IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 September 2023, accepted 18 September 2023, date of publication 22 September 2023,
date of current version 27 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3317956

RESEARCH ARTICLE

Electromyography Based Gesture Decoding Employing Few-Shot Learning, Transfer Learning, and Training From Scratch

RICARDO V. GODOY[1], **(Graduate Student Member, IEEE),**
BONNIE GUAN[1], **(Graduate Student Member, IEEE), FELIPE SANCHES**[1], **ANANY DWIVEDI**[2],
AND MINAS LIAROKAPIS[1], **(Senior Member, IEEE)**
[1]New Dexterity Research Group, Department of Mechanical and Mechatronics Engineering, The University of Auckland, Auckland 1010, New Zealand
[2]Artificial Intelligence (AI) Institute, Division of Health, Engineering, Computing and Science, University of Waikato, Hamilton 3216, New Zealand

Corresponding author: Ricardo V. Godoy (rdeg264@aucklanduni.ac.nz)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Auckland Human Participants Ethics Committee (UAHPEC) under Reference No. 019043.

ABSTRACT Over the last decade several machine learning (ML) based data-driven approaches have been used for Electromyography (EMG) based control of prosthetic hands. However, the performance of EMG-based frameworks can be affected by: i) the onset of fatigue due to long data collection sessions, ii) musculoskeletal differences between individuals, and iii) sensor position drifting between different sessions with the same user. To evaluate these aspects, in this work, we compare the performance of EMG-based hand gesture decoding models developed using three approaches. This comparison allows for future works in EMG-based Human-Machine Interfaces development to make more informed ML decisions. First, we trained from scratch a Transformer-based architecture, called Temporal Multi-Channel Vision Transformer (TMC-ViT). For our second approach, we utilized a pre-trained and fine-tuned TMC-ViT model (a transfer learning approach). Finally, for our third approach, we developed a Prototypical Network (a few-shot learning approach). The models are trained in a subject-specific and subject-generic manner for eight subjects and validated employing the 10-fold cross-validation procedure. This study shows that training a deep learning decoding model from scratch in a subject-specific manner leads to higher decoding accuracies when a larger dataset is available. For smaller datasets, subject-generic models, or inter-session models, the few-shot learning approach produces more robust results with better performance, and is more suited to applications where long data collection scenarios are not possible, or where multiple users are intended for the interface. Our findings show that the few-shot learning approach can outperform training a model from scratch in different scenarios.

INDEX TERMS Electromyography, gesture decoding, deep learning, few-shot learning, transfer learning.

I. INTRODUCTION

The human hand is a powerful and dexterous end-effector, allowing humans to learn and explore their environment through complex interactions. It enables us to perform these interactions executing a wide range of tasks, from grasping and moving objects used in everyday life to executing various gestures in social settings [1]. Therefore, in case of a limb loss, people experience a tremendous loss of dexterity, which is detrimental to their quality of life [2]. According to [3], approximately 540,000 amputees suffer from upper limb loss in the US, while the numbers are expected to be doubled by 2050. Europe has approximately 4.66 million limb amputees, with up to 431,000 amputations performed each year [4]. Recent technological advancements have resulted in the development of prosthetic hands that are

The associate editor coordinating the review of this manuscript and approving it for publication was Shovan Barma.

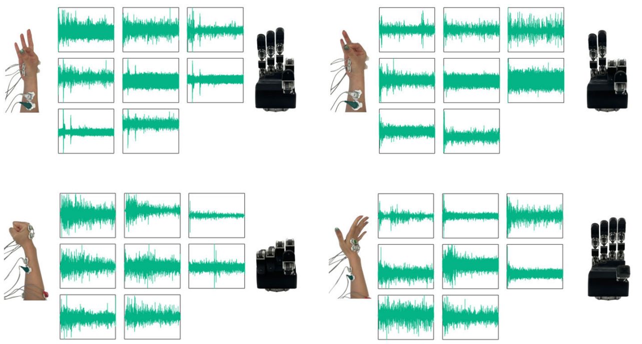

FIGURE 1. Myoelectric activations of different gestures. During the execution of the pinch, tripod, power, and extension gestures, the muscles produce differing muscle activations. Machine learning-based models can be developed to decode these signals for the control of a robot hand.

becoming increasingly dexterous for amputees. However, for a natural control of the prosthetic hand, the user's intention needs to be efficiently and accurately decoded. Therefore, there is a need for intuitive interfaces that facilitate accurate control of prosthetic devices.

For intuitively controlling such devices, researchers have proposed using biosignal-based human-machine interfaces (HMIs). Different sensing modalities can be used to develop HMIs. Some frequently used modalities include: electromyography (EMG) [5], [6], [7], ultrasonography [8], and mechanomyography [9]. However, the most common biosignals-based HMIs are EMG-based, as they are easy to use due to their non-invasive nature and high temporal resolution [10], [11]. Such EMG-based human-machine interfaces (HMI) can be developed using analytic or machine learning methods. The analytic methods simulate the activation of the muscles using models of the physical characteristics of the muscles [12]. These models are complex to develop as they depend on parameters such as the muscle fibre length and muscle contraction velocity that vary for different muscle types and individuals.

To overcome the issues with the analytic models, machine learning-based data-driven approaches have been employed. Machine learning methods are a powerful tool for analyzing and decoding EMG signals, as well as for applications in various fields, such as computer vision, natural language processing, and robotics [13], [14], [15], [16], [17], [18]. Machine learning methods facilitate the classification of hand gestures and movement, as well as the continuous decoding of dexterous and complex motions [19], [20]. Previous works have used EMG-based interfaces for teleoperating robotic arm-hand systems [5], [21], rehabilitation using robotic exoskeletons [22], entertainment (myo-games) [23], and for developing muscle-computer interfaces [24].

However, EMG-based HMIs have limitations: myoelectric signals are affected by fatigue and depend on sensor placement, and they vary among individuals due to musculoskeletal differences. Consequently, studies often create subject-specific machine learning models through supervised learning with datasets containing multiple gesture repetitions. Even though creating subject-specific models can reach accuracies higher than 90% even when decoding several gestures [25], [26], long data collection sessions may be required each time a new decoding model needs to be developed (e.g., for a new user or for the same user when wearing the interface again after a break). Data management and a lack of data prove to be an existing challenge in human-computer interfacing [27]. As such, transfer learning approaches are a viable alternative for addressing this issue by using learned features from a pre-trained model trained on a larger dataset of other subjects to improve the performance of a model for a new user where the data for this new user/subject is scarce. Studies have shown that transfer learning approaches can improve EMG-based gesture decoding [28], [29], [30] using popular publicly available datasets such as the Ninapro dataset [31]. However, transfer learning approaches are prone to overfitting (especially when the target dataset is small), they rely on the quality and representativeness of the source dataset, and they can lead to extra computational complexity (since these models are usually deeper) or even be more time-consuming (due to the inherent pre-training process on a large dataset).

Another machine learning approach to avoid long data collection sessions is to use few-shot learning techniques. These methods can use only a few samples to generalize the predictions to new unseen data. Studies have used few-shot learning techniques to produce inter and intra-session models to decode hand gestures using EMG signals as input, using techniques such as siamese networks or meta-learning approaches [32], [33]. Siamese networks may require more training time and extra model complexity due to the duplicate nature of such architectures, shared weights, and simultaneous training compounded by the need to learn quadratic pairs. Conversely, training most gradient-based meta-learning approaches can also be challenging due to the presence of two levels of training, leading to increased complexity and resource requirements in hyper-parameter search, in which the several hyper-parameters need to be carefully tuned for optimal performance. Prototypical networks can offer a more efficient alternative to meta-learning algorithms. This approach is significantly simpler and more efficient than some of the meta-learning approaches, yet it can achieve state-of-the-art results [34].

In this paper, we compare the performance of three different training approaches for developing EMG-based gesture decoding models, allowing for future works in EMG-based HMI development to make more informed decisions in implementing machine learning frameworks to decode the EMG signals. In our first approach, we develop a Transformer-based architecture called

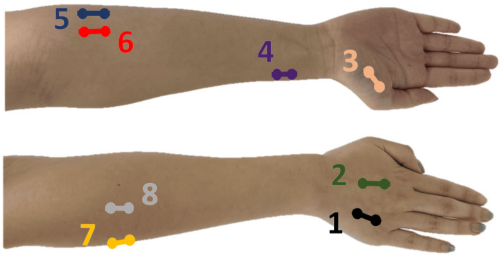

FIGURE 2. Electrodes' placement on the forearm of the participants. Eight bipolar channels are used to collect the EMG signals of the user in order to train the machine learning models employed to perform the EMG-based gesture decoding.

Temporal Multi-Channel Vision Transformer (TMC-ViT) and train it from scratch. The second approach uses a pre-trained and fine-tuned TMC-ViT model, and lastly, a Prototypical Network is developed. The performance of these models is evaluated in different scenarios for decoding four hand gestures, shown in Fig. 1, plus the rest gesture (making a total of 5 classes) using myoelectric activations from the human forearm. The machine learning models are trained in a subject-specific and subject-generic manner, and also in an inter and intra-session manner. We also assess the performance of the three approaches for decoding samples of fatigued EMG signals through datasets with an increased number of gesture repetitions. In the case of smaller datasets, subject-generic models, or inter-session models, the few-shot learning method yields more robust outcomes with enhanced performance. This approach is particularly well-suited for situations where extended data collection is impractical or when multiple users are anticipated to use the HMI. This paper demonstrates that, in different scenarios, few-shot learning can surpass the effectiveness of training a model from scratch.

The rest of the paper is organized as follows. Section II presents the dataset collected for this study, how the signals are processed, and the classification models developed in this paper, as well as how these models were trained and evaluated. Section III presents the results obtained in this paper, which are discussed in detail in Section IV. Finally, Section V concludes the paper and presents potential future directions.

II. METHODS

A. DATASET

The dataset used in this study was collected from eight non-disabled subjects. More information regarding the subjects can be found in Table 1. The study was approved by the University of Auckland Human Participants Ethics Committee (UAHPEC), reference number #019043. All experiments were performed in accordance with relevant guidelines and regulations. Prior to the study, participants provided written and informed consent to the experimental procedures.

TABLE 1. Information of the participants. *M* stands for male, *F* for female, *R* for right, and *L* for left.

Subj.	1	2	3	4	5	6	7	8
Age	25	24	28	23	23	29	34	33
Gender	M	M	M	F	M	F	M	M
Handedness	R	L	L	R	R	R	R	R

In this study, eight bipolar EMG electrodes were employed to measure the myoelectric activations and perform EMG data acquisition (see Fig. 2). An informed decision was made for the selection of the muscle sites for decoding hand gestures and motions based on our previous research [6], [19], [35], [36], [37]. Since the majority information in the EMG signals is contained within the frequency band of 0 Hz to 500 Hz [38], [39], the EMG signals were acquired with a sampling rate of 1,200 Hz using a g.tec's g.USBamp bioamplifier. The acquired data was bandpass filtered using a Butterworth filter (5 Hz - 500 Hz). The electric line noise was filtered out using a notch filter of 50 Hz. Each subject performed five gestures: pinch, tripod, power, finger extension, and rest. These gestures were selected based on the most common grasps identified by Bullock et al. [40]. For each gesture, the participant started with 10 seconds of rest, during which the hand was completely relaxed, followed by 10 seconds of gesture execution. Visual cues were provided to the participants as a three-second counter on a computer screen to switch between the gesture state and the rest state. This included a software trigger to label the two states for creating the ground truth data for supervised learning algorithms. This procedure was repeated nine more times for each gesture, resulting in ten repetitions in total.

A second session of EMG data collection was performed for the first five subjects based on availability, with the objective of analyzing the effects on the gesture decoding accuracies for a subject as a result of variations arising in the data from the same subject participating in the experiments on different days. To do this, the decoding models are trained on data from one session and tested on the data from another session. This session followed the same data collection procedure as the first one.

B. PREPROCESSING

In this subsection, we present the data preprocessing methods employed for the EMG-based gesture classification.

1) WINDOWS SIZE

To provide input to the supervised machine learning models during training, a sliding window with a duration of 200 ms and a step size of 20 ms was used. The choice of a larger window size (more than 125 ms) was aimed to prevent significant biases and variances [41]. However, considering the real-time constraints required for the efficient control of prostheses and robotic devices, the window size was kept smaller than 300 ms [42].

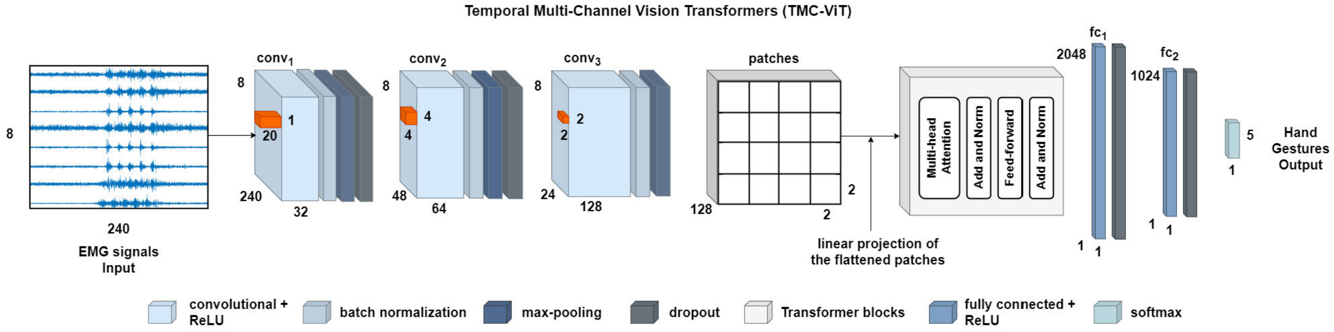

FIGURE 3. TMC-ViT model developed for both training from scratch and transfer learning. The EMG signals are 8 × 240 matrices, in which the lines are the eight electrode channels, and the columns are 240 time-steps. The two max-pooling layers reduce the input dimension while maintaining relevant input formation. The filters are shown in orange. A dropout of 0.2 is employed after each batch normalization layer. After the convolutional block, 2 × 2 patches are extracted from the convolutional blocks' output and provided to the transformer block.

2) DATA BALANCE

To address the prevalence of the rest class at the end of the data collection, a balancing approach was implemented. Since the rest class becomes predominant after five gestures, it was important to ensure equal representation of each gesture sample in order to avoid any class bias. To achieve this, the random undersampling technique was employed.

3) DATA TYPE

Minimal data preprocessing was employed for the EMG data to develop the deep learning models, eliminating the need for manual feature engineering. The utilized approach exploits the inherent capability of deep learning methods to automatically extract meaningful features from raw data. To enhance the learning capabilities of the models during the training process and to remove the requirement for normalization in the data preprocessing stage, batch normalization layers [43] were incorporated into the architectures. This integration not only accelerated the training speed, but also handled the normalization process within the models themselves.

C. TRAIN AND TEST SETS

In this study, we trained our models in both a subject-specific and subject-generic manner. For the subject-specific case, one model is trained for each subject. We trained each model two times, using data from the same session (intra-session) and from different sessions (inter-session).

1) SUBJECT-SPECIFIC SETS, INTRA-SESSION

When data from the same session is used, one out of ten gesture repetitions is used to test the model, while the others are used for training, leading to a 10-fold cross-validation. The subject-specific models for the same session were further evaluated in terms of the number of gesture repetitions used for training. The test set is comprised of the last two gesture repetitions in order to evaluate how the model would perform when being evaluated on the data where likely fatigue was present, representing the most challenging ones to be decoded. The root mean square (RMS) value and the median frequency (MDF) value of the first and last repetition were calculated in order to compare the fatigue between these gesture repetitions. The RMS value of the last repetition shows a percent increase of ∼ 29.8% across all the muscles compared to the first repetition, while MDF shows a percent decrease of ∼ 12.3%. An increase in the RMS value and a decrease in the MDF value indicates the onset of fatigue [44], showing that fatigue is more evident in the last gesture repetitions. Each model was trained for only the first repetition, then using only the first two gesture repetitions, and so on until eight gesture repetitions used for training were reached, while all being tested in the last two gesture repetitions.

2) SUBJECT-SPECIFIC SETS, INTER-SESSION

When data from a different session is used, the model is trained on all the gesture repetitions from the first session and validated in one repetition at a time from the second session, comprising the 10-fold cross-validation.

3) SUBJECT-GENERIC SETS

For the subject-generic case, a model is trained on the data from all other subjects except the testing subject, from which one repetition will be used per fold for testing.

D. CLASSIFICATION MODELS

1) TEMPORAL MULTI-CHANNEL VISION TRANSFORMERS

The Transformers networks [13] marked a significant advancement in natural language processing (NLP). Transformers are designed to process sequential data without suffering from vanishing gradients like the recurrent neural networks and the impossibility of parallelization inherent to these recurrent techniques. These architectures are based only on attention mechanisms, which creates an attention-based representation for each element in the input sequence. The attention mechanism used by Vaswani et al. [13] was the Scaled Dot-Product Attention, given by

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V, \qquad (1)$$

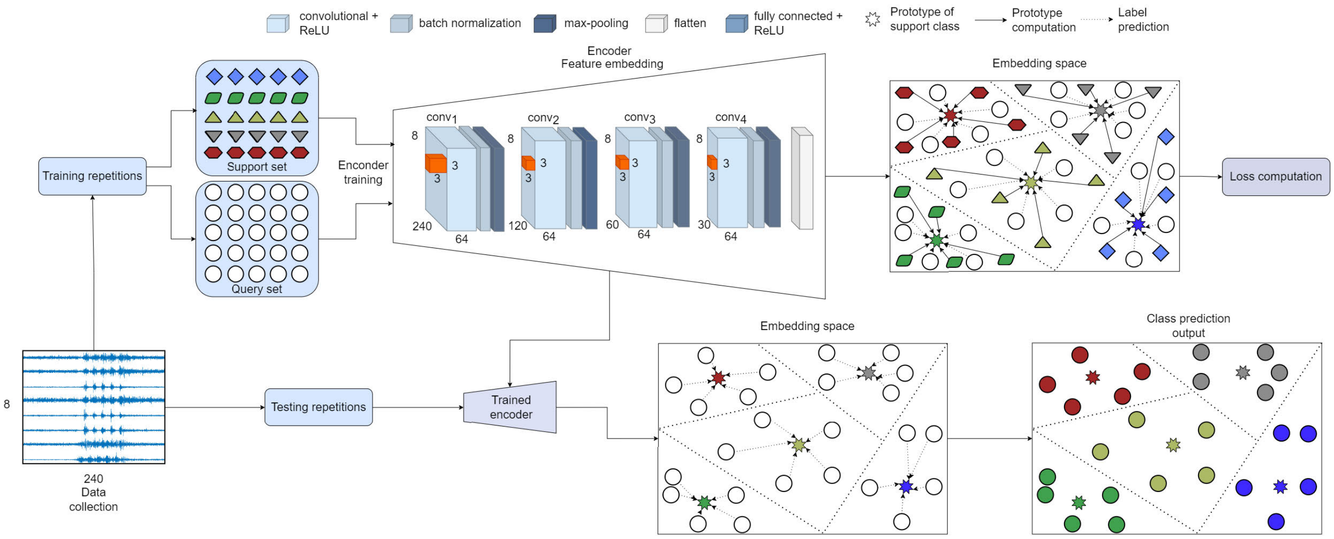

FIGURE 4. **Prototypical network for EMG-based gesture decoding. The support and query sets are provided to an encoder that embeds the input EMG data. Based on the distance between the embedded vector of queries and class prototypes, the model predicts the hand gesture type.**

where $\sqrt{d_k}$ is the so-called scale factor, and Q, K, and V are vectors called query, key, and value, respectively, that are going to be used inside attention layers in order to compute the attention value for each element.

Vaswani et al. employed attention in different positions of different representations of input subspaces through a mechanism called Multi-Head Attention, which allows parallel computation and calculates a richer representation of the input sequence. In the Multi-Head Attention, the same Q, K, and V vectors are multiplied by learned weight matrices. Hence, the attention is calculated for each head h, and the concatenation of these three values is multiplied by a matrix W_O to generate the output of the Multi-Head Attention, as follows

$$MultiHead(Q, K, V) = concat(head_1, \ldots, head_h)W_O$$
$$head_i = Attention(Q\ W_i^Q, K\ W_i^K, V\ W_i^V), \tag{2}$$

where W_i^Q, W_i^K e W_i^V are the learned weight matrices, one for each head.

Vision Transformer (ViT) [14] is a Transformer model adapted to use images as input. Thus, instead of processing 1D sequential data, ViT uses 2D images as input. In a first step, the ViT subdivides the input image $x \in \mathbb{R}^{H \times W \times C}$ into a sequence of flattened 2D patches $x_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$, where (H, W) is the resolution of the original image, C is the number of channels, (P, P) is the resolution of each image patch, and $N = HW/P^2$ is the resulting number of patches. A linear embedding sequence of these patches and position embeddings are then provided as input to a Transformer encoder. While the position embedding adds input topology information, the ViT processes the image with a linear projection of the flattened patches, whose components indicate low-dimensional correlations in the patches, and the Multi-Head Attention mechanism aggregates image information across all layers.

In this study, we employed a ViT adaptation called Temporal Multi-Channel Vision Transformer (TMC-ViT) [26], shown in Fig. 3. The TMC-ViT is a Transformer-based model that adapts the ViT to process temporal data with multiple channels, such as EMG signals. From our previous works, we found that TMC-ViT outperforms well-established deep learning techniques, such as CNNs, or classic machine learning techniques, such as Random Forest, in both classification and regression tasks [19], [20]. The decoding accuracy of the TMC-ViT is even higher when raw EMG data is used as input [26]. This model employs convolutional, batch normalization [43], dropout [45], and max-pooling layers to extract embeddings while reducing the input dimension and maintaining important information. Three convolutional blocks are used before the data is supplied to a ViT that extracts 2 × 2 patches and provides the output to a Transformer encoder composed of eight Multi-head Attention layers [13] with four heads each. After the Transformer blocks, two fully-connected layers with 2048 and 1024 neurons, respectively, followed by a softmax layer with five neurons, perform the gesture class prediction.

2) PROTOTYPICAL NETWORK

Prototypical Network [46] is a few-shot classification approach based on the concept that exists an embedding space where data points cluster around a central prototype representation for each class. A prototype is an M-dimensional representation $\mathbf{c}_k \in \mathbb{R}^M$ of each class k. To achieve this, a neural network-based encoder is employed to learn a

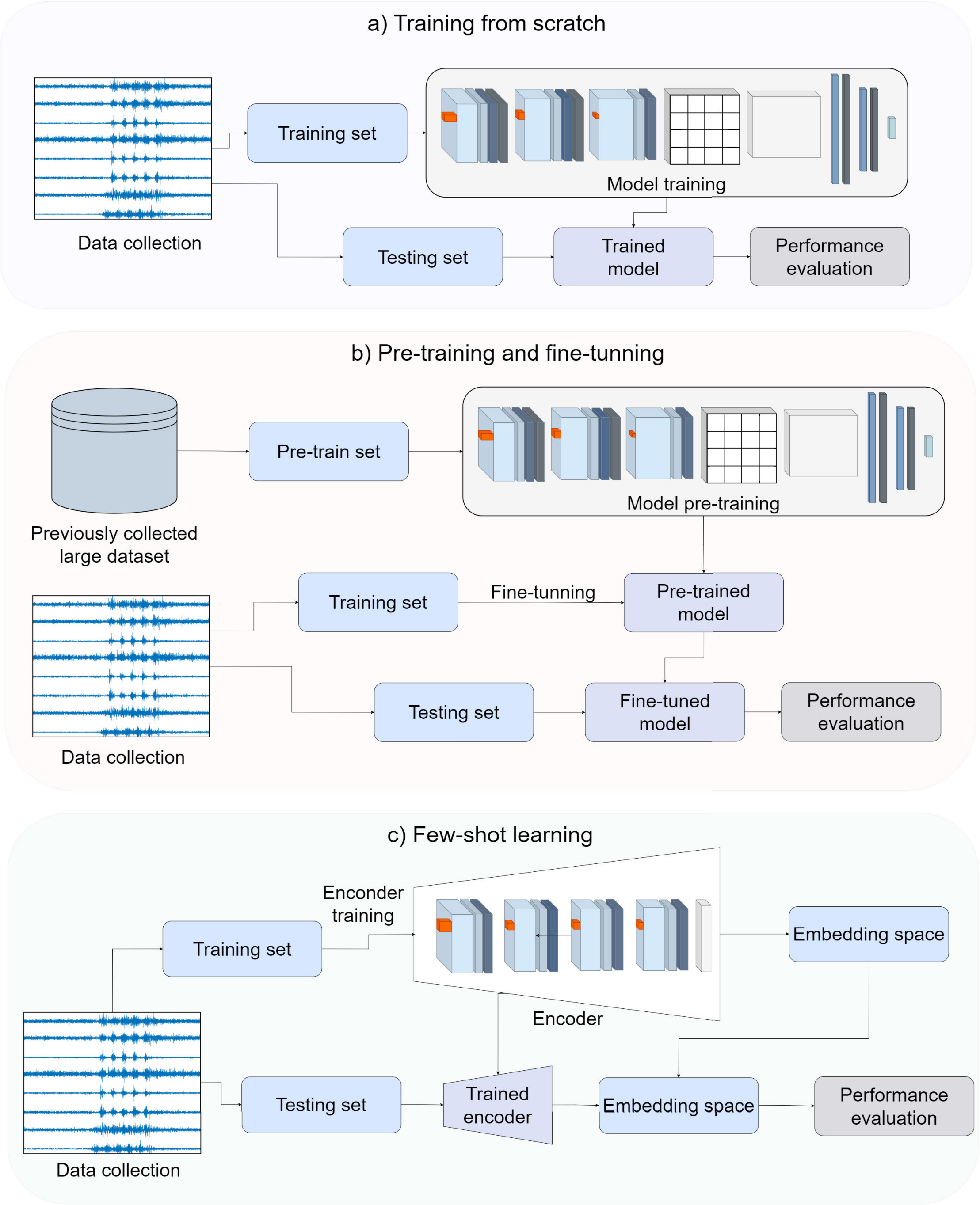

FIGURE 5. **Three learning approaches employed in this study. a) Training from scratch approach. For subject-specific models, one repetition is used for testing, while the others are used during the training phase. The model is trained on the training set and then evaluated on the test set for decoding accuracy. b) Transfer learning training approach. The model is first pre-trained on a larger and previously collected dataset. For the subject-specific model, the pre-train set comprises data from other subjects. After the pre-train step, the model is fine-tuned on nine gesture repetitions from a subject, and the performance of this fine-tuned model is evaluated on the test repetition. c) 5-way 5-shot learning training approach. Five samples for each class are randomly picked from the training set to comprise the support and query set. These sets are provided to an encoder that embeds the input data. The embedded vector of the support data is averaged to form the class prototype. The distance between queries and prototypes is calculated, and the loss is computed and backpropagated. The model is then evaluated on the testing set, also comprised of five samples per class for the support and query set.**

non-linear mapping from the input data to the embedding space through an embedding function $f_\phi : \mathbb{R}^D \rightarrow \mathbb{R}^M$ with learnable parameters ϕ, where D is the dimension of the feature vector of a labeled example in the few-shot problem. The prototype for a particular class is determined by calculating the mean of its support set within the embedding space:

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\phi(x_i), \tag{3}$$

where S_k denotes the set of examples with class k and $x_i \in \mathbb{R}^D$ is the D-dimensional feature vector of an example, and $y_i \in \{1, \ldots, K\}$ is the corresponding label.

Classification of an embedded query point is accomplished via a softmax over distances to class prototypes, i.e., by identifying the nearest class prototype in the embedding space by computing the distance between each unlabelled image and the prototype. So, given a distance function $d : \mathbb{R}^M \times \mathbb{R}^M \rightarrow [0, +\infty)$, the Prototypical Network produces a distribution over classes for a query point x based on a softmax over distances to the prototypes in the embedding space:

$$p_\phi(y = k \mid x) = \frac{exp(-d(f_\phi(x), c_k))}{\sum_{k'} exp(-d(f_\phi(x), c_{k'}))} \tag{4}$$

Learning is done by minimizing the negative log-probability $J(\phi) = -log\, p_\phi(y = k \mid x)$ of the true class k via SGD. The distance is computed using the Euclidean distance metric. This model can be used to generalize to classes not seen in the training set, given only a small number of samples of each new class. In our case, all the classes are known since the training phase, and only the generalization capabilities of the prototypical network will be exploited based on only a few samples of each class. The Prototypical Network employed in this paper for EMG-based hand gesture decoding is shown in Fig. 4.

E. TRAINING AND EVALUATION

Our models were developed in Python. Each model was trained and evaluated on the New Zealand eScience Infrastructure (NeSI) using NVIDIA HGX A100 80Â GB memory GPUs. Three different training approaches were employed in this study: 1) training a model from scratch, 2) using a larger dataset to pre-train and then fine-tune the model in a target dataset, and 3) using few-shot learning. These training approaches are shown in Fig. 5. For 1) and 3), we trained the models in both subject-specific and subject-generic ways. 2) was only trained for subject-specific, since in the case of subject-generic, 1) and 2) would result in the same model. The efficiency of all models was assessed using accuracy. The description of the three training approaches employed in this study is presented next.

1) TRAINING FROM SCRATCH

To train a full-scale TMC-ViT model from scratch, we employed early stopping on validation loss to determine the optimal number of epochs for training, using Adam as the optimizer [47]. During training, the loss function was sparse categorical cross-entropy. The procedure for TMC-ViT training from scratch is illustrated in Fig. 5 - 1).

2) TRANSFER LEARNING

Transfer learning involves utilizing knowledge acquired from solving one problem and applying it to a different but similar problem. In our specific scenario, we aim to use the features learned by a model that can interpret gestures performed by various users and employ those features to interpret EMG signals from a different individual. Typically, transfer learning is employed when the available dataset is insufficient to train a complete model from scratch for a given task or user. The typical approach to transfer learning in the context of deep learning involves the following steps. 1) loading a pre-trained model that has been trained on a larger dataset for a similar task. 2) preserving the valuable information contained within, either freezing the entire model or certain parts of it for subsequent training iterations. 3) adding new trainable layers on top of the frozen layers or retraining the last ones. These newly trained layers will learn to utilize the existing features to generate predictions on a new dataset. 4) training the new layers using a new specific dataset, allowing them to adapt and optimize their predictions based on the learned features. 5) a final and optional step known as fine-tuning can be employed. This involves unfreezing the entire model obtained thus far and re-training it using the new data, employing a very low learning rate (to avoid overfitting and/or losing all the learned information from the previous training). Fine-tuning has the potential to achieve significant enhancements by incrementally adjusting the pre-trained features to better align with the characteristics of the new data, i.e., the data from the new subject.

In our study, a TMC-ViT model is pre-trained on the data of the subjects, except the testing subject, using early stopping on validation loss and Adam as the optimizer. Again, the sparse categorical cross-entropy was employed. After the model is pre-trained and the base model is frozen, the last fully-connected layers are unfrozen and trained for 20 epochs on the training set for the testing subject. Then, the whole model is unfrozen for fine-tuning, trained for ten epochs, and learning rate of 10^{-5}. This procedure is shown in Fig. 5 - 2).

3) FEW-SHOT LEARNING

In the case of the few-shot classification, we have only a small labeled set of examples (support set) to predict classes for the unlabeled samples (query set), defining an N-way K-shot problem, where N stands for the number of classes, and K for the number of samples from each class. In this study, we performed a 5-way 5-shot by randomly selecting five samples for each of the five classes from the dataset to form the support set for the model. Subsequently, five samples are chosen from the same set of five classes to constitute the query set.

The prototypical models were trained via SGD with Adam, and followed the procedure in the original paper [46]. An initial learning rate of 10^{-1} was used and divided in half every 2000 episodes. The models were trained using Euclidean distance in a 5-shot scenario, with training

TABLE 2. Subject-specific models' performance trained on the same session, using three different learning approaches.

Fold	Subj 1	Subj 2	Subj 3	Subj 4	Subj 5	Subj 6	Subj 7	Subj 8
TMC-ViT trained from scratch								
0	95.10	98.10	76.10	97.60	98.60	95.70	80.80	96.60
1	95.10	95.50	73.50	98.60	96.20	98.40	97.30	97.00
2	98.80	96.80	92.60	98.60	99.20	95.60	95.90	98.80
3	97.90	98.60	89.90	97.10	99.00	97.30	97.00	97.60
4	98.20	98.10	85.20	97.80	99.00	96.60	98.30	97.70
5	98.40	98.60	95.20	97.70	98.90	97.90	97.70	94.40
6	97.70	98.60	94.40	98.50	98.90	98.50	97.60	85.10
7	98.80	98.00	92.10	98.10	98.50	97.70	97.90	95.70
8	98.30	98.20	95.90	97.60	97.40	98.50	96.10	94.90
9	86.70	100	75.00	93.30	93.30	100	80.00	92.00
AVG	**97.59**	**97.76**	**88.32**	**97.96**	**98.41**	**97.36**	**95.40**	**95.31**
SD	**3.71**	**1.19**	**8.91**	**1.55**	**1.86**	**1.37**	**7.14**	**3.98**
Pre-trained and fine-tuned TMC-ViT								
0	84.10	97.80	73.90	95.30	97.30	93.80	77.30	97.00
1	94.80	97.20	55.90	98.60	98.20	98.60	97.60	96.70
2	96.40	98.50	94.00	98.60	98.90	97.80	96.30	98.50
3	97.70	98.80	86.90	98.00	98.90	97.90	97.00	98.40
4	98.60	98.30	92.20	97.90	98.80	98.50	98.20	99.00
5	98.60	98.70	95.00	98.60	98.90	98.30	98.50	97.20
6	98.10	98.50	93.50	99.10	98.90	98.90	98.80	88.50
7	99.00	98.90	95.00	98.80	99.00	98.50	98.40	97.80
8	98.60	98.90	94.50	98.90	98.60	99.50	98.50	98.90
9	86.70	90.00	85.00	86.70	86.70	90.00	70.00	92.00
AVG	**95.26**	**97.56**	**86.59**	**97.05**	**97.42**	**97.18**	**93.06**	**96.40**
SD	**5.38**	**2.71**	**12.67**	**3.80**	**3.80**	**2.96**	**10.40**	**3.44**
Prototypical Network								
0	92.76	97.65	89.12	95.72	96.74	95.10	95.55	96.42
1	94.30	96.19	85.45	97.33	97.67	97.52	94.71	94.82
2	97.34	96.24	90.15	98.04	98.72	87.42	96.82	98.36
3	95.61	97.36	89.16	93.50	98.08	94.88	95.33	96.86
4	96.09	98.01	83.64	96.80	97.68	95.43	97.83	97.38
5	97.88	98.15	92.48	97.81	97.82	95.84	94.85	91.45
6	97.54	96.44	92.35	97.97	95.43	95.40	97.37	86.43
7	96.90	96.34	89.92	96.59	96.26	96.39	96.85	94.70
8	97.70	97.34	81.97	94.31	94.92	97.84	92.35	96.07
9	93.00	96.00	86.00	93.00	94.00	95.00	94.00	91.00
AVG	**95.91**	**97.15**	**88.02**	**96.11**	**96.73**	**95.08**	**95.57**	**94.35**
SD	**1.94**	**0.81**	**3.59**	**1.89**	**1.54**	**2.88**	**1.69**	**3.67**

episodes containing five classes and query points per class. The loss was the negative log-probability. The 5-way 5-shot procedure we followed is shown in Fig. 5 - 3).

III. RESULTS

In this section, we present the results obtained by the models training on the subject-specific and subject-generic sets using the three different training approaches. The tables presented in this section show the decoding accuracy achieved by the models in percentages. *Subj.* stands for tested subject, *AVG* for average, and *SD* for standard deviation.

TABLE 3. Subject-specific models' performance trained on the same session, with an increasing number of gesture repetitions used for training. *rep* stands for gesture repetition.

# of rep	1	2	3	4	5	6	7	8
TMC-ViT trained from scratch								
Subj 1	74.60	97.10	97.90	98.90	98.80	98.90	99.10	99.10
Subj 2	95.20	94.80	98.20	98.50	99.20	99.10	99.20	99.10
Subj 3	61.80	54.70	57.10	59.00	78.30	79.40	84.50	96.40
Subj 4	90.70	88.00	93.60	78.00	85.00	89.40	82.50	98.00
Subj 5	95.90	94.70	89.10	94.50	98.30	98.00	97.80	98.10
Subj 6	84.60	94.40	96.70	87.70	94.40	97.70	98.60	98.50
Subj 7	45.30	53.70	69.30	92.50	87.30	95.80	96.30	96.20
Subj 8	96.00	96.40	97.00	96.90	97.30	97.80	97.20	96.70
AVG	**80.51**	**84.23**	**87.36**	**88.25**	**92.33**	**94.51**	**94.40**	**97.76**
SD	**18.66**	**18.73**	**15.55**	**13.69**	**7.83**	**6.86**	**6.82**	**1.18**
Pre-trained and fine-tuned TMC-ViT								
Subj 1	73.50	80.80	96.80	98.20	98.20	98.30	98.30	98.60
Subj 2	90.50	92.20	95.50	95.40	96.00	96.20	95.80	96.10
Subj 3	35.20	35.20	34.30	55.10	61.60	62.30	80.70	95.80
Subj 4	95.00	93.40	87.50	82.50	86.80	86.50	86.70	94.40
Subj 5	86.50	81.50	83.30	87.40	93.00	92.90	93.70	94.60
Subj 6	80.30	84.50	93.00	83.60	80.50	92.80	98.10	98.20
Subj 7	47.10	44.70	55.80	75.40	75.80	89.80	96.50	96.50
Subj 8	93.50	95.90	95.80	96.20	96.40	96.60	96.40	96.60
AVG	**75.20**	**76.03**	**80.25**	**84.23**	**86.04**	**89.43**	**93.28**	**96.35**
SD	**22.38**	**23.09**	**22.93**	**14.16**	**12.74**	**11.61**	**6.29**	**1.50**
Prototypical Network								
Subj 1	84.20	91.60	93.80	96.00	97.00	97.40	97.50	98.10
Subj 2	91.60	97.10	97.60	98.00	96.40	97.70	98.40	97.10
Subj 3	78.50	76.00	67.90	67.40	73.60	73.50	80.10	83.70
Subj 4	91.20	90.70	89.90	91.10	91.40	90.70	93.20	95.00
Subj 5	86.70	90.00	94.20	94.80	93.30	93.20	93.70	94.00
Subj 6	88.12	85.62	92.02	92.29	97.36	97.35	97.71	97.11
Subj 7	84.40	85.20	86.80	88.10	91.90	90.60	92.20	91.30
Subj 8	94.30	92.60	94.30	95.90	96.10	96.00	95.50	96.20
AVG	**87.38**	**88.60**	**89.56**	**90.45**	**92.13**	**92.06**	**93.54**	**94.06**
SD	**5.06**	**6.36**	**9.33**	**9.83**	**7.84**	**8.05**	**5.89**	**4.71**

A. SUBJECT-SPECIFIC MODELS, INTRA-SESSION

Table 2 presents the performance of the subject-specific models trained on only one subject at a time for the data collected in the same session.

The analysis of Table 2 shows that high accuracies can be achieved for classifying five gestures using raw EMG data as input when training models in a subject-specific manner. Training a TMC-ViT from scratch resulted in the highest accuracies, outperforming the other two training approaches for 6 out of 8 tested subjects, achieving accuracies as high as 98.41%. It can be noticed that some gesture repetitions can even achieve 100% decoding accuracy, e.g., repetition number 9 for subject 2. The pre-trained TMC-ViT and Prototypical networks performed better for one subject each compared to the other approaches. The Prototypical

TABLE 4. Subject-specific models' performance trained on different sessions.

Fold	Subj 1	Subj 2	Subj 3	Subj 4	Subj 5
	TMC-ViT trained from scratch				
0	64.29	66.80	64.45	64.73	63.31
1	63.75	67.60	65.54	63.91	62.52
2	64.09	69.00	64.08	64.49	62.65
3	64.98	61.30	63.96	64.45	62.83
4	63.92	64.10	63.89	64.58	63.00
5	63.86	68.60	64.08	64.25	63.37
6	64.39	59.00	63.77	63.73	63.16
7	64.10	70.10	64.14	64.68	60.00
8	63.70	69.40	65.11	64.42	61.79
9	64.40	56.00	60.00	63.00	62.00
AVG	**64.15**	**65.19**	**63.90**	**64.22**	**62.46**
SD	**0.39**	**4.89**	**1.48**	**0.54**	**1.01**
	Prototypical Network				
0	81.17	81.42	65.29	83.44	65.19
1	79.63	79.55	79.74	85.07	79.44
2	82.82	78.94	72.54	86.76	74.03
3	79.75	79.52	77.78	85.48	73.56
4	75.51	79.81	65.07	83.96	80.67
5	79.60	79.73	72.34	82.35	82.08
6	74.17	79.52	72.48	85.75	78.84
7	77.40	79.85	68.62	84.14	83.64
8	75.80	79.82	69.04	83.72	81.95
9	75.41	78.99	65.01	82.04	67.98
AVG	**78.12**	**79.72**	**70.79**	**84.27**	**76.74**
SD	**2.87**	**0.68**	**5.18**	**1.50**	**6.31**

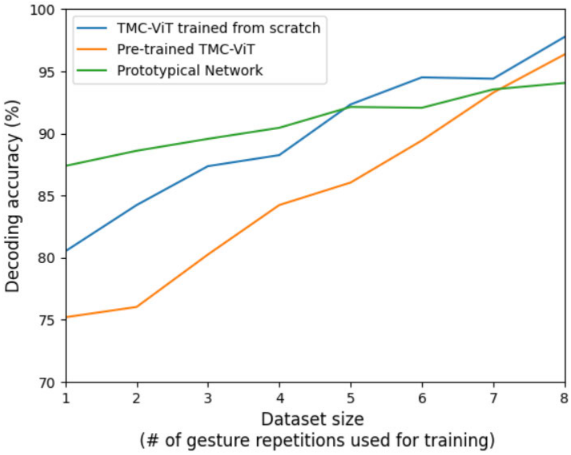

FIGURE 6. Gesture decoding accuracy. As the number of gesture repetitions used for training increases, the TMC-ViT models tend to outperform the Prototypical network model.

TABLE 5. Subject-generic models' performance.

Fold	Subj 1	Subj 2	Subj 3	Subj 4	Subj 5	Subj 6	Subj 7	Subj 8
	TMC-ViT trained from scratch							
0	81.00	81.50	79.50	74.20	83.10	86.80	86.60	91.00
1	77.40	88.30	86.40	89.80	85.00	85.20	80.10	84.10
2	80.60	70.30	78.70	86.90	80.20	85.30	84.70	80.20
3	81.20	82.70	78.50	88.30	79.10	78.90	84.50	84.50
4	84.70	83.30	77.90	71.50	76.40	79.80	83.60	89.20
5	76.30	90.60	82.50	77.10	83.00	80.40	75.90	71.60
6	80.70	71.30	80.30	83.10	81.30	78.20	76.20	56.50
7	83.10	72.00	79.80	81.90	86.70	80.80	76.60	88.40
8	77.70	88.40	84.60	80.40	79.10	85.00	76.50	89.00
9	80.00	80.00	80.00	86.70	80.00	80.00	80.00	80.00
AVG	**80.27**	**80.84**	**80.82**	**81.99**	**81.39**	**82.04**	**80.47**	**81.45**
SD	**2.58**	**7.45**	**2.80**	**6.20**	**3.08**	**3.16**	**4.11**	**10.52**
	Prototypical Network							
0	82.98	82.96	82.69	85.51	83.12	89.94	80.74	92.18
1	82.16	90.96	76.86	90.10	86.14	92.51	84.00	91.13
2	82.68	88.83	81.90	91.46	85.19	85.65	76.87	94.52
3	82.63	88.80	72.31	82.32	89.63	92.46	75.19	96.26
4	81.08	88.15	63.98	87.17	84.93	93.87	84.66	95.99
5	82.28	91.73	74.29	87.18	84.46	93.18	77.16	87.48
6	86.52	90.04	79.73	85.82	87.74	92.03	88.53	77.69
7	80.54	87.29	80.40	83.80	84.12	91.02	86.74	82.77
8	87.29	90.72	81.79	87.09	87.98	92.44	89.28	79.55
9	82.77	92.40	82.82	85.01	83.11	89.10	80.01	92.49
AVG	**83.09**	**89.19**	**77.68**	**86.55**	**85.64**	**91.22**	**82.32**	**89.01**
SD	**2.16**	**2.72**	**6.02**	**2.73**	**2.19**	**2.44**	**5.05**	**6.81**

Network achieved the lowest standard deviation for most of the subjects (subjects 1, 2, 3, 5, 7, and 8), demonstrating a good generalization among gesture repetitions for the same subject. It is also interesting to notice that, in general, the models present a higher decoding accuracy when tested in the intermediary gesture repetitions. This is explained by the fact that in the last gesture repetitions, the EMG signal changes from the initial gesture repetitions due to fatigue and, in the first gesture repetitions, usually the participant is still adapting to the data collection procedure, resulting in a not so accurate gesture performance and hence not so precise repetition, leading to a more noisy EMG signal or even wrongly labeled data. Therefore, training a TMC-ViT from scratch for subject-specific and intra-session models shows the best performance for most of the tested subjects, meaning these models are best suited when developing EMG-based interfaces to achieve optimal gesture decoding in such scenarios.

In order to explore the performance of the training techniques based on the dataset size, we trained the same models with an increasing number of gesture repetitions used for training, leaving the last two gesture repetitions for training. The performance achieved by these models is shown in Table 3. All the tested learning approaches show a decoding accuracy improvement the bigger the number of gesture repetitions used for training, as shown in Fig. 6. The approach that presents the best performance for a smaller number of gesture repetitions is the Prototypical Network, as expected for a few-shot learning approach. At the mark of five training gesture repetitions, the TMC-ViT trained

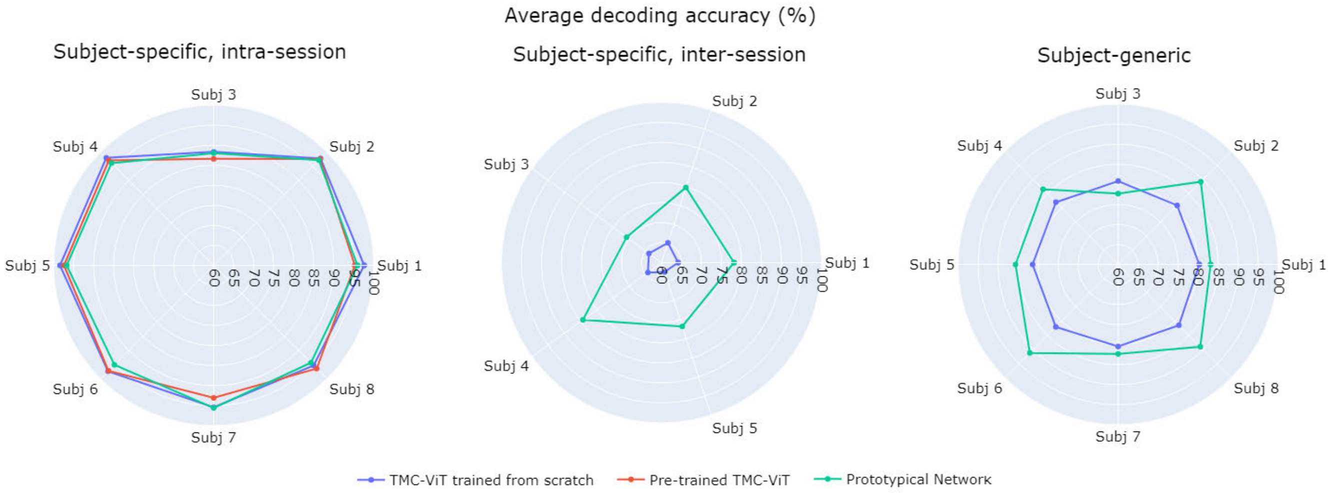

FIGURE 7. Average accuracy between folds for each subject from the different sets, i.e., subject-specific intra-session, subject-specific inter-session, and subject-generic sets. The subject-specific models trained for intra-session data achieve higher accuracy than those trained and evaluated for the other training sets. The subject-specific inter-session models achieve the worst average accuracies, showing how EMG data changes from session to session, even when collected from the same subjects. Even though the subject-generic models perform worse than the subject-specific intra-session models, they still present consistent decoding accuracies among the tested subjects. In general terms, training a TMC-ViT from scratch achieves higher performance for the subject-specific intra-session set, while the Prototypical Network achieves better performance for the subject-specific inter-session and subject-generic sets.

from scratch outperforms the Prototypical Network, while the pre-trained model begins to outperform the few-shot learning approach when seven gesture repetitions are used for training. These results highlight that in selecting the training model and learning technique for a given task, taking into account dataset size is paramount to choosing the most appropriate framework in order to achieve best performance (i.e, good gesture decoding).

B. SUBJECT-SPECIFIC MODELS, INTER-SESSION

Table 4 shows the results achieved by the TMC-ViT trained from scratch and the Prototypical Network in decoding EMG signals while the model was trained on a different data collection session. As seen in Table 4, the Prototypical Network outperformed the TMC-ViT for all tested subjects. The few-shot learning approach achieves from 70.79% to 84.27%, while the TMC-ViT achieves decoding accuracies as high as 65.19%, demonstrating that when the EMG-based interface is deployed using data collected from another session for the same subject, the Prototypical Network may present higher gesture decoding capabilities.

C. SUBJECT-GENERIC MODELS

The final set of experiments was composed of training the TMC-ViT and the Prototypical Network from scratch. The results achieved for these two approaches are shown in Table 5. For this training set, the Prototypical Network outperformed the TMC-ViT from 7 out of 8 tested subjects, with accuracies ranging from 83.03% to 91.22%. The TMC-ViT achieved accuracies as high as 82.04%, showing the potential of this few-shot learning approach in decoding hand gestures from EMG data of an unseen user, demonstrated by its higher accuracies for 7 out of 8 tested subjects.

IV. DISCUSSION

In this section, we discuss the results in Section III. These results are summarized in Fig. 7.

A. SUBJECT-SPECIFIC MODELS, INTRA-SESSION

Table 2 shows the results obtained with the three training approaches when trained on the subject-specific set of data collected in the same session. This training set can represent the ideal scenario when deploying an HMI, i.e., when a considerably large dataset for the subject that is going to use the interface has been collected and is available and when the interface is used (or tested) in the same session as the data is collected (or when the machine learning decoding model is trained). Typically, in this scenario a deep learning model trained from scratch is expected to achieve the best decoding accuracies, and as observed in Table 2, the TMC-ViT model trained from scratch outperformed the other training approaches for the majority of tested subjects. The decreased performance achieved by the same architecture when pre-trained on other subjects is explained by the fact the EMG signals differ considerably from subject to subject, even when the electrodes are placed on the same muscles while the user performs the same gestures. Precise electrode placement is often required for surface EMG and even a slight deviation in electrode placement will alter the signals, although not dramatically. Again, since the subject-specific models use several gesture repetitions for training (9 out of 10), the training set comprises of a dataset, large enough, to train the deep learning models' weights, achieving high performance. The high performance of the TMC-ViT model when trained from scratch with larger datasets is evident when we analyze Table 3 and Fig. 6. It can be seen that a dataset with 5 repetitions of a gesture is enough to produce a model that outperforms the Prototypical Network. Still, the performance of the few-shot learning approach is notable, achieving 87.38% decoding accuracy using only one repetition for training. This result is even more impressive considering the fact that the model is trained on the first collected repetition (minimal fatigue) and evaluated in the last two gesture repetitions when fatigue is most prevalent. However, training a model from scratch remains the best approach when collecting more data from the same subject is feasible.

B. SUBJECT-SPECIFIC MODELS, INTER-SESSION

When the data is collected in a different session, the Prototypical Network outperforms the TMC-ViT, as shown in Table 4. In that case, the TMC-ViT achieves an accuracy as high as 65.19%, seeing a dramatic drop from the one observed for the TMC-ViT subject-specific model trained on data from the same session, that led to a decoding accuracy as high as 98.41% (see Table 2). The fact that EMG signals change drastically when collected in different sessions is one factor that hinders the applicability of HMI in general. However, if the HMI needs to operate for data collected in different sessions, for example, when removing a prosthetic and putting it on again, the Prototypical Network can be a viable option to be employed for EMG decoding, achieving satisfactory accuracies as high as 84.27%.

C. SUBJECT-SPECIFIC MODELS

The results achieved by the subject-generic models, shown in Table 4, show that the Prototypical Network outperforms the TMC-ViT for most of the tested subjects. The analysis of this table also makes it clear that data from the same subject but collected in different sessions can be treated as data from different subjects, given the similar results achieved by the subject-specific models trained on data from different sessions and the subject-generic models (see Table 4 and 5). Therefore, based on the results from Table 5, a few-shot learning approach capable of predicting classes based on only a small set of labeled examples seems to generalize better for unseen data from new subjects (or different sessions), representing the best approach for this scenario. One example where this may be suitable is in making EMG-based controllers for computers, gaming, and robotic control and telemanipulation purposes.

V. CONCLUSION

In this work, we compare three learning techniques in the context of developing EMG-based HMI applications to decode five hand gestures. We trained a TMC-ViT, a novel deep learning architecture, from scratch; we pre-trained this same architecture in a larger dataset; and we trained a few-shot learning technique called Prototypical Network. These models were trained in a subject-specific and subject-generic way. We further explored the influence of the number of gesture repetitions used during training in decoding accuracy and how the models perform when evaluated on data from the same user but collected during a different session.

Training a model from scratch resulted in the best decoding performance among the subject-specific models for six of eight tested subjects, achieving an accuracy as high as 98.41% when data from the same session and nine gesture repetitions are used for training (one is left for testing). However, if less than five gesture repetitions are used for training or when data from different sessions are used for testing, the Prototypical Network outperforms the TMC-ViT models. Regarding subject-generic models, the Prototypical Network achieves the highest accuracy compared to the other training techniques, with accuracies as high as 84.27%. Therefore, in cases where a large training set from the same session is available, training a deep learning model from scratch is advised and will result in the best decoding accuracy. However, a few-shot learning approach can deliver the best performance for scenarios where data is scarce, collected in different sessions, or collected for several users.

Future work will explore the same scenarios but for different biological signals, such as lightmyography [48] or forcemyography, in order to systematically evaluate different human-machine interfaces. Moreover, we may explore unsupervised approaches to learning meaningful correspondences between different users' EMG data, as even though few-shot learning proved to be a good approach to address how to learn EMG features between multiple subjects and/or unseen subjects, the performance of such models is still considerable lower than subject-specific models trained from scratch.

REFERENCES

[1] A. Dwivedi, "Analysis, development, and evaluation of muscle machine interfaces for the intuitive control of robotic devices," Ph.D. dissertation, Dept. Mech. Eng., Univ. Auckland, 2021.

[2] A. C. Roşca, C. C. Baciu, V. Burtăverde, and A. Mateizer, "Psychological consequences in patients with amputation of a limb. An interpretative-phenomenological analysis," *Frontiers Psychol.*, vol. 12, May 2021, Art. no. 537493.

[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] M. Bumbasirevic, A. Lesic, T. Palibrk, D. Milovanovic, M. Zoka, T. Kravic-Stevovic, and S. Raspopovic, "The current state of bionic limbs from the surgeon's viewpoint," *EFORT Open Rev.*, vol. 5, no. 2, pp. 65–72, Feb. 2020.

[5] J. Vogel, C. Castellini, and P. van der Smagt, "EMG-based teleoperation and manipulation with the DLR LWR-III," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2011, pp. 672–678.

[6] Y. Kwon, A. Dwivedi, A. J. McDaid, and M. Liarokapis, "On muscle selection for EMG based decoding of dexterous, in-hand manipulation motions," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 1672–1675.

[7] C. Castellini and P. van der Smagt, "Surface EMG in advanced hand prosthetics," *Biol. Cybern.*, vol. 100, no. 1, pp. 35–47, Jan. 2009.

[8] C. Castellini and G. Passig, "Ultrasound image features of the wrist are linearly related to finger positions," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2011, pp. 2108–2114.

[9] M. O. Ibitoye, N. A. Hamzaid, J. M. Zuniga, and A. K. A. Wahab, ''Mechanomyography and muscle function assessment: A review of current state and prospects,'' *Clin. Biomechanics*, vol. 29, no. 6, pp. 691–704, Jun. 2014.

[10] A. Dwivedi, H. Groll, and P. Beckerle, ''A systematic review of sensor fusion methods using peripheral bio-signals for human intention decoding,'' *Sensors*, vol. 22, no. 17, p. 6319, Aug. 2022.

[11] M. Perusquia-Hernandez, M. Hirokawa, and K. Suzuki, ''Spontaneous and posed smile recognition based on spatial and temporal patterns of facial emg,'' in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, 2017, pp. 537–541.

[12] A. V. Hill, ''The heat of shortening and the dynamic constants of muscle,'' *Proc. R. Soc. London. Ser. B, Biol. Sci.*, vol. 126, no. 843, pp. 136–195, 1938.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, ''Attention is all you need,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, ''An image is worth 16×16 words: Transformers for image recognition at scale,'' 2020, *arXiv:2010.11929*.

[15] M. Moradi, Y. Weng, and Y.-C. Lai, ''Defending smart electrical power grids against cyberattacks with deep Q-learning,'' *PRX Energy*, vol. 1, no. 3, Nov. 2022, Art. no. 033005.

[16] Z.-M. Zhai, M. Moradi, L.-W. Kong, and Y.-C. Lai, ''Detecting weak physical signal from noise: A machine-learning approach with applications to magnetic-anomaly-guided navigation,'' *Phys. Rev. Appl.*, vol. 19, no. 3, Mar. 2023, Art. no. 034030.

[17] S. E. Razavi, M. A. Moradi, S. Shamaghdari, and M. B. Menhaj, ''Adaptive optimal control of unknown discrete-time linear systems with guaranteed prescribed degree of stability using reinforcement learning,'' *Int. J. Dyn. Control*, vol. 10, no. 3, pp. 870–878, Jun. 2022.

[18] M. Yang, ''A survey on few-shot learning in natural language processing,'' in *Proc. Int. Conf. Artif. Intell. Electromechanical Autom. (AIEA)*, May 2021, pp. 294–297.

[19] R. V. Godoy, A. Dwivedi, B. Guan, A. Turner, D. Shieff, and M. Liarokapis, ''On EMG based dexterous robotic telemanipulation: Assessing machine learning techniques, feature extraction methods, and shared control schemes,'' *IEEE Access*, vol. 10, pp. 99661–99674, 2022.

[20] R. V. Godoy, A. Dwivedi, and M. Liarokapis, ''Electromyography based decoding of dexterous, in-hand manipulation motions with temporal multichannel vision transformers,'' *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2207–2216, 2022.

[21] P. K. Artemiadis and K. J. Kyriakopoulos, ''EMG-based control of a robot arm using low-dimensional embeddings,'' *IEEE Trans. Robot.*, vol. 26, no. 2, pp. 393–398, Apr. 2010.

[22] K. Kiguchi and Y. Hayashi, ''An EMG-based control for an upper-limb power-assist exoskeleton robot,'' *IEEE Trans. Syst., Man, Cybern., B*, vol. 42, no. 4, pp. 1064–1071, Aug. 2012.

[23] L. van Dijk, C. K. van der Sluis, H. W. van Dijk, and R. M. Bongers, ''Learning an EMG controlled game: Task-specific adaptations and transfer,'' *PLoS ONE*, vol. 11, no. 8, Aug. 2016, Art. no. e0160817.

[24] T. S. Saponas, D. S. Tan, D. Morris, J. Turner, and J. A. Landay, ''Making muscle-computer interfaces more practical,'' in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2010, pp. 851–854.

[25] M. Simão, P. Neto, and O. Gibaru, ''EMG-based online classification of gestures with recurrent neural networks,'' *Pattern Recognit. Lett.*, vol. 128, pp. 45–51, Dec. 2019.

[26] R. V. Godoy, G. J. G. Lahr, A. Dwivedi, T. J. S. Reis, P. H. Polegato, M. Becker, G. A. P. Caurin, and M. Liarokapis, ''Electromyography-based, robust hand motion classification employing temporal multi-channel vision transformers,'' *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 10200–10207, Oct. 2022.

[27] Z. Dou, Y. Sun, Z. Wu, T. Wang, S. Fan, and Y. Zhang, ''The architecture of mass customization-social Internet of Things system: Current research profile,'' *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 10, p. 653, Sep. 2021.

[28] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, ''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.

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

[30] R. Soroushmojdehi, S. Javadzadeh, A. Pedrocchi, and M. Gandolla, ''Transfer learning in hand movement intention detection based on surface electromyography signals,'' *Frontiers Neurosci.*, vol. 16, pp. 1–18, Nov. 2022.

[31] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G.-M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, ''Electromyography data for non-invasive naturally-controlled robotic hand prostheses,'' *Sci. Data*, vol. 1, no. 1, pp. 1–13, Dec. 2014.

[32] 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.

[33] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, ''Siamese convolutional neural network and few-shot learning for embedded gesture recognition,'' in *Proc. 20th IEEE Interregional NEWCAS Conf. (NEWCAS)*, Jun. 2022, pp. 114–118.

[34] J. Snell, K. Swersky, and R. S. Zemel, ''Prototypical networks for few-shot learning,'' 2017, *arXiv:1703.05175*.

[35] A. Dwivedi, Y. Kwon, A. J. McDaid, and M. Liarokapis, ''EMG based decoding of object motion in dexterous, in-hand manipulation tasks,'' in *Proc. 7th IEEE Int. Conf. Biomed. Robot. Biomechatronics (Biorob)*, Aug. 2018, pp. 1025–1031.

[36] A. Dwivedi, Y. Kwon, A. J. McDaid, and M. Liarokapis, ''A learning scheme for EMG based decoding of dexterous, in-hand manipulation motions,'' *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2205–2215, Oct. 2019.

[37] A. Turner, D. Shieff, A. Dwivedi, and M. Liarokapis, ''Comparing machine learning methods and feature extraction techniques for the EMG based decoding of human intention,'' in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 4738–4743.

[38] C. Ngo, C. Munoz, M. Lueken, A. Hülkenberg, C. Bollheimer, A. Briko, A. Kobelev, S. Shchukin, and S. Leonhardt, ''A wearable, multi-frequency device to measure muscle activity combining simultaneous electromyography and electrical impedance myography,'' *Sensors*, vol. 22, no. 5, p. 1941, Mar. 2022.

[39] R. Merletti and P. J. Parker, *Electromyography: Physiology, Engineering, and Non-Invasive Applications*, vol. 11. Hoboken, NJ, USA: Wiley, 2004.

[40] I. M. Bullock, J. Z. Zheng, S. De La Rosa, C. Guertler, and A. M. Dollar, ''Grasp frequency and usage in daily household and machine shop tasks,'' *IEEE Trans. Haptics*, vol. 6, no. 3, pp. 296–308, Jul. 2013.

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

[42] K. Englehart, B. Hudgin, and P. A. Parker, ''A wavelet-based continuous classification scheme for multifunction myoelectric control,'' *IEEE Trans. Biomed. Eng.*, vol. 48, no. 3, pp. 302–311, Mar. 2001.

[43] S. Ioffe and C. Szegedy, ''Batch normalization: Accelerating deep network training by reducing internal covariate shift,'' in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[44] W. Guo, X. Sheng, and X. Zhu, ''Assessment of muscle fatigue by simultaneous sEMG and NIRS: From the perspective of electrophysiology and hemodynamics,'' in *Proc. 8th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, May 2017, pp. 33–36.

[45] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, ''Dropout: A simple way to prevent neural networks from overfitting,'' *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[46] J. Snell, K. Swersky, and R. S. Zemel, ''Prototypical networks for few-shot learning,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[47] D. P. Kingma and J. L. Ba, ''Adam: A method for stochastic optimization,'' in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015, pp. 1–15.

[48] M. Shahmohammadi, B. Guan, R. V. Godoy, A. Dwivedi, P. Nielsen, and M. Liarokapis, ''On lightmyography based muscle-machine interfaces for the efficient decoding of human gestures and forces,'' *Sci. Rep.*, vol. 13, no. 1, p. 327, Jan. 2023.

RICARDO V. GODOY (Graduate Student Member, IEEE) received the Bachelor of Engineering degree in mechatronics engineering, in 2019, and the M.Sc. degree in mechanical engineering from the University of Sáo Paulo, Sáo Carlos, Brazil, in 2021. He is currently pursuing the Ph.D. degree in mechatronics engineering with the New Dexterity Research Group, The University of Auckland, New Zealand. He works on the analysis and development of novel human–machine interfaces (HMI) for the control of robotic and bionic devices while focusing on the challenges and limitations in the use of HMI for robust grasping and decoding of dexterous, in-hand manipulation tasks. He also works on the development of novel and robust deep learning-based solutions for bio-signals analyses and processing.

BONNIE GUAN (Graduate Student Member, IEEE) received the Bachelor of Engineering degree (Hons.) in mechatronics engineering from The University of Auckland, New Zealand, in 2022, where she is currently pursuing the Ph.D. degree with the New Dexterity Research Group. Her work focuses on exploring novel methods for developing human–machine interfaces to facilitate an intuitive and dexterous control of robotic and prosthetic devices.

FELIPE SANCHES received the Bachelor of Science and M.Sc. degrees in computer science from the University of Sáo Paulo, Sáo Carlos, Brazil, in 2016 and 2021, respectively. He is currently pursuing the Ph.D. degree in mechatronics engineering with the New Dexterity Research Group, The University of Auckland, New Zealand. His work focuses on transferring dexterous manipulation skills from humans to robots. He is also interested in the application of machine learning and reinforcement learning for solving robotics tasks.

ANANY DWIVEDI received the B.Tech. degree in electronics and communication engineering from the LNM Institute of Information Technology, Jaipur, India, in 2015, and the M.S. degree in robotics engineering from the Worcester Polytechnic Institute (WPI), Worcester, MA, USA, in 2017. He is currently pursuing the Ph.D. degree with the New Dexterity Research Group, The University of Auckland, New Zealand, where his work focused on deciphering the myoelectric activity of the muscles of the human forearm and hand to decode dexterous manipulation motions in real as well as virtual environments. He joined the Chair of Autonomous Systems and Mechatronics, Friedrich-Alexander-Universität-Erlangen–Nürnberg, Germany, as a Postdoctoral Researcher, after the Ph.D. degree, where his work focused on development of bi-directional human–machine interfaces to increase presence and embodiment of the user during interaction with virtual environments while considering human-centered design principles. He is a Postdoctoral Researcher with the Artificial Intelligence Institute, University of Waikato, New Zealand, with a focus on the applications of machine learning techniques in data streams adaptation and interpretation for weak signals and extreme events.

MINAS LIAROKAPIS (Senior Member, IEEE) received the Diploma degree in computer engineering from the University of Patras, Patras, Greece, the M.Sc. degree in information technologies in medicine and biology from the National Kapodistrian University of Athens, Athens, Greece, and the Ph.D. degree in mechanical engineering from the National Technical University of Athens, Athens. He is currently an Associate Professor with the Department of Mechanical and Mechatronics Engineering, The University of Auckland, New Zealand, and the Director of the New Dexterity Research Group (www.newdexterity.org). Previously, he was a Postdoctoral Associate with the GRAB Laboratory, Yale University, USA. He is the Founder of the OpenBionics initiative (www.openbionics.org) and the Co-Founder of OpenRobotHardware (www.openrobothardware.org) and HandCorpus (www.handcorpus.org). His research interests include providing robotics solutions to everyday life problems, modeling, designing, and controlling novel robotics and bionics hardware.