

Unveiling EMG Semantics: A Prototype-Learning Approach to Generalizable Gesture Classification

Hunmin Lee¹, Ming Jiang¹, Jinhui Yang¹, Zhi Yang² and Qi Zhao¹

¹ Department of Computer Science and Engineering, University of Minnesota, Twin Cities, USA

² Department of Biomedical and Engineering, University of Minnesota, Twin Cities, USA

E-mail: lee03915@umn.edu, mjiang@umn.edu, yang7004@umn.edu, yang5029@umn.edu, qzhao@cs.umn.edu

Abstract. Upper limb loss can profoundly impact an individual's quality of life, posing challenges to both physical capabilities and emotional well-being. To restore limb function by decoding electromyography (EMG) signals, in this paper, we present a novel deep prototype learning method for accurate and generalizable EMG-based gesture classification. Existing methods suffer from limitations in generalization across subjects due to the diverse nature of individual muscle responses, impeding seamless applicability in broader populations. By leveraging deep prototype learning, we introduce a method that goes beyond direct output prediction. Instead, it matches new EMG inputs to a set of learned prototypes and predicts the corresponding labels. This novel methodology significantly enhances the model's classification performance and generalizability by discriminating subtle differences between gestures, making it more reliable and precise in real-world applications. Our experiments on four Ninapro datasets suggest that our deep prototype learning classifier outperforms state-of-the-art methods in terms of intra-subject and inter-subject classification accuracy in gesture prediction. The results from our experiments validate the effectiveness of the proposed method and pave the way for future advancements in the field of EMG gesture classification for upper limb prosthetics.

Keywords: Generalizability, Hand Gesture Classification, Prototype Learning, Surface Electromyography

Submitted to: *J. Neural Eng.*

1. Introduction

Over two million people in the USA alone suffer from the devastating disability of upper limb loss, and millions more are at risk of amputation [1]. Amputees face severe difficulties in their daily activities due to the loss of a limb, especially for those missing an upper limb [2]. To restore function and independence for these patients, there has been ongoing development of prosthetic hands, arms, and other assistive devices that serve both rehabilitation and robotic purposes. Powered myoelectric prostheses, designed for rehabilitation, aim to provide amputees with intuitive motor control and sensory feedback. These prosthetic limbs are controlled by electromyographic (EMG) signals from the user’s muscles, enabling basic hand grasp and arm movements. However, current myoelectric prostheses offer limited dexterity, which hinders the full restoration of natural hand function. As a result, accurate and generalizable EMG pattern recognition techniques are needed to enhance the functionality of these prostheses, allowing for nuanced and precise movements. Emphasizing the generalizability of these techniques across diverse user populations is essential for real-world application, ensuring the benefits reach individuals with varying physiological characteristics. By employing advanced machine learning algorithms and signal processing methods, there is an opportunity to refine EMG-based pattern recognition, enabling a more comprehensive range of gestures and intuitive control, thus bridging the gap between amputees’ needs and current prosthetic capabilities.

Smooth and responsive control of a multi-functional myoelectric hand requires advanced signal processing to make accurate predictions of user intent from surface EMG (sEMG) recordings. Conventional EMG-based gesture classification typically involves multiple stages of handcrafted feature extraction and pattern recognition. In particular, raw EMG signals are segmented into discrete gestures, then handcrafted features are manually engineered to characterize relevant patterns of each segment. Finally, these features are used by classifiers such as Linear Discriminant Analysis [3], Hidden Markov Models [4, 5], Gaussian Mixture Models [6, 7], or Support Vector Machines [8, 9] to predict gesture labels. While these standard approaches have enabled basic myoelectric control, they lack the flexibility to discover sophisticated patterns in the data and often lead to lower prediction accuracy. Therefore, achieving highly accurate and generalizable EMG-based control requires machine learning methods that can model the complex and non-linear relationships in the EMG signals.

In recent years, EMG pattern recognition for gesture classification using deep neural networks (DNNs) has gained increasing attention. These deep learning methods can discover complex patterns in the data that enable accurate predictions of user intent. State-of-the-art DNN-based approaches have achieved over 90% accuracy in

classifying up to 53 different hand gestures on public benchmark datasets [10, 11]. However, a key limitation prohibits the practical use of these methods in applications.

Generalizability in EMG-gesture classification for prosthetic devices using deep learning is a vital challenge that needs to be resolved. It ensures adaptability across diverse users and environments, facilitating wider applicability without extensively customized, individualized training. A model with strong generalization and credible classification accuracy performs reliably on new data, which is crucial for real-world scenarios where users have varied muscle activities. This ease of deployment simplifies fitting the device to different users, improving the overall user experience by offering consistent control. Additionally, it drives technological advancements, potentially leading to more natural movements and increased functionality, ultimately enhancing users’ quality of life. Existing methods perform well when training and testing models on the same subject, while most do not perform well when tested on new subjects. Thus, the high intra-subject performance on benchmark EMG datasets does not guarantee general and dependable control that can translate into practical prosthetic devices, and whether these devices can be generalized to different users remains an open question.

To address these challenges, we propose a novel approach to learning generalizable representations based on prototype learning. The proposed method allows the DNN model to characterize the relationships between EMG sensor data by clustering similar gesture examples into prototypes. The deep prototype learning classifier is trained on a large dataset of digital glove data to group similar EMG patterns into a set of gesture prototypes corresponding to joint movements. To perform prediction on new EMG inputs, the classifier first measures its similarity to each prototype and then predicts its gesture class based on the similarity measures. This two-stage approach, from learning the prototypes to predicting its label, enables the model to be highly general while still discriminating subtle differences between gestures. The set of prototypes essentially encodes a generalizable representation of EMG patterns for each gesture. They can be visualized or analyzed to gain insight into the features that are important for the classification of a particular class.

With this new approach, we conduct comprehensive experiments to compare different EMG features and prototypes. The results show that deep prototype learning classifiers outperform direct state-of-the-art methods consistently across multiple public datasets. Most importantly, while previous methods struggle to generalize across subjects due to inter-subject variability, our prototype approach successfully achieves improved accuracy for new subjects by matching its gestural prototypes. In sum, the contributions of this paper are summarized as follows:

- We propose a deep prototype learning method for EMG gesture classification. By predicting

the gesture class of the input EMG signal based on its closest gestural prototypes, this approach demonstrates generalizability to new subjects.

- Comprehensive experiments across four public datasets with varying numbers of gestures and participants are conducted to demonstrate the notable classification performance and generalizability of our deep prototype learning approach.
- Our results show high intra-subject accuracy and improvement in inter-subject evaluation performance. Our qualitative results show that our approach can generalize across subjects and various gesture predictions.

The rest of the paper is organized as follows. Section 2 reviews the related work on EMG-based gesture classification of deep learning approaches. Section 3 describes the methodology used in this paper, including the data collection and pre-processing, feature extraction and selection, and classification algorithms. Section 4 presents the results of the experiments, including the accuracy and generalizability of our proposed algorithm compared to existing approaches. Finally, Section 6 summarizes the main findings of the paper and discusses future research directions.

2. Related Work

In recent years, there has been growing interest in developing machine learning methods for EMG-based hand gesture classification. Below, we summarize the related works with an outline of the EMG feature extraction methods. Additionally, we provide a summary of both conventional machine learning methods as well as deep learning-based approaches.

2.1. EMG Feature Extraction

Many studies [12–14] have extensively compared the quality of the processed features, usually based on the performance of the machine learning or deep learning gesture classifiers. The feature selection process can significantly affect the classifier performance, as each processed feature creates a different learning space, which ultimately changes the learning of the optimal decision boundaries for the given gestures. Furthermore, it is crucial to note that feature selection has implications for computational cost, as redundant features will introduce unnecessary overload. Thus, selecting high-quality feature candidates is one of the most important aspects when designing an effective and efficient EMG-based system. Feature extraction involves capturing relevant information from the raw EMG signals and representing them as feature vectors that can be processed by the classifiers. The feature vectors can be obtained in different domains, including the time domain [15–17], frequency domain [16–18],

time-frequency domain [17, 19], and spatial domain [17, 20]. In the time domain, features such as mean, variance, root mean square, and waveform length provide information about the amplitude and shape of the signal over time. In the frequency domain, features derived from the power spectrum, such as spectral moments or the spectral centroid, capture the distribution of signal energy across different frequency components. In the time-frequency domain, techniques such as the short-time Fourier transform (STFT), or wavelet transform can be employed to obtain features that reveal how the frequency content of the signal changes over time. In the spatial domain, features related to electrode placement or muscle activation patterns are often calculated.

2.2. Machine Learning to EMG-Based Gesture Classification

In the field of EMG signal classification, various machine learning algorithms have been used [21], including Naive Bayes (NB), k-nearest neighbor (k-NN), support vector machines (SVMs), decision trees, and other models that developed from those basic models [21–23]. These classifiers take the EMG signal as input and predict the gesture classes. However, these models are often lower in classification accuracy for EMG-based gesture classification compared to deep learning-based approaches [17]. In recent years, the integration of deep learning approaches into EMG-based gesture classification systems has gained considerable attention. For instance, Atzori *et al* [10] compared a Convolutional Neural Network (CNN) with conventional classifiers on sparse multichannel Ninapro databases, achieving gesture classification accuracies of 66.6% and 60.3% on the first and second subdatabases, respectively, which were lower than those achieved by conventional classifiers. To improve CNN-based gesture classification, Olsson *et al* [24] proposed a CNN-based multi-labeled classification scheme using HD-EMG (high-density EMG), achieving an accuracy of 78.7% in 14 healthy subjects. Zhai *et al* [25] introduced a CNN-based strategy with self-recalibrating capacity, outperforming uncalibrated classifiers by 10.18% for 50 hand movements. Chen *et al* [26] employed a 3D CNN for HD-EMG-based gesture classification, achieving superior results compared to instant EMG-based methods, albeit with higher computation costs. While CNNs have demonstrated effectiveness in gesture classification using HD-EMG, their performance with sparse multichannel EMG remains suboptimal.

2.3. Prototype Learning

Prototype learning has emerged as a powerful tool for efficient and interpretable classification, leveraging representative data points (“prototypes”) to capture the essence of different classes within a dataset. Several notable

works have demonstrated its effectiveness in various domains [27–30]. Prototype learning methodologies have recently found application in the field of EMG-gesture classification, primarily serving two overarching objectives: mitigating computational expenses [31, 32] and detecting unknown gestures [33, 34]. A seminal study by Sziburis *et al* [31, 32] marked the inception of employing prototype concepts in EMG-gesture classification. Here, the study employed a k-NN model to classify gestures utilizing prototypes derived from EMG signals. Taking advantage of distinctive gesture prototypes, studies were proposed to detect unknown gestures and further exclude them from the end-to-end classification process. Utilizing a prototype-based CNN model, Wu *et al* [33] investigated the delineation of classical prototypes for various gesture types and computed threshold-based distances to discern signals indicative of unknown gestures. Similarly, Liu *et al* [34] leveraged a generative encoder model to curate prototypes for each gesture class, specifically for unknown gesture detection. While these studies prioritize recognizing unknown gesture classes within subjects, our goal and method are distinct. In this work, we tackle two broader challenges: achieving generalizability across subjects and improving model interpretability for upper-limb gesture classification. By factoring high-level gestures into low-level features, we capture core prototypes that work across diverse subjects, addressing the limitations of prior works. Additionally, translating prototypes into joint activation space offers a clear view of the model’s decision-making process, exceeding the limited interpretability of existing methods.

3. Methodology

Our proposed deep prototype learning method consists of prototype learning and gesture classification processes. Prototype learning aims to learn the prototypes with a multi-label classification paradigm to characterize the movements of hand joints. These prototypes encode joint-level kinematic activation related to the gestures and serve as critical components of gesture classification. We train a gesture classifier by transferring prototype parameters and relevant knowledge from the prototype learning stage. This section outlines the EMG-based gesture classification problem and presents our proposed prototype learning and prototype-based classification methodologies.

3.1. Problem Statement

Let X_s be the 2D vector of raw EMG signals directly curated from the electrode s . Generally, the standard deep learning paradigm involves the application of preprocessed data as illustrated in equation (1), where the function f_1 preprocesses the X_s to $\hat{X}_{(\delta,s)}$ based on various techniques (e.g., Denoising using filters, downsampling) and the time window-unit computations where δ refers to a total number

of window size. Subsequently, the objective function f_2 in equation (2) maps the $\hat{X}_{(\delta,s)}$ to a set of features λ , which reflects unique significance such as statistical measures (e.g., Variance), frequency characteristics (e.g., Average power spectrum), or general knowledge (e.g., Zero crossing). f_2 returns the 4D tensor output, where each tensor is the union of single window features: $\tilde{X}_{(\forall\lambda,\forall\delta,\forall s)}$.

$$f_1 : X_{\forall s} \rightarrow \hat{X}_{(\forall\delta,\forall s)} \quad (1)$$

$$f_2 : \hat{X}_{(\forall\delta,\forall s)} \rightarrow \tilde{X}_{(\forall\lambda,\forall\delta,\forall s)} \quad (2)$$

To predict the classification label of the processed features, a classifier $M(\cdot)$ is defined as:

$$\hat{y} = \sigma(M(\mathbf{W}, \bigcup_{\forall f, \forall \delta, \forall s} \tilde{X}_{(\lambda, \delta, s)})) \quad (3)$$

Here, σ is the softmax activation function for returning the soft probability of each gesture class, \mathbf{W} represents the trainable parameters, and \hat{y} refers to the final output. The objective of gesture classification is to minimize the loss function, as defined in equation (4), where i denotes the index of samples and y_i represents the class label of the sample with index i .

$$\min_{\mathbf{W} \in \mathbb{R}^d} \sum_{\forall i} \mathbb{L}(y_i, \hat{y}_{(i)}) \quad (4)$$

Existing methods for upper-limb gesture classification based on EMG signals [4–13, 15] mainly employ end-to-end models that implicitly learn the correlations between EMG-based input features $\hat{X}_{(\delta,s)}$ and gesture labels \hat{y} . The main challenges with black-box networks are their limited ability to generalize to new data and their complexity, which hinders adaptability and transparency. This compromises interpretability, making it difficult to address overfitting or bias. To overcome these challenges and improve generalization, we propose a deep prototype learning framework. This framework decomposes EMG patterns into atomic prototypes, enhancing interpretability and aiding generalization across different conditions and datasets for practical applications.

3.2. Prototype Learning Method

Establishing a coherent connection between high-level semantics and low-level EMG features is crucial for generalization and interpretation in gesture classification. Integrating abstract gesture concepts with detailed muscle activity enhances the model’s ability to generalize across various scenarios and provides deeper insight into learned representations. To achieve this, we employ the use of semantic prototypes. These prototypes embody meaningful and distinctive gesture patterns, making them valuable in deciphering the underlying intentions of EMG signals. For example, they may represent EMG signals encoding

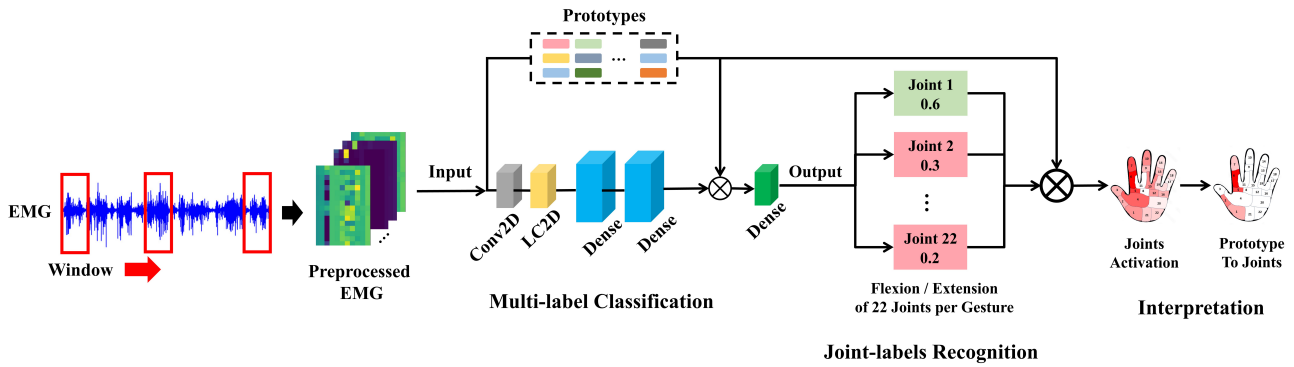


Figure 1. Overview of the proposed prototype learning method. The prototypes are learned through the multi-label classification of binary-coded joint movements. This approach establishes an association between prototypes and joint movements, thereby facilitating the interpretable classification of gestures using joint-kinematic information. By leveraging this transferred knowledge, our method not only learns representative prototypes but also provides insight into the underlying joint movements associated with each gesture type.

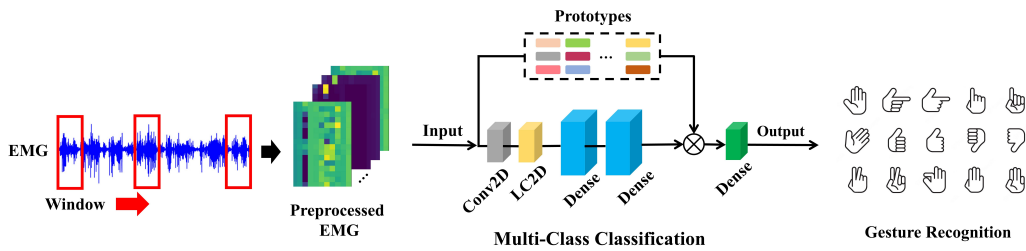


Figure 2. Overview of the proposed gesture classification model based on the learned prototypes. It measures the similarity between the input features and the prototypes to predict the gesture label. By leveraging the concept of prototype-based classification, our model effectively captures the key features of gestures, offering robust classification outcomes.

patterns of hand movements such as the flexion or extension of a finger joint. Prototype-based gesture classification is then achieved by matching EMG features to prototypes of different gestures.

As shown in Figure 1, our prototype learning method is based on a baseline CNN architecture, with a novel prototype layer that learns to factorize gestures into prototypical bases. Different from the conventional methods that classify gestures based on the penultimate layer features O , we decompose gestures with trainable prototypes P and utilize the combinations of their matched prototypes for classification. To achieve this, we compute the similarity between the deep features O and the prototypes P as

$$\alpha_p = \delta(O \cdot P) \quad (5)$$

where δ is the sigmoid activation function for normalization. Therefore, by strategically delving into the penultimate layer’s output of the baseline, the prototypes can encapsulate semantic-level insights rather than low-level details of the EMG signals.

To obtain the semantic meanings of the prototypes, we leverage the Cyberglove [35] data synchronously collected with EMG signals, and perform multi-label classification [36, 37] to learn prototypes from the data. The goal is to

extract meaningful information from the fine-grained joint movements and learn discriminative prototypes that capture the relationship between joint movements and performed gestures. Specifically, the multi-label classification outputs y_{multi} is predicted based on the α_p in equation (5), to fit the ground-truth joint movement labels obtained from the digital glove data, where the classification labels (i.e., binary multi-labels) indicate the flexion or extension of each joint decided by comparing the joint angles between the resting state and the gesture state:

$$y_{multi} = \text{Cls}(\alpha_p) \quad (6)$$

We train the network with a standard binary cross-entropy (BCE) loss for multi-label classification and extract prototypes P after the training. This method factorizes gestures encoded in the EMG signals into a set of bases formed by prototypes, so the learned prototypes encode important semantics representing a variety of joint movements.

3.3. Gesture Classification with Prototypes

With our prototypes constructing the joint-level semantic space for bridging different gestures, we further leverage

Table 1. Joints and abbreviations of Cyberglove sensors [38]. The F and A in the ‘Notation’ column refer to Flexion and Abduction, and the numbers 1 ~ 5 indicate the following: 1 (Thumb), 2 (Index), 3 (Middle), 4 (Ring), and 5 (Pinky). This joint notation denotes specific movement captured by the Cyberglove sensors, aiding in the interpretation of the kinematics of each joint during the qualitative analysis in Section 4.5.

Sensor	Name	Finger	Notation
1	Carpometacarpal joint	Thumb	CMC1_F
2	Metacarpophalangeal joint	Thumb	MCP1_F
3	Interphalangeal joint	Thumb	IP1_F
4	Carpometacarpal joint	Thumb	CMC1_A
5	Metacarpophalangeal joint	Index	MCP2_F
6	Proximal interphalangeal joint	Index	PIP2_F
7	Interphalangeal joint	Index	IP2_F
8	Metacarpophalangeal joint	Middle	MCP3_F
9	Proximal interphalangeal joint	Middle	PIP3_F
10	Interphalangeal joint	Middle	IP3_F
11	Metacarpophalangeal joint	Between	MCP2-3_A
12	Metacarpophalangeal joint	Ring	MCP4_F
13	Proximal interphalangeal joint	Ring	PIP4_F
14	Interphalangeal joint	Ring	IP4_F
15	Metacarpophalangeal joint	Between	MCP3-4_F
16	Metacarpophalangeal joint	Pinky	MCP5_F
17	Proximal interphalangeal joint	Pinky	PIP5_F
18	Interphalangeal joint	Pinky	IP5_F
19	Metacarpophalangeal joint	Between	MCP4-5_F
20	Metacarpophalangeal joint	Between	MCP4-5_A2
21	Wrist joint	Wrist	WRIST
22	Carpometacarpal joint (palmar arching)	Pinky	CMC5_F

the learned prototypes to adaptively incorporate the distinctive correlation between joint movements and gesture classes. Figure 2 provides an overview of the proposed gesture classification model with prototype matching. Instead of predicting the gesture labels using the deep features O , it takes advantage of joint-level patterns encoded in the learned prototypes by computing the similarity measure α_p as described in equation (5). It then predicts the gesture label \hat{y} with α_p . The network architecture is similar to the one used in prototype learning, with the final classification layer changed to predict the gesture labels. The model parameters were initialized with those from the prototype learning network and fine-tuned using the gesture labels, while the prototypes remained static. With this approach, the network associates joint movements based on their relationships with distinct prototypes and adaptively integrates relevant prototypes to predict the gestures.

4. Experiments

In this section, we present the implementation details and carry out experiments to analyze the proposed method. First, we present details of the experiment setups, including datasets, compared methods, evaluation, and implementation details. Second, we evaluate the

gesture classification performance of our proposed deep prototype learning method by comparing it to state-of-the-art methods, the baseline network, and different variants of our method. Next, we assess the generalizability of our proposed methods by conducting inter-subject evaluation. Finally, we present qualitative examples to visualize and interpret the learned semantic prototypes.

4.1. Experiment Setups

4.1.1. Datasets The primary objective of our experiments is to evaluate the performance and generalizability of our method for EMG-based gesture classification. To achieve this, we conducted experiments using four well-established and publicly available EMG datasets: Ninapro DB1, DB2, DB3, and DB5. These datasets cover a wide range of experimental settings, including both basic and complex gestures recorded from intact and amputee subjects. They provide synchronized information from multiple data sources, including surface EMG data collected from up to 16 locations on the users’ dominant forearm and biceps. The gesture categories within each dataset include basic finger movements, simple hand movements, and grips on the wrist and forearm, as well as more intricate grasping motions involving various objects.

All four datasets provide hand gesture data collected

Table 2. Details of the four Ninapro databases used in this study.

	DB1	DB2	DB3	DB5
Gestures	53	50	50	53
Subjects	27	40	11	10
Trials	10	6	6	6
Training trials	1,3,4,6,7,8,9	1,3,4,6	1,3,4,6	1,3,4,6
Testing trials	2,5,10	2,5	2,5	2,5
EMG Channels	10	12	12	16
Cyberglove sensors	22	22	22	22
Sampling Rate	100 Hz	2k Hz	2k Hz	200 Hz

using the Cyberglove device, an effective tool for capturing kinematic information of hand gestures [39]. The Cyberglove data comprises 22 joint-angle sensors that record the rotation trajectory of distinct joints. This includes the kinematic data of the joints acquired from the Cyberglove, synchronized with the EMG signals. Table 1 provides the terminology for specific joints measured with Cyberglove [38]. We specifically excluded subjects other than 5 and 6 for the multi-label classification in DB3 due to noisy sensor data. Furthermore, we excluded the Cyberglove sensor 11 data due to similar noise issues [40].

Detailed specifications of these datasets are provided in Table 2. To maintain consistency, we split the training and test trials following the identical settings used in MV-CNN [41]. On these diverse datasets and employing the described exclusions and partial implementations when necessary, our experiments comprehensively evaluate the effectiveness and applicability of our proposed method for EMG-based gesture classification across various scenarios and challenges.

4.1.2. Compared Methods We conduct a comprehensive comparison of our proposed gesture classification method with existing CNN-based gesture classification models, including GengNet [42], Cheng *et al* [43], Wei *et al* [44], E2CNN [45], Yang *et al* [46], Zhai *et al* [25], Ding *et al* [49], Chen *et al* [26], Vitale *et al* [50], Peng *et al* [51], AtzoriNet [52], CNNLM [53], EVCNN [24], Hu *et al* [54], Pizzolato *et al* [55], MSCNet [56], DVMSCNN [57], and MV-CNN [41]. These models, like ours, are CNN-based classifiers that independently classify each frame of EMG data. Many of them have achieved state-of-the-art performances on the Ninapro datasets. For example, E2CNN [45] and Yang *et al* [46] have both achieved over 90% classification accuracy on the Ninapro DB1 dataset. Particular emphasis is given to the comparison with MV-CNN [41], as it has been extensively evaluated on diverse public datasets and demonstrated a consistently high accuracy over all datasets. The comparison with these models aims to highlight the superiority of our deep

Table 3. Neural network architecture of the proposed method.

Layer	Filters/Strides (Dropout Rate)	Output Shape
Input	-	(None, 16, 11, 1)
BatchNorm	-	(None, 14, 9, 64)
Conv2D	3x3/1	(None, 14, 9, 64)
BatchNorm	-	(None, 12, 7, 64)
Conv2D	3x3/1	(None, 12, 7, 64)
BatchNorm	-	(None, 12, 7, 64)
LC2D	1x1/1	(None, 12, 7, 64)
BatchNorm	-	(None, 12, 7, 64)
LC2D	1x1/1	(None, 12, 7, 64)
BatchNorm	-	(None, 12, 7, 64)
Dropout	0.5	(None, 12, 7, 64)
Flatten	-	(None, 5376)
Dense	-	(None, 512)
BatchNorm	-	(None, 512)
Dropout	0.65	(None, 512)
Dense	-	(None, 512)
BatchNorm	-	(None, 512)
Dropout	0.65	(None, 512)
Dense	-	(None, classes)

prototype learning method over the state-of-the-art methods and showcase the added value of the proposed prototype learning in improving gesture classification accuracy and generalizability.

Furthermore, to provide a meaningful evaluation, we compare the performance of our deep prototype learning method with two baseline models. The first baseline, referred to as Baseline-CNN, is a plain CNN model without the prototype layer, as demonstrated in Figure 3. The network starts with a batch normalization (BN) layer, which normalizes the input data to accelerate the training process and improve the model’s generalization. This is followed by two layers of 2D convolution (Conv2D), each followed by

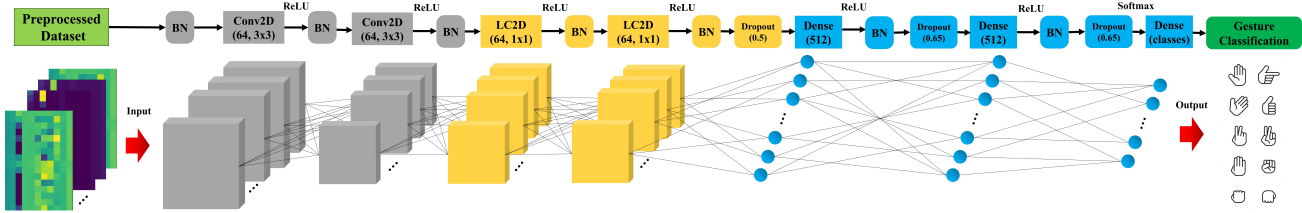


Figure 3. The proposed deep prototype learning classifier adopts a baseline CNN architecture consisting of two 2D convolution layers, followed by two locally-connected layers and three fully-connected layers dedicated to gesture classification. It is designed to effectively capture and extract hierarchical patterns from the preprocessed EMG features, facilitating accurate and robust gesture classification.

a batch normalization layer. The network further processes the data using two locally connected 2D layers (LC2D). This allows the model to capture more fine-grained spatial information. Additionally, a batch normalization layer follows each locally connected layer. To prevent overfitting and enhance the model’s generalization, a dropout layer with a rate of 0.5 is used after the batch normalization layer. This is followed by two fully-connected layers with 512 neurons that learn high-level abstract representations of the input data. A dropout layer with a rate of 0.65 is applied to regularize each fully-connected layer. Feature outputs of the fully-connected layers, denoted as O , focus on high-level gesture patterns instead of low-level details in the EMG data [47, 48]. Finally, the last fully-connected layer computes the predicted output from the features O . Table 3 presents a comprehensive list of the baseline network layers. The second baseline, Baseline-Proto, follows our proposed architecture but is trained directly on the gesture labels, rather than learned with the joint movement labels. These baselines help demonstrate the effectiveness of incorporating joint movement labels into the learning process and its impact on the overall performance of the model.

4.1.3. Evaluation To assess the performance of our models, we conducted both intra-subject and inter-subject evaluations. For the *intra-subject evaluation*, we followed the experimental settings in [41], where the training and test datasets were split according to the details presented in Table 2. For Ninapro DB1, we trained the models using the 1st, 3rd, 4th, 6th, 7th, 8th, and 9th trials, and then tested them on the remaining trials for each subject. Similarly, for Ninapro DB2, DB3, and DB5, the models were trained using the 1st, 3rd, 4th, and 6th trials, and tested on the remaining trials for each subject. The overall gesture classification accuracy was then averaged across all subjects in each database.

For the *inter-subject evaluation* in DB1, we employed Leave-One-Subject-Out Cross-Validation (LOSOCV). In this approach, the data from each subject were used as the test set, and the models were trained using data from all other subjects. This process was repeated for each subject, and the gesture classification accuracy was

averaged across all subjects to obtain a comprehensive evaluation of model performance. For the inter-subject experiments in DB2, DB3, and DB5, due to their increased number of subjects, we employed four-fold cross-validation by randomly splitting the subjects into four different blocks and using three blocks for training and the other for testing. These settings also follow the experimental protocol of the MV-CNN study [41].

To quantify the performance of our models, we utilized gesture classification accuracy as the primary evaluation metric. This metric indicates the overall percentage of correctly predicted gestures from the test set, providing a reliable measure of model effectiveness. In addition to accuracy, we also reported the confusion matrix for each dataset. The confusion matrix is based on the normalized sum of intra-subject gesture classification results. It offers valuable insights into the model’s performance in distinguishing between different gesture classes and helps identify any patterns of misclassifications. These evaluation methods and metrics comprehensively assess the accuracy and generalization capabilities of our proposed models across different databases and subjects, providing a robust analysis of their performance in EMG-based gesture classification.

4.1.4. Dataset Preprocessing During the experimental procedures, the EMG feature sets denoted as Phin_FS1 [58] were derived from the Ninapro DB1, DB2, DB3, and DB5 datasets. The conventional sEMG-based gesture recognition frameworks [3, 6, 9–11, 14–19, 21–23] using machine learning utilizes a sliding window data processing strategy to reduce high-dimensional signals into discriminative features, aiming to decrease computational complexity, emphasize signal characteristics, attenuate noise, and enhance generalization capabilities for building a practical gesture classification model. To ensure uniformity and adhere to established protocols, each dataset underwent feature extraction utilizing 200 ms time windows with an incremental size of 10 ms, and preprocessing using a low-pass filter with a cut-off frequency of 2 kHz, as well as a downsampling process following the experimental protocol outlined in [41]. Phin_FS1 feature set represents a widely acknowledged feature extraction method within the

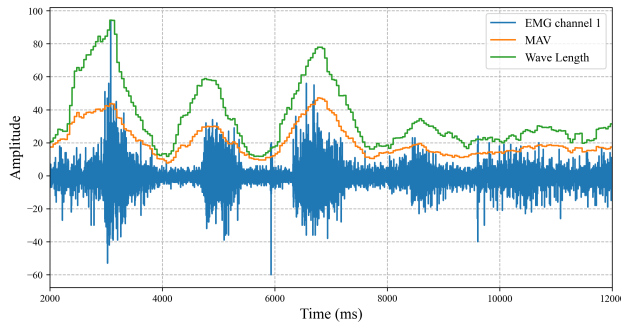


Figure 4. The similarity in temporal patterns between the extracted features and the original EMG patterns. The features were extended by the reversed length of the down-sampling process, utilizing data from subject 1 of the Ninapro DB5 dataset. To provide effective visualization, the amplitude of MAV was enlarged by three times, whereas the waveform length values were reduced to 10% from the original amplitude. The temporal span considered in this representation ranges from the initial 2000 ms to 12000 ms.

classical limb gesture recognition domain, incorporating effective time and frequency domain analyses [59] as illustrated in Table 1 of the prior work [41]. The resultant feature set comprises 11 distinct features: mean absolute value (MAV), MAV slope, waveform length, Willison amplitude, zero crossings, four autoregression coefficients, mean frequency, and power spectrum ratio. They were extracted from each EMG channel, resulting in 11 total features per time window. To prepare the data for input to the compared models for gesture classification, we transformed the extracted features into signal images. By arranging the 11 features into a 2D matrix for each time window, the resulting signal images effectively captured internal temporal patterns. For instance, MAV variations, waveform length, and preprocessed EMG signal for sensor channel 1 were visually represented in Figure 4, expressing similar patterns in the time domain, aiding in the identification of nuanced temporal dynamics.

Another essential aspect of employing window-based feature image representation lies in capturing cross-channel and cross-feature correlations. This transformation emphasizes the encapsulated correlations among different channels and features. During kernel processing in the convolution layer, features extracted from the designated region of interest are integrated, consolidating feature and sensor elements and facilitating the learning of local to global features throughout the convolution computation. This enhances the model’s capacity to discern relationships among different features and channels, thereby augmenting its comprehension of intricate patterns across multiple EMG channels and preprocessed features.

4.1.5. Model Hyperparameter Setting For our experiments, we set the number of prototypes to 100, aligning it with the number of fine-grained joint movements for effective representation. The network was trained using the

Adam optimizer [60] for a varying number of epochs, with 100 epochs for DB1 and DB5 and 400 epochs for DB2 and DB3. The batch size was set to 128 for DB1 and DB5, and 32 for DB2 and DB3. The learning rate was reduced by 50% every 20 epochs in DB1 and DB5, and similarly, with a reduction every 40 epochs in experiments involving DB2 and DB3. The initial learning rate for all datasets was set to 0.002. All our experiments were performed in a GPU environment with an RTX 3080 Ti GPU equipped with 16GB of VRAM and an RTX 3070 Ti GPU equipped with 8GB of VRAM.

4.2. Intra-Subject Performance

We begin by demonstrating the effectiveness of our proposed method in the intra-subject evaluation. The results presented in Table 4 clearly show that our method outperforms all these methods in terms of predicting gestures for the same subject. Even when compared with the state-of-the-art MV-CNN [41], which leverages multiple feature sets and a multi-view architecture, our method surpasses MV-CNN by a significant margin. It is worth noting that DB3 poses a challenge for all methods, as evident by the relatively lower accuracy scores compared to other databases. However, our method still demonstrates superior performance in this database, indicating its capability to handle the complexities of the gestures present in DB3. This indicates the effectiveness of our proposed approach in accurately recognizing gestures from EMG signals.

Comparing our method with the baseline CNN and prototype learning methods, we observe significant improvements in classification accuracy. First, without the prototype layer introduced in the proposed method, the Baseline-CNN method performs better than most state-of-the-art methods on DB1, DB2, and DB5. However, it exhibits lower accuracy on DB3. The comparison between Baseline-CNN and Baseline-Proto demonstrates that the addition of a prototype layer improves the overall classification accuracy by 2-7%. This result suggests that the prototype layer effectively captures unique joint patterns, serving as key indicators for accurate classification. Furthermore, our method surpasses Baseline-Proto, particularly notable in the substantial improvement of classification accuracy from 63.25% to 76.78% on DB3. In essence, our model’s proficiency in capturing discriminative joint movement patterns in EMG signals significantly contributes to its elevated classification performance, as evidenced by the consistent improvements across diverse databases.

The confusion matrices generated by our methodology are presented in Figure 5, offering a comprehensive analysis of the relationships between predicted and ground-truth gesture classes within the initial subject of each dataset for the intra-subject classification task. While the diagonal pattern in the confusion matrix suggests good overall

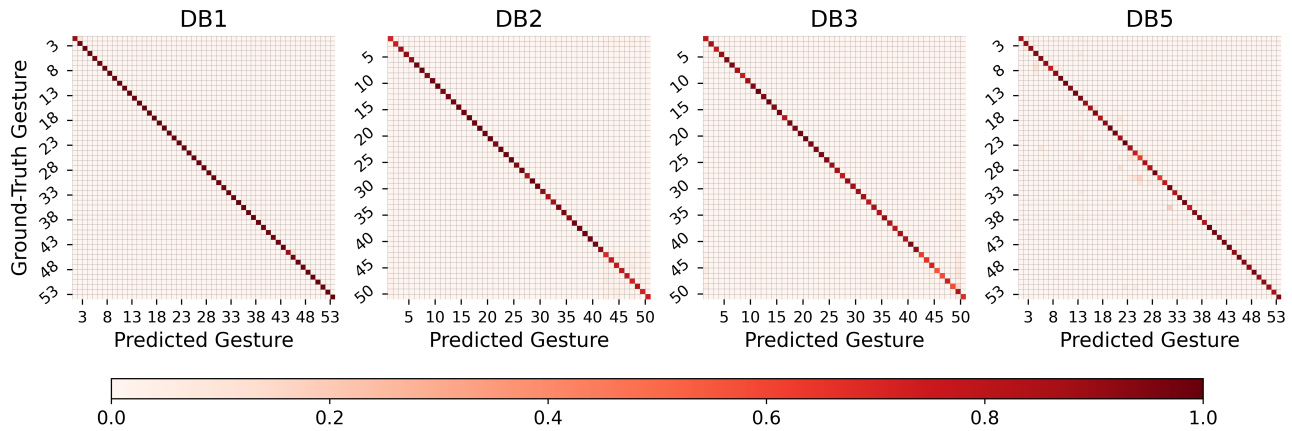


Figure 5. Illustrations of the confusion matrices of intra-subject result utilizing the initial subject for Ninapro DB1, DB2, DB3 and DB5. These matrices offer a visual representation of the performance of classification models in discerning between different hand gestures across the four databases. It is notable that the performance, as indicated by the overall classification accuracy and the distribution of classification errors, demonstrates consistently high levels of accuracy across the databases, underscoring the efficacy of the classification algorithms employed in this study. Note that the values presented are normalized within the range of 0 to 1, which corresponds to a percentage scale of 0 to 100%.

Table 4. Intra-subject gesture classification accuracy (%).

Method	DB1	DB2	DB3	DB5
GengNet [42]	77.80	-	-	-
Cheng <i>et al</i> [43]	82.54	-	-	-
Wei <i>et al</i> [44]	85.00	-	-	-
E2CNN [45]	91.27	-	-	-
Yang <i>et al</i> [46]	93.52	-	-	-
Zhai <i>et al</i> [25]	-	78.71	-	-
Ding <i>et al</i> [49]	-	78.86	-	-
Chen <i>et al</i> [26]	-	-	-	69.62
Vitale <i>et al</i> [50]	-	-	-	74.00
Peng <i>et al</i> [51]	-	-	-	77.90
AtzoriNet [52]	66.60	75.27	-	-
CNNLM [53]	79.26	78.71	-	-
EVCNN [24]	81.57	66.64	-	-
Hu <i>et al</i> [54]	87.00	82.20	-	-
Pizzolato <i>et al</i> [55]	69.45	65.49	60.42	-
MSCNet [56]	83.24	82.94	69.15	-
DVMSCNN [57]	85.72	83.29	70.58	-
MV-CNN [41]	88.20	83.70	64.30	85.40
Baseline-CNN	90.11	83.52	56.37	85.46
Baseline-Proto	93.80	85.14	63.25	88.39
Ours	96.25	88.97	76.78	90.80

accuracy across different gesture classes, a closer look reveals that specific gestures pose more difficulty for the model than others. Specifically, in the case of DB2, gestures 42 through 50 (‘Tip pinch grasp’ to ‘Cut with knife’) were more common to experience classification errors than others. For DB3, gestures 42 through 46 and

Table 5. Inter-subject gesture classification accuracy (%).

Method	DB1	DB2	DB3	DB5
MV-CNN [41]	19.00	16.00	7.00	20.00
Baseline-CNN	16.85	12.63	5.19	18.09
Baseline-Proto	20.67	18.41	8.65	20.18
Ours	23.15	21.07	9.06	23.27

48 (‘Tip pinch grasp’ to ‘Extension type grasp’ and ‘tripod grasp’) were prone to being misclassified to gesture 49 (‘Turn a screw’), while DB5, gestures 25 (‘Wrist flexion’) and 29 (‘Wrist extension with closed hands’) manifest considerably elevated error rates. In the context of DB1, marked by notably higher classification accuracies, no consistent misclassification patterns are observed. This analysis demonstrates the overall high accuracy of gesture classification within subjects, and highlights areas for further model refinement.

4.3. Inter-Subject Performance

Assessing the generalizability of a gesture classification model is of paramount importance as it ensures its effectiveness in recognizing diverse patterns exhibited by different subjects. In this study, we conducted an inter-subject evaluation to determine the extent to which our proposed deep prototype learning model can effectively generalize its learned knowledge to new subjects.

The results presented in Table 5 demonstrate the superiority of our proposed method over state-of-the-art alternatives such as MV-CNN and the two baseline methods. The MV-CNN and the Baseline-CNN methods achieve relatively low accuracy on all databases, which can

be attributed to the challenges in generalizing the model’s knowledge to new subjects. The low accuracy may be attributed to individual differences in muscle size, shape, and electrode placement affecting EMG signal generation, making models learned on one subject poorly generalizable to others. Variations in skin impedance, sweating, and electrode-skin contact can also introduce noise and artifacts, impacting classification accuracy. Despite the challenges, the Baseline-Proto method exhibits improved accuracy on all datasets. The introduction of the prototype layer helps boost performance compared to the baseline CNN, particularly on DB1 and DB3, indicating its ability to handle individual variability.

4.4. Further Performance Analyses

In this section, we delve further into the additional performance metrics to enhance the comprehensive understanding of our prototype learning model. We introduce additional metrics for both intra- and inter-subject classification tasks utilizing the same dataset, including precision, recall, F1 score, total training time, epoch required to train the model until convergence, and the inference time in Table 6. Note that the ‘Training time’ in the table measures the time required to train each data sample in a single epoch, and the ‘Inference time’ measures the time required to predict class using a single test data sample. The prolonged duration of training for inter-subject DB1 can be attributed to the utilization of the leave-one-subject-out strategy and larger epochs trained during the execution of the experimental protocol. This strategy necessitated the training of all subjects, with the exception of the subjects designated for testing. Considering the specified epochs and training duration, it is observed that DB1 and DB5 exhibit a more rapid convergence in intra-subject results compared to other databases. This empirical finding suggests that prototype learning effectively demonstrates adaptive behavior on these datasets, facilitating a relatively straightforward extraction of noteworthy prototypes. This corresponds with DB1 and DB5 having no observed consistent classification error, indicative of a well-trained model in these instances. These metrics, collectively, offer a nuanced evaluation of the classification outcomes and contribute valuable insights for refining the method’s efficacy, particularly in addressing the identified challenges associated with specific gestures.

4.5. Qualitative Results

The results presented in the previous sections highlight the effectiveness of our method in learning semantic prototypes that encapsulate a wide range of fine-grained gesture patterns, leading to improved generalizability across diverse subjects and datasets. By providing visualizations of important joint movements in gesture classification, we aim to gain further insights into the reasoning behind the model’s decisions. As shown in Figure 6, we integrate

prototype knowledge from models associated with each gesture class into the joint movements. This transformation involves computing averages across all subjects within the Ninapro DB5 dataset, and the resultant predictions are compared with the ground-truth joint movements for analysis. This analysis facilitates a comprehensive understanding of the model’s interpretative capacity and proficiency in capturing fine-grained patterns in gesture classification.

For instance, in the case of gestures 1 (index finger flexion) and 2 (index finger extension) unveil a notable alignment between the joint movements extracted from the multi-label classification model and the joint degrees measured from the raw Cyberglove dataset. Specifically, Cyberglove sensors 6 (PIP2_F) and 7 (IP2_F) consistently exhibit high activation (indicated in red color), emphasizing significant contribution in the joints of the index finger. Additionally, sensors 2 (MCPI_F), 3 (IP1_F), and 4 (CMC1_A), covering the thumb joints, display simultaneous high degrees of contribution due to the influence of proximate index finger activity. The prototypes employed in the second stage of gesture classification consistently emphasize similar joints, indicating their significant relevance in gesture identification. The consistent activation patterns of critical joints underscore the model’s ability to discern critical joints crucial for distinguishing between gestures, which further reinforces the model’s accuracy in gesture identification. These findings offer the potential for enhancing the development of more precise gesture classification algorithms, thus contributing to ongoing advancements in gesture classification systems.

5. Discussions

This section delves into the key insights garnered from our deep prototype learning approach for EMG-based gesture classification. Focusing on generalizability and interpretability, we discuss the technical novelty and advantages of our method. We then explore the real-world implications of these advancements and discuss the limitations of this work and potential future directions to push the boundaries of gesture classification in EMG applications.

5.1. Technical Novelty and Advantages

Conventional deep learning methods in EMG-based gesture classification often rely on black-box models, learning abstracted representations only from gesture classes. These methods, while effective, can struggle with inter-subject generalization and lack interpretability. To address this challenge, our study introduces a novel aspect of prototype learning by decomposing gestures into biomedically relevant joint components, which benefits the generalizability and interpretability of models. The advantages of our

Table 6. Extended evaluation of our prototype learning model on the Ninapro datasets. The metrics, including precision, recall, F1 score, training time, epoch for convergence, and inference time, provide insights into the model’s performance and adaptability across diverse tasks. Notably, datasets like DB1 and DB5 showcase rapid convergence and minimal errors, suggesting efficient training and effective prototype extraction.

	Intra-subject				Inter-subject			
	DB1	DB2	DB3	DB5	DB1	DB2	DB3	DB5
Accuracy (%)	99.33	88.69	81.33	90.04	24.69	25.83	10.13	20.91
Precision (%)	99.34	89.16	82.81	90.52	23.66	25.55	10.48	24.91
Recall (%)	99.33	88.75	80.95	90.05	24.68	26.02	9.71	20.91
F1 Score (%)	99.33	88.77	81.39	90.06	20.96	25.26	9.86	19.25
Training time (ms)	1.60	2.73	1.26	2.23	1.82	2.85	1.33	2.25
Epochs required	46	223	267	80	250	300	300	300
Inference time (ms)	1.37	2.44	1.08	2.05	1.54	2.46	1.10	1.84

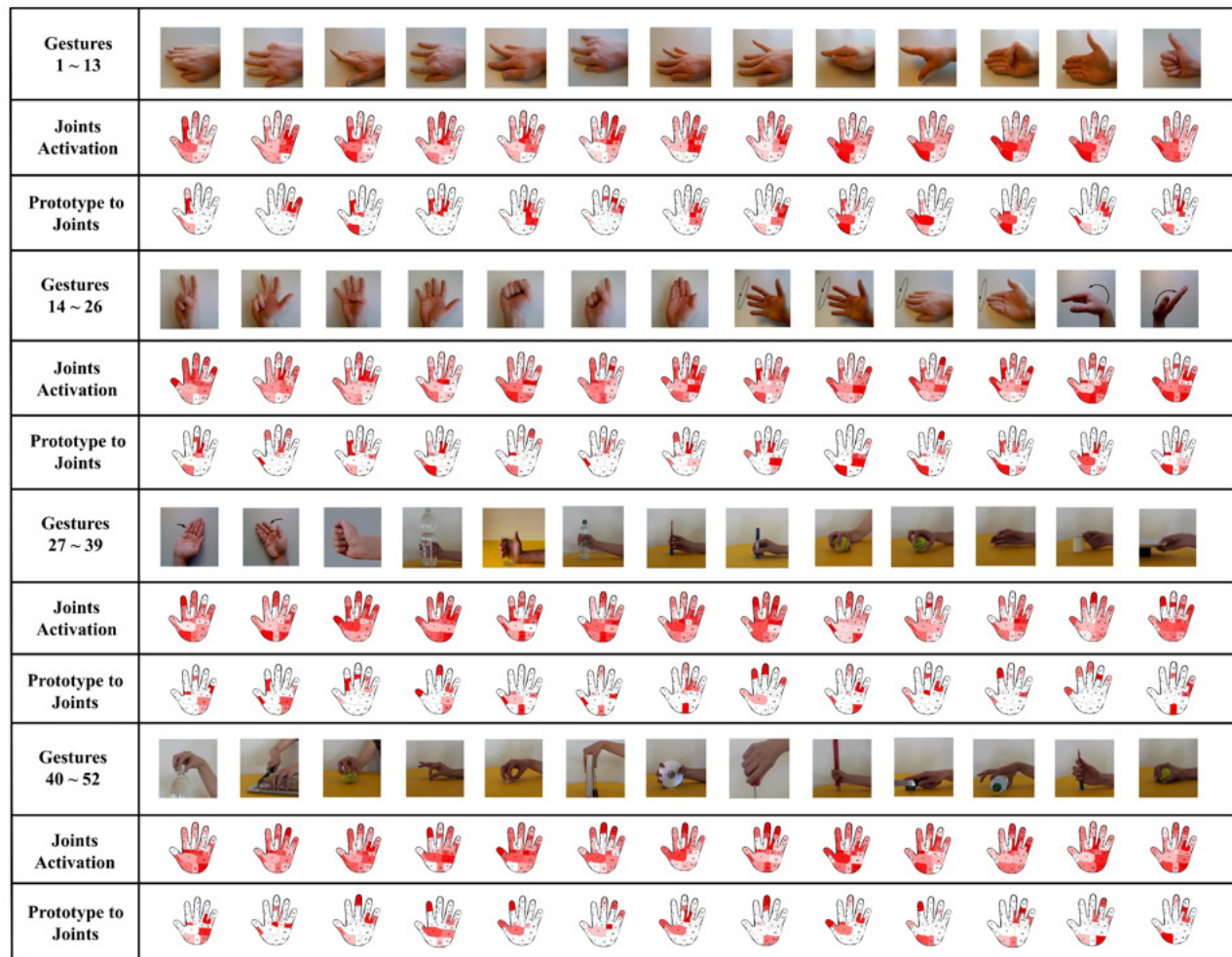


Figure 6. Qualitative visualization of the classified gestures in three aspects: gestures themselves (row 1), underlying joint movements (row 2), and the top five influential prototypes (row 3, represented by their dominant joint). Colors indicate joint activation level, aiding the interpretation of how prototypes contribute to gesture classification.

method are two-fold: (1) Our method leverages biomechanically relevant joint components, capturing core features less susceptible to inter-subject variations. This translates to superior performance across diverse subjects, as shown

in Table 5 and Table 6). (2) Unlike black-box approaches, we enable visualization of joint activation patterns for each prototype, as shown in Figure 6. This unveils the contribution of individual joints to gesture classification, offering

valuable insights for model improvement and personalization. These distinct advantages – enhanced generalizability and actionable interpretability – position our novel prototype learning approach as a significant step forward for EMG-based gesture classification. In the following, we discuss the broader impacts of generalizable and interpretable models, the limitations of this work, and future directions.

5.2. Generalizability and Real-World Significance

Inter-subject variability remains a significant bottleneck in achieving robust gesture classification for real-world applications [17]. Anatomical differences, like muscle size and electrode placement, lead to diverse EMG signal patterns even for identical gestures across individuals [66, 67]. Additionally, personalized movement strategies and health conditions further complicate the issue [16, 17]. These factors significantly impact model generalizability, often rendering them ineffective when applied to users outside the training data. Aligning with research advocating for generalizable models in prosthetics [63–65], our work presents a novel deep prototype learning approach that tackles these challenges head-on. Supported by the experimental results reported in this paper (see Table 5 and Table 6), our deep prototype learning approach offers a significant step toward real-world applications. This improvement in generalizability has a profound impact on real-world applications. For example, in prosthetics, users with varying anatomies or muscle control can experience more reliable and intuitive control of their devices [17, 68]. Beyond prosthetics, generalizable gesture classification can benefit various human-computer interfaces [69]. Individuals with diverse needs can seamlessly interact with their surroundings using intuitive gestures.

5.3. Interpretability Impacts

A major challenge in deep learning is the opacity of complex models. Users struggle to understand how these models arrive at decisions, hindering trust and limiting potential improvements. This lack of interpretability is particularly concerning in healthcare and medical domains, where transparency is crucial [61, 62]. While previous EMG-based gesture classification models often prioritized optimizing intra-subject performance through complex architectures [3, 9–11, 15–19, 21–26, 42–46, 49–57], our study emphasizes the crucial role of interpretability. Beyond accuracy, it offers significant real-world benefits. By visualizing how each joint contributes to the overall gesture classification (as shown in Figure 6), clinicians and researchers can directly understand the rationale behind our model’s decisions. In prosthetics, for example, identifying specific joint activations associated with misinterpreted gestures can guide targeted rehabilitation interventions [70, 71]. This allows therapists to focus on areas where users

might be struggling with muscle control or movement patterns. Additionally, interpretability can help highlight potential biases in the training data, ensuring the model is generalizable and effective across diverse user populations. The ability to visualize joint activation patterns can also be a valuable tool for researchers studying human movement and neuromuscular control [72]. By understanding how different gestures activate various muscle groups, researchers can gain deeper insights into motor function and rehabilitation strategies.

5.4. Limitations and Future Directions

While significantly better than existing methods, as shown in Table 5, our method experiences a noticeable decline in the inter-subject performance when trained on data from amputee subjects (DB3) compared to intact subjects (DB1, DB2, DB5). This reduction can be attributed to a variety of factors inherent to amputee populations, such as variations in the residual forearm after amputation, diverse experiences and adaptations accumulated over the years since amputation, individual health conditions affecting EMG signal measurements, and the historical extent of utilization of the amputated forearm, including differing levels of experience with prosthetic and other assistive devices. These factors collectively contribute to a highly non-IID (Independent and Identically Distributed) dataset across amputee subjects, where data distributions differ significantly between individuals. This inherent heterogeneity presents a major challenge for model generalization across the broader amputee population. Our future work will explore advanced techniques to address this issue, aiming to generalize our model more effectively across diverse amputee individuals and improve its real-world applicability. Future work could also explore personalized training protocols based on individual joint activation patterns identified through our interpretability methods. This holds promise for tailoring rehabilitation strategies to specific patient needs and improving prosthetic control efficacy.

6. Conclusion

In conclusion, this paper addresses the challenges in EMG-based gesture classification by proposing a novel deep prototype learning method. The approach achieves improved generalizability to new subjects, allowing the model to classify gestures based on a set of learned prototypes representing fine-grained joint movement patterns encoded in the EMG data. Comprehensive experiments across multiple public datasets demonstrate the classification performance and generalizability of the proposed approach. The proposed method’s capacity for generalization and prototype learning empowers both clinical practitioners and users to grasp the underlying rationale driving the model’s predictions. This broad applicability is crucial for building trust

in the technology and facilitating its adoption in real-world applications. However, it is important to acknowledge that while the results demonstrate high intra-subject accuracy, there is a significant performance gap between intra-subject performance and inter-subject performance.

7. Acknowledgement

This research is supported by Fasikl.

References

- [1] Hanger Clinic (n.d.). Limb Loss and Limb Difference: Facts, Statistics, and Resources. 2022 April, 1, 2022, from <https://hangerclinic.com/blog/prosthetics/limb-loss-and-limb-difference-facts-statistics-resources/>
- [2] Ribak S, Oliveira E-J N-D, Rosolino G-P, Orru Neto P and Tietzmann A 2018 Epidemiology of traumatic injuries of the upper limbs in a university hospital *Acta ortop. bras.* **26** 370–373
- [3] Antuvan C-W and Ma L 2019 An LDA-based approach for real-time simultaneous classification of movements using surface electromyography *IEEE Trans. Neural Syst Rehabil. Eng.* **27** 552–61
- [4] Hu Y and Wang Q 2020 A comprehensive evaluation of hidden Markov model for hand movement recognition with surface electromyography *Proc. 2020 2nd Int. Conf. on Robot., Intel. Cont. and Artificial Intelligence* pp 85–91
- [5] Sigalingging X K, Budiarsa A-P B, Leu J-S and Takada J-I 2019 Electromyography-based gesture recognition for quadriplegic users using hidden Markov model with improved particle swarm optimization *Int. J. Distrib. Sens. Netw.* **15** 1550147719862219
- [6] Chen Y, Zhang H, Wang C, Ang K-K, Ng S-H, Jin H and Lin Z 2023 A hierarchical dynamic Bayesian learning network for EMG-based early prediction of voluntary movement intention *Sci. Rep.* **13** 4730
- [7] Kusuru D, Turlapaty A-C and Thakur M 2021 A Laplacian-Gaussian Mixture Model for Surface EMG Signals from Upper Limbs *2021 43rd Annual Int. Conf. of the IEEE Eng. in Medicine & Biology Society (EMBC)* pp 681–5
- [8] Toledo-Pérez D C, Rodríguez-Reséndiz J, Gómez-Loenzo R-A and Jauregui-Correa J-C 2019 Support vector machine-based EMG signal classification techniques: A review *Appl. Sci.* **9** 4420
- [9] Zhang Z, Tang Y, Zhao S and Zhang X 2019 Real-Time Surface EMG Pattern Recognition for Hand Gestures Based on Support Vector Machine *2019 IEEE Int. Conf. on Robot. and Biomimetics (ROBIO)* pp 1258–62
- [10] Atzori M, Gijsberts A, Castellini C, Caputo B, Hager A-G M, Elsig S, Giatsidis G, Bassetto F and Müller H 2014 Electromyography data for non-invasive naturally-controlled robotic hand prostheses *Sci. Data* **1** 1–13
- [11] Du Y, Jin W, Wei W, Hu Y and Geng W 2017 Surface EMG-Based Inter-Session Gesture Recognition Enhanced by Deep Domain Adaptation *Sens., MDPI* **17** 458
- [12] Arief Z, Sulistijono I-A and Ardiansyah R-A 2015 Comparison of five time series EMG features extractions using Myo Armband *2015 int. electronics symposium (IES)* pp 11–4
- [13] Junior J-J A-M, Freitas M-L, Siqueira H-V, Lazzaretti A-E, Pichorim S-F and Stevan Jr S-L 2020 Feature selection and dimensionality reduction: An extensive comparison in hand gesture classification by sEMG in eight channels armband approach *Biomed. Signal Process. Control.* **59** 101920
- [14] Côté-Allard U, Campbell E, Phinyomark A, Laviolette F, Gosselin B and Scheme E 2020 Interpreting deep learning features for myoelectric control: A comparison with handcrafted features *Front. bioeng. biotech.* **8** 158
- [15] Parajuli N, Sreenivasan N, Bifulco P, Cesarelli M, Savino S, Niola V, Esposito D, Hamilton T-J, Naik G-R, Gunawardana U and Gargiulo G-D 2019 Real-time EMG based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation *Sens., MDPI* **19** 4596
- [16] Samuel O-W, Asogbon M-G, Geng Y, Al-Timemy A-H, Pirbhulal S, Ji N, Chen S, Fang P and Li G 2019 Intelligent EMG pattern recognition control method for upper-limb multifunctional prostheses: advances, current challenges, and future prospects *IEEE Access* **7** 10150–65
- [17] Li W, Shi P and Yu H 2021 Gesture recognition using surface electromyography and deep learning for prostheses hand: state-of-the-art, challenges, and future *Front. Neurosci.* **15** 621885
- [18] Asghar A, Khan S-J, Azim F, Shakeel C-S, Hussain A and Niazi I-K 2022 Review on electromyography-based intention for upper limb control using pattern recognition for human-machine interaction *Proc. Inst. of Mech. Eng., Part H: J. of Eng. in Med.* **236** pp 628–45
- [19] Mahmood N-T, Al-Muifraje M-H, Saeed T-R and Kaittan A-H 2020 Upper prosthetic design based on EMG: a systematic review *IOP Conf. Ser. Mater. Sci. Eng.* **978** 012025
- [20] Xiong D, Zhang D, Zhao X and Zhao Y 2021 Deep Learning for EMG-based Human-Machine Interaction: A Review *IEEE/CAA J. Autom. Sin.* **8** 512–33
- [21] Jaramillo-Yáñez A, Benalcázar M-E and Mena-Maldonado E 2020 Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review *Sens., MDPI* **20** 2467
- [22] Dhiman R, Luthra P and Singh N-T 2023 Different Categories of Feature Extraction and Machine Learning Classification Models Used for Hand Gesture Recognition Systems: A Review *2023 IEEE 8th Int. Conf. for Convergence in Technology (I2CT)* **9** pp 1–7
- [23] Bhushan S, Alshehri M, Keshta I, Chakraverti A-K, Rajpurohit J and Abugabah A 2022 An experimental analysis of various machine learning algorithms for hand gesture recognition *Electronics, MDPI* **11** 968
- [24] Olsson A-E, Björkman A and Antfolk C 2022 Automatic discovery of resource-restricted convolutional neural network topologies for myoelectric pattern recognition *Comput. Biol. Med.* **120** 103723
- [25] Zhai X, Jelfs B, Chan R-H-M and Tin C 2017 Self-Recalibrating Surface EMG Pattern Recognition for Neuroprosthesis Control Based on Convolutional Neural Network *Front. Neurosci.* **11** 379
- [26] Chen L, Fu J, Wu Y, Li H and Zheng B 2020 Hand gesture recognition using compact CNN via surface electromyography signals *Sens., MDPI* **20** 672
- [27] Yang H M, Zhang X Y, Yin F and Liu C L 2018 Robust classification with convolutional prototype learning *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 3474–82
- [28] Li O, Liu H, Chen C and Rudin C 2018 Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions *Proc. AAAI Conf. on Artificial Intelligence* **32**
- [29] Wang B, Yang K, Zhao Y, Long T and Li X 2023 Prototype-based intent perception *IEEE Trans. Multimedia.* **25** 8308–19
- [30] Chen S and Zhao Q 2023 Divide and Conquer: Answering Questions with Object Factorization and Compositional Reasoning *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 6736–45
- [31] Sziburi T, Markus N and Davide B 2021 Prototype Reduction on sEMG Data for Instance-based Gesture Learning towards Real-time Prosthetic Control *BIO SIGNALS* 299–305
- [32] Sziburi T, Markus N and Davide B 2023 Instance-based learning with prototype reduction for real-time proportional myocontrol: a randomized user study demonstrating accuracy-preserving data reduction for prosthetic embedded systems *Med. Biol. Eng. Comput.* 1–31
- [33] Wu L, Liu A, Zhang X, Chen X and Chen X 2022 Unknown Motion Rejection in Myoelectric Pattern Recognition Using Convolutional Prototype Network *IEEE Sens. J.* **22** 4305–14
- [34] Liu C, Han C, Zhou C, Cai C, Xiang S, Ni H and Qian

- D 2023 Towards Open-set Gesture Recognition via Feature Activation Enhancement and Orthogonal Prototype Learning (arXiv:2312.02535)
- [35] Atzori M, Gijsberts A, Kuzborskij I, Elsig S, Hager A G M, Deriaz O, Deriaz O, Castellini C, Müller H and Caputo B 2014 Characterization of a benchmark database for myoelectric movement classification *IEEE Trans. Neural Syst. Rehabil. Eng.* **23** 73–83
- [36] Chen Z, Wei X, Wang P and Guo Y 2019 Multi-label image recognition with graph convolutional networks *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* pp 5177–86
- [37] Lapin M, Hein M and Schiele B 2017 Analysis and Optimization of Loss Functions for Multiclass, Top-k, and Multilabel Classification *IEEE Trans. Pattern Anal. Mach. Intell.* **40** 1533–54
- [38] Jarque-Bou N J, Scano A, Atzori M and Müller H 2019 Kinematic synergies of hand grasp: a comprehensive study on a large publicly available dataset *J. Neuroeng. Rehabil.* **16** 1–14
- [39] Atzori, M, Gijsberts A, Heynen S, Hager A. G. M, Deriaz O, van P der Smagt, Castellini C, Caputo B and Müller H 2012 Building the Ninapro database: A resource for the robotics community *2012 4th IEEE RAS & EMBS Int. Conf. on Biomed. Robot. and Biomechatronics (BioRob)*, **9** pp 1258–65
- [40] Jarque-Bou N-J, Atzori M and Müller H 2020 A large calibrated database of hand movements and grasps kinematics *Sci. Data* **7** 12
- [41] Wei W, Dai Q, Wong Y, Hu Y, Kankanhalli M and Geng W 2019 Surface-Electromyography-Based Gesture Recognition by Multi-View Deep Learning *IEEE Trans. Biomed. Eng.* **66** 2964–73
- [42] Geng W, Du Y, Jin W, Wei W, Hu Y and Li J 2016 Gesture recognition by instantaneous surface EMG images *Sci. Rep.* **6** 36571
- [43] Cheng Y, Li G, Yu M, Jiang D, Yun J, Liu Y, Liu Y and Chen D 2021 Gesture recognition based on surface electromyography-feature image *Concurr. Comput. Pract. Exp.* **33** e6051
- [44] Wei W, Wong Y, Du Y, Hu Y, Kankanhalli M and Geng W 2017 A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface *Pattern Recognit. Lett.* **119** 131–8
- [45] Qureshi M-F, Mushtaq Z, Rehman M Z-U and Kamavuako E-N 2023 E2CNN: An Efficient Concatenated CNN for Classification of Surface EMG Extracted From Upper Limb *IEEE Sens. J.* **23** 8989–96
- [46] Yang Z *et al.* 2021. Dynamic gesture recognition using surface EMG signals based on multi-stream residual network *Front. bioeng. biotechnol.* **9** 779353
- [47] Bellitto G, Proietto Salanitri F, Palazzo S, Rundo F, Giordano D and Spampinato C 2021 Hierarchical domain-adapted feature learning for video saliency prediction *Int. J. Comput. Vis.* **129** 3216–32
- [48] Borys K, Schmitt Y-A, Nauta M, Seifert C, Krämer N, Friedrich C-M and Nensa F 2023 Explainable AI in medical imaging: An overview for clinical practitioners—Beyond saliency-based XAI approaches *Eur. J. Radiol.* **9** 110786
- [49] Ding Z, Yang C, Tian Z, Yi C, Fu Y and Jiang F 2022 sEMG-based gesture recognition with convolution neural networks *Sustainability, MDPI* **10** 1865
- [50] Vitale A, Donati E, Germann R and Magno M 2022 Neuromorphic edge computing for biomedical applications: Gesture classification using EMG signals *IEEE Sens. J.*, **22** 19490–9
- [51] Peng F, Chen C, Lv D, Zhang N, Wang X, Zhang X and Wang Z 2022 Gesture recognition by ensemble extreme learning machine based on surface electromyography signals *Front. Hum. Neurosci.* **16** 911204
- [52] Atzori M, Cognolato M, and Müller H 2016 Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands *Front. neurorobot.* **10** 9
- [53] Chen X, Li Y, Hu R, Zhang X and Chen X 2020 Hand gesture recognition based on surface electromyography using convolutional neural network with transfer learning method *IEEE J. Biomed. Inform. and Health Inform.* **25** 1292–304
- [54] Hu Y, Wong Y, Wei W, Du Y, Kankanhalli M, and Geng W 2018 A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition *PLoS ONE* **13** e0206049
- [55] Pizzolato S, Tagliapietra L, Cognolato M, Reggiani M, Müller H and Atzori M 2017 Comparison of six electromyography acquisition setups on hand movement classification tasks *PLoS ONE* **12** e0186132
- [56] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition (arXiv:1409.1556)
- [57] Zhang Y, Yang F, Fan Q, Yang A and Li X 2022 Research on sEMG-based gesture recognition by dual-view deep learning *IEEE Access* **10** 32928–37
- [58] Phinyomark A, Phukpattaranont P and Limsakul C 2012 Feature reduction and selection for EMG signal classification *Expert Syst. Appl.* **39** 7420–31
- [59] Chowdhury R H, Reaz M B, Ali M A B M, Bakar A A, Chellappan K and Chang T G 2013 Surface electromyography signal processing and classification techniques *Sens.* **13** 12431–66
- [60] Kingma D P and BA J 2014 Adam: A method for stochastic optimization (arXiv:1412.6980)
- [61] Vellido A 2020 The importance of interpretability and visualization in machine learning for applications in medicine and health care *Neural. Comput. Appl.* **32** 18069–83
- [62] Antoniadi A M, Du Y, Guendouz Y, Wei L, Mazo C, Becker B A and Mooney C 2021 Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review *Appl. Sci.* **11** 5088
- [63] Gehlhar R 2023 Model-Based Lower-Limb Powered Prosthesis Control: Developing and Realizing Nonlinear Subsystem Control Methods for Generalizable Prosthesis Control. Doctoral dissertation (Pasadena, CA: California Institute of Technology)
- [64] Hutchinson K, Reyes I, Li Z and Alemzadeh H 2023 Evaluating the Task Generalization of Temporal Convolutional Networks for Surgical Gesture and Motion Recognition using Kinematic Data *IEEE Robot. Autom. Lett.* **8** 5132–39
- [65] Gulati P, Hu Q and Atashzar S F 2021 Toward deep generalization of peripheral emg-based human-robot interfacing: A hybrid explainable solution for neurobotic systems *IEEE Robot. Autom. Lett.* **6** 2650–57
- [66] Botros F S, Phinyomark A and Scheme E J 2020 Electromyography-based gesture recognition: Is it time to change focus from the forearm to the wrist? *IEEE Trans. Industr. Inform.*, **18** 1 174–84
- [67] Lara J E, Cheng L K, Röhrle O, Paskaranandavivel N 2021 Muscle-specific high-density electromyography arrays for hand gesture classification *IEEE Trans. Biomed. Eng.* **69** 5 1758–66
- [68] Erik W, Cruz T, Emondi A, Langhals N, Naufel S, Peng G, Schulz B and Wolfson M 2020 Advanced technologies for intuitive control and sensation of prosthetics *Biomed. Eng. Lett.* **10** 119–28
- [69] Vuletic T, Duffy A, Hay L, McTeague C, Campbell G and Grealy M 2019 Systematic literature review of hand gestures used in human computer interaction interfaces *Int. J. Hum. Comput. Stud.* **129** 74–94
- [70] Kang P, Li J, Jiang S and Shull P B 2022 Reduce system redundancy and optimize sensor disposition for EMG–IMU multimodal fusion human–machine interfaces with XAI *IEEE Trans. Instrum. Meas.* **72** 1–9
- [71] Gozzi N, Malandri L, Mercorio F and Pedrocchi A 2022 XAI for myo-controlled prosthesis: Explaining EMG data for hand gesture classification *Knowl.-Based Syst.* **240** 108053
- [72] Yoo J, Ilhan Y, Ina Y, Kim S, Yu R, Kim K, Kim K and Lee S 2022 Residual one-dimensional convolutional neural network for neuromuscular disorder classification from needle electromyography signals with explainability *Comput. Methods Programs Biomed.* **226** 107079