focus on subject-specific solutions. Deep learning methods, which often have complex structures, are particularly data-hungry and can be time-consuming to train, making them less practical for subject-specific applications. The main contribution of this article is to propose and develop a generalizable, sequential decoder of transient high-density sEMG (HD-sEMG) that achieves 73% average accuracy on 65 gestures for partially-observed subjects through subject-embedded transfer learning (TL), leveraging pre-knowledge of HGR acquired during pretraining. The use of transient HD-sEMG before gesture stabilization allows us to predict gestures with the ultimate goal of counterbalancing system control delays. The results show that the proposed generalized models significantly outperform subject-specific approaches, especially when the training data is limited and there is a significant number of gesture classes. By building on pre-knowledge and incorporating a multiplicative subject-embedded structure, our method comparatively achieves more than 13% average accuracy across partially-observed subjects with minimal data availability. This work highlights the potential of HD-sEMG and demonstrates the benefits of modeling common patterns across users to reduce the need for large amounts of data for new users, enhancing practicality.**



***Index Terms*— Gesture recognition, high-density EMG, human–computer interface (HCI), transfer learning (TL).**

I. Introduction

THE increasing use of Internet of Things (IoT) and investment in commercial augmented and virtual reality (AR/VR) applications suggest a growing demand for human–computer interfaces (HCIs) [1], [2]. This demand is further highlighted by the growing population of people with disability and amputees in the United States [3] that underlines the importance of neurorobotic systems (e.g., exoskeletons and prosthetics) equipped with HCI.

Surface electromyography (sEMG) has been commonly used to register the activation of the peripheral nervous system and as part of noninvasive neural interfaces [4], [5], [6], [7], [8], [9], [10], [11], [12]. High-density surface EMG (HD-sEMG) is a variant of noninvasive sEMG, collected through arrays of densely located electrodes to provide a more detailed scan of the propagation of neural drive over space and a high-resolution representation of the muscle activity [11], [13], [14], [15], [16]. The aforementioned modalities have a wide range of applications, including gesture classification and tracking in HCI [6], [7], [8], [9], [10], [17], and beyond, such as the assessment of muscle function [18], the diagnosis of neuromuscular disorders [5], and the evaluation of muscle fatigue [4], [19]. Machine learning and deep learning (DL) have enabled the development of robust decoders for sEMG

Manuscript received 20 January 2024; revised 5 March 2024; accepted 6 March 2024. Date of publication 20 March 2024; date of current version 1 May 2024. This work was supported by the U.S. National Science Foundation under Grant 2121391 and Grant 2229697. The associate editor coordinating the review of this article and approving it for publication was Prof. Yuedong Xie. *(Golara Ahmadi Azar and Qin Hu are co-first authors.) (Corresponding author: S. Farokh Atashzar.)*

Please see the Acknowledgment section of this article for the author affiliations.

Digital Object Identifier 10.1109/JSEN.2024.3377247

and HD-sEMG to detect the intended motor commands of users in various contexts. As described in Section II, recent advances in DL have obtained remarkably accurate decoding of gestures from sEMG signals [11], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37].

Despite the progress, a remaining key challenge of using DL for decoding sEMG is that sEMG signals vary significantly among various subjects [9], [38]. Thus, individual subject characteristics such as age, muscle composition, skinfold thickness, and gesture styles and habits can all influence the mapping between the sEMG space and intended motions [6], [7], [8], [9]. This variability has made it unlikely for a single model to accurately predict a high number of gestures across multiple subjects without retraining and fine-tuning [7]. For example, [20] showed that a convolutional neural network (CNN) used to decode 15 hand gestures achieved a high validation accuracy of 91.26% in a subject-specific experiment but a low test accuracy of 48.40% when tested on unseen subjects.

The poor performance on unseen subjects has necessitated the development of subject-specific models. As a result, a complex DL model needs to be fully retrained for each new subject using sizable labeled data from that individual. This exhaustive training and tedious data collection process raise questions about such DL techniques in terms of effectiveness and translation beyond research labs [9]. The aforementioned challenge has motivated the research on potential generalizable approaches to decode sEMG. Early attempts at generalization (see [6], [7], [9], [36], [37], [39], [40], [41] described in Section II), mostly using multichannel bipolar sEMG, have demonstrated successes only on tasks with limited number of gestures (in the range of five to 18 gestures) and limited gesture complexity (focusing on gestures with highly distinguishable patterns). In contrast, this work will attempt generalization for decoding 65 gestures, including complex and similar motions. We will investigate the power of HD-sEMG with 128 channels that can capture muscle activity with high spatiotemporal resolution. This is done with the goal of detecting underlying patterns of muscle activation propagated in the space and time that can possibly be used for generalization.

A. Contributions of This Work

The goal of this study is to push the boundaries of generalizable and agile gesture decoding by attempting to achieve high classification accuracy for a large number of gestures across different subjects via minimal HD-sEMG data while focusing on the transient phase of gesture conduction (to reduce decoding latency in the resulting HCI). We propose a dilated bidirectional long short-term memory (d-biLSTM) model that combines the advantages of temporal dilation and a bidirectional structure. At root, our model is designed with the goal of overcoming the decoding complexity and inherent variability in sEMG signals among subjects. Addressing this problem can significantly impact clinical and practical applications of sEMG in HCI. For this purpose, our training approach (i.e., subject-embedded transfer learning (TL) for gesture prediction using transient HD-sEMG) is composed of two phases: 1) training the base model to capture the common neurophysiological patterns of gesture performance from a limited number of subjects through "common parameters" and 2) retraining the model on a new subject with limited available HD-sEMG data (referred to as "partially-observed subject" in the rest of the article) to find the subject-specific projection through common and "subject-specific" parameters. In this work, the mapping of each subject index to the subject-specific projection is considered an embedding. In the first phase, the base model is trained with data from few subjects (referred to as pretraining subjects). Both common and subject-specific parameters are trained from scratch in this phase. During the retraining phase on a new subject, the common parameters learned in the first phase are used as the initial condition, and an embedding vector for the new subject is initialized by the average of embedding vectors corresponding to pretraining subjects. Unlike traditional TL, the proposed method enables subject-specificity in the pretraining set by incorporating multiplicative subject embedding. We demonstrate that the proposed method has several significant advantages over pure subject-specific models and previous traditional TL. The main contribution of this article is the development of a generalizable, lightweight sequential decoder that can achieve 73% average accuracy on 65 gestures using only the transient phase of high-density sEMG (HD-sEMG) of partially-observed subjects through subject-embedded TL. This contribution is described in an itemized format as follows.

1) *Generalization With HD-sEMG Signals and Large Numbers of Gestures:* As described in Section II, earlier efforts on generalizing to new subjects have mostly demonstrated success with limited number of gestures. In contrast, this article presents generalization performance on 65 gestures. This is motivated as we use HD-sEMG which introduces higher information rate for finding common patterns. Most of the prior works (with the exception of [40]) attempted generalization only using sparsely located bipolar sEMG rather than HD-sEMG.
2) *Generalization With Transient-Phase HD-sEMG:* The prior work on the topic of generalization uses the plateau phase of sEMG during gesture conduction when the signal is mostly stable. This, in general, can introduce extra latency in the HCI system and inaccuracy during transition from one gesture to another. In this article, for the first time, we approach the more challenging problem of generalization on transient phase of HD-sEMG with the goal of predicting the upcoming gestures and reducing the latency in HCI.
3) *Generalization With Minimal New Data:* The proposed model reaches an average accuracy of 73% across partially-observed subjects when having access to a limited number of repetitions per gesture during the retraining phase. More specifically, in this study, the challenging problem of *single-repetition* decoding has been addressed, requiring retraining our model for each subject using one repetition of data. This means that we can achieve 13% more accuracy compared to the state-of-the-art subject-specific counterpart while having access to only 25% of data.

4) *Lightweight Bidirectional LSTM:* In addition to the multiplicative embedding, the proposed model is compact with 79 K trainable parameters (e.g., compared to 1.7 million parameters of CapsNet [41]). It should be noted that for cloud computing, the compactness of the model structure saves substantial computational resources for rapid upgrades of the models on the cloud and individual devices. Moreover, model compactness enhances the practicality in terms of the implementation of portable hardware, such as HCI controllers.

The results of this article support the hypothesis that subject-embedded TL can indeed improve the HGR accuracy on new subjects with limited calibration. We have observed that the proposed generalized model consistently and significantly outperforms both purely subject-specific models as well as traditional TL-based models for all rates of data availability. The accuracy improvement gap is particularly large when sufficient training data (i.e., sufficient repetitions for each gesture) from the partially-observed subject is not available.

The remainder of this article is organized as follows. Section II reviews the prior studies considering the problem of HGR from sEMG signals, including subject-specific models and attempts toward generalization to new subjects. In Section III, our proposed subject-embedded TL strategy and the intuition it relies on are explained. Section IV introduces the dataset and pre-processing scheme used in this study. Section V provides an overview of the proposed model architecture. In Section VI, we describe the training and retraining configurations, the experiments conducted, and the corresponding results. Section VII is dedicated to comparing our proposed model to other LSTM-based architectures and analyzing the benefits of the proposed embedding-based generalization approach. We also evaluate our method on the steady-state (i.e., plateau) phase of HD-sEMG signals. Finally, in Section VIII, we summarize our observations and draw conclusions.

II. Prior Works: From Subject-Specific to Generalization

A. Subject-Specific Models

Subject-specific models for sEMG are trained for each individual and require complete retraining before being used on a new subject. Previous subject-specific studies have primarily used feature extraction methods and traditional machine learning techniques. These are just a few examples of the classic efforts in this field [42]. Linear discriminant analysis (LDA) [10], [43], Gaussian Naive Bayes [44], clustering-based algorithms [45], decision trees, and hidden Markov models [46] have been used to decode hand gestures from sEMG signals with high accuracies. sEMG signals are complex and variable, with a non-stationary and nonlinear relationship to muscle contractions [47]. These characteristics make it difficult to model the relationship between the signal and the gesture explicitly, especially for a large number of gestures. To address these issues (complexity, variability, and non-stationary nature of sEMG signals), DL algorithms have been increasingly used to decode sEMG signals into gesture classes [21], [22]. In the following, we provide some examples of the use of DL for this purpose.

CNNs [23], [24], [25], [26] and CNN-inspired architectures such as temporal convolutional networks (TCNs) [30], [31], compact CNNs (EMGNet) [28], 3-D CNNs [27], and dilated CNNs [29] have been studied extensively toward myoelectric control and pattern recognition in the past few years. These models have reached high accuracies (e.g., 97%) depending on the number of gestures during subject-specific studies. Selection of the convolution kernels, number of model layers, and dilation order are examples of model parameters in CNNs. Long short-term memory (LSTM) networks have demonstrated satisfactory performances as well. Such architectures have been significantly improved by introducing temporal dilation [11], [32], reaching an accuracy of 83% for decoding 65 gestures from HD-sEMG signals. Number of layers, dimension of hidden units, and dilation order are among the model parameters for LSTM-based structures. Hybrid architectures, combining CNNs and LSTMs, are shown to be accurate as well [33]. DL algorithms require a large amount of labeled data for training in order to achieve satisfactory accuracy, and obtaining such data can be impractical in many cases [6]. Transformer-based and few shot learning (FSL)-based frameworks have been proposed to address the elongated training time and the limited data availability problems of DL models, respectively [34], [35], [36], [37].

B. Efforts at Generalization and Their Limitations

Several works have attempted to develop generalized models, but remain limited in various aspects. Matsubara and Morimoto [39] propose a bilinear model that can detect five gestures with an accuracy of 73% by an adaptation process. In addition, in [7], an unsupervised domain adaptation (UDA) is proposed to classify six gestures with an average accuracy of 90.41%. TL can be used to improve model performance in a target domain through knowledge from the source domain. It is especially useful in the HGR context due to the subject variability of sEMG signals mentioned earlier. Yu et al. [40] propose a TL strategy with majority voting that reaches an average accuracy of 95.97% for 12 basic finger movements in CapgMyo-DBc (a HD-sEMG database [48]). These three studies [7], [39], [40] only consider a few number of gestures. The effect of TL on improving the accuracy of a convolutional network architecture to detect 18 gestures for new subjects is studied in [6]. They report an accuracy of 68.98% when given four repetitions of new data. Besides the limited number of gestures, more repetitions are required to calibrate this model for new subjects, in comparison to our proposed model. The authors of [41] propose the dilated efficient capsular neural network (CapsNet) that can predict 17 gestures from the transient phase of sEMG signals with an accuracy of 78.3%. The disadvantages of this model include the large number of trainable parameters and low number of gestures. Shi et al. [9] report both subject-specific and generalized (inter-subject) accuracy for static and dynamic gestures. Shi et al. [9] have introduced a CNN model named the multitask dual-stream supervised domain adaptation network (MDSDA) that exhibits long-term robustness and adaptability

to multiple subjects, reporting an inter-subject accuracy of 97.2% in detecting ten gestures. The low number of gestures and high complexity of the model are limitations of the study in [9]. Rahimian et al. [36], [37] use FSL to improve accuracy on new subjects. They report accuracies in the range of 76.39%–81.29% based on different architectures they use for five-way five-shot experiments. Few number of gestures (five-way) and requirement for more repetitions (five-shot) of new data are the limitations of these studies. In a very recent study [49], domain generalization and UDA were integrated into a single framework that detects seven gestures from HD-sEMG signals. Similarly, the low number of gestures is the limitation of this study besides the low-complexity of the targeted gestures. Table I summarizes these studies and highlights the existing research gaps and limitations. Based on the comprehensive literature review conducted in this article, it can be observed that the major research gaps include the lack of studies focusing on generalization over subjects for a large number of complicated gestures and securing reliability of prediction when enough calibration data is not available (all using compact DL techniques). These issues have been addressed in this study.

III. Subject-Embedded Transfer Learning

We briefly describe the general principle of the proposed subject-embedded TL and how it contrasts to other methods. Consider a general problem of predicting some target y from an input x. In the sEMG problem, y will be the gesture index and x will be an array representing the multichannel data collected over some time interval. Let u denote a subject index. One simple predictor would be of the form

$$\widehat{y} = f\,(x, \theta) \tag{1}$$

where $\widehat{y}$ is the prediction of the target y, and $f(x, \theta)$ is a function with parameters θ. For example, $f(x, \theta)$ could be a neural network with input x and θ would be weights and biases. By a **common model**, we mean that we learn a single common parameter θ for all subjects u. The obvious drawback with a common model is that it cannot capture subject-specific characteristics of the mapping. The other extreme would be a **subject-specific model** where one set of parameters θ is learned for each subject u. As mentioned in Section I, the challenge of subject-specific models is that they require significant training data for each subject.

One approach to reduce the data required for subject-specific models is to use what we will call **standard TL**. In this method, one typically first selects one or more *pretraining subjects* and learns a *common base model* $\widehat{y} = f(x, \theta^0)$ for these pretraining subjects where θ^0 represents the *base parameters*. Then, given a new subject u, the parameters θ^0 are finely tuned to obtain a new subject-specific parameter $\theta(u)$. The simplest method is to divide the parameters into components $\theta = (\theta_1, \theta_2)$. For example, θ_1 are the weight and biases for the initial layers, and θ_2 are the parameters for the final [generally fully connected (FC)] layers. In the pretraining phase, we learn base parameters $\theta^0 = (\theta_1^0, \theta_2^0)$. For the new subject, we freeze θ_1^0 and only learn a subject-specific component, $\theta_2(u)$, thereby reducing the parameters to be learned. The problem in this method, is that the base model is not subject-specific, and therefore, may not be able to provide a good fit over a large pretraining set.

For the proposed subject-embedded TL, we similarly divide the parameters into two components, $\theta = (\theta_1, \theta_2(u))$. In the pretraining phase, we learn a parameter $\theta^0 = (\theta_1^0, \theta_2(u))$, where the first component, θ_1^0, is common to all pretraining subjects. However, unlike standard TL, the second parameter, $\theta_2(u)$, is dependent on the subject index u within the pretraining set. The mapping of the subject index u to the parameters $\theta_2(u)$ can thus be seen as an embedding of the subject in some parameter space. This embedding enables the base model to have a subject-specific component.

For a new subject, u', not in the pretraining set, we run traditional gradient-descent learning on the data from a new subject where: 1) we initialize the first component $\theta_1 = \theta_1^0$, the common parameters in the base model and 2) we initialize the second component, $\theta_2(u')$ to the average of $\theta_2(u)$ for u in the pretraining set. The initialization of $\theta_1 = \theta_1^0$ implicitly captures the common aspects of the model from the pretraining set, while the search over $\theta_2(u')$ helps capture the subject-specific characteristics of the new subject.

IV. Database

A. Data Acquisition

Our work aims to develop a robust, multifunctional HCI control system capable of supporting a diverse range of control tasks through HD-sEMG data. To this end, the study uses a publicly available open-source HD-sEMG database containing 65 isometric hand gestures [47]. HD-sEMG data provide rich spatiotemporal information about underlying muscle activity and are particularly useful in recognizing a large number of gestures. The database includes 16 gestures with one degree of freedom (DoF), 41 gestures with two DoFs, and eight gestures with three DoFs, encompassing a range of finger and wrist movements such as bending, stretching, rotating, grasping, pointing, and pinching. The three DoF gestures are essential to maintain an ordinary daily life, whereas the one and two DoF gestures are the basic components for more complicated gestures. Fig. 1 shows two example gestures with their corresponding muscle-activity heatmaps. The signals were collected by a Quattrocento (OT Bioelettronica, Turin, Italy) biomedical amplifier through two 8×8 electrode grids (128 sensors in total) positioned on the volar and dorsal aspects of the forearm at a sampling rate of 2048 Hz. Please note that Fig. 2 is taken from our experimental setup for which we re-created the electrode placement similar to that of the dataset [47] used in this article. There might be some misplacement. The purpose of Fig. 2 is only for visualization. This database was collected from 20 non-disabled subjects (14 men and six women with average age of 35) who were instructed to perform each gesture for five repetitions, each lasting 5 s, with a 5-s inter-repetition rest period. The plateau phase of the repetitions is often used in gesture recognition due to the stability of muscle contraction during gesture maintenance, introducing control delay in practical applications. However, this study focuses only on transient-phase signals, which include the

TABLE I
COMPARISON BETWEEN THE PROPOSED MODEL WITH THE STATE-OF-THE-ART EFFORTS IN sEMG-BASED SUBJECT GENERALIZATION FOR HGR

Paper	# Sub.	# Ges.	Window Length	Signal Type	Method	Accuracy %	Limiations
[39]	11 (LOO)	5	128 ms	N/A	bilinear model	73	low # gestures, feature extraction
[7]	9 (LOO)	6	256 ms	plateau	UDA	90.41	low # gestures, need for reliable data pairs
[40]	10 (LOO)	12	timestamp	N/A	TL + majority voting	95.97	low # gestures
[6]	10 (LOO)	18	260 ms	N/A	CWT+TL	68.98	low # gestures, high # repetitions
[41]	40	17	300 ms	transient	TL	78.3	low # gestures, high model complexity
[9]	12 (LOO)	10	200 ms	N/A	domain adaptation	97.2	low # gestures, high model complexity
[36], [37]	8*	5	200 ms	N/A	Few shot learning	76.39-81.29	low # gestures, high # repetitions
[49]	8 (LOO)	7	256 ms	N/A	domain generalization + UDA	95.71	low # gestures and complexity of gestures, feature extraction
Our work	15*	65	200 ms	transient	subject-embedded TL	73.22	Requirement for pre-training subjects

Note: #: Number; ms: Millisecond; Sub.: Subjects; Ges: Gestures; LOO: Leave-one-out strategy for selecting training and testing subjects; N/A: unspecified; UDA: Unsupervised domain adaptation; TL: Transfer Learning; timestamp: inputs are vectors including all channels, not windows of time series; CWT: continuous wavelet transform; ∗: we refer to testing subjects only.

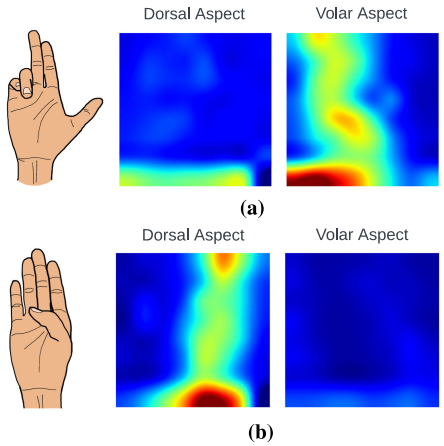


Fig. 1. Two example gestures with corresponding heatmaps that are the root mean square of a 200 ms window. (a) Little and ring fingers bend. (b) All fingers extension (without thumb).

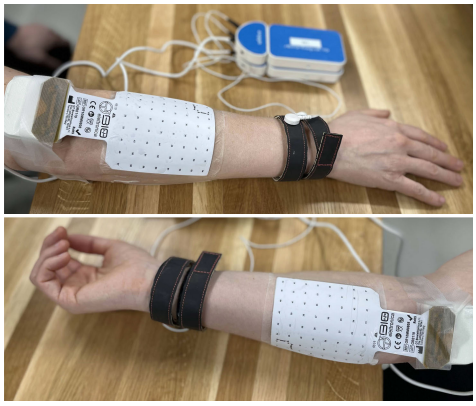


Fig. 2. Placement of two 8 × 8 electrode grids, with one on the dorsal aspect (outer forearm) and the other on the volar (inner forearm) aspect of the forearm [50].

most dynamic muscle activity, to design an agile HCI control system that can begin recognizing gestures as soon as a user initiates one. HGR on transient data transforms gesture detection into gesture prediction, minimizing control delay.

B. Data Pre-Processing

The raw HD-sEMG signals from the two 8 × 8 electrode grids are flattened and concatenated to form 128-channel signals suitable to our proposed sequential model. The total data can then be represented as a tensor, $x[n, i, t]$, where n is the sample index, i is the channel, and t is the time index within the sample. Each sample is the data from one repetition, so we will use the terms repetition and sample interchangeably. Magnitudes of muscle signals such as sEMG vary according to muscle type, length, and velocity [51]. sEMG can benefit from normalization in the sense that the signals collected from each electrode can contribute equally during model training [52]. We normalize the raw signals using z-score transformation via the means and standard deviations from only the training data. The standardization method converts the raw signals to a common scale such that the standardized signals have a zero mean and unit standard deviation. We then obtain scaled data

$$v[n, i, t] = \frac{x[n, i, t] - \mu[i]}{\sigma[i]} \tag{2}$$

where

$$\mu[i] := \frac{\sum_{n \in N_{\text{tr}}, t} x[n, i, t]}{N} \tag{3a}$$

$$\sigma[i] := \sqrt{\frac{\sum_{n \in N_{\text{tr}}, t} (x[n, i, t] - \mu[i])^2}{N}}. \tag{3b}$$

Following the previous work [11], we define the duration of the transient phase as the first 0.5 s of each repetition, according to the average force signals of each gesture. Windowing is a commonly used data augmentation technique for model performance and generalization enhancement in sEMG-based gesture detection. A repetition of the normalized signals is segmented into multiple overlapping windows before being fed into the proposed model. Implementing overlapping windows is a standard convention for sEMG-based HGR. It should be noted that the stride should be larger than the idle time needed by the processor for signal conditioning (e.g., normalization) and model inference [53]. Regarding the method implemented in this article for testing the accuracy of the system, it should be noted that overlapping windows do not cause any data leakage since the testing data and training data are separated by different gesture repetitions. We use a window size of 200 ms and a stride of 10 ms to meet the requirement of real-time control based on the standards followed by the literature [54], [55], [56], [57]. Based on our measurements, the average over 15 subjects for combined signal conditioning and inference time is 1.6 ms on CPU (NYU Greene HPC; 24-core Intel Cascade Lake Platinum 8268 chips) which allows for short stride length of 10 ms used in this article. This means that our proposed model has the potential to be employed for real-time inference, assuming that it has been retrained on a few samples for a new subject. The data pre-processing can be visualized as the upper box in Fig. 3.

In the pretraining phase, the train-test division is determined by assigning repetitions 1, 3, and 5 to the training set, and the remaining repetitions (2 and 4) to the testing set. To evaluate the capability of the generalized models on partially-observed subjects given different data availability in retraining, we calibrate the generalized models on any selection of one (33%), two (67%), and all three (100%) of the training repetitions. The train-test split for subject-specific model training follows the setups in the retraining phase.

Remark 1: Please note that this article proposes a generalized hand gesture recognition (HGR) system that requires minimum data from new subjects by acknowledging the between-subject variabilities and in combination with challenges imposed due to the loosely controlled data acquisition environment. It should be mentioned that variations between data from different subjects may be due to differences in neurophysiological characteristics, experimental variations, and environmental variations, which are all common issues challenging the re-utilization of inference models on new subjects. This is the focus of the article where we propose a deep-learning model and training approach to target these issues. However, specifically targeting the problem of electrode misplacement and displacement within one session or between different sessions is out of the scope of the current study. Readers may refer to our recent works [50], [58] for more information regarding electrode misplacement and displacement challenges.

V. Model Architecture

We propose the d-biLSTM model that combines the advantages of temporal dilation and a bidirectional structure. The model consists of three components: a three-layer d-biLSTM, a classifier with FC layers and dropout, and an embedding layer that captures subject dependencies in the generalized model. We will now provide a brief overview of these components. Table II summarizes all the model parameters and their corresponding values.

1) biLSTM: The exploding and vanishing gradient issue in recurrent neural networks (RNNs) has been well-studied in the literature and is often addressed through the use of LSTM cells [59], [60]. LSTM introduces additional gating mechanisms that enable the model to selectively retain or forget information, allowing it to better capture long-range dependencies in the input data. Inspired by the work in [32], we introduce temporal dilation in the LSTM architecture to further improve its ability to capture long-term dependencies in the input data. Temporal dilation allows for an expansion of the temporal receptive field (in the time series) without increasing the number of parameters and indeed, reducing the computational cost, making it an effective method for capturing longer-range dependencies and complex input sequences. In addition, in order to fully utilize all of the past and future information available within a specific signal window, we utilize a bidirectional LSTM (biLSTM) structure instead of a standard LSTM. A biLSTM processes the input sequence in both the forward and backward directions, allowing it to integrate contextual information from both past and future time steps within the processing window. Our experiments demonstrate that the biLSTM can achieve comparable accuracy to the LSTM while requiring fewer trainable parameters. The model consists of three d-biLSTM layers, each containing 32 hidden units and dilated with a factor of three, such that the next d-biLSTM layer has (1/8) connected LSTM cells compared to the current layer (details of homogeneous temporal dilation can be referred to [32]). The combination of temporal dilation and bidirectional processing enables the d-biLSTM model to effectively learn and classify the complex sequential sEMG data. Each d-biLSTM layer has a set of forward and backward outputs which are added before being passed as the input to the next layer. The final forward and backward hidden states are concatenated (yielding a 64-dimensional vector) before being fed into the classifier module.

2) Embedding: The weights of an embedding layer create a matrix that serves as an encoder of subject-specific information, resembling a lookup mechanism. The dimension of the embedding matrix can be adjusted based on the number of subjects and the specific model architecture in which it will be used. When given a subject index, a row from the embedding matrix corresponding to that specific subject is extracted and used in the model. In this study, we use a multiplicative embedding structure, where the extracted row is multiplied with the output of the first FC layer in the classifier module. The embedding rows have a dimension of 32 to match the structure of the classifier. This allows the model to effectively capture subject dependencies in the input data and improve performance on the classification task.

3) Classifier: As part of the classifier, first, an FC layer with the hyperbolic tangent activation function is used to decrease

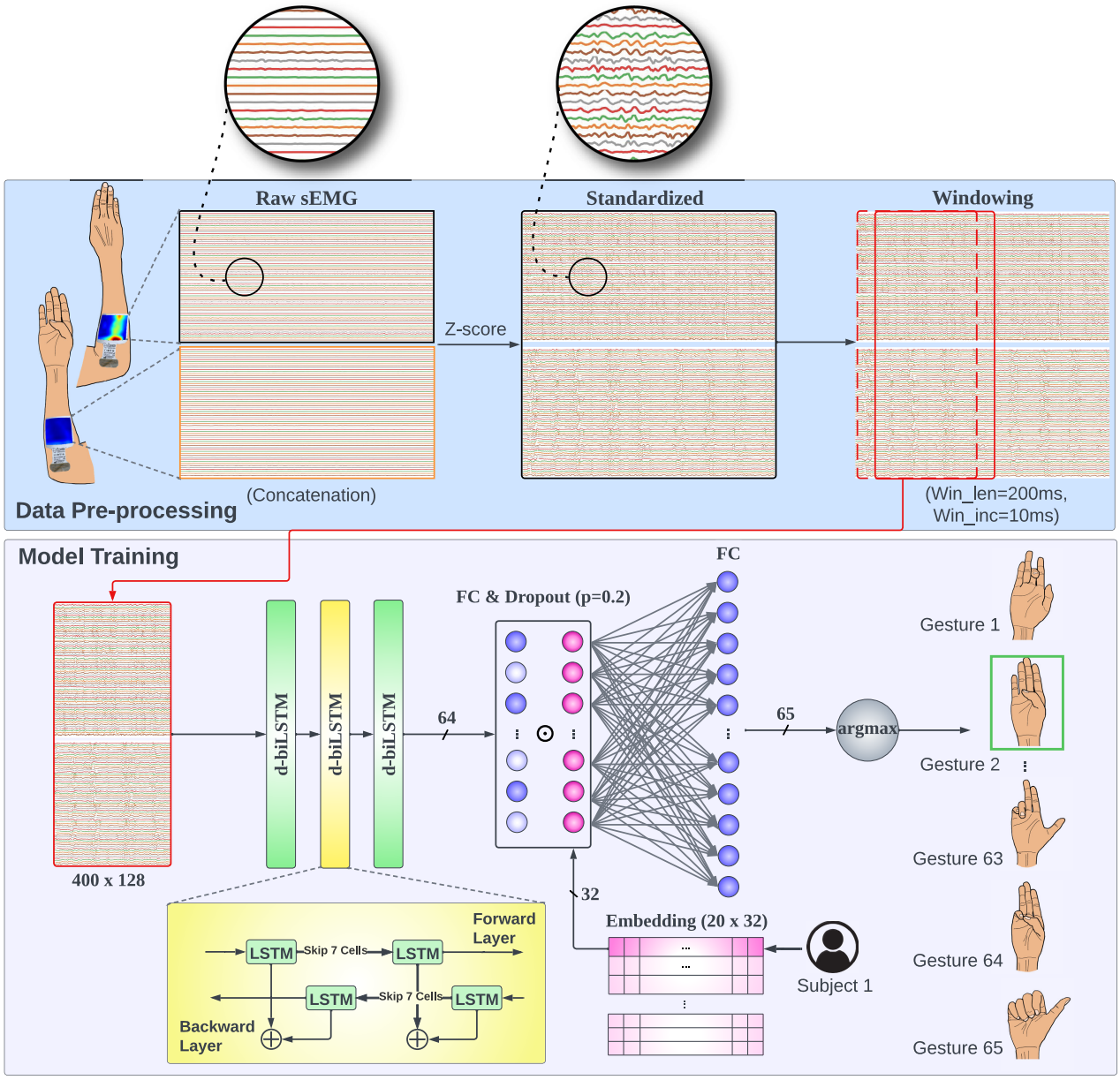


Fig. 3. Figure illustrates the proposed pipeline for gesture prediction using HD-sEMG data. The blue area represents the data pre-processing steps, including sensor flattening and windowing. The purple area shows the proposed d-biLSTM model, which takes each window of normalized HD-sEMG data as input and predicts the 65 gestures according to the maximum predicted probability output from the softmax function. The generalized model (shown in the figure) includes a subject embedding layer (shown in magenta), which captures subject-specific dependencies in the data, while the subject-specific counterpart excludes this layer. This allows the model to effectively learn and classify sequential data for a wide range of subjects.

the output dimension of the d-biLSTM module from 64 to 32. It is followed by a dropout layer with a rate of 0.2 to avoid overfitting the training data. The resulting vector is multiplied by the embedded vector extracted from the embedding module, given the subject index. A final FC layer with the softmax activation function is used to assign probabilities to various gesture classes. The predicted label corresponds to the gesture with the highest output probability.

In this study, two-phase generalization is conducted on the model with the embedding layer (named generalized model), whereas for the conventional subject-specific model that is only trained in one phase, no embedding layer is considered. Fig. 3 demonstrates the data pre-processing and model training pipelines. The lower box (i.e., model training) shows the proposed model structure.

VI. Experiments and Results

In this section, experimental results are reported regarding various phases of training for the proposed generalized models in comparison with the performance of the corresponding subject-specific models. Also, additional results will be reported regarding the effect of data reduction and task complexity (i.e., number of gestures) on the performance. It will be shown that

1) The proposed generalized model outperforms its subject-specific counterpart more significantly when pre-

TABLE II
PROPOSED MODEL PARAMETERS. PLEASE NOTE THAT #, FC, AND TANH REFER TO NUMBER, FC LAYER, AND HYPERBOLIC TANGENT, RESPECTIVELY

Component	parameter	value
biLSTM	layers	3
biLSTM	hidden units	32
biLSTM	dilation order	3
Embedding	dimension	32
Embedding	type	multiplicative
First FC	activation	tanh
First FC	output shape	32
Dropout	rate	0.2
Second FC	activation	Softmax
Second FC	output shape	# gestures

TABLE III
COMPARISON OF THE NUMBER OF TRAINABLE PARAMETERS AND TRAINING EPOCHS WHEN PREDICTING 65 GESTURES

	Subj. training	Gen. pre-training	Gen. retraining
Parameters	78,721	78,881	78,753
Epochs	200	200	100

dicting a higher number of gestures with fewer available data, benefiting from the proposed TL which includes the multiplicative embedding layer through weight initialization in retraining.

2) The pretrained weights represent the HGR pre-knowledge captured from known subjects, resulting in faster convergence for a new subject (100 epochs compared to 200) and reducing the chances of ending up at local minima.

In all experiments, Adam optimizer with learning rate $1e-3$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1e-08$ is used. The proposed generalized model is pretrained for 200 epochs with patience 40 on a number of subjects (referred to as the pretraining subjects). It is later retrained for only 100 epochs on a new subject (i.e., partially-observed subject). Moreover, the embedding vector of the partially-observed subject is initialized as the average of vector values for the pretraining subjects before the retraining phase. Categorical cross-entropy is used as the loss function and validation categorical accuracy is monitored. The number of training parameters is about 79 000 for the generalized and subject-specific models, depending on the number of gestures. Table III summarizes the number of parameters and training epochs for two phases of the generalized model and one phase of subject-specific model training when detecting 65 gestures. The subject-specific model in each gesture detection task is trained from scratch for 200 epochs.

A. Pretraining

As an initial step, the proposed generalized model needs to be pretrained on a number of subjects. At this phase, we have access to sufficient recordings of gestures from multiple subjects. To determine the optimal number of subjects to use in the pretraining phase of the generalized model, we conduct a series of experiments in which the model is pretrained on one to seven subjects for the task of detecting 65 gestures. We select a diverse group of subjects in terms of subject-specific accuracy to mimic the real-world scenario in which a variety of subjects might be chosen for pretraining. Table IV shows the generalized average accuracy of detecting 65 gestures evaluated on the remaining 13 subjects on three levels of data availability (33%, 67%, and 100% which correspond to one, two, and three retraining repetitions per gesture, respectively). It can be seen that increasing the number of pretraining subjects improves the generalized accuracy, and thus, the generalized model consistently outperforms the subject-specific model, especially when insufficient data is available for a new subject. Fig. 4 shows the accuracy improvement of the proposed generalized model when compared with the subject-specific model with respect to the number of subjects used in the pretraining phase. It can be observed that the accuracy improvement from adding the sixth and seventh subjects is less significant. Given the limited number of subjects available in the dataset (20 subjects), we use a maximum of five subjects to pretrain the generalized model in all other experiments in order to have sufficient test subjects for statistical analysis. The pretraining step is performed in an offline manner and takes about 6 h on average over 30 225 sample windows. All experiments are conducted on CPU (NYU Greene HPC; 24-core Intel Cascade Lake Platinum 8268 chips).

B. Data Reduction Experiments

As mentioned earlier, a main challenge in developing PR models is the requirement of a large dataset to train complex models for each new subject. In this section, we analyze the effect of available training data on the final accuracy. This reduction is based on the available training repetitions and will be analyzed in the following setting.

1) The generalized model is adequately pretrained on five random subjects using all three repetitions {1,3,5}.
2) After proper initialization of the embedding vector corresponding to a new subject, the generalized model is retrained on a subset of {1,3,5} repetitions of that subject (i.e., the partially-observed subject). The subsets roughly measure to 100% (all three repetitions), 67% (two repetitions), and 33% (one repetition) of the data.
3) The subject-specific model is trained from scratch on the same subset as the retraining of the generalized model.
4) Final accuracies of both models are evaluated on repetitions {2,4} of that subject.

The results of these experiments for the task of predicting 65 gestures are demonstrated in the right-most column of Fig. 5. Noting that retraining the proposed generalized model converges faster than training a subject-specific model from scratch (Table III), we also observe that for a high number of gestures, without having access to sufficient samples from the new subject, the proposed model more significantly outperforms the subject-specific counterpart. In fact, given only one repetition of each of the 65 gestures for a new subject, our proposed model achieves a prediction accuracy that is about 13% higher than the subject-specific accuracy. To further emphasize the superiority of our model to the subject-specific

TABLE IV
COMPARISON OF AVERAGE ACCURACY FOR PARTIALLY-OBSERVED SUBJECTS (EVALUATED ON 13 SUBJECTS) FOR PREDICTING 65 GESTURES

Data access	Number of pre-training subjects for generalized model							Subject
(%)	1	2	3	4	5	6	7	specific
33	43.28	43.46	47.00	49.58	52.95	54.67	55.47	40.20
67	61.49	60.86	62.67	64.99	67.03	68.13	68.83	58.53
100	71.16	70.87	71.04	73.26	73.85	74.53	75.96	70.08

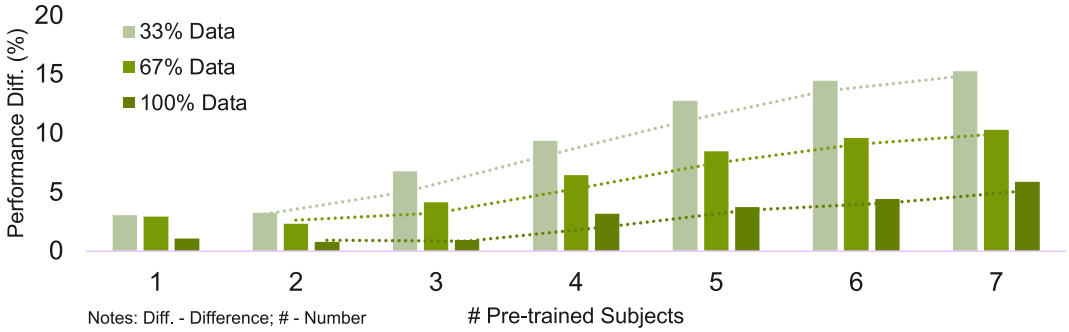


Fig. 4. Performance difference when given a varying number of pretraining subjects. The performance difference (*y*-axis) is the average accuracy difference between a generalized model and its subject-specific counterpart across partially-observed subjects based on the same percentage of retraining data. The lighter the color shade, the less the retraining data. The dotted lines indicate the moving averages as the trend of the ascending performance gap as more known subjects are included in pretraining.

TABLE V
COMPARISON OF RUN-TIME FOR TRAINING THE SUBJECT-SPECIFIC MODEL (SUBJ.) FROM SCRATCH AND RETRAINING OUR PROPOSED MODEL (GEN.) FOR THE TASK OF DETECTING 65 GESTURES. RUN-TIME IS MEASURED IN SECONDS (S). ALL EXPERIMENTS HAVE BEEN CONDUCTED ON CPU (NYU GREENE HPC; 24-CORE INTEL CASCADE LAKE PLATINUM 8268 CHIPS)

Data access	Run-time (s)	
(%)	subj.	gen.
33	1046.54	1038.66
67	1714.73	1684.83
100	2756.25	2139.56

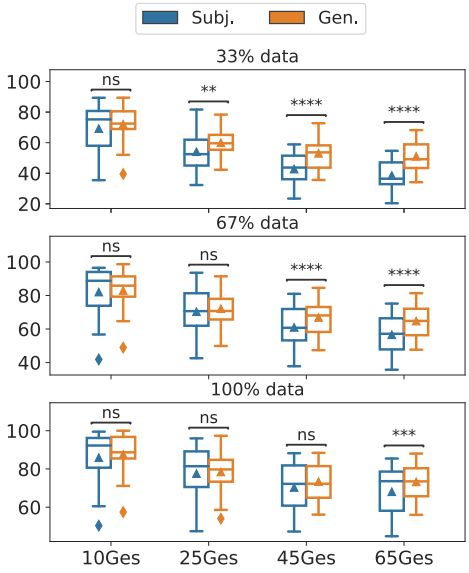


Fig. 5. Comparison of models during data reduction experiments when the generalized model is pretrained on five random subjects. The compared distribution pairs are dependent, as they are the accuracies of the same subjects derived from the generalized and subject-specific models. Also, the accuracy differences are not normally distributed. Thus, the Wilcoxon signed-rank test with $\alpha = 0.05$ is employed as the statistical hypothesis test. *X*- and *Y*-axes denote the number of gestures and accuracy, respectively. Each box plot includes 15 data points that represent 15 partially-observed subjects.

counterpart, we compare the average run-time of retraining the generalized model with the average run-time of training of the subject-specific model for the task of detecting 65 gestures. All experiments are conducted on CPU (NYU Greene HPC; 24-core Intel Cascade Lake Platinum 8268 chips) and we provide the results for various data availability scenarios. Table V shows this comparison.

C. Comparing to Traditional TL

In order to emphasize the advantage of using a subject-embedded structure, we also implement a traditional TL scheme for detecting 65 gestures by pretraining the model excluding the embedding layer on five random subjects (subjects 1, 6, 10, 11, and 14), and then after freezing all layers except the FC layers of the classifier, we retrain the model on various amounts of data from a new subject. The training configurations match those of the generalized model for a fair comparison. The results are reflected in the last column of Table VI. The results show that even by 100% retraining on a new subject, the traditional TL model has a lower accuracy (around 9% less) compared to the generalized model that is only retrained with 33% data.

D. Task Complexity Experiments

To investigate the effect of task complexity on model performance, we conduct experiments on the proposed generalized model on various numbers of gestures, ranging from 10 to 65.

TABLE VI

COMPARISON OF AVERAGE ACCURACY FOR PARTIALLY-OBSERVED SUBJECTS (EVALUATED ON 15 SUBJECTS) EVALUATED FOR TASKS OF DETECTING 10, 25, 45, AND 65 GESTURES. PRETRAINED ON FIVE RANDOM SUBJECTS. SUBJ., GEN., AND TL REFER TO SUBJECT-SPECIFIC, GENERALIZED, AND TRADITIONAL TL MODELS, RESPECTIVELY. THE COMPARISON WITH TL MODELS IS CONDUCTED ONLY WHEN CLASSIFYING 65 GESTURES

Data access	10 Gestures		25 Gestures		45 Gestures		65 Gestures		
(%)	subj.	gen.	subj.	gen.	subj.	gen.	subj.	gen.	TL
33	69.20	72.10	54.07	59.94	42.72	53.01	38.56	51.05	32.01
67	82.02	82.94	70.30	72.11	60.94	66.87	56.58	64.70	38.34
100	85.88	87.61	77.56	78.42	70.29	73.36	68.04	73.22	41.87

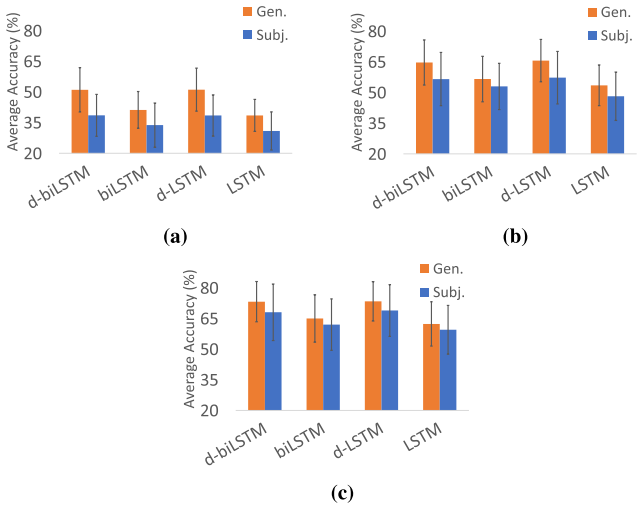


Fig. 6. Comparative results on 65 gestures when given different data availability. The error bars represent one standard deviation away from the mean values. (a) 33% Data. (b) 67% Data. (c) 100% Data.

For each experiment, five random subjects (subjects 1, 6, 10, 11, and 14) are selected for the pretraining phase and the remaining subjects are used for testing. Table VI shows the final test accuracy for different amounts of retraining data. It is important to note that the testing subjects for Table VI are different from those in Table IV, so the final average accuracy for 65 gestures is not directly comparable. The results in Table VI and Fig. 5 show that the performance gap between the generalized and subject-specific models increases as the task complexity increases. The largest gap was observed when the models are asked to predict 65 gestures using only 33% of data from a partially-observed subject. This suggests that the generalized model becomes more effective as the number of gestures to be detected increases. In this case, the subject-specific model would require more data for accurate classification. We would like to highlight the significance of the 73.2% performance over 65 gestures noting that the "chance threshold" (i.e., random guess baseline) for a problem of this scale is around 1.5%. This performance is also remarkable when comparing to various aspects of existing literature (e.g., number of subjects, gestures, window length, and signal phase outlined in Table I). In order to see which gestures are the most difficult to predict correctly, we analyze the confusion matrices for various subjects. Our observations show that gestures including multi-DoF movements are harder to detect. Specifically, multi-DoF gestures that include rotation of the wrist, moving the ring or little finger are more frequently confused with other gestures. This could be caused by the fact that many individuals cannot move specific fingers (e.g., the ring finger) in an isolated manner from other fingers, resulting in involuntary movement of other digits that can cause the model to classify the gestures incorrectly. A more detailed explanation of the factors affecting model performance is as follows.

1) **The number of gestures:** The probability threshold of the correct prediction of a random guess from a model decreases as the number of classified gestures increases. Additionally, the larger number of classified gestures usually include multi-DoF gestures (especially complex wrist gestures) that have higher requirements for gesture distinguishment on the proposed model. Compared to the most recent generalized HGR works in the literature, our proposed model classifies 3.6× to 13× more number of gestures.
2) **The number of subjects:** For "Generalized" HGR, the larger the number of involved subjects in the studies, the higher the generalizability and practicality of the proposed systems. However, the recognition accuracy of an unseen/partially unseen subject can be different from another due to factors like noise, even under the same experimental setup for data collection. The accuracy variance can increase as the number of involved subjects increases.
3) **Window length:** There is a trade-off between window length and model performance; generally, the longer the window length, the more information has been included in a single input, and the higher the recognition accuracy, but the slower and less intuitive/practical the control. In this article, we use a window size of 200 ms that meets the requirement of real-time control.
4) **The phase of input sEMG signals:** To achieve high performance in sEMG-based HGR, most of the existing works in the literature (either conventional subject-specific or generalized HGR) used steady-state-phase (or plateau-phase) sEMG signals as model inputs, which will introduce control delay. Also, separating the transition phase is not always practical. This article includes the transient-phase sEMG, which accounts for only 10% of the entire repetition of gesture performance, turning the gesture recognition task into gesture prediction and counterbalancing/reducing the control delay.

TABLE VII
PERFORMANCE COMPARISON BETWEEN GENERALIZED AND SUBJECT-SPECIFIC MODELS

Model	Average Accuracy (Generalized)	Average Accuracy (Subject-specific)
33% Data		
d-biLSTM (Ours)	51.05%	38.56%
Regular biLSTM	41.22%	33.79%
d-LSTM	51.14%	38.49%
Regular LSTM	38.52%	30.94%
67% Data		
d-biLSTM (Ours)	64.70%	56.58%
Regular biLSTM	56.62%	52.99%
d-LSTM	65.62%	57.29%
Regular LSTM	53.52%	48.21%
100% Data		
d-biLSTM	73.22%	68.04%
Regular biLSTM	65.00%	62.01%
d-LSTM	73.39%	68.89%
Regular LSTM	62.32%	59.47%

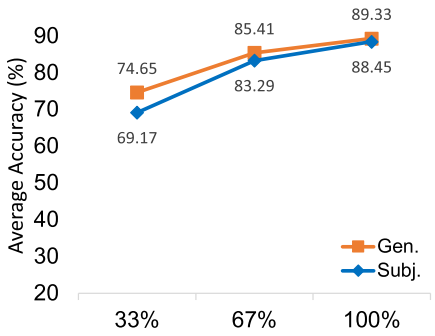


Fig. 7. Average results of d-biLSTM on plateau-phase signals given any data availability.

VII. COMPARATIVE STUDY

This work presents a novel method for transferring knowledge learned from known subjects to new ones using subject-embedded TL to the best of our knowledge. To demonstrate the superiority of our generalized d-biLSTM model over other state-of-the-art sequential DL models, we apply the same subject-embedded TL to a regular LSTM, a dilated LSTM (d-LSTM), and a regular biLSTM. These models are pretrained on the same five random subjects (subjects 1, 6, 10, 11, and 14) used by our proposed d-biLSTM. The comparative study is based on 33%, 67%, and 100% of retraining repetitions for all 65 gestures for the remaining 15 subjects. The results are shown in Fig. 6 and Table VII. It is worth noting that although the d-LSTM model performs similar to our proposed model, it requires an additional 40 960 trainable parameters (51.9% more complex than the proposed model) in both generalized and subject-specific models. In addition, we evaluate the subject-embedded TL technique of our proposed d-biLSTM on signals from the plateau phase (determined as 3 s after the transient phase). The results (shown in Fig. 7) demonstrate that our proposed method can be applied to signals from any phase of a repetition. Noting the higher accuracy when using plateau phase signals compared to transient phase, we would like to emphasize the significance of using transient phase signals for mitigating control delay and turning the gesture recognition task into gesture prediction.

The results demonstrate the impact and potential of our proposed subject-embedded TL approach for generalized muscle activity classification. Analyzing the results of our experiments reveals three main observations.

Observation 1: Dilated models, including the proposed d-biLSTM model, significantly outperform their non-dilated counterparts in terms of classification accuracy across all levels of training data availability. Moreover, the proposed generalized d-biLSTM model exhibits comparable accuracy to the generalized d-LSTM model while requiring fewer trainable parameters, making it a more efficient choice for classification.

Observation 2: The generalized models consistently outperform the subject-specific models in both transient and plateau phase experiments for all data availability conditions. This demonstrates the effectiveness of our proposed approach in transferring pre-learned knowledge from known subjects to classify muscle activity from new subjects.

Observation 3: The performance gap between the generalized d-biLSTM model and its subject-specific counterpart significantly increases with every 33% reduction in training data availability. This demonstrates the robustness and generalizability of the proposed approach, even in situations where limited data is available for training.

Overall, these findings suggest that the proposed subject-embedded TL approach using the d-biLSTM model is promising for accurate and efficient muscle activity classification in real-world scenarios involving partially-observed subjects with limited data availability.

VIII. CONCLUSION

In this study, we introduce a subject-embedded TL approach to mitigate the challenge of insufficient training data in DL-based HGR. Our proposed d-biLSTM model incorporates a multiplicative embedding layer that encodes subject-specific information, enabling the model to capture subject-specific neurophysiological features while learning HGR pre-knowledge from multiple subjects during pretraining. The resulting generalized models, retrained based on this pre-knowledge, demonstrate superior performance compared to subject-specific counterparts trained from scratch. This performance advantage is particularly evident when data availability is limited, or the number of gestures is large. Additionally, our approach uses transient-phase HD-sEMG signals, corresponding to muscle contraction prior to the maintenance of gestures, to minimize control delay in practical applications. It should also be noted that to the best of our knowledge, our proposed generalized model is the first that includes an embedding layer in a d-biLSTM structure for HGR. For future work, we aim to scale up the evaluation of our proposed HGR by testing our proposed model on multiple open-source databases combined with a separate dataset we will collect. This can help test the model on a larger number of subjects. Collecting and releasing data for such a combined evaluation is out of the scope of this current paper but an important future step of this work.

ACKNOWLEDGMENT

Golara Ahmadi Azar is with the Department of Electrical and Computer Engineering, University of California at Los Angeles (UCLA), Los Angeles, CA 90095 USA.

Qin Hu is with the Department of Electrical and Computer Engineering, New York University (NYU), New York, NY 11201 USA.

Melika Emami was with the Department of Electrical and Computer Engineering, University of California at Los Angeles (UCLA), Los Angeles, CA 90095 USA. He is now with Optum AI Labs, Eden Prairie, MN 55344 USA.

Alyson Fletcher is with the Department of Electrical and Computer Engineering and Department of Statistics, Mathematics, and Computer Science, University of California at Los Angeles (UCLA), Los Angeles, CA 90095 USA.

Sundeep Rangan is with the Department of Electrical and Computer Engineering, New York University (NYU), New York, NY 11201 USA, and also with NYU WIRELESS, New York, NY 11201 USA.

S. Farokh Atashzar is with the Department of Electrical and Computer Engineering, New York University (NYU), New York, NY 11201 USA, also with the Department of Mechanical and Aerospace Engineering, Biomedical Engineering, NYU WIRELESS, New York, NY 11201 USA, and also with the NYU Center for Urban Science and Progress (CUSP), New York, NY 11201 USA (e-mail: f.atashzar@nyu.edu).

REFERENCES

[1] L. S. Vailshery. (2018). *IoT Connections North America 2018, 2019 and 2025*. [Online]. Available: https://www.statista.com/statistics/933099/iot-connections-north-america/

[2] T. Alsop. (2024). *Investment in Augmented and Virtual Reality (AR/VR) Technology Worldwide in 2024, by Use Case*. [Online]. Available: https://www.statista.com/statistics/1098345/worldwide-ar-vr-investment-use-case/

[3] K. Ziegler-Graham, E. J. MacKenzie, P. L. Ephraim, T. G. Travison, and R. Brookmeyer, "Estimating the prevalence of limb loss in the United States: 2005 to 2050," *Arch. Phys. Med. Rehabil.*, vol. 89, no. 3, pp. 422–429, Mar. 2008.

[4] R. Merletti and D. Farina, Eds., *Surface Electromyography* (IEEE Press Series on Biomedical Engineering). Nashville, TN, USA: Wiley, Apr. 2016.

[5] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: Detection, processing, classification and applications," *Biol. Proced. Online*, vol. 8, no. 1, pp. 11–35, 2006.

[6] 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, Apr. 2019.

[7] X. Zhang, X. Zhang, L. Wu, C. Li, X. Chen, and X. Chen, "Domain adaptation with self-guided adaptive sampling strategy: Feature alignment for cross-user myoelectric pattern recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1374–1383, 2022.

[8] M. R. Ahsan, M. I. Ibrahimy, and O. O. Khalifa, "EMG signal classification for human computer interaction a review," *Eur. J. Sci. Res.*, vol. 33, pp. 480–501, Jan. 2009.

[9] P. Shi, X. Zhang, W. Li, and H. Yu, "Improving the robustness and adaptability of sEMG-based pattern recognition using deep domain adaptation," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 11, pp. 5450–5460, Nov. 2022.

[10] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau, "EMG feature evaluation for improving myoelectric pattern recognition robustness," *Exp. Syst. Appl.*, vol. 40, no. 12, pp. 4832–4840, Sep. 2013.

[11] T. Sun, Q. Hu, J. Libby, and S. F. Atashzar, "Deep heterogeneous dilation of LSTM for transient-phase gesture prediction through high-density electromyography: Towards application in neurorobotics," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 2851–2858, Apr. 2022.

[12] S. Padhy, "A tensor-based approach using multilinear SVD for hand gesture recognition from sEMG signals," *IEEE Sensors J.*, vol. 21, no. 5, pp. 6634–6642, Mar. 2021.

[13] 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.

[14] A. Bahador, M. Yousefi, M. Marashi, and O. Bahador, "High accurate lightweight deep learning method for gesture recognition based on surface electromyography," *Comput. Methods Programs Biomed.*, vol. 195, Oct. 2020, Art. no. 105643.

[15] A. Boschmann and M. Platzner, "Reducing classification accuracy degradation of pattern recognition based myoelectric control caused by electrode shift using a high density electrode array," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 4324–4327.

[16] A. Pradhan, J. He, and N. Jiang, "Performance optimization of surface electromyography based biometric sensing system for both verification and identification," *IEEE Sensors J.*, vol. 21, no. 19, pp. 21718–21729, Oct. 2021.

[17] X. Zhang, Z. Yang, T. Chen, D. Chen, and M. Huang, "Cooperative sensing and wearable computing for sequential hand gesture recognition," *IEEE Sensors J.*, vol. 19, no. 14, pp. 5775–5783, Jul. 2019.

[18] L. Bi, A.-G. Feleke, and C. Guan, "A review on EMG-based motor intention prediction of continuous human upper limb motion for human-robot collaboration," *Biomed. Signal Process. Control*, vol. 51, pp. 113–127, May 2019.

[19] H. A. Yousif et al., "Assessment of muscles fatigue based on surface EMG signals using machine learning and statistical approaches: A review," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 705, no. 1, Nov. 2019, Art. no. 012010, doi: 10.1088/1757-899x/705/1/012010.

[20] M. Jafarzadeh, D. C. Hussey, and Y. Tadesse, "Deep learning approach to control of prosthetic hands with electromyography signals," in *Proc. IEEE Int. Symp. Meas. Control Robot. (ISMCR)*, Sep. 2019, pp. A1-4-1-A1-4-11.

[21] D. Xiong, D. Zhang, X. Zhao, and Y. Zhao, "Deep learning for EMG-based human-machine interaction: A review," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 3, pp. 512–533, Mar. 2021.

[22] 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, doi: 10.3389/fnins.2021.621885.

[23] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "Real-time, simultaneous myoelectric control using a convolutional neural network," *PLoS ONE*, vol. 13, no. 9, Sep. 2018, Art. no. e0203835.

[24] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "Regression convolutional neural network for improved simultaneous EMG control," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 036015.

[25] A. R. Asif et al., "Performance evaluation of convolutional neural network for hand gesture recognition using EMG," *Sensors*, vol. 20, no. 6, p. 1642, Mar. 2020.

[26] Z. Yang, A. B. Clark, D. Chappell, and N. Rojas, "Instinctive real-time sEMG-based control of prosthetic hand with reduced data acquisition and embedded deep learning training," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2022, pp. 5666–5672.

[27] J. Chen, S. Bi, G. Zhang, and G. Cao, "High-density surface EMG-based gesture recognition using a 3D convolutional neural network," *Sensors*, vol. 20, no. 4, p. 1201, Feb. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/4/1201

[28] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, "Hand gesture recognition using compact CNN via surface electromyography signals," *Sensors*, vol. 20, no. 3, p. 672, Jan. 2020.

[29] E. Rahimian, S. Zabihi, S. F. Atashzar, A. Asif, and A. Mohammadi, "Semg-based hand gesture recognition via dilated convolutional neural networks," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2019, pp. 1–5.

[30] J. L. Betthauser, J. T. Krall, R. R. Kaliki, M. S. Fifer, and N. V. Thakor, "Stable electromyographic sequence prediction during movement transitions using temporal convolutional networks," in *Proc. 9th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Mar. 2019, pp. 1046–1049.

[31] J. L. Betthauser et al., "Stable responsive EMG sequence prediction and adaptive reinforcement with temporal convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 6, pp. 1707–1717, Jun. 2020.

[32] T. Sun, Q. Hu, P. Gulati, and S. F. Atashzar, "Temporal dilation of deep LSTM for agile decoding of sEMG: Application in prediction of upper-limb motor intention in NeuroRobotics," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 6212–6219, Oct. 2021.

[33] E. Rahimian, S. Zabihi, A. Asif, and A. Mohammadi, "Hybrid deep neural networks for sparse surface EMG-based hand gesture recognition," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 371–374.

[34] M. Montazerin, S. Zabihi, E. Rahimian, A. Mohammadi, and F. Naderkhani, "ViT-HGR: Vision transformer-based hand gesture recognition from high density surface EMG signals," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 5115–5119.

[35] E. Rahimian, S. Zabihi, A. Asif, S. F. Atashzar, and A. Mohammadi, "Trustworthy adaptation with few-shot learning for hand gesture recognition," in *Proc. IEEE Int. Conf. Auto. Syst. (ICAS)*, Aug. 2021, pp. 1–5.

[36] 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.

[37] E. Rahimian, S. Zabihi, A. Asif, S. F. Atashzar, and A. Mohammadi, "Few-shot learning for decoding surface electromyography for hand gesture recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 1300–1304.

[38] H. Xu and A. Xiong, "Advances and disturbances in sEMG-based intentions and movements recognition: A review," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13019–13028, Jun. 2021.

[39] T. Matsubara and J. Morimoto, "Bilinear modeling of EMG signals to extract user-independent features for multiuser myoelectric interface," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2205–2213, Aug. 2013.

[40] Z. Yu, J. Zhao, Y. Wang, L. He, and S. Wang, "Surface EMG-based instantaneous hand gesture recognition using convolutional neural network with the transfer learning method," *Sensors*, vol. 21, no. 7, p. 2540, Apr. 2021.

[41] E. Tyacke et al., "Hand gesture recognition via transient sEMG using transfer learning of dilated efficient CapsNet: Towards generalization for neurorobotics," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 9216–9223, Oct. 2022.

[42] M. Sim ao, 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.

[43] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Exp. Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012.

[44] R. S. Byfield, R. Weng, M. Miller, Y. Xie, J.-W. Su, and J. Lin, "Real-time classification of hand motions using electromyography collected from minimal electrodes for robotic control," *Int. J. Robot. Control*, vol. 3, no. 1, p. 13, Aug. 2021.

[45] K. Momen, S. Krishnan, and T. Chau, "Real-time classification of forearm electromyographic signals corresponding to user-selected intentional movements for multifunction prosthesis control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 535–542, Dec. 2007.

[46] 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. Hum.*, vol. 41, no. 6, pp. 1064–1076, Nov. 2011.

[47] N. Malešević et al., "A database of high-density surface electromyogram signals comprising 65 isometric hand gestures," *Sci. Data*, vol. 8, no. 1, p. 63, Feb. 2021.

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

[49] X. Li, X. Zhang, X. Chen, X. Chen, and L. Zhang, "A unified user-generic framework for myoelectric pattern recognition: Mix-up and adversarial training for domain generalization and adaptation," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 8, pp. 2248–2257, Aug. 2023.

[50] T. Sun, J. Libby, J. Rizzo, and S. F. Atashzar, "Deep augmentation for electrode shift compensation in transient high-density sEMG: Towards application in neurorobotics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Kyoto, Japan, Dec. 2022, pp. 6148–6153, doi: 10.1109/IROS47612.2022.9981786.

[51] T. J. Roberts and A. M. Gabaldon, "Interpreting muscle function from EMG: Lessons learned from direct measurements of muscle force," *Integrative Comparative Biol.*, vol. 48, no. 2, pp. 312–320, Jun. 2008.

[52] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 105524.

[53] M. A. Oskoei and H. Hu, "Myoelectric control systems—A survey," *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275–294, Oct. 2007. [Online]. Available: https://api.semanticscholar.org/CorpusID:32104730

[54] H. F. Hassan, S. J. Abou-Loukh, and I. K. Ibraheem, "Teleoperated robotic arm movement using electromyography signal with wearable myo armband," *J. King Saud Univ. Eng. Sci.*, vol. 32, no. 6, pp. 378–387, Sep. 2020.

[55] W.-T. Shi, Z.-J. Lyu, S.-T. Tang, T.-L. Chia, and C.-Y. Yang, "A bionic hand controlled by hand gesture recognition based on surface EMG signals: A preliminary study," *Biocybern. Biomed. Eng.*, vol. 38, no. 1, pp. 126–135, 2018.

[56] A. Aranceta-Garza and B. A. Conway, "Differentiating variations in thumb position from recordings of the surface electromyogram in adults performing static grips, a proof of concept study," *Frontiers Bioeng. Biotechnol.*, vol. 7, p. 123, May 2019.

[57] M. F. Wahid, R. Tafreshi, and R. Langari, "A multi-window majority voting strategy to improve hand gesture recognition accuracies using electromyography signal," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 427–436, Feb. 2020.

[58] E. Tyacke, K. Gupta, J. Patel, R. Katoch, and S. F. Atashzar, "From unstable contacts to stable control: A deep learning paradigm for HD-sEMG in neurorobotics," New York Univ., Brooklyn, NY, USA, Tech. Rep., Sep. 2023.

[59] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[60] F. Quivira, T. Koike-Akino, Y. Wang, and D. Erdogmus, "Translating sEMG signals to continuous hand poses using recurrent neural networks," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Mar. 2018, pp. 166–169.

Golara Ahmadi Azar received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2020, and the M.Sc. degree in electrical and computer engineering from the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, in December 2022, where she is currently pursuing the Ph.D. degree with the Electrical and Computer Engineering Department.

Her research is focused on theoretical machine learning with a focus on asymptotic learning and generalization with applications in high-dimensional biosignal processing.

Qin Hu (Graduate Student Member, IEEE) received the M.Sc. degree in computer engineering from New York University (NYU), New York, in 2019, and the M.Sc. degree in quantitative methods and modeling from the City University of New York, New York, NY, USA, in 2020, where she is currently pursuing the Ph.D. degree with the Electrical and Computer Engineering Department, working with Medical Robotics and Interactive Intelligent Technologies Laboratory (MERIIT Lab).

Her research focuses on explainable, human-centered, and trustworthy AI in biosignal processing, specifically for neural interfaces.

Melika Emami received the B.Sc. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2016, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, in 2018 and 2022, respectively.

She is currently a Machine Learning Scientist with Optum AI Labs, Eden Prairie, MN, USA, working on advancing AI solutions that improve healthcare outcomes. Her earlier research has been on theoretical machine learning with a focus on the asymptotics for learning and generalization in neural networks.

Alyson Fletcher received the M.S. degree in electrical engineering from the University of Iowa, Iowa City, IA, USA, in May 2002, and the M.S. degree in mathematics and the Ph.D. degree in electrical engineering and computer sciences (EECS) from the University of California at Berkeley, Berkeley, CA, USA, in May 2005 and January 2006, respectively.

Since 2016, she has been on the faculty of the Departments of Statistics, Mathematics, EE, and Computer Science, University of California at Los Angeles, Los Angeles, CA, USA.

Sundeep Rangan (Fellow, IEEE) received the B.A.Sc. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in May 1992, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley, Berkeley, CA, USA, in 1995 and December 1997, respectively.

He has held postdoctoral appointments with the University of Michigan, Ann Arbor, MI, USA, and Bell Labs, Murray Hill, NJ, USA. In 2000, he Co-Founded (with four others) Flarion Technologies, Bedminster, NJ, USA, a spin-off of Bell Labs, that developed Flash OFDM, the first cellular OFDM data system and pre-cursor to 4G cellular systems including LTE and WiMAX. In 2006, Flarion was acquired by Qualcomm Technologies, San Diego, CA, USA. He was a Senior Director of Engineering with Qualcomm involved in OFDM infrastructure products. He joined NYU Tandon (formerly NYU Polytechnic), Brooklyn, NY, USA, in 2010, where he is currently a Professor of Electrical and Computer Engineering. He is the Associate Director with NYU WIRELESS, New York, NY, USA, an industry-academic research center on next-generation wireless systems.

S. Farokh Atashzar (Senior Member, IEEE) is currently an Assistant Professor with New York University (NYU), New York, NY, USA, jointly appointed with the Department of Electrical and Computer Engineering and the Department of Mechanical and Aerospace Engineering. He is also affiliated with the Department of Biomedical Engineering, NYU, also, NYU WIRELESS, New York, and NYU Center for Urban Science and Progress, New York. Before joining NYU, he was a Postdoctoral Scientist with Imperial College London, London, U.K. At NYU, he is the Director of the Medical Robotics and Interactive Intelligent Technologies (MERIIT) Laboratory. The Laboratory is mainly funded by the U.S. National Science Foundation. His research interests include a human–machine interface, human-centered robotics, neural interfacing, deep learning, and nonlinear control.

Prof. Atashzar was a recipient of several awards, including the 2021 Outstanding Associate Editor of IEEE ROBOTICS AND ROBOTICS. He is currently an Associate Editor for IEEE TRANSACTIONS ON ROBOTICS and IEEE ROBOTICS AND AUTOMATION LETTERS.