



## Review

# Literature survey on machine learning techniques for enhancing accuracy of myoelectric hand gesture recognition in real-world prosthetic hand control

Hongquan Le<sup>a</sup>, Marc in Het Panhuis<sup>b</sup>, Gursel Alici<sup>a,\*</sup>

<sup>a</sup> School of Mechanical, Materials, Mechatronic and Biomedical Engineering, University of Wollongong, NSW 2522, Australia

<sup>b</sup> School of Science, University of Wollongong, NSW 2522, Australia

## ARTICLE INFO

## Article history:

Received 23 March 2025

Revised 10 June 2025

Accepted 30 June 2025

Available online 11 July 2025

## Keywords:

Muscle-computer interface

Surface electromyography

Prosthetic hand

Myoelectric control

Deep learning

Domain adaptation

Transfer learning

## ABSTRACT

The human hand, essential for performing daily tasks and facilitating social interaction, is indispensable to everyday life. Millions worldwide experience varying levels of amputation, profoundly affecting their physical, emotional, and psychological well-being, limiting independence, and reducing quality of life. Myoelectric prosthetics, the most advanced active prosthetic hands, use surface electromyography (sEMG) signals and pattern recognition to translate user intentions into control signals. Despite these advancements, high rejection rates persist due to the non-stationarity of sEMG signals, leading to inconsistent and often frustrating user experiences. As a result, clinical and academic research has increasingly focused on improving myoelectric hand gesture recognition under real-world conditions to reduce rejection rates and enhance user acceptance of myoelectric prostheses. Given the vast and diverse range of methods applied in previous research, this survey aims to systematically highlight key studies and provide an overview of the field's current achievements. Furthermore, research on machine learning for myoelectric hand gesture recognition has been largely influenced by unrelated fields of computer science, such as computer vision and natural language processing. However, myoelectric hand gesture recognition presents unique challenges, particularly severe and unpredictable covariate shifts in sEMG signals, which require specialized approaches. To address these challenges, we propose a new taxonomy for categorizing machine learning models based on feature extraction methods and decision boundary strategies. Additionally, this paper highlights the need for benchmark datasets that accurately reflect real-world conditions and emphasizes the importance of re-evaluating real-time performance, particularly when using long temporal contextual windows. This study concludes with research challenges and future research directions to enhance the accuracy of myoelectric hand gesture recognition using machine learning techniques.

© 2025 The Author(s). Published by Elsevier B.V. on behalf of Shandong University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The human hand has an extraordinary and intricate anatomical structure. With sophisticated structure featuring more than 25 degrees of freedom, 18 tendons across the wrist, 19 articulations, and opposable thumbs [1,2], it enables activities essential for daily life, social interaction, and participation, making it an indispensable part of human functionality. The loss of hands has profound physical, emotional, and psychological consequences. Physically, it limits an individual's ability to perform essential tasks [3], affecting independence and overall quality of life. Emotionally, it can impact self-esteem, body image, and social interactions, often resulting in significant psychological

distress, including depression and post-traumatic stress disorders [4]. Individuals who undergo amputation explore various types of prosthetic devices to restore the functionality of their lost segments [5].

Myoelectric-powered devices (myoelectric hands) represent a more advanced type of active prosthetic. These devices utilize electromyography signals, along with other sensory modalities such as accelerometers and gyroscopes, to interpret the user's intentions and control the terminal device [6]. Myoelectric prostheses are significantly more expensive than body-powered prosthetics due to the higher costs associated with manufacturing, fitting, training, and maintenance [7,8]. Myoelectric hands are currently the focus of extensive research in both commercial and academic fields. While body-powered prosthetics typically offer only one or two degrees of freedom, advanced myoelectric hands are designed with multiple degrees of freedom, closely mimicking the movements of a natural human hand [9,10].

\* Corresponding author.

E-mail address: [gursel@uow.edu.au](mailto:gursel@uow.edu.au) (G. Alici).

Despite rapid advances in the mechanical design of high-dexterity artificial hands [11–14], modern prosthetic hands still face significant challenges in replicating natural hand function, primarily due to control limitations [15,16]. However, full dexterity is often unnecessary for activities of daily living (ADLs) such as eating, dressing, using cutlery, and expressing body language, which can be performed with a small set of discrete hand gestures [3,17]. For this reason, hand gesture recognition has become the preferred control method for myoelectric prosthetic hands. In this approach, muscle co-activation patterns (muscle synergies) detected through EMG signals are assigned to specific hand gestures, leveraging the phenomenon of phantom hand sensation – the ability to perceive and control missing limbs [18]. This control method significantly enhances functionality compared to its predecessor, single-DOF proportional control [19]. Additionally, myoelectric hand gesture recognition can integrate components like muscle activity detection and proportional regression [20,21]. Muscle activity detection, also known as EMG onset and offset detection [22], filters out transient components of muscle activity, while proportional regression allows users to control the force, position, or movement speed of a degree of freedom by modulating the intensity of their muscle activity [19].

Despite significant advancements in the field, myoelectric prosthetic systems continue to experience high rejection rate [3, 19]. While many studies report ideal performance with recognition accuracies exceeding 90% [23–26], these results are typically achieved under controlled laboratory conditions. In real-world applications, myoelectric control faces multiple sources of covariate shift, leading to performance degradation. Recognition accuracies for hand gesture recognition can drop to below 50% under real world conditions [26–28]. The causes of this degradation can be broadly classified into two categories:

- **Physiological and Environmental Shifts:** These include changes in electrode placement, variations in skin-electrode impedance [29], and limb positioning [30,31].
- **Behavioral Shifts:** These involve variations in contraction force [26], muscle fatigue [32], and natural concept drift caused by inconsistencies in how users perform hand movements [33].

These factors introduce substantial variability, making it challenging for myoelectric systems to maintain consistent performance in real-world conditions.

Addressing the non-stationarity of myoelectric hand gesture recognition has been a central focus of recent research. This survey aims to systematically, analyze, and consolidate existing strategies for improving recognition accuracy in real-world applications. The highlight main contributions of this study are:

- We propose a new taxonomy for categorizing myoelectric hand gesture recognition models based on feature extraction methods and decision boundaries, emphasizing its significance in designing effective machine learning models for myoelectric control.
- We provide an in-depth review of the field's advancements, covering traditional machine learning, deep learning, and various strategies for overcoming the challenges posed by non-stationarity in myoelectric hand gesture recognition.
- We identify immediate research challenges and directions in myoelectric hand gesture recognition, highlighting the need for a benchmarking dataset that accurately reflects real-world conditions, improving model interpretability, reevaluating real-time delay, and supervised transfer learning of temporal features.

The remaining sections are organized as follows: Section 2 provides an overview of fundamental concepts related to myoelectric hand gesture recognition, including the generation of sEMG signals, definitions of spatial and temporal aspects in this context, and our newly proposed taxonomy for classifying myoelectric hand gesture recognition approaches in terms of feature extraction methods and decision boundaries. Section 3 focuses on advancements and notable work in classical machine learning for myoelectric hand gesture recognition, covering topics such as feature engineering, the impact of decision boundaries, and automatic feature and channel selection. Section 4 reviews major developments in deep learning-based models, including convolutional neural networks, recurrent neural networks, and transformer-based architectures. Section 5 examines various strategies for addressing sEMG non-stationarities, analyzing their respective advantages and limitations. Section 6 outlines current challenges and immediate research directions in the field, aiming to improve recognition accuracy and user acceptance. Finally, Section 7 presents concluding remarks.

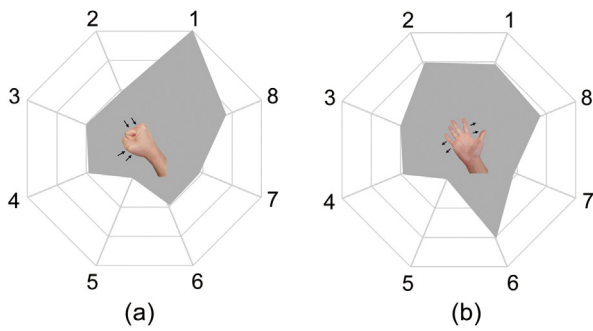
## 2. Overview of machine learning approach for myoelectric hand gesture recognition

### 2.1. Generation, collection of sEMG signal

The EMG signal contains valuable information that reflects muscle activity. It results from the electrical signals generated by the central nervous system, which are transmitted to the neuromuscular junctions and then propagate along muscle fibers to the tendon ends. These electrical signals trigger muscle contraction by causing sarcomere filaments to slide over one another. The central nervous system regulates contraction force and velocity by adjusting the number of recruited motor units and their firing rates. sEMG collected from the surface of the skin represents the summation of all motor unit action potential (MUAP) trains within the recording area [19,34]. There are two primary methods for recording EMG signals: invasive intramuscular EMG and non-invasive surface EMG. Intramuscular EMG involves inserting needle electrodes deep into the muscle tissue to collect data. In contrast, surface EMG is a non-invasive and harmless method of signal acquisition. However, compared to intramuscular EMG, surface EMG is more susceptible to noise, as the signal must pass through layers of fat and skin tissues. These layers introduce Gaussian noise and impedance, altering the signal's characteristics [35].

### 2.2. Definition of temporal and spatial information in myoelectric control systems

Action recognition, including myoelectric hand gesture recognition, is inherently a spatiotemporal problem because it requires interpreting actions that unfold across both spatial and temporal dimensions [36,37]. In myoelectric hand gesture recognition, spatial information refers to localized muscle activity and muscle synergies, which denote the coordinated activation of groups of muscles working together to produce controlled movements [38]. Spider plot is one method used to visualize spatial information in sparse EMG systems (Fig. 1). The use of spider plot visualization has been featured in commercial products as a visual guide for both the prosthetic user and the prosthetist during the user training phase, and showing significant information in a single plot. Clearly distinguished hand gestures should exhibit distinct spatial activation patterns, as shown in Fig. 1 with two representative gestures: closed hand and the open hand. Temporal information in surface electromyography relates to the time-dependent



**Fig. 1.** Spider plot visualizing myoelectric spatial activation of (a) closed and (b) opened hand gesture in sparse EMG systems with 8 data points or sEMG electrode readings.

characteristics of the electrical activity generated by muscle contractions. This includes features such as signal amplitude, information entropy [25], and frequency spectrum [30], which are linked to motor unit firing rates and recruitment patterns during contractions [39].

### 2.3. Taxonomies for categorizing myoelectric hand gesture recognition models

Machine learning models in myoelectric hand gesture recognition are commonly categorized as either traditional machine learning or deep learning approaches [40]. Traditional machine learning employs classifiers such as Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Random Forests (RF), and shallow Feedforward Neural Networks. These models are typically simpler and have relatively fewer trainable parameters. In contrast, deep learning leverages complex architectures like deep convolutional neural networks and recurrent neural networks. While these models are more expressive due to their larger number of trainable parameters, they are also more prone to overfitting. Deep learning models are capable of directly extracting discriminative features from raw sEMG signals [29,41]. Deep learning models can be divided into two distinct phases: feature extraction and classification, where the classification phase is often performed using feedforward neural networks.

However, the designation of traditional machine learning and deep learning might not be useful as it does not adequately capture the overlap between the two approaches. Alternatively, we propose a new taxonomy machine learning models for myoelectric hand gesture recognition based on feature extraction approach and decision boundaries. Compared to the categorization of models as simply traditional or deep learning, our proposed taxonomy draws more attention to the specific mechanisms by which each model operates. This categorization framework offers greater flexibility and provides more practical guidance for designing novel machine learning models for myoelectric hand systems.

#### 2.3.1. Categorization based on the feature extraction approach

In terms of categorization based on **feature extraction approach**, myoelectric hand gesture recognition models can be classified into three categories:

- **Spatially Agnostic Feature Extraction (Fig. 2a):** These models extract the same set of temporal features across all channels. This is the most commonly used approach in myoelectric hand gesture recognition due to its simplicity and minimal computational requirements [25]. It also strictly enforces a temporal-spatial divide-and-conquer strategy [24],

as the temporal features are extracted independently of the spatial arrangement of the channels. Furthermore, because classification is based solely on spatial information, specifically, muscle synergies, this approach significantly enhances model interpretability.

- **Spatially Specific Feature Extraction (Fig. 2b):** These models extract a different set of features for each channel. In this configuration, only temporal features are extracted during the feature extraction stage. However, since the features extracted still depend on their respective spatial locations, the temporal-spatial separation is not as strict as in the Spatially Agnostic Feature Extraction configuration [24]. Moreover, most work in this configuration has appeared in the form of channel and feature selection approaches (discussed in detail in Section 3.3), with some deep CNN configurations also falling under this configuration [24].
- **Spatiotemporal Feature Extraction (Fig. 2c):** These models extract both spatial and temporal features during the feature extraction stage. Examples include most configurations of deep learning models [23,41,42] or cross-channel correlation measures [30]. Compared to Spatially Agnostic and Spatially Specific Feature Extraction, this configuration is the least interpretable because it combines spatial and temporal information in a way that makes it difficult to disentangle their individual contributions during inference.

#### 2.3.2. Categorization based on the decision boundary

Machine learning models can also be categorized based on their decision boundaries into linear and non-linear types. Linear decision boundaries separate different classes in feature space using hyperplanes. Their primary advantage lies in their simplicity, which makes them easier to understand and visualize. Furthermore, because they involve fewer parameters, models with linear decision boundaries are less prone to overfitting. In contrast, non-linear decision boundaries use complex, non-linear shapes to separate classes in feature space. These boundaries are particularly well-suited for classifying data that cannot be linearly separated, making them robust for handling complex datasets. However, their complexity can increase the risk of overfitting, potentially degrading model performance when classifying out-of-distribution samples. The significance of decision boundaries is often overlooked in the context of myoelectric hand gesture recognition, especially when machine learning models are exposed to complex and unpredictable covariate shifts caused by physiological, environmental, and behavioral factors.

In Fig. 3, we showcase the decision boundaries of 16 popular machine learning classifiers applied to the same 3-class classification dataset. These classifiers include Neural Network (or multilayer perceptron MLP), LDA, QDA, SVM, RF, k-NN, Nearest Centroid, and Logistic Regression. It is important to note that SVM and LR do not natively support multiclass classification. Therefore, for these two classifiers, the one-vs-one multiclass classification scheme is visualized. In general, non-linear decision boundaries demonstrate a high diversity in their configurations, with some being highly segmented and fragmented, as illustrated in Fig. 3(b)(i)(j). In contrast, linear decision boundaries produce relatively consistent results across different classification methods, as shown in Fig. 3(a)(c)(e)(k)(l) in our simulated experiment. In high-dimensional feature spaces, linear classifiers such as MLP, SVM, Nearest Centroid, and Logistic Regression derive decision boundaries as hyperplanes in the feature space. In contrast, LDA takes a different approach by first identifying a discriminant subspace of dimension  $n_{\text{classes}} - 1$ , where it maximizes between-class variance and minimizes within-class variance [43], subsequently deriving the decision boundary on this discriminant subspace.

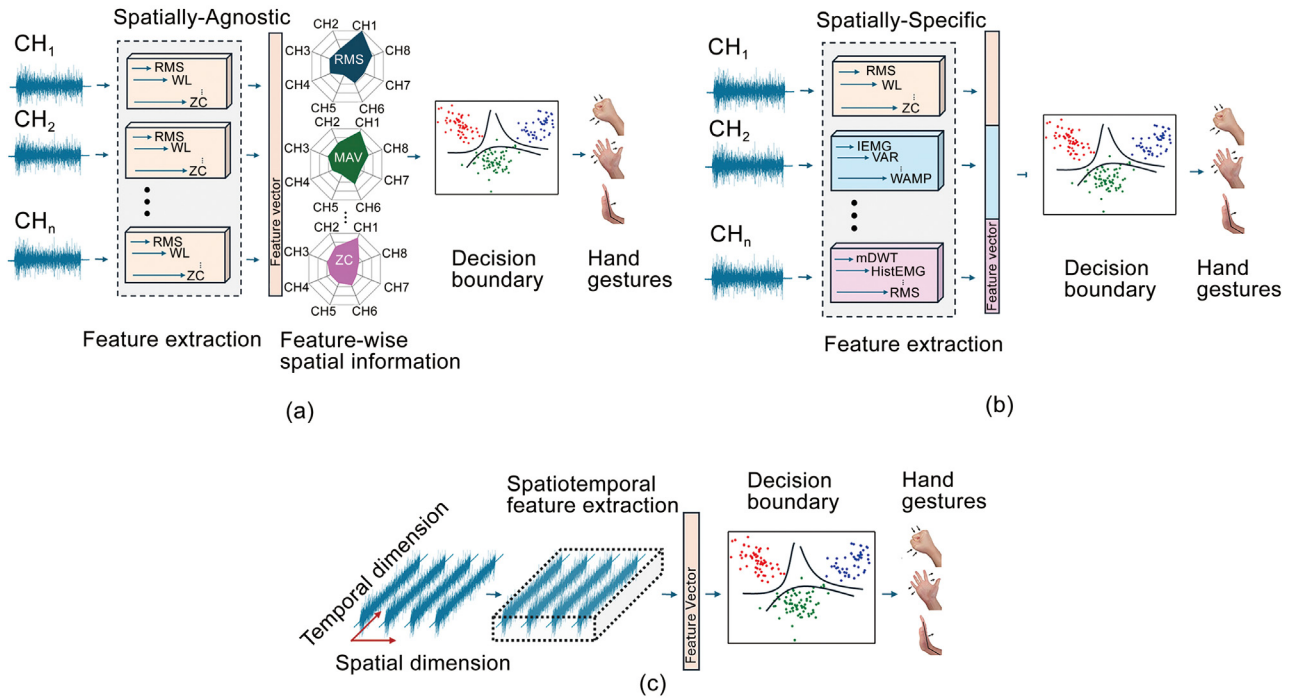


Fig. 2. Three categories of myoelectric hand gesture recognition: (a) Spatially Agnostic, (b) Spatially Specific, and (c) Spatiotemporal Feature Extraction.

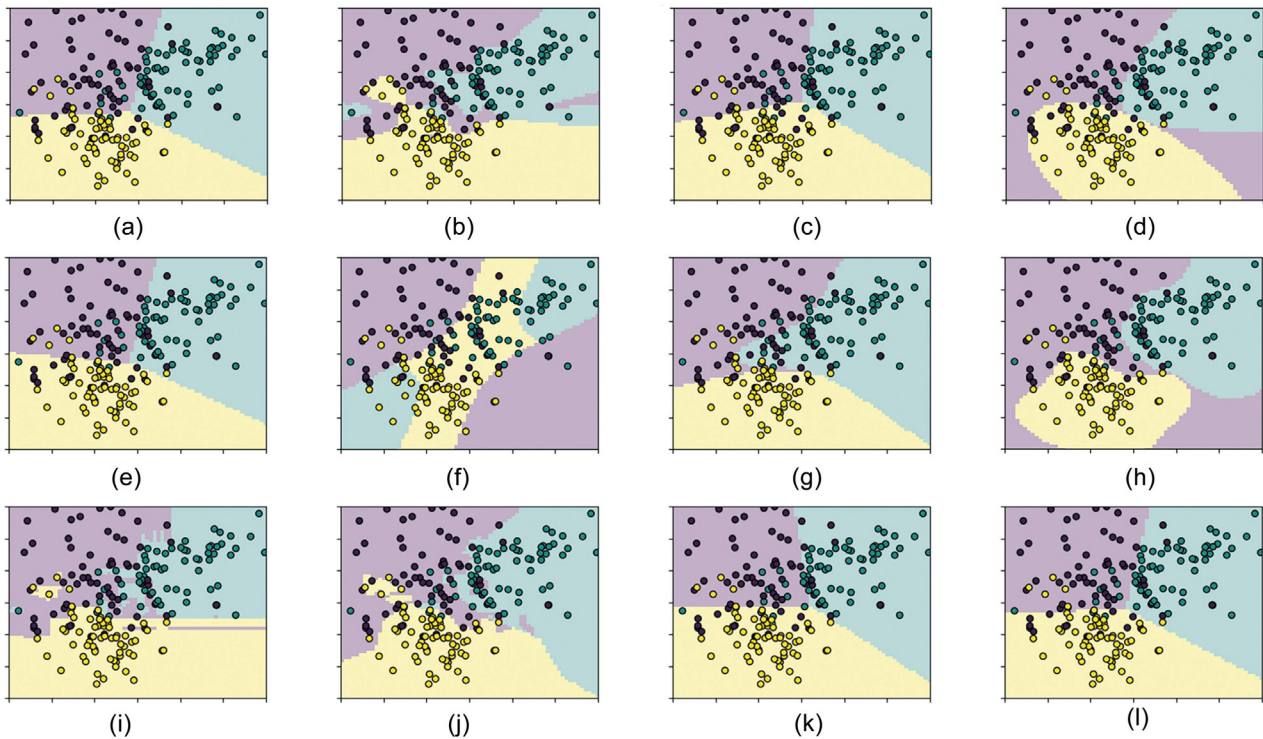


Fig. 3. Illustration of the decision boundaries of 12 commonly used classifiers for myoelectric hand gesture recognition on an example three-class classification problem represented by three colors. (a) Multilayer perception without activation function. (b) Multilayer perception with ReLU activation. (c) Linear discriminant analysis. (d) Quadratic discriminant analysis. (e) Linear support vector machine. (f) 2nd Polynomial kernel support vector machine. (g) 3rd polynomial kernel support vector machine. (h) Radial basis function support vector machine. (i) Random forest. (j) k-Nearest neighbor. (k) Nearest centroid. (l) Logistic regression.

**Table 1**  
Commonly used feature sets for myoelectric hand gesture recognition.

Name	Feature set
AR	Autoregressive coefficient [46]
ATZORI's	rms, mdwt, hemg, mav, wl, ssc, zc [47]
DU's	iav, var, wamp, wl, ssc, zc [47]
HUDGIN's	mav, wl, ssc, zc [47]
PHIN's_1	mav, wl, wamp, zc, mavs, ar, mnf, psr [47]
PHIN's_2	SampEn, cc, rms, wl [25]
TD5	mav, wl, var, ssc, zc [48]
TD5AR	mav, wl, var, ssc, zc, ar [48]
TD5AR_ALT	mav, ssc, wl, var, wamp, arc, zc [49]
CWTC	Continuous Wavelet Transform Coefficients [50]
DWPTC	Discrete Wavelet Packet Transform Coefficients [51]
DWTC	Discrete wavelet transform coefficients (DWTC) [52]
TD_PSD	Temporal spectral descriptor [53]
DOSWALD's	Hilbert-Huang Transform coefficient and mav, wl, wamp, zc, mavs, ar, mnf, psr
FFT	Fast Fourier Transform coefficient [54]
STFT	Short time Fourier Transform [55]

**Feature Acronyms and Definitions:** Average Amplitude Change (aac), Auto-Regressive Coefficient (ar), Cepstrum Coefficient (cc), Difference Absolute Standard Deviation Value (dasdv), Histogram of EMG (hemg), Integral Absolute Value (iav), Log Detector (log), Log-variance (logvar), Marginal of Discrete Wavelet Transform (mdwt), Maximum Fractal Length (mfl), Mean Absolute Value (mav), Mean Absolute Value Slope (mavs), Mean Frequency (mnf), Myopulse Percentage Rate (myop), Power Spectrum Ratio (psr), Root Mean Square (rms), Sample Entropy (sampen), Slope Sign Change (ssc), Time-Domain Power Spectral Moments (TD-PSR), Variance (var), V-Order (v), Waveform Length (wl), Willison Amplitude (wamp), and Zero Crossing (zc).

### 3. Classical machine learning for myoelectric hand gesture recognition

#### 3.1. Feature engineering of sEMG signal

Feature extraction and feature selection are crucial steps for developing a successful machine learning model. In addition, feature extraction helps remove unwanted noise and isolate relevant, discriminative information. In the context of prosthetic control, extracting features from EMG signals involves quantitatively describing signal characteristics that accurately convey the muscle contraction process [44]. Table 1 highlights commonly used feature sets for sEMG machine learning models. Among these, features representing the low-frequency envelope of sEMG signals, such as the mav and rms, are widely used in myoelectric human-machine interfaces [15,45].

Statistical features such as variance (VAR), standard deviation (STD), and kurtosis are popular for myoelectric analysis. These features indicate the probability density function of the signal, from which the contraction level can be approximated [56,57]. Sample entropy (SampEn), which measures the complexity of time-series signals, is a good indicator of muscle activity. SampEn has been shown to correlate with muscle contraction level and EMG amplitude [58,59].

Time-series modeling features, such as autoregressive coefficients (AR), have been shown to help reduce classification error rates [60]. An autoregressive model represents a signal as a linear combination of its past values and a stochastic error term. Mathematically, it can be expressed as [46]:

$$x[n] = \sum_{i=1}^p a_i x[n-i] + e[n], \quad (1)$$

Where  $p$  represents the order of the model,  $x[n]$  is the current value of the time series,  $x[n-i]$  denotes the  $i$ th past value of the signal  $x$ ,  $a_i$  are the AR coefficients, and  $e[n]$  is the white noise error term. Additionally, the AR coefficients have direct relationship

with the power spectral density of signal, which can be calculated as [46]:

$$P(f) = \frac{1}{|1 + \sum_{i=1}^p a_i e^{-j2\pi f i}|^2} \quad (2)$$

Frequency-domain features provide information about muscle fatigue [61]. They are primarily extracted using Fast Fourier transform. The most common frequency features found in the literature are mean frequency (MNF) and median frequency (MDF), which are [44]:

$$MNF = \frac{\sum_{j=1}^M f_j P_j}{\sum_{j=1}^M P_j}, \quad (3)$$

$$\sum_{j=1}^{MDF} P_j = \sum_{j=MDF}^M P_j = \frac{1}{2} \sum_{j=1}^M P_j, \quad (4)$$

where  $f_j$  and  $P_j$  is the frequency value and power spectrum at the  $j$ th bin of sEMG spectrum.

Compared to conventional time-domain and frequency-domain features, time-frequency features offer greater benefits for controlling neuroprosthetics and analyzing muscle activity. Time-frequency domain features provide richer two-dimensional information, revealing patterns in both time and frequency [55]. Time-frequency domain features can be generated using Short-Time Fourier Transform (STFT) or Wavelet Transform (Fig. 4). STFT operates by performing consecutive Fourier transforms on windowed data. However, the size of the window significantly affects the results of STFT [55]. Furthermore, because the Fourier transform decomposes signals into constituent sinusoidal components, it is less effective for analyzing the non-stationary and stochastic characteristics of EMG signals [55,62]. In contrast, Wavelet Transform overcomes the drawbacks of STFT. It does not require selecting appropriate window size [55], making it more adaptable for non-stationary signal analysis. Additionally, for hand gesture classification, wavelet-based time-frequency features have been shown to improve accuracy rates compared to those derived from STFT [63,64].

#### 3.2. Impact of decision boundaries on myoelectric hand gesture recognition

In this section, we summarize how the choice of decision boundaries impacts myoelectric hand gesture recognition performance. Specifically, we focus on spatially agnostic feature extraction models, where the same feature set is extracted for all sEMG channels (Fig. 2a).

In scenarios where the number of hand gestures is relatively small, the linear decision boundary of LDA has been shown to be more robust under covariate shifts in sEMG data. Phinyomark, et al. [25] conducted a multi-day study on feature selection with dataset containing seven hand gestures and four electrodes. To achieve consistent cross-day hand gesture recognition setup, skin-safe marker was used to mark the precise location of sEMG electrode. They found that LDA outperformed RF and QDA in cross-day scenarios without retraining. Using only the sample entropy feature, LDA achieved a recognition accuracy of 93.37%, compared to 88% with RF and 86% with QDA. With the SampEn, CC, RMS, and WL feature set, LDA's accuracy rose to 97.75%, while RF achieved 89% and QDA dropped to 78%. In a similar cross-day experimental setup involving eight electrodes and eight hand gestures, Vidovic, et al. [65] reported that same-session recognition accuracy was comparable between LDA and QDA with the logarithm of signal variance for both intact and amputee subjects. However, in cross-day scenarios without domain adaptation, LDA consistently outperformed QDA. Specifically, for

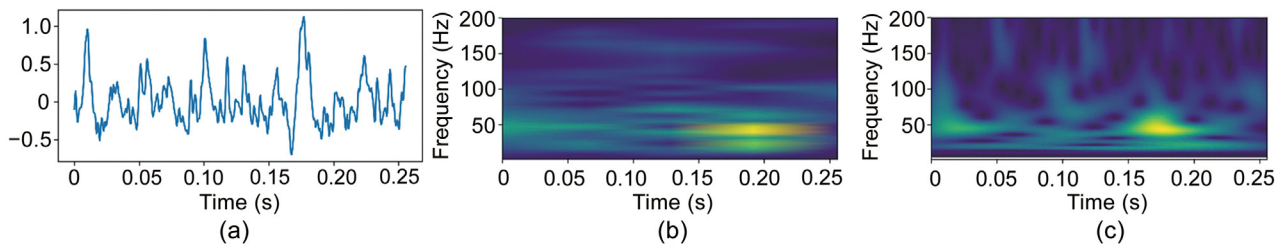


Fig. 4. Comparison between (a) Raw sEMG Input, (b) Short-Time Fourier Transform and (c) Continuous Wavelet Transform.

amputee subjects, LDA achieved a cross-day performance of over 60%, while QDA was approximately 57%.

Al-Timemy, et al. [26] conducted an experiment on sEMG data involving variations in contraction force on amputee subjects. They found that when the training dataset included hand gestures recorded across multiple force levels (low, medium, and high), LDA achieved the highest recognition accuracy at 82.58%, outperforming RF and k-NN on multiple feature sets, which achieved accuracy of 82.03% and 80.86%, respectively. Similar results were observed in “train-one-force, test-all-forces” evaluation scheme, where LDA continued to return superior performance. Lorrain, et al. [66] investigated dynamic sEMG hand gesture recognition, including both stationary and transient segments with 6 pairs of electrodes and 9 classes of hand motion. Using a threshold-based method with a time-domain 5 feature set, they found that LDA achieved better recognition accuracy (92.13%) compared to RBF SVM, which achieved 90.81% for the one-versus-one variant and 78.5% for the one-versus-rest variant.

### 3.3. Automatic feature and channel selection in traditional myoelectric hand gesture classification

Feature and channel selection are critical in traditional classifiers as they directly impact recognition accuracy and computational efficiency. First, it helps minimize computational time, as reducing the number of features or channels decreases the dimensionality of the data, leading to faster processing and reduced resource consumption during training and inference [67]. Second, feature selection and channel selection help to remove unwanted inferences, which may compromise recognition accuracy. For example, Liao et al. [11] observed that their proposed envelope feature set significantly helped maintain cross-day, uncalibrated recognition accuracy compared to the TD5 feature set. Moreover, it has been observed that the majority of sEMG manual feature extractions are highly redundant and satisfactory recognition performance can be achieved with a small number of electrodes [67,68]. Dimensionality reduction techniques like Principal Component Analysis (PCA) [69] and feature selection both reduce dimensionality but differ in several key aspects. PCA, as an unsupervised method, does not consider discriminative relationships with output labels and may remove low-variance directions that are crucial for class separability [70]. Additionally, PCA still requires full computation for every dimension in the feature space, leading to computationally expensive inference, even after dimensionality is reduced prior to classification. In contrast, feature or channel selection methods completely disable the computation of the excluded dimensions.

Feature selection, in its simplest form, can be achieved by identifying highly discriminative features using metrics such as the ANOVA F-score, a statistical measure that quantifies the correlation between continuous input features and their ability to separate output classes. Liang, et al. [71] explored the interaction between feature selection methods and classifiers on a dataset of sEMG signals recorded from 2 channels across 20 subjects

performing 9 hand gestures. A total of 76 features were extracted, with 52 derived from raw sEMG signals and 24 from the sEMG envelope. They found that optimal recognition performance of 95% was achieved using a subset of 33 features, selected via the Relief-F metric and classified using the k-NN algorithm. Miah, et al. [72] introduced a feature selection metric based on the  $p$ -value from logistic regression. The proposed method was evaluated on two public datasets: the Myo Armband 21 Gesture (MA21) dataset and the UC2018 Dual Myo (UC8) dataset. On the MA21 dataset, the feature selection approach consistently outperformed the original feature set, achieving a high recognition accuracy of 97.26% with the Extra Trees Classifier, 0.3% improvement over the accuracy without feature selection. Similarly, on the UC8 dataset, their method achieved accuracy of 97.33% with feature selection compared to 96.61% without it. Al-Angari, et al. [73] investigated feature selection in scenarios involving changes in upper limb positions. They employed two methods: distance-based feature selection approach using Mahalanobis distance to distinguish between classes, and correlation-based feature selection method that utilized mutual information between features and classes. These methods were evaluated on a dataset comprising 5 upper limb positions and 9 distinct hand gestures. The results showed that both feature selection methods outperformed the commonly used 5-feature time-domain set and the complete 10-feature time-domain set across all subjects and upper limb positions.

However, previous methods relying solely on class separation as the metric for feature selection often result in highly redundant feature sets. To optimize recognition performance while minimizing the number of required features and channels, the Minimum Redundancy Maximum Relevance (mRMR) automatic feature selection method is employed [74]. Relevance is typically measured using the ANOVA F-score, while redundancy is evaluated through Pearson correlation, which quantifies the linear relationships between input features [74].

Khushaba and Al-Jumaily [67] proposed an optimization method for feature selection based on the mRMR criterion. Using a privately recorded dataset of 10 hand gestures from 6 subjects, with signals collected from 16 electrodes, they demonstrated that satisfactory recognition performance (97% accuracy) could be achieved using only 3 electrodes. Furthermore, by applying the same technique for spatially specific feature selection (Fig. 2b), they found that using 10 features per channel resulted in a recognition accuracy of 99%. Liu, et al. [75] evaluated two feature reduction methods, namely Minimum Redundancy Maximum Relevance (mRMR) and Markov Random Field (MRF), to optimize myoelectric pattern recognition for individuals with spinal cord injuries. They found that MRF was more effective in selecting features compared to mRMR and the conventional Sequential Forward Selection (SFS) method. Using a HD-EMG system with 57 channels, they demonstrated that a subset of 10 features, selected from a total of 570 features, achieved satisfactory recognition performance with 95% accuracy.

Mesa, et al. [76] introduced a multivariate variable selection method, mRMR-F-Test Correlation Out (mRMR-FCO), to identify

**Table 2**  
Summary of automatic feature selection methods for myoelectric hand gesture recognition.

Method	Spatially specific
Liang, et al. [71] Feature significant ranking by Pearson Correlation, F-test, Chi-square test, Relief-F	Yes
Miah, et al. [72] Logistic Regression $P$ -value	Yes
Al-Angari, et al. [73] correlation-based method and distance-based method	Yes
Khushaba and Al-Jumaily [67] Particle Swarm Optimization and mRMR	Yes
Liu, et al. [75] Markov random field and mRMR	Yes
Mesa, et al. [76] mRMR-F-test Correlation Out	No

the most informative and least redundant combinations of sEMG channels and signal features for hand-gesture recognition. By evaluating 32 electrode positions and 86 signal features across 17 subjects, they demonstrated that a reduced set of 7 channels and 7 features could achieve high classification rates (86.4%) for 14 static hand gestures. Optimal recognition performance was achieved with a combination of sEMG electrodes placed both across and along the forearm. Unlike previous studies, their proposed automatic feature selection method returned optimal spatially agnostic feature set (see Table 2).

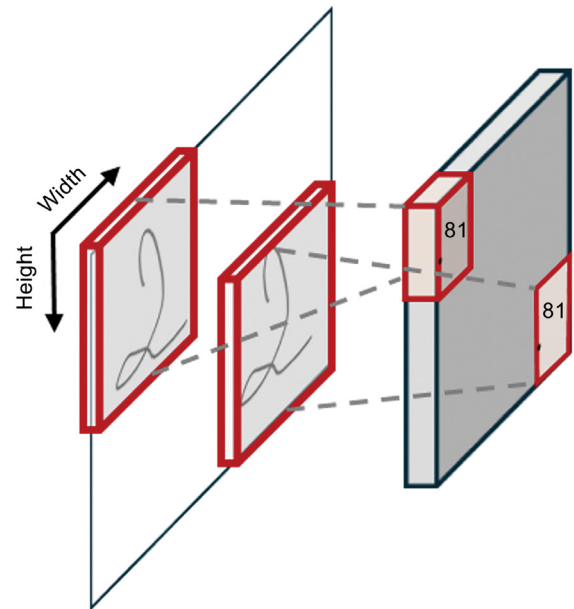
#### 4. Deep learning for myoelectric hand gesture recognition

In the previous section, we have critically analyzed the application of classical machine learning in myoelectric hand gesture recognition, emphasizing the importance of feature selection [25] and the impact of decision boundary selection [65,77]. Traditional manual feature extraction relies heavily on expert knowledge and a deep understanding of the nature of sEMG signals. In contrast, advancements in deep learning have driven significant interest in its application to sEMG hand gesture recognition. Unlike traditional machine learning, deep learning offers the advantage of data-driven feature extraction, where features are adaptively learned directly from raw sEMG data, reducing reliance on manual engineering. Given the extensive literature available on deep learning for myoelectric hand gesture recognition, in this section, we focus on highlighting notable studies while placing significant emphasis on the underlying mechanisms of deep learning models to serve as a foundation for future research.

##### 4.1. Convolutional Neural Network models for learning translation-invariant features

Convolutional Neural Networks (CNNs) marked a significant breakthrough in image recognition with the introduction of LeNet in the 1990s, primarily due to their inherent property of translational invariance, also referred to as translational-invariant inductive bias [78]. This property enables CNNs to efficiently identify spatial hierarchies and patterns within images, regardless of their location in the frame [78,79] (Fig. 5). Since then, CNNs have become a cornerstone for achieving state-of-the-art performance across various domains, extending beyond computer vision into fields such as audio processing [80]. Leveraging this success, CNNs have also been widely applied to surface electromyography (sEMG) for hand gesture recognition. Notable early applications include Atzori, et al. [41], who adapted a modified LeNet architecture for gesture recognition using the public Ninapro dataset, and Geng, et al. [81], who employed CNNs to classify hand gestures from HD-sEMG data.

In the early work by Atzori, et al. [41], they proposed the use of a 2D-CNN (Fig. 6) based on LeNet and tested this model



**Fig. 5.** Illustration of a 2D convolution operation applied to a 2D image containing the number '2' in two different locations. The same receptive filter slides over the input image, resulting in the same activation values in the corresponding feature map for both locations.

on the Ninapro dataset. Similar to the classical machine learning pipeline, raw sEMG was first segmented using a window size of 150 ms. Their model achieved recognition accuracy comparable to that of classical machine learning approaches. Specifically, the accuracy was 66.59% on DB1, 60.28% on DB2, and 38.09% for amputee subjects in DB. Due to the convolutive kernels spanning across both spatial and temporal dimensions, the CNN model's architecture led to the extraction of spatiotemporal features.

While Atzori, et al. [41] work focused on sparse sEMG systems, Geng, et al. [81] (Fig. 7) experimented with both HD-sEMG and sparse sEMG systems. The input to their model was instantaneous sEMG images. For HD-sEMG data, this image was structured as a matrix corresponding to the rows and columns of the HD-sEMG array. In the case of sparse sEMG systems, the sEMG image was reshaped into  $1 \times \text{number-of-channels}$  format. Additionally, majority voting was applied to improve recognition accuracy. On their privately recorded dataset of 8 hand gestures, their proposed method achieved 89.3% recognition accuracy using only instantaneous HD-sEMG images, which improved to 99.0% when majority voting with 149 samples was employed. On the sparse Ninapro DB1 dataset with 8 hand gestures, the method achieved 78.1% recognition accuracy using instantaneous images, which improved to 84.5% when majority voting with a 200 ms window was applied. It is important to note that in the model proposed by Geng, et al. [81], the CNN focused solely on extracting spatial features (Fig. 7).

Translational invariance is a fundamental principle of CNN architecture (Fig. 5). This property is achieved through the convolution operation, where a filter slides over the input data to detect features regardless of their position in the input. As CNNs were originally designed for image processing, adapting these techniques for myoelectric hand gesture recognition requires careful consideration, particularly in the arrangement of sEMG channels. Both [41,81] utilized CNNs to classify hand gestures, but their approaches differ, especially in how they handle the spatial arrangement of HD-EMG and sparse sEMG data. For HD-EMG, the spatial location and adjacency of each channel were preserved to reflect their physical arrangement on the sensor grid (Fig. 7).

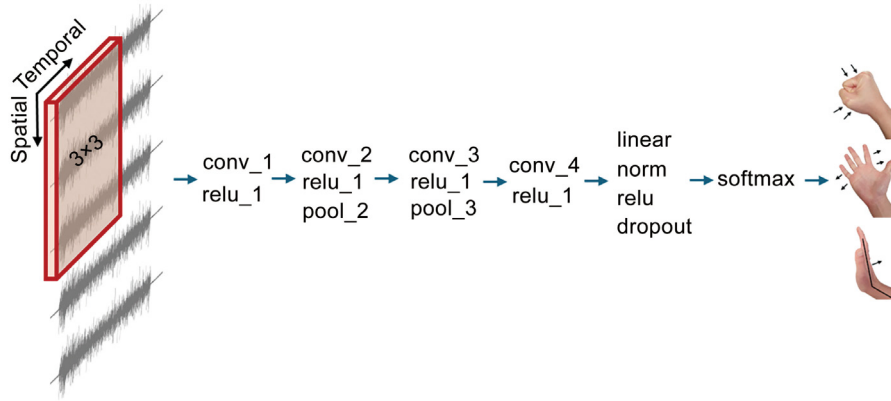


Fig. 6. LeNet-inspired 2D-CNN architecture for sEMG-based hand gesture recognition, where CNN filters learn spatiotemporal features [41].

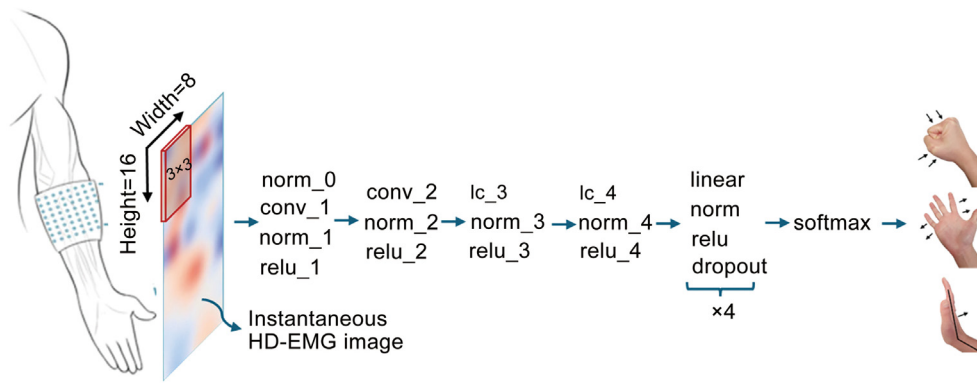


Fig. 7. Myoelectric hand gesture recognition using instantaneous high-density EMG images, where CNN filters are limited to learning spatial information exclusively. Source: Adapted from [81] CC BY 4.0

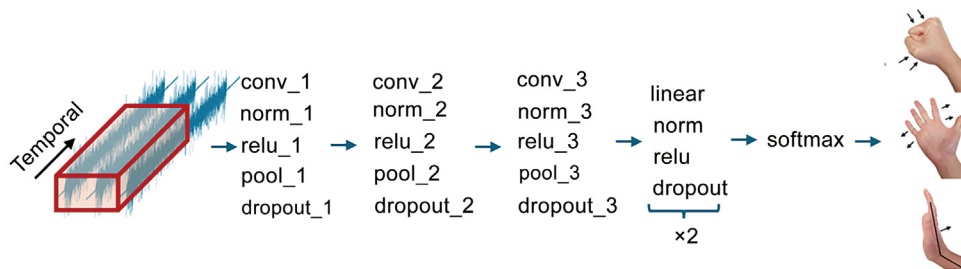
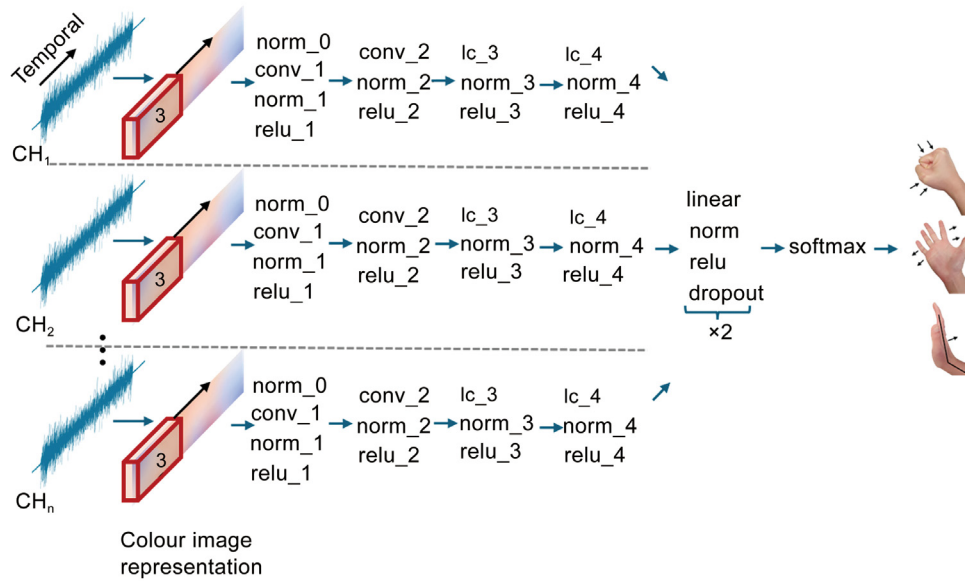


Fig. 8. sEMG hand gesture recognition using 1D-CNN architecture [42].

In contrast, when sparse sEMG data from the Ninapro dataset was used, the adjacency between channels did not fully align with their actual placements (Fig. 6). Additionally, while both studies used two-dimensional CNNs, the interpretation of these dimensions varied. In [41], one dimension represented temporal features, and the other represented spatial features (Fig. 6). Conversely, in [81], both convolutive dimensions were spatial (Fig. 7). In the first layer of the 2D-CNN model proposed by [41], the convolutive kernel is limited to operating on 3 channels at a time (Fig. 6). Alternatively, 1D-CNN, as featured in Soroushmojdehi, et al. [42] (Fig. 8), can be used. The 1D-CNN offers the advantage of making no assumptions about spatial information, as the output of each calculation is summed across all channels (Fig. 8). The use of spatial convolution assumes spatial dependency between muscles. However, during movement generation, different muscle groups can be controlled independently, leading to sEMG

signals from different locations being statistically independent of one another [24]. To address this issue, Wei, et al. [24] proposed a divide-and-conquer strategy using a multi-stream CNN. In this approach, each stream processes inputs from single or multiple channels positioned near one another. In the specific case of a single-channel-per-grouping multi-stream model (Fig. 9), the model performs spatially aware feature extraction with convolution layers, focusing on learning purely temporal features. [24] evaluated this configuration on sparse sEMG data from the Ninapro D1 dataset, which includes 52 hand gestures, achieving a recognition accuracy of 85.0%. This approach outperformed single-stream models, other multi-stream grouping configurations, and classical machine learning methods. Additionally, Wei, et al. [24] tested the multi-stream CNN on the HD-EMG dataset from CSL-HDEMG, which includes 27 finger-hand gestures. By dividing the HD-EMG data into patches of  $7 \times 8$  channels, they



**Fig. 9.** Divide-and-conquer multi-stream CNN architecture for myoelectric hand gesture recognition, where each stream corresponds to a single channel [24]. Source: Color stripe is reused from [81] CC BY 4.0

achieved an optimal recognition accuracy of 90.3% for instantaneous HD-EMG images. When a majority voting scheme was applied over a 150 ms window, the performance further improved to 93.6%. However, a limitation of their evaluation on CSL-HDEMG was that it focused solely on spatial information, similar to the approach used by Geng, et al. [81].

#### 4.2. Sequential modeling with Recurrent Neural Networks

While CNNs assume that data is translationally invariant, RNNs assume that the order of the input matters and learn to capture dependencies across time or sequence steps. Fig. 10b illustrates a basic RNN model in unrolled format for clarity, where  $\mathbf{x}^t$ ,  $\mathbf{h}^t$  and  $\mathbf{o}^t$  denotes input vector, hidden state and output vector at time step  $t$  respectively and  $\mathbf{U}$ ,  $\mathbf{W}$ , and  $\mathbf{V}$  are trainable parameters. RNN assumes that the current hidden state depends on both the current input and the previous hidden state:

$$\mathbf{h}_t = \sigma(\mathbf{W}\mathbf{h}_{t-1} + \mathbf{U}\mathbf{x}_t + \mathbf{b}) \quad (5)$$

Where  $\sigma$  denotes the activation function. Notice that, in scenarios where the input contains multiple channel configurations (as shown in Fig. 10b), RNN models capture both spatial and temporal information, encoded as  $\mathbf{U}$ ,  $\mathbf{W}$ , and  $\mathbf{V}$ . Additionally, the non-linearity introduced by their activation functions further enhances their ability to learn intricate sequential patterns.

Koch, et al. [82] were among the first to apply GRU-variant RNNs for myoelectric hand gesture recognition. They proposed a novel sequence analysis approach leveraging RNNs to classify sequences of arbitrary lengths. Specifically, with each incoming data window, the current window was used to update the GRU model's hidden states. Additionally, their model performed classification using manually extracted sEMG features. Their system achieved recognition accuracies of 79.2% and 76.7% on Ninapro DB I and DB II, respectively, and 53.8% on the amputee dataset of Ninapro DB III. Notably, they observed that their model could reliably predict future hand gestures up to 200 ms in advance.

While the model by Koch, et al. [82] involved manual feature extraction, Hu, et al. [23] introduced a hybrid method combining attention-based CNN and LSTM. In this approach, CNN feature

extractor processes sEMG images structured identically to those used by Geng, et al. [81]. Using the same sEMG image representation as input, the hybrid CNN-LSTM model outperformed basic CNN models across all datasets. While Hu, et al. [23] utilized sEMG images and a 2D-CNN in their classification scheme, Karnam, et al. [83] explored hybrid models that combined a 1D-CNN and LSTM. When tested on sparse sEMG datasets from Ninapro DB I and DB II, Karnam, et al. [83] model outperformed Hu, et al. [23] by 4% and 10% in recognition accuracy, respectively.

In another study, Wu, et al. [84] investigated the potential of using LSTM layers preceding CNN layers for myoelectric hand gesture recognition. They evaluated three models: conventional CNN-LSTM, novel LSTM-CNN, and basic LSTM. When tested on the MyoDataset, which includes five hand gestures, their findings demonstrated that the LSTM-CNN model outperformed both the CNN-LSTM and the basic LSTM, highlighting the potential of LSTM for extracting meaningful features directly from raw sEMG signals. Le, et al. [85] extended the multi-stream CNN architecture proposed by [24] by incorporating LSTM layers at the end of each stream. Unlike most prior work using recurrent neural networks, the multi-stream CNN-LSTM architecture proposed by Le, et al. [85] falls under the category of spatially specific feature extraction (Fig. 2b). Their approach achieved improved performance in both same-session recognition accuracy and cross-day uncalibrated recognition accuracy when evaluated on the Ninapro DB6 dataset [77].

#### 4.3. Learning global information with Transformer-based models

While CNNs and RNNs have been widely used for various machine learning tasks, including myoelectric hand gesture recognition, with varying degrees of success, both models primarily focus on localized patterns and struggle with capturing long-range dependencies efficiently. Specifically, CNNs extract features by applying convolutional filters over small local receptive fields, making them excellent for spatial hierarchies but unsuitable for handling sequential, order-important data. RNNs, on the other hand, process sequences step by step, making them capable of handling ordered data but often inefficient for capturing long-range dependencies due to issues like vanishing gradients. To

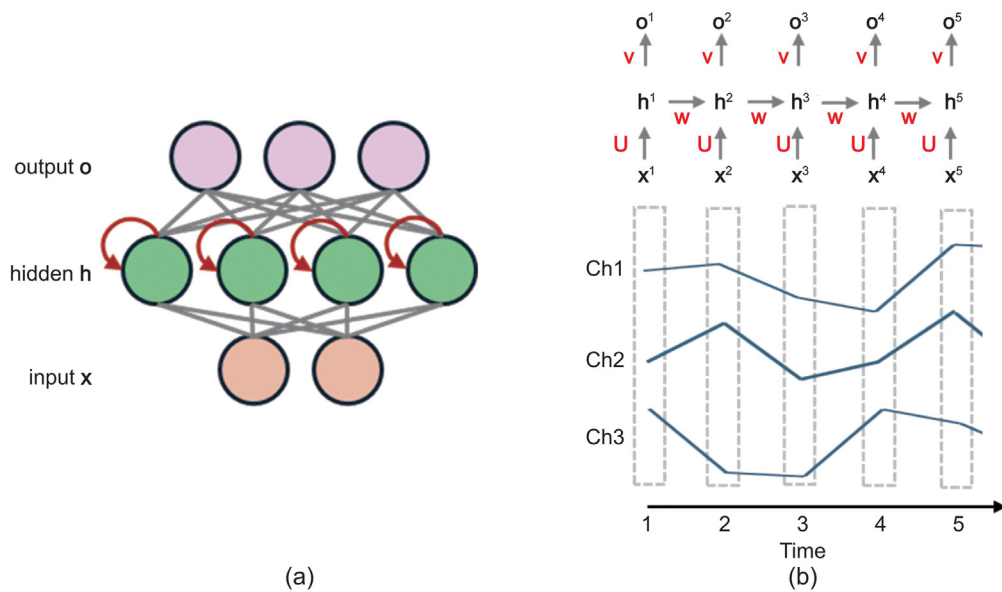


Fig. 10. Illustration of basic recurrent neural network in (a) unfolded and (b) unfolded form with  $U$ ,  $W$ , and  $V$  as trainable parameters [79] (bias term  $b$  omitted for clarity).

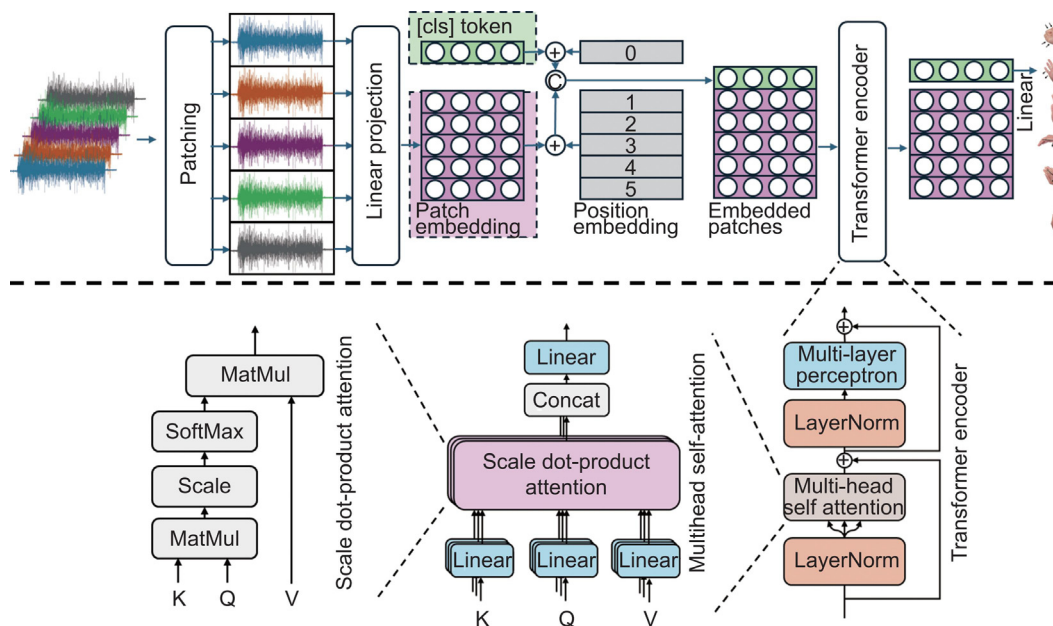
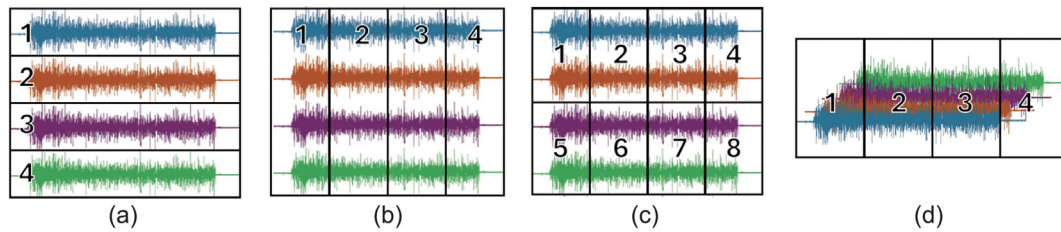


Fig. 11. Illustration of the TraHGR-TNET model for hand gesture recognition Zabihi, et al. [90].

overcome these limitations, Transformer models, introduced by Vaswani, et al. [86] in the field of Natural Language Processing (NLP), eliminate the reliance on solely local information in CNNs and RNNs through the use of self-attention mechanisms. These models have been a key component of many state-of-the-art solutions in both NLP and computer vision, including BERT [87], GPT [88], and ViTs (Vision Transformers) [89]. The self-attention mechanism enables the model to compute relationships between subsegments of input features, capturing both local and long-range dependencies. However, self-attention alone does not inherently provide a way to process sequential data. To address this, positional encoding is used, where the position of each subsegment is represented as a combination of sine and cosine functions. This encoding informs the model about the order of each subsegment, allowing it to handle sequential input effectively [86].

Zabihi, et al. [90] proposed a two-stream Transformer-based network for myoelectric hand gesture recognition (TraHGR), consisting of a temporal stream (TraHGR-TNET Fig. 11) and a featural stream (TraHGR-TNET). The temporal stream segments the input into individual patches, where each patch represents a single electrode input (Fig. 12a). In contrast, the featural stream divides the input into multiple spatiotemporal patches (Fig. 12b). Despite their differences, both streams effectively learn spatiotemporal features, as the extracted patches are concatenated before being fed into the Transformer encoder. A special classification token ([CLS], shown in Fig. 10) is added to capture global context information. Final classification is performed using only the [CLS] token. To fuse information from both streams, their respective [CLS] tokens are summed before classification. This framework achieved a recognition accuracy of 88.72% on the Ninapro



**Fig. 12.** Different patching schemes for transformer-based myoelectric hand gesture recognition: (a) Temporal patching, (b) Featural patching, (c) Alternative featural patching, (d) 1D-multichannel featural patching.

DB2 dataset, with the temporal stream demonstrating superior effectiveness compared to the featural stream.

Shen, et al. [91] proposed using ViTs for multi-view myoelectric hand gesture recognition, with raw sEMG and frequency power density used as input. Similar to Geng, et al. [81], the input data was first converted into a colored image, where one dimension represented channels and the other either time or frequency. Subsequently, the input-colored image was first fed into convolutional models to obtain deep embeddings. Then, the deep latent embedding was segmented using alternative feature patching scheme (Fig. 12c) into multiple patches both horizontally and vertically. Their proposed method achieved state-of-the-art performance on Ninapro DB2 (80.02% with 49 hand gestures and 83.47% with 17 hand gestures) and on Ninapro DB5 (76.83% with 12 hand gestures and 73.23% with 17 hand gestures). This model makes predictions based on spatiotemporal features due to its configuration of convolution layers and self-attention layers.

Zabihi, et al. [92] proposed lightweight, hybrid CNN and attention mechanism models for myoelectric hand gesture recognition. Similar to Shen, et al. [91], the model proposed by Zabihi, et al. [92] also presents a spatiotemporal feature extraction approach due to the use of the Hierarchical Depthwise Convolution operation, where the input representation is split into groups of closely related channels, and each group influences the next. Additionally, 1D multichannel feature patching (Fig. 12d) was applied before feeding the data into a multi-head self-attention mechanism. Their proposed framework achieved state-of-the-art performance on Ninapro DB2, attaining 82.91% accuracy while using 18.87 times fewer parameters compared to previous high-performing models. Liu, et al. [93] proposed a sophisticated yet effective hybrid CNN-Transformer model for myoelectric hand gesture recognition. Their approach integrates multiple attention mechanisms, including the Attention Fusion Block (combining channel and spatial attention), Multi-Head Attention Block, and Multiscale Feature Attention (inspired by Hierarchical Depthwise Convolution by Zabihi, et al. [92]). This model achieved an impressive recognition accuracy of 99.02% while maintaining a minimal inference time of just 14.7 ms.

Wang, et al. [94] enhanced the basic Transformer model proposed by [86] by introducing Temporal Depthwise Convolutional Multi-Head Self-Attention. In the original method, the Query and Key tokens were derived through a simple linear transformation. To improve upon this, Wang, et al. [94] used dilated causal convolution combined with a non-linear ReLU activation function to transform raw sEMG input data to obtain the Query and Key tokens. The Value token, however, continued to be obtained through a simple linear transformation. This approach outperformed several previously proposed Transformer-based networks across multiple datasets, achieving recognition accuracy of 94.65% on the OYDB dataset.

Duan, et al. [95] proposed a transformer-based global-local fusion framework for multi-modal hand gesture recognition using sEMG and acceleration data. Unlike most conventional machine

learning architectures that extract features using a single window size, their approach extracts handcrafted features using multiple window sizes to enhance global-local granularity. For each modality, global information is captured through a transformer encoder, while local information is learned using time-based convolution, channel-based convolution, and time-channel-based convolution. The proposed framework demonstrated improved recognition accuracy compared to other methods, achieving results ranging from 83.93% to 97.77% on the NinaPro DB2, DB3, DB6, and DB7 datasets.

## 5. Strategies to overcome non-stationarity in sEMG for improved hand gesture recognition performance

Many previous studies report high recognition accuracy; however, these evaluations are often conducted under controlled conditions where the testing environment closely matches the training environment. As sEMG signals are highly susceptible to external noise, replicating such ideal performance in real-world scenarios remains challenging. Furthermore, recognition performance often achieves satisfactory results only under trained conditions. When exposed to new conditions, intrinsic distributional changes in the data frequently lead to performance degradation [26,30,48].

To enhance the performance of prosthetic hand control systems in real-world conditions, development efforts should focus beyond achieving high accuracy under conditions identical to the training environment. Instead, they should emphasize generalization accuracy. Specifically, training recognition accuracy refers to the accuracy achieved under conditions identical to those used during training (e.g., the same arm posture or contraction force). Unseen recognition accuracy, on the other hand, measures performance under new conditions (e.g., training on one arm posture and testing on a different posture). Generalization accuracy, which evaluates performance across all conditions, both training and unseen, provides the most reliable metric for assessing myoelectric control systems. As a result, recent research on myoelectric hand gesture recognition has shifted its focus toward improving sEMG pattern recognition in the presence of external disturbances. To address the challenges posed by these disturbances, various approaches and solutions have been proposed in the literature.

### 5.1. Data abundance approach

The simplest solution to address sEMG non-stationarity is to increase data abundance by collecting large, diverse datasets that capture variations in signal conditions. However, this approach has unintended consequences, such as increasing the user training burden, which contributes to the already high rejection rate among myoelectric prosthetic users. Furthermore, when condition-specific performance, defined as the accuracy achieved

when training and testing occur under the same limited conditions (e.g., the same limb position), is considered, studies have shown that increased data abundance can potentially reduce performance compared to the baseline performance of condition-specific models.

Hargrove, et al. [96] investigated the effect of electrode displacement on pattern recognition performance. Four displacement positions were tested, and an LDA classifier was trained to recognize ten different gestures. The study found that classification accuracy dropped by more than 40% when electrodes were displaced, with recognition rates falling as low as 50%. When data from all potential displacements were included in the training set, recognition accuracy increased across all testing positions, with the lowest accuracy reaching approximately 80%. However, this research was limited to evaluating seen recognition accuracy. In follow-up research, Hargrove, et al. [97] employed a high-density monopolar sEMG system arranged in four constellations around the forearm circumference to account for a wider range of potential electrode shifts. Their findings indicated that grouping sEMG data from multiple electrode shift scenarios generally improved seen recognition performance. However, when compared to shift-specific models, where the classifier is trained and tested on data from a single electrode position configuration, reduction in accuracy was observed. This reduction became more pronounced as the number of hand gestures increased.

Scheme, et al. [98] investigated the impact of varying limb positions on classification accuracy. Eight distinct limb positions were selected, and for each position, eight different gestures were recorded at a medium force level. An LDA classifier was trained using time-domain features. While intra-position accuracy rates were high (exceeding 89%), inter-position classification accuracy dropped significantly, falling to as low as 50%. Grouping datasets from all limb positions during training improved inter-position accuracy from 65% (achieved with single-position training) to 92.6%. However, this approach resulted in lower accuracy compared to limb position-specific training, which performed better within individual positions. While Scheme, et al. [98] focused solely on static upper limb postures, Liu, et al. [99] extended the analysis to include results from four static and three dynamic (moving) upper limb postures, aligning with the findings of [98]. Liu, et al. [99] found that limb position-specific models (intra-position) achieved a recognition accuracy of 94.73%. Although grouping data from all static and dynamic limb positions increased recognition accuracy from 81.27% to 91.8%, this result was still 2.93% lower than that of limb position-specific models.

All results from [97–99] indicate that while data abundance improves recognition performance under training conditions, none of these studies evaluated scenarios involving unseen conditions. Furthermore, despite the improvements over single-condition training, the results still showed reduced accuracy compared to condition-specific models. To investigate the cause of this phenomenon, metrics for measuring hand gesture cluster consistency in feature space, as proposed by Bunderson and Kuiken [33], were utilized by Liu, et al. [99] and Radmand, et al. [100]. In Radmand, et al. [100] investigation of 16 selected upper limb postures, they found that increasing the number of training limb positions led to reduced gesture repeatability (indicated by increased RI distance) and reduced gesture separability (indicated by decreased SI metric). This phenomenon is illustrated in Fig. 13. Additionally, Radmand, et al. [100] observed that data collected from closely similar arm postures was highly consistent and did not introduce variability. Similarly, Liu, et al. [99] found that hand gestures were highly repeatable within the same arm position but not across different arm positions, with dynamic limb positions resulting in high global variance as indicated by the mean semi-principal axis. Moreover, as Bunderson and

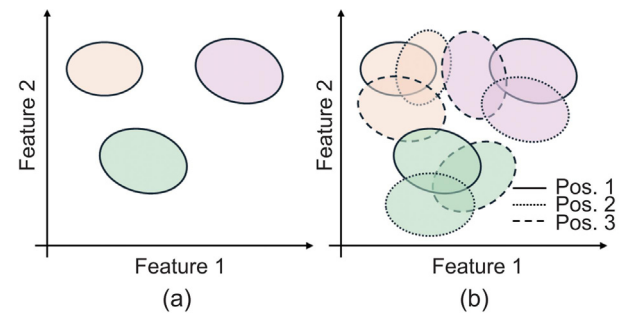


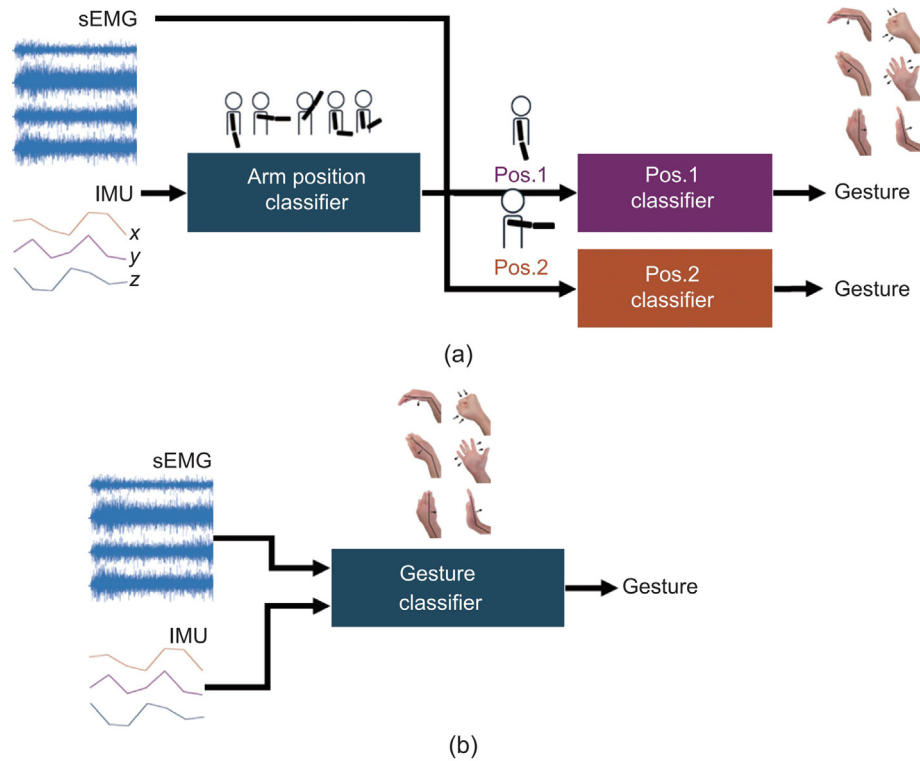
Fig. 13. Illustration of the feature space of a dataset recorded under (a) a single limb position and (b) multiple limb positions, highlighting the decreased repeatability and separability as the number of limb positions increases [100].

Kuiken [33] demonstrated, user expertise influenced the repeatability and separability of hand gesture clusters in feature space. Changes in upper limb posture induced covariate shift due to both physiological factors (how the hand gesture is executed) and environmental factors (electrode shifts caused by changes in muscle shape).

## 5.2. Multi-modal sensory fusion approach

In myoelectric hand gesture recognition, multi-modal sensory fusion is a feasible solution commonly employed to introduce complementary information and enhance system performance [101,102]. Drawing from the observations of Radmand, et al. [100], many studies have explored the potential of IMU sensory modalities as a means of recognizing current arm posture to leverage limb-position-specific classifiers. Specifically, Scheme, et al. [98] investigated a dual-stage classification scheme in which the first stage classifies arm position, followed by hand gesture recognition in the second stage (Fig. 14a). Their results demonstrated that arm position could be classified with nearly 100% accuracy across most positions using two accelerometers placed on the forearm and upper arm. Subsequently, hand gestures were recognized using arm-position-specific classifiers. However, this dual-stage approach has a significant drawback: it is heavily reliant on the accuracy of the arm position classifier. Furthermore, in the best-case scenarios, the overall recognition accuracy is constrained by the performance of the arm-position-specific classifier. In contrast, a multi-modal fusion scheme (Fig. 14b), which combines sEMG and accelerometer data as input, consistently improved recognition accuracy across all tested limb positions. In follow-up research, Fougner, et al. [31] provided a more extensive report on solutions to mitigate the effects of limb position. Their findings highlighted that increasing the number of trained limb positions significantly enhances recognition accuracy and emphasized the potential benefits of incorporating accelerometer measurements as part of the solution. While both Scheme, et al. [98] and Fougner, et al. [31] relied on manual feature extraction and LDA, Williams, et al. [103] proposed a deep hybrid CNN-RNN model for end-to-end myoelectric hand gesture recognition under varying arm positions. Their findings showed that the CNN-RNN model, when trained using multi-modal data in a dual-stage configuration (Fig. 14a), achieved 99% accuracy in recognizing hand gestures under trained limb positions, slightly surpassing the recognition accuracy of LDA, which was 97.67%.

Samuel, et al. [104] built upon the work of Scheme, et al. [98] by evaluating the potential of a multi-modal approach combining sEMG and accelerometer data under dynamic scenarios, including sitting, walking, and ascending or descending stairs. Their results revealed that accelerometers achieved higher recognition



**Fig. 14.** Two strategies for integrating accelerometer data into myoelectric hand gesture recognition across multiple limb positions: (a) Dual-stage classification, and (b) Multi-modal fusion.

accuracy in intra-scenario cases. However, in inter-scenario cases, while the classification accuracy of sEMG remained around 80%, the accuracy of accelerometers dropped to less than 40%. This decline was attributed to the high sensitivity of accelerometers to body movements, causing hand gesture clusters recorded in different scenarios to diverge significantly. Due to this limitation, Samuel, et al. [104] concluded that the dual-stage approach outperforms multi-modal fusion strategies, achieving classification accuracies of 89% and 87%, respectively. Both methods also demonstrated superior performance compared to simply concatenating sEMG data from multiple scenarios. Notably, Wei, et al. [105] demonstrated improved recognition accuracy in unimodal myoelectric hand gesture classification by using generative adversarial networks (GANs) to model the relationship between sEMG and inertial measurement unit (IMU) data. This approach enabled the generation of virtual IMU measurements from sEMG alone, highlighting a strong intrinsic link between the two modalities.

Zhang, et al. [106] explored a fusion strategy using co-located sEMG data and pressure forcemyography (pFMG). Individually, sEMG and pFMG achieved “seen” recognition accuracies (trained and tested on the same set of limb positions) of 59.56% and 81.98%, respectively, using a deep CNN. With late fusion strategies, the “seen” recognition accuracy improved significantly to 87.79%. A key advantage of the pFMG sensor over IMU is its insensitivity to body motion, making it more robust in dynamic scenarios. The robustness of EMG-FMG fusion against limb position and grasp load effects was further confirmed by Young, et al. [107], who demonstrated that EMG-FMG fusion achieved a recognition accuracy of 97.34%, representing a 15% improvement over unimodal sEMG recognition. In another research, incorporation of sEMG and FMG was also shown to improve recognition accuracy of transient sEMG data, as demonstrated by the cross-modal attention CNN proposed by Oyemakinde, et al. [108]. Additionally, fusion of sEMG and FMG has been shown to enhance recognition accuracy for previously unseen subjects [109].

### 5.3. Sparse representation approach

Bethausser, et al. [48] proposed a method based on sparse representation using the sparsity  $l_1$  constraint. Given a dataset of  $C$  classes represented as  $\mathbf{A} = [\mathbf{A}_1 | \mathbf{A}_2 | \dots | \mathbf{A}_C]$ , a new data sample  $\mathbf{y}$  can be represented as a linear combination:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \quad (6)$$

To enforce sparsity, the  $l_1$  constraint is applied to  $\mathbf{x}$ , resulting in the following optimization problem:

$$\mathbf{x} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_1 \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{y} \quad (7)$$

It has been observed that data from the same class of hand gestures tend to be reconstructed more effectively using data from the training dataset of the same class with  $l_1$  constraint. To enhance inference speed for real-time applications, Bethausser, et al. [48] proposed the use of extreme machine learning to reduce the number of hand gesture candidates prior to solving the optimization problem (7). In contrast to [31,98,103], Bethausser, et al. [48] focused on reporting recognition accuracy in scenarios where the training and testing limb positions differed. In the “train 1, test 1 other” evaluation scheme, their proposed method based on sparse representation demonstrated high accuracy compared to conventional LDA models, even when testing limb positions were substantially different from training positions. This increased recognition performance was further evident in the “train  $k$ , test others” evaluation scheme, which was tested on both amputee and able-bodied subjects. Additionally, they demonstrated that the increased efficiency of using deep extreme learning in conjunction with the sparse representation model resulted in higher throughput and reduced overshoot in online experiments. In follow-up research, Bethausser, et al. [110] showed that the combination of sparse representation and extreme machine learning was highly effective in recognizing hand

gestures under untrained electrode shifts. The method achieved recognition accuracy above 85% for electrode shifts of up to 1.5 cm.

#### 5.4. Overcoming switching class problems with post-processing and extended temporal context

It has been observed that using a small window size helps reduce input lag [111]. However, this often results in decreased recognition accuracy [112], particularly manifesting as class-switching errors. Class switching refers to instability in the recognition output, where the system fails to maintain consistent classifications, resulting in frequent and erroneous transitions between hand gestures. One of the most used post-processing strategies to address this issue is majority voting. In this approach, the final recognition result is determined by identifying the mode of decision outcomes accumulated across multiple windows [21,113]. Barona López, et al. [114] proposed an alternative strategy for post-processing hand gesture predictions by replacing single spurious outputs based on results from adjacent windows. From a processing perspective, combining results from multiple windows during post-processing is functionally like increasing the window size, as it extends the temporal context.

Betthausen, et al. [115] proposed the use of a Temporal Convolutional Neural Network (TCNN) to overcome the problem of class-switching errors. In their work, MAV feature was extracted from each raw sEMG window with a window size of 200 ms and a step size of 25 ms, and 60 windows were concatenated to produce an input array with a temporal context of 1600 ms. TCNN outperformed both frame-wise classification using traditional classifiers and temporal LSTM models with the same temporal context size in terms of recognition accuracy and stability score. In subsequent research, Betthausen, et al. [116] further improved these results by incorporating encoder-decoder TCNN models. This approach not only enhanced recognition accuracy and stability scores but also resulted in lower delays when switching from rest to active hand gestures on the Ninapro DB2 dataset.

Zhang, et al. [21] proposed the use of muscle activity detection to lock the output hand gesture during active segments. In their study, Zhang, et al. [21] investigated four different post-processing strategies: majority voting, muscle activity detection using the sEMG Mean Absolute Value, the sEMG Teager-Kaiser Energy Operator, and Force Myography. They found that incorporating muscle activity detection generally resulted in higher recognition accuracy compared to frame-wise classification or majority voting. Among the approaches, FMG-based muscle activity detection was found to be more stable than sEMG-based methods.

#### 5.5. Improving data consistency through user training and leveraging shift invariant properties

Improving data consistency across multiple conditions can be achieved by enhancing user proficiency in executing hand gestures. Stuttaford, et al. [117] explored this approach by investigating the impact of different types of feedback on user performance in hand gesture tasks. In their study, participants engaged in an interactive protocol to control a cursor and hit a target using hand gestures performed in various arm positions. Three types of feedback were provided: (i) Concurrent feedback, where the cursor was displayed in real-time; (ii) Delayed feedback, where the cursor appeared only after the movement was completed; and (iii) No feedback. Their findings showed that delayed feedback training significantly improved performance in untrained limb positions.

Better data consistency can also be achieved through effective feature design. For example, Khushaba, et al. [30] and Al-Timemy, et al. [26] proposed a novel feature set called the Time-Dependent Spectral Descriptor, designed to be invariant to amplitude translation in sEMG. This feature set effectively addresses challenges posed by variations in sEMG recording conditions, such as changes in limb position and contraction force. Khushaba, et al. [30] evaluated the performance of TD-PSD in conjunction with spectrum correlation for myoelectric hand gesture recognition under varying arm positions. Spectrum correlation, which measures the similarity between electrode pairs, serves as a spatial coherence metric. This approach enables hand gesture classification using spatiotemporal information (Fig. 15). The proposed feature extraction method was assessed under limb position changes using various classifiers. In a “train on one limb position, test on all positions” scenario, SVM with RBF kernel achieved optimal recognition accuracy, with 98.32% for intra-limb position and 77.42% for inter-limb position. The LDA classifier yielded slightly lower results, with 98.21% for intra-limb position and 77.51% for inter-limb position. In “train on all limb positions, test on all positions” scenario, their proposed feature extraction method significantly improved recognition accuracy across all tested limb positions compared to baseline feature sets. Al-Timemy, et al. [26] further evaluated the TD-PSD feature set using hand gesture data recorded from multiple amputee subjects under varying contraction forces. When trained on one contraction force, the TD-PSD feature set consistently achieved higher recognition accuracy compared to baseline feature sets across multiple classifiers, including LDA, random forest, naïve Bayes, and k-NN, for both intra- and inter-contraction force scenarios. In the “train on all forces, test on all forces” evaluation scheme, the LDA classifier with the TD-PSD feature set achieved superior recognition accuracy (82.58%) compared to other classifiers.

He, et al. [118] proposed using Fast Fourier Transform with channel-wise and global normalization to improve hand gesture recognition under varying contraction forces. The sEMG signal spectrum was first extracted using FFT and divided into multiple sub-bands. For channel-wise normalization, FFT values in each sub-band were normalized against the corresponding sub-band across all channels. For global normalization, FFT values were normalized across all channels and sub-bands. This FFT-based feature extraction method significantly improved recognition accuracy for both seen and unseen contraction forces.

In cases of covariate shift due to changes in contraction force, the concept of muscle synergy, defined as the co-activation of multiple muscles to facilitate the elicitation of hand gestures, is often utilized to maintain classification accuracy. Onay and Mert [119] proposed Phasor Represented EMG, which employs complex numbers to represent muscle synergy, thereby enhancing classification accuracy under varying muscle contraction forces. Phasor Represented EMG is derived from two temporal sEMG features, RMS and WL, with spatial information encoded using a cylindrical forearm model. When tested on the same dataset as Al-Timemy, et al. [26], Phasor Represented EMG demonstrated improved recognition accuracy with an LDA classifier when tested on unseen contraction forces. Teng, et al. [38] proposed a method for hand gesture recognition based on the minimum reconstruction error of task-specific synergies. The process begins with manual feature extraction to identify relevant features from raw sEMG data. For each hand gesture in the training dataset, an alternating non-negativity constrained least squares algorithm is applied to map the gesture into task-specific synergies. New hand gesture data is then classified by comparing its similarity to the task-specific synergies of each hand gesture. The study found that while the proposed method achieved recognition accuracy comparable to conventional LDA with manual feature extraction, it significantly outperformed LDA in scenarios involving inter-contraction force, demonstrating notably higher recognition accuracy.

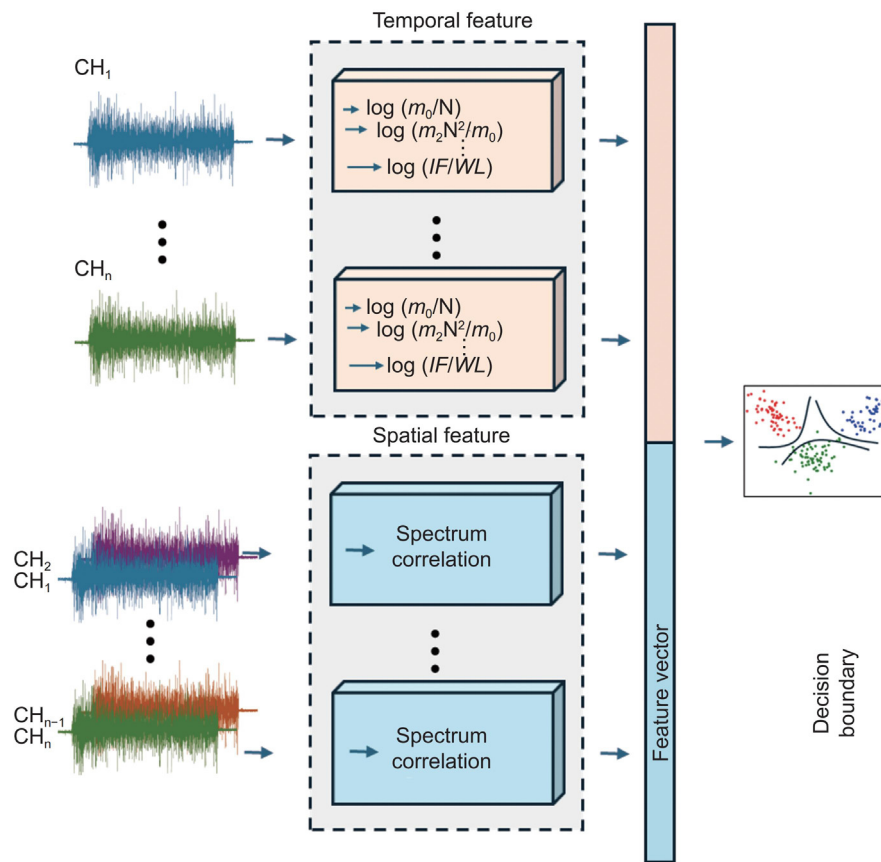


Fig. 15. Illustration of spatiotemporal TD-PSD feature extraction for enhancing robustness in recognition accuracy to variations in arm position [30].

5.6. Supervised and unsupervised transfer learning approaches

5.6.1. Supervised myoelectric hand gesture recognition with labeled calibration data

Supervised transfer learning is a machine learning technique designed to leverage a larger, more comprehensive source dataset alongside a smaller target dataset to enhance recognition performance on the target dataset [120]. This approach has practical applications in myoelectric control, as it reduces the training required for the target user, thereby lessening their training burden. Additionally, this technique is frequently employed for tasks such as cross-day calibration [65] and cross-subject transfer [64]. At its simplest, transfer learning can be performed by combining the source and target datasets into a single dataset. However, this approach may decrease performance on the target dataset due to increased variance in hand gesture clusters [100]. In hand gesture recognition, both spatial and spatiotemporal transfer learning have been explored in previous studies. **Spatial transfer learning** in myoelectric hand gesture recognition typically refers to the use of traditional classifiers. Since traditional machine learning models often extract the same set of temporal features from every sEMG channel, transfer learning in this context is limited to the classification stage. This involves adjusting the decision boundary to balance the influence of the source and target datasets. In contrast, **spatiotemporal transfer learning** is more commonly found in deep learning models, where spatiotemporal convolutive kernels trained on the source dataset are repurposed for the target dataset. From another perspective, supervised transfer learning for sEMG hand gesture recognition can also be achieved through feature alignment or by transferring model parameters. **Feature alignment** involves mapping the feature distributions of the source and target datasets into a shared representation, while

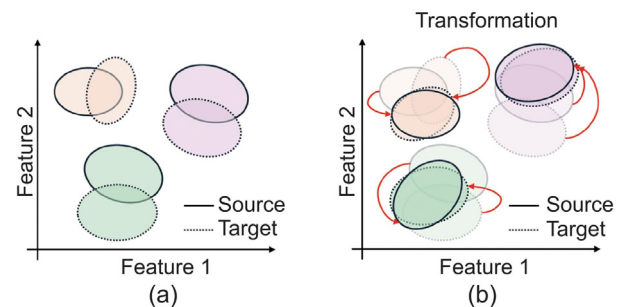


Fig. 16. Illustration of the feature space alignment method for transfer learning: (a) representation of the source and target datasets in the original feature space, and (b) their alignment after transformation.

**parameter transfer involves** adapting pre-trained model weights from the source dataset to the target dataset (see Table 3).

To address the challenges of data aggregation, feature alignment methods are employed to transform source and target datasets, minimizing differences between them (Fig. 16). Khushaba [121] proposed using Canonical Correlation Analysis (CCA) to transform the source and target datasets into a unified representation, enabling classification to be performed on the transformed data. This method was tested in two scenarios involving amputee subjects: (i) bilateral finger movement subspace alignment and (ii) cross-subject subspace alignment. In the bilateral finger movement experiment, the goal was to improve performance on the amputated arm by leveraging hand gestures from the intact limb. Using CCA for alignment, the method achieved a recognition accuracy of 86.3%, surpassing baseline

**Table 3**  
Key differences between Supervised and Unsupervised Transfer Learning for myoelectric hand gesture recognition.

	Supervised transfer learning	Unsupervised transfer learning
Targeted Labeled Hand Gesture	- Required	- Not required
User training burden	- Increased	- Not increased
Techniques	- Feature alignment of both marginal distribution $P(\mathbf{x})$ and conditional distribution $P(y   \mathbf{x})$ . - Parameter transfer learning	- Feature alignment of only marginal distribution $P(\mathbf{x})$ . - Conversion to supervised transfer learning through pseudo-label generation.
Main application	- Cross-day supervised calibration	- Maintain recognition accuracy in prolonged usage

methods that trained and tested on the same arm (86%) or aggregated data from both arms without alignment (83.52%). In the cross-subject subspace alignment scenario, data from multiple amputee subjects was aggregated and aligned with a normal-limb expert subject using CCA. Similarly, sEMG features extracted from the testing subject were aligned with those of the expert subject using CCA. The transformed, aggregated dataset was then used to train a hand gesture classifier. This approach achieved a recognition accuracy of 82.96%, outperforming the baseline method of training and testing on the same subject, which achieved 80.53%. These results demonstrate the effectiveness of leveraging a source dataset through CCA transformation to enhance recognition accuracy. Kulwa, et al. [122] proposed an alignment method on the Riemannian manifold, a manifold of symmetric positive-definite matrices, to improve myoelectric hand gesture recognition under varying contraction forces. Specifically, their approach classified hand gestures solely based on the covariance matrix computed from each window of multichannel raw sEMG data. Their results demonstrated that this alignment method not only improved performance for seen contraction forces but also generalized effectively to unseen contraction forces, outperforming traditional classifiers that rely on manual feature extraction.

The limitation of feature alignment methods used by Khushaba [121] and Kulwa, et al. [122] lies in their implicit assumption that the global means of the source and target datasets are equal. To address this issue, Hoshino, et al. [123] proposed an alternative approach that incorporates affine transformations in the context of multiple cross-subject transfer learning. Their method uses ensemble models, where separate models are trained on different pairs of source and target subjects and then combined to improve overall performance. Their findings revealed that, for scenarios involving a single source subject, a model trained solely on the target subject provided a strong baseline, outperforming both data aggregation and affine transformation-based transfer learning models. However, as the number of source subjects increased, affine transformation-based transfer learning began to demonstrate superior performance. The study achieved peak recognition accuracy with seven source subjects, where the SVM-RBF model reached 91.51% and the MLP model achieved 92.67% in recognition accuracy.

Model parameter-space transfer learning aims to repurpose previously trained source model parameters for adapting to a new target dataset. This type of transfer learning can be implemented using traditional classifiers, such as LDA, QDA, MLP, or deep learning models. Vidovic, et al. [65] introduced a method for transfer learning with LDA and QDA classifiers to maintain cross-day classification accuracy. Since LDA and QDA rely on the class-wise mean vector  $\mu_j$  and covariance matrix  $S_j$ , the decision boundaries can be adjusted to control the influence of the source and target datasets using the following formulas:

$$\tilde{\mu}_j = (1 - \tau)\mu_j^{source} + \tau\mu_j^{target} \quad (8)$$

$$\tilde{S}_j = (1 - \lambda)S_j^{source} + \lambda S_j^{target} \quad (9)$$

Here,  $\tau$  and  $\lambda$  are hyperparameters that balance the influence of the source and target datasets (as illustrated in Fig. 17). Vidovic, et al. [65] demonstrated that the proposed supervised transfer learning scheme successfully maintained recognition accuracy above 92% over a five-day period for both able-bodied and amputee subjects, representing a 20% improvement compared to not using the transfer learning scheme. Furthermore, online experiments showed 25% increase in recognition accuracy when transfer learning was applied. Notably, they observed that transfer learning with LDA outperformed QDA in terms of recognition accuracy. Building on this work, [124] extended the approach by investigating the effects of calibration datasets and their influence on classification performance over new days. While Vidovic, et al. [65] focused on transfer learning models trained using a single source subject, Liu, et al. [124] expanded the method to multi-source transfer learning. In addition to LDA-based transfer learning, they also proposed using transfer learning with polynomial classifier. Consistent with Vidovic, et al. [65], Liu, et al. [124] confirmed that the linear decision boundaries of LDA provided better recognition accuracy compared to the non-linear boundaries of the polynomial classifier, both in transfer learning and baseline models. Moreover, with just 2 s of calibration data, transfer learning with LDA achieved an accuracy of 86.72% for able-bodied subjects, which was 7.67% higher than the baseline LDA model trained solely on calibration data. Similarly, for amputee subjects, transfer learning with LDA achieved an accuracy of 82.88% with 2 s of calibration data, further increasing to 85% when 5 s of calibration data were used.

Hoshino, et al. [123] explored transfer learning using a 3-layer MLP, where adaptation to new subject data was achieved by fine-tuning only the weights between the third hidden layer and the output layer. This fine-tuning process utilized a single repetition of the source subjects' data. Their results showed that transfer learning with the MLP outperformed feature alignment methods, achieving a recognition accuracy of 93% with one source subject. Furthermore, the accuracy increased to 95% when five source subjects were used.

Transfer learning for deep learning models, in its simplest form, involves reusing a pretrained source model as the starting point for training the target model. This method is commonly referred to as fine-tuning transfer learning. Ameri, et al. [29] proposed a transfer learning approach using convolutional neural networks to address performance degradation caused by electrode shifts. The approach involved utilizing a pretrained CNN model, initially trained on pre-shift data, and fine-tuning the convolutional layers with minimal post-shift data. In an experiment involving 13 able-bodied participants with 10 electrodes placed circumferentially around the forearm, electrode shifts were simulated by altering the electrode order either clockwise or counterclockwise. The results demonstrated that the proposed fine-tuning transfer learning framework achieved a recognition accuracy of 93%, outperforming strategies that relied on data aggregation or training exclusively with post-shift data. In another

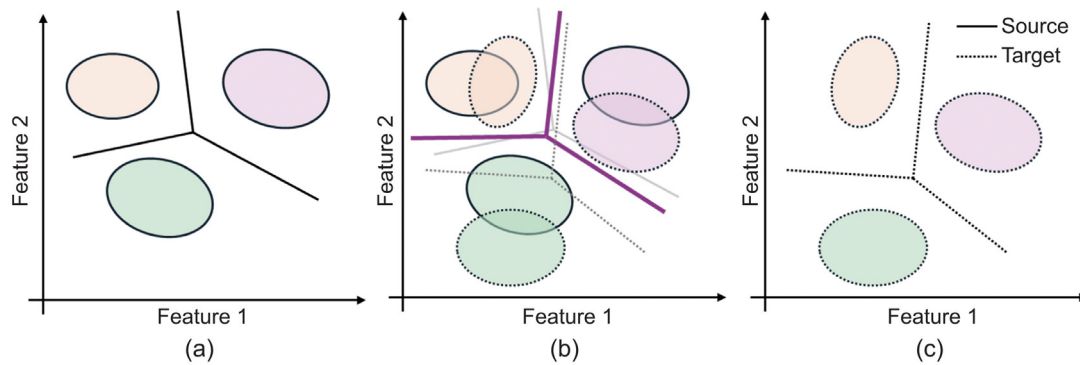


Fig. 17. Illustration of decision boundary from (a) source dataset only, (b) model parameter-space transfer learning and (c) target dataset only.

study on subject transfer learning, Kim, et al. [125] proposed using a source subject model as the starting point for training the target subject model. Similar to the approach of Hoshino, et al. [123], Kim, et al. [125] explored combining predictions from multiple source subject models through an ensemble model. This transfer learning and ensemble method significantly improved recognition accuracy compared to the baseline model trained solely on the target subject, for both able-bodied and amputee participants.

However, using fine-tuning in transfer learning can be unpredictable due to a phenomenon called catastrophic forgetting, where the model entirely forgets information learned from the source dataset. To mitigate this issue, it is often better to freeze certain layers of the pretrained model, as these layers encode the learned information. Soroushmojehi, et al. [42] proposed a two-stream CNN architecture for cross-subject transfer learning. One stream is trained on a large dataset from multiple subjects, while the other is trained on a smaller dataset from the target subjects. The convolutional layers in each stream act as spatiotemporal feature extractors, and the outputs from the source and target streams are concatenated to predict hand gestures. Cote-Allard, et al. [64] introduced transfer learning using progressive neural networks [126]. In this approach, a convolutional model is first trained on source subjects, and its weights are frozen. To adapt the model to the target subject, new convolutional layers are added and connected to the pretrained layers. This setup allows the model to learn from the target subject while preserving patterns learned from the source subjects. Alternatively, Chen, et al. [127] proposed a CNN-LSTM-based transfer learning model. In this approach, CNN layers are first trained on the source subjects and then used as spatiotemporal feature extractors. The features extracted by the CNN layers are subsequently fed into LSTM layers, which are trained exclusively on the target subject's data. This approach leverages the spatiotemporal representations learned by the pretrained CNN layers to improve recognition performance on the target dataset. Ketykó, et al. [28] proposed a two-stage LSTM-based subject transfer learning framework. In this framework, spatiotemporal information is encoded directly into the LSTM weights, unlike other methods that rely on CNNs for feature extraction. This transfer learning process consists of two stages. In the pretraining stage, the matrix  $M$  is initialized as an identity matrix, and the LSTM weights are pretrained using the source subject dataset. To adapt to new target subjects, the LSTM weights are frozen while the weights of  $M$  are retrained. Here,  $M$  acts as a spatial filter, aligning spatial information between the source and target subjects, ensuring better adaptation and improved recognition accuracy.

Besides leveraging pretrained weights from source subjects, knowledge transfer can also be achieved by aligning the source

and target domains in the feature space using deep neural networks. Zou and Cheng [128] proposed a multi-scale 2D-CNN framework for cross-subject transfer learning. Multi-scale 2D-CNNs utilize convolutional layers with receptive fields of varying sizes, enabling the model to capture a broad spectrum of spatiotemporal features, ranging from short-range to long-range dependencies. Unlike traditional weight transfer learning models, Zou and Cheng [128] approach performs transfer learning simultaneously by leveraging data from both source and target subjects concurrently. To enhance alignment in the feature space, their method incorporates additional loss functions, including Jensen-Shannon divergence and center-loss, alongside the standard cross-entropy loss. On the public Ninapro DB6 dataset, this supervised transfer learning framework achieved a cross-day recognition accuracy of 90.3%, which is 5.5% higher than without transfer learning. For cross-subject transfer learning, it achieved a recognition accuracy of 74.48%, reflecting a 12% improvement over non-transfer learning models. Similarly, Shi, et al. [129] explored cross-subject and cross-day transfer learning using a 2D-CNN framework. Their approach used maximum mean discrepancy (MMD) loss to reduce differences between the source and target datasets. This method significantly outperformed fine-tuning approaches, achieving an outstanding recognition accuracy of 97.2% in multi-day and cross-subject transfer learning scenarios. Yang, et al. [130] proposed a hybrid CNN-LSTM architecture that incorporates subject-aware contrastive loss for cross-subject knowledge transfer. This contrastive loss function is designed to minimize the distance between data points belonging to the same subject and gesture (strong pairs), while maximizing the distance from dissimilar samples (weak pairs). Their method demonstrated high recognition performance, achieving accuracies of 96.23% on the Ninapro DB1 dataset and 97.24% on Ninapro DB2.

Additionally, it is worth noting that transfer learning with classical machine learning models [65, 121, 122, 124] represents a form of spatial information transfer, as the transfer occurs exclusively during the classification stage. In these studies, temporal features are extracted consistently across electrode locations. In contrast, most studies on transfer learning using deep neural networks for sEMG hand gesture recognition, such as [64, 125, 128, 129], involve spatiotemporal transfer learning, where both spatial and temporal knowledge are jointly transferred. This distinction arises primarily from the architecture of the models employed, rather than any fundamental limitation of deep learning itself.

While transfer learning is a potentially effective approach to improving gesture recognition accuracy, privacy concerns have been raised because sEMG data can contain biometric information [131]. To address these concerns, especially in cross-subject transfer learning scenarios with data aggregation approach, federated learning has been proposed. Federated learning is a decentralized machine learning approach where data remains on local

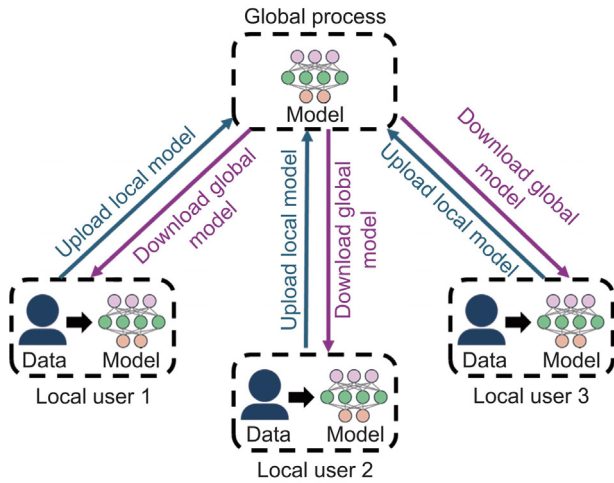


Fig. 18. Federated learning framework to prevent biometric data leakage.

devices and only model updates are shared. This enhances privacy by preventing the leakage of sensitive biometric data (Fig. 18). The most widely used method in federated learning is federated averaging, where the global model weight  $\mathbf{W}_{global}^{(t+1)}$  at iteration  $t+1$  is computed as the average of the local model weights  $\mathbf{W}_k^{(t+1)}$

$$\mathbf{W}_g^{(t+1)} = \frac{1}{K} \sum_{k=1}^K \mathbf{W}_k^{(t)} \quad (10)$$

where  $K$  is the number of local nodes [132]. Zafar, et al. [132] conducted experiments using federated learning for transfer learning on a private dataset comprising 8 hand gestures across 10 subjects. Using their proposed Spatial Attention-Embedded Residual Network, they reported that federated averaging achieved a recognition accuracy of 91.72%. Zhang, et al. [133] proposed using federated learning as a pretraining framework with convolutional neural network with self-attention. After federated pretraining, the global model is fine-tuned on a target subject. Testing on the NinaPro DB5 dataset, the proposed federated transfer learning framework achieved a recognition accuracy of 74.58%. Lee, et al. [134] proposed a novel federated learning framework named FedAssist, designed to address the problem of non-independent and identically distributed (non-IID) sEMG datasets. The FedAssist framework improved recognition accuracy when tested on the NinaPro DB6 dataset, outperforming both federated averaging and traditional data-naive aggregation approaches across multiple non-IID scenarios.

### 5.6.2. Unsupervised transfer learning and self-adaptation

Supervised machine learning requires labeled data from both the source and target domains. However, obtaining these labels often involves active system recalibration by the user, which is both time-consuming and inconvenient. To mitigate the challenges posed by this lengthy supervised calibration process, unsupervised transfer learning and self-adaptation techniques have been developed. These approaches aim to maintain classification accuracy by compensating for gradually changing confounding factors in sEMG systems without the need for extensive manual intervention. Unsupervised machine learning typically involves aligning the source and target domains using unsupervised metrics. Some notable techniques in unsupervised machine learning include:

- **Canonical Correlation Analysis:** Resolves a subspace that maximizes the correlation between the source and target datasets [135].

- **Subspace Alignment:** Finds a rotation transformation that aligns the principal components of the source and target datasets [136].
- **Geodesic Flow Kernel (GFK):** Identifies intermediate PCA subspaces between the source and target on the Grassmannian manifold [137].
- **Correlation Alignment (CORAL):** Transforms the target dataset into the source dataset by maximizing their correlation [138].
- **Transfer Component Analysis (TCA):** Finds a transformation that minimizes the distribution shift measured using maximum mean discrepancy (MMD) between the source and target datasets [139].

Shi, et al. [140] proposed the use of unsupervised domain adaptation with CNN and Correlation Alignment. This approach achieved a recognition accuracy of 90.19% when the model was adapted using just two repetitions of each hand gesture without labels, representing a 10% improvement compared to models without domain adaptation. Similarly, Zou and Cheng [128] found that unsupervised domain adaptation, achieved by aligning the global distribution between the target and source datasets, led to an 11% improvement in cross-subject scenarios and a 2% improvement in cross-day scenarios. Zhang, et al. [141] proposed a more selective procedure for target sample selection in unsupervised domain adaptation, termed the self-guided adaptive sampling (SGAS) strategy. Specifically, only pairs of samples with small Maximum Mean Discrepancy distances were considered for domain adaptation, as pairs with large distances could negatively impact the adaptation process. With the proposed SGAS approach, recognition accuracy improved to 90.41%, which was 22% higher than models without unsupervised domain adaptation and 10% higher than models using domain adaptation without sample selection.

Unsupervised domain adaptation can be reformulated as a supervised problem by incorporating a pseudo-label generation process. In this approach, sample points and pseudo-labels are utilized to adjust decision boundaries. Chen, et al. [142] proposed generating pseudo-labels by adapting existing models using newly incoming data. Their self-adapting scheme improved recognition accuracy by 2.2% and 1.6% for QDA and LDA classifiers, respectively. However, a drawback of Chen, et al. [142] method is that instantaneous predictions can be inaccurate, leading to unpredictable long-term effects. To address this issue, Zhai, et al. [113] proposed recalibrating hand gesture recognition models by retraining them using pseudo-labels generated from the majority voting of median likelihood. This method achieved a 10% improvement in intact subjects and a 2.99% improvement in amputee subjects, compared to models without a recalibration scheme. Alternatively, Amsüss, et al. [143] introduced a method that incorporates an MLP, using both sEMG features and the LDA prediction confidence score as input to determine whether a model should be updated. This self-enhancing mechanism resulted in a 31.6% improvement in recognition accuracy compared to models without such mechanisms. Wang, et al. [144] proposed a novel pseudo-label generation method using K-Means clustering. Class-wise mean vectors from the source dataset were used as initialization values to determine class-wise mean vectors for the target dataset. Final pseudo-labels were assigned by considering both the source dataset's class-wise means and the estimated class-wise means from the target dataset. These pseudo-labels were then used for cross-domain error minimization, resulting in improved recognition accuracy compared to pure unsupervised domain adaptation on HD-EMG datasets, as well as Ninapro DB4 and DB5.

**Table 4**

Summary of challenges in current myoelectric hand gesture recognition techniques and proposed future research directions.

Challenges	Research directions
<ul style="list-style-type: none"> <li>• Leveraging cross-session, cross-subject data.</li> <li>• Increased user training burden.</li> <li>• Negative knowledge transfer.</li> </ul>	<ul style="list-style-type: none"> <li>• Dynamic benchmarking datasets.</li> <li>• Enhance model interpretability.</li> <li>• Evaluation of real-time response delay on hand gesture recognition.</li> </ul>
<ul style="list-style-type: none"> <li>• Generalization to unseen conditions.</li> <li>• Low interpretability.</li> </ul>	<ul style="list-style-type: none"> <li>• Supervised transfer learning of temporal features.</li> </ul>
<ul style="list-style-type: none"> <li>• Minimizing inference time.</li> <li>• Wearable inference hardware.</li> </ul>	<ul style="list-style-type: none"> <li>• Lightweight deep learning model.</li> </ul>

## 6. Discussions

In Table 4, we have identified some key challenges for myoelectric hand gesture recognition and potential research directions. Moreover, to provide a comprehensive overview of machine learning techniques for myoelectric hand gesture recognition, Table 5 highlights the application, advantages and drawbacks of each technique.

### 6.1. Current research challenges of myoelectric hand gesture recognition

The most challenging aspect of myoelectric hand gesture recognition for prosthetic control is balancing the trade-off between data abundance and user training burden. At present, data abundance remains the primary strategy to improve recognition performance in the presence of non-stationarities of sEMG signals [16]. However, this approach comes at the cost of increased training demands on users, which in turn contributes to higher prosthetic rejection rates. In the case of changes in arm position, increasing the number of gesture repetitions during the training phase does not necessarily lead to improved recognition accuracy on unseen limb positions [145]. Moreover, the effectiveness of data abundance is constrained by factors such as electrode shift [29] and variations in participant behavior [33]. These challenges are exacerbated when the data collection process is split across multiple sessions, as this can introduce unwanted, compounded covariate shifts in the recorded data. Even when the arm position remains constant, in calibrated cross-day hand gesture recognition, naively combining previously recorded data with new calibration gestures can lead to negative knowledge transfer, resulting in decreased recognition accuracy [146]. As a result, despite the availability of multiple public datasets, myoelectric hand gesture recognition models are still typically trained in a subject-specific manner.

Generalization performance, *i.e.*, the ability to accurately predict newly encountered hand gestures despite sEMG non-stationarities—is a fundamental goal in the field. However, this objective is not fully addressed, as many proposed methods are evaluated on the Ninapro dataset, which was recorded under fixed and controlled conditions. Additionally, hand gesture recognition approaches are typically categorized into two groups: traditional machine learning methods, which rely primarily on spatially agnostic feature extraction (Fig. 2a), and deep learning methods, which predominantly employ spatiotemporal feature extraction configurations (Fig. 2c). Despite a few studies exploring spatially specific deep learning models [24,85] and spatially agnostic deep learning models [147] and highlighting their superior performance in myoelectric hand gesture recognition, these deep learning configurations remain largely underexplored in

the literature. Furthermore, while several studies have emphasized the importance of model selection in traditional machine learning – where LDA has been shown to achieve the highest recognition accuracy in many cases [25,26,65] – the impact of decision boundaries on recognition performance has received little attention. Le, et al. [145] found that when trained on data recorded across multiple arm positions, a 1D-CNN outperformed LDA on seen arm positions. However, on unseen arm positions, the performance of both models was comparable.

Controlling prosthetic hands is a primary application of myoelectric hand gesture recognition. For commercial deployment, machine learning models must be implemented on wearable hardware that supports fast inference [111,148], while also being power-efficient for prolonged use and cost-effective for large-scale production. However, most recent research, particularly in the area of deep learning, remains in the experimental phase. One major barrier is that efficient inference with deep learning models typically requires specialized hardware, such as GPUs or embedded accelerators [149,150], which may not be practical for integration into wearable prosthetic devices. As a result, traditional machine learning models remain the default strategy for control systems in commercial prosthetic hands [3].

### 6.2. Immediate future research directions

#### 6.2.1. Benchmarking datasets reflecting real-world conditions

In many previous studies, researchers often evaluated their proposed machine learning models on privately recorded datasets. However, recognition performance is influenced by various factors in the experimental setup, with participants' expertise being the most significant. Furthermore, when tested on public benchmarking datasets, there are often discrepancies in the train-test cross-validation schemes. Many studies fail to reserve full gesture repetitions for testing, thereby violating the fundamental goal of hand gesture recognition, to accurately predict newly encountered hand gestures. Consequently, we did not compare recognition performance across studies due to the lack of consistency.

The Ninapro dataset [77] is often regarded as the standard benchmarking dataset for myoelectric hand gesture recognition due to its large number of participants, extensive range of hand gestures, and diverse recording setups. However, its main limitation is the lack of diversity in hand gesture recognition conditions, as it was primarily recorded with the arm in fixed postures. Additionally, the high diversity of hand gestures may not be essential for myoelectric hand gesture recognition research, as several surveys [3,17] have highlighted that the quality of grasp control, such as recognition accuracy under disturbances, is more important than the number of grasp types. Furthermore, evaluating recognition performance on new hand gestures within the same controlled conditions does not accurately reflect real-world scenarios. Instead, evaluation should include not only training conditions but also novel ones. The Biosignals Repository [26,27,30] addresses some of these issues by providing three datasets with gestures recorded under varied contraction forces, different arm and hand orientations, and including both able-bodied and amputee participants. This diversity makes it a strong candidate for becoming a standard benchmarking dataset. However, the Biosignals Repository is not without drawbacks. Most of the recorded gestures involve wrist movements rather than a comprehensive range of hand gestures, and only stationary states of hand gestures are captured. To advance neuroprosthetic research and improve recognition accuracy under real-world conditions, future sEMG hand gesture datasets should not only incorporate commonly used hand gestures, but also capture gestures under a broader range of recording conditions (such as contraction forces, arm position, usage conditions, skin-electrode conditions), ensuring better generalizability and real-world applicability.

**Table 5**

Key machine learning techniques used for myoelectric hand gesture recognition, and their applications, advantages, and challenges.

Machine learning techniques	Application	Advantages	Challenges
-Spatially agnostic feature extraction (Fig. 2a).	-Feature Extraction.	-Enforces a strict temporal-spatial divide. -Low computational cost. -High interpretability.	-Limited to temporal features only.
-Spatially Specific Feature Extraction (Fig. 2b).	-Feature Extraction.	-Different features based on electrode position.	-Less interpretable than spatially agnostic methods. - Not commonly used.
-Spatiotemporal Feature Extraction (Fig. 2c).	-Feature Extraction.	-Simultaneously captures spatial and temporal information. -Fast inference, can run on low-cost portable hardware. -Subspace method, generative model, less sensitive to outliers [43]. -Proven effectiveness under non-stationary conditions of sEMG [25,65].	- Reduced interpretability - Prone to overfitting -Limited scalability due to eigenvalue decomposition, singular value decomposition solution [43].
-LDA Linear Decision Boundary.	-Classification.	-Fast inference, can run on low-cost portable hardware. -Scalable due to use of quadratic programming [151].	- Limited to binary classification. -Sensitive to outliers due to large margin loss [151,152].
-Linear SVM Linear Decision Boundary.	-Classification.	-Scalable through parallelizable computation and mini-batch training [79].	-Require specialized hardware for inference. -Tends to favor easily classified examples due to cross-entropy loss [153]. -Not commonly used.
-Neural Network without Activation (Linear Decision Boundary).	-Classification.	-Highly flexible decision boundary.	-Prone to overfitting and poor generalization under sEMG non-stationarity [25,65].
-Non-Linear Decision Boundary.	-Classification.	-Learns hierarchical spatiotemporal features [36]. -Scalable to large datasets. -Improved recognition performance over spatiotemporal variants [24]. -Learns sequential dependencies effectively [79].	-Require specialized hardware for inference -Risk of overfitting with small datasets [145]. -Require specialized hardware for inference. -Limited to hierarchical learning of temporal features only. -Require specialized hardware for inference. -Prone to gradient issues with long sequences [79]. -Often configured to learn spatiotemporal information (Fig. 10b).
-Convolutional Neural Network (Spatiotemporal Configuration Fig. 6, Fig. 7 and Fig. 8)	-Feature Extraction and Classification.	-Captures global information [89]. -Processes long sequences via positional encoding and multi-head attention [86]. -Improves recognition under dynamic conditions [26,27,103,154].	-Requires large datasets for effective training [89]. -Often configured to learn spatiotemporal information (Fig. 12). -Increases user-training burden. -Not guaranteed performance improvement under new conditions [145]. -Increases system complexity.
-Convolutional Neural Network (Temporal Configuration Fig. 9).	-Feature Extraction and Classification.	-Enhances recognition performance under dynamic conditions [103,106]. -Improves recognition performance under electrode shifts and changes in arm position [48,110].	-High computational complexity.
-Recurrent Neural Network (Fig. 10).	-Feature Extraction and Classification.	-Improves hand gesture output stability [112,115,116]. -Improves hand gesture recognition under changes in contraction force and limb position [26,30,119].	-May increase output latency [111]; further evaluation required. -Feature design requires sEMG domain expertise.
-Transformer (Fig. 11 [90]).	-Feature Extraction and Classification.	-Improves cross-session [65], cross-subject [64] and cross-condition [121,146] recognition performance.	-Requires a small amount of labeled calibration gestures. -Limited to spatial transfer [65] and spatiotemporal transfer [29]. -Potential negative transfer [120]. -Potential negative transfer [120].
-Data Abundance.	-Data Collection.	-Maintains recognition accuracy during prolonged usage [142]. -Does not require labeled data.	
-Multi-Modal Sensory Fusion.	-Data Collection.		
-Sparse Representation [48,110].	-Classification.		
-Increasing Temporal Context[112,115,116].	-Feature Extraction.		
-Leveraging Shift-Invariant Features [26,30,119].	-Feature Extraction.		
-Supervised Transfer Learning [29,65].	-Feature Extraction and Classification.		
-Unsupervised Transfer Learning [140].	-Feature Extraction and Classification.		

### 6.2.2. Strategies to enhance interpretability and recognition accuracy

The lack of interpretability in deep models remains a significant drawback, potentially impeding the widespread adoption of deep learning solutions for prosthetic hand control. Specifically, hand gesture consistency, which reflects user expertise, is commonly measured by the distance between gesture clusters [33]. Among various feature extraction methods, spatially agnostic configurations are considered the most interpretable (Figure 2a) due to their ability to preserve spatial information [147]. Conversely, most deep learning models are inadvertently designed to focus exclusively on either spatial information (Fig. 7 [81]) or spatiotemporal information (Fig. 6 [41], Fig. 8 [42], Fig. 10b). This tendency often leads to the formation of highly abstract representations, which increases the risk of overfitting. However, it is possible to exclude spatial information during training through careful network design. For example, restricting convolutional receptive fields to a single channel can prevent the unintended incorporation of spatial information [24,147] (Fig. 9).

Numerous studies on myoelectric hand gesture recognition have highlighted the robustness of linear decision boundaries in addressing sEMG non-stationarity, particularly when using the LDA classifier [25,26,65,66]. Furthermore, research on myoelectric proportional control [155] has concluded that simple linear regression performs comparably to highly complex deep neural networks in online settings. This is because of human motor learning and adaptation, where sufficient training enables users to become proficient with the system. As illustrated in Fig. 3, linear decision boundaries also reduce the likelihood of overfitting compared to complex nonlinear decision boundaries.

For the reasons highlighted above, we recommend that future research on myoelectric hand gesture recognition focus on spatially agnostic feature extraction combined with linear decision boundaries. Additionally, in the context of highly divergent datasets (Fig. 13), using linear decision boundaries instead of nonlinear ones shifts the responsibility for discriminative power to the feature extraction stage, facilitating the learning of shift-invariant representations. Such shift-invariant representations can be achieved through divergence minimization loss functions, such as Maximum Mean Discrepancy [129]. While current literature features many new models that utilize the latest techniques in machine learning, such as Generative Adversarial Networks [40], we make a strategic choice to exclude such research from this review due to the lack of model interpretability resulting from their architecture.

### 6.2.3. Further examination of effect of increasing window size on real-time performance

Due to the highly stochastic nature of sEMG signals, instantaneous values have limited utility. Instead, moving window is commonly employed to extract temporal features from the signal. Increasing the window size generally improves recognition accuracy. For example, Menon, et al. [112] conducted a comprehensive study and found that classification accuracy increased with larger window sizes. In experiments with able-bodied participants using 16 electrodes, accuracy rose from 80% at a window size of 50 ms to 90% at 300 ms, with minimal improvements beyond 300 ms. A similar trend was observed in individuals with partial-hand amputations. Phinyomark, et al. [25] reported similar findings using a feature set comprising Sample Entropy, Cepstral Coefficients, Root Mean Square, and Waveform length. Their study demonstrated recognition accuracy increasing from 95.71% at a 125 ms window size to 98.79% at 500 ms. Furthermore, Betthausen, et al. [116] demonstrated that an encoder-decoder temporal convolutional neural network with temporal context size of 1600 ms improved performance stability, even in online

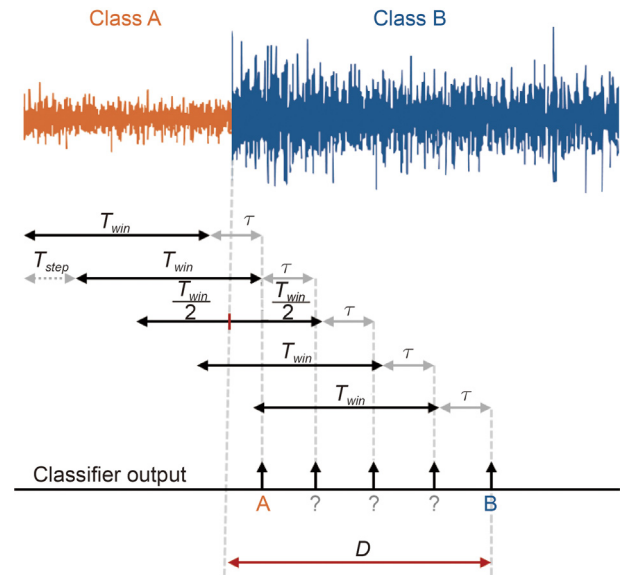


Fig. 19. Maximum delay between a movement class change and the controller update in overlapping moving window scheme.

Source: Adapted from [111].

control tasks. This improvement is attributed to the ability of longer window sizes to capture long-term dependencies relevant to human movement while suppressing instantaneous noise. In contrast, the degree of overlap (step size) between consecutive windows does not significantly affect recognition accuracy, regardless of window size or the number of channels used [112]. This is because overlapping primarily functions as a method of data augmentation, increasing variation in the training dataset. In real-time control applications, the step size is typically chosen to be greater than the processing time (*i.e.*, the time required for feature extraction and machine learning computations) [148].

However, using long temporal contexts is generally discouraged in the field of myoelectric hand gesture recognition due to the issue of increased perceived delay in real-time control, even when computation time is negligible. Controller delay refers to the time difference between an input action and the corresponding output response (Fig. 19) [156]. This unintended effect has been previously discussed in [111,148]. Let  $T_{win}$  denote the window size (or analysis window size),  $T_{step}$  represent the step size, and  $\tau$  indicate the processing time. The worst-case scenario for controller delay is given by [111,148]:

$$D \approx \frac{1}{2}T_{win} + T_{step} + \tau \quad (11)$$

Controller delay is widely regarded as the primary bottleneck in the usability of most interactive systems [157]. Increased input latency can make prosthetic devices feel unintuitive, leading to a diminished sense of embodiment. Physiological research indicates that the electromechanical delay of human muscles ranges from 30 ms to 100 ms [158]. Thus, keeping the controller delay below this range allows users to maintain intuitive control.

It is worth noting that in Farrell [148], it was assumed that when the current analysis falls between two classes, the prediction is assigned to the previous class. This led to discrepancies between the theoretical perceived delay and the experimental delay observed in their validation experiments. As recent findings in the literature highlight significant improvements in recognition performance with increased temporal context size [25,112,116], they suggest that the relationship between temporal context size and perceived delay requires further reevaluation in future research.

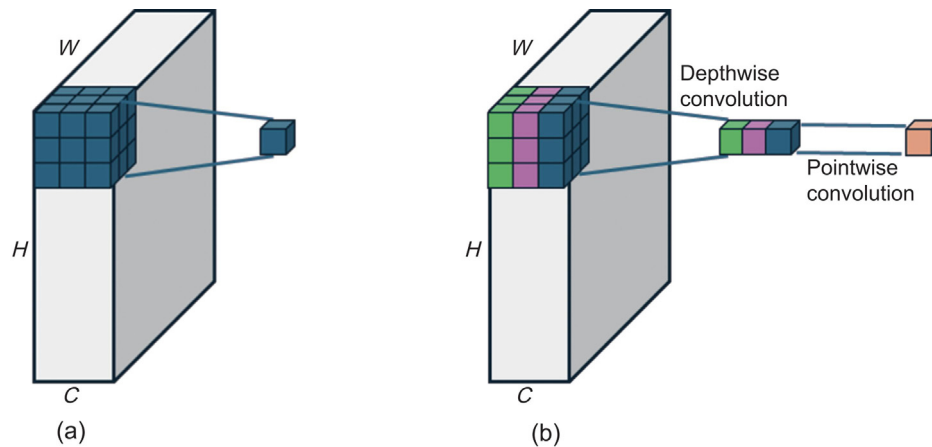


Fig. 20. Illustration of (a) standard convolution operation and (b) depthwise separable convolution operation.

#### 6.2.4. Supervised transfer learning of temporal features

In machine learning applications such as computer vision, audio signal processing, and natural language processing, success is largely driven by the effective utilization of large datasets. Although numerous public myoelectric hand gesture recognition datasets are available, effectively leveraging these datasets is challenging. These challenges arise due to variations in recording instruments and sensor placement. Naively combining these datasets, where possible, can result in highly divergent data (as shown in Fig. 13), leading to suboptimal recognition performance [145,146]. Additionally, there are cases where the number of channels is not consistent across datasets, making data aggregation impossible.

Supervised transfer learning in myoelectric hand gesture recognition is often employed to enhance the recognition performance of a target dataset by effectively leveraging a source dataset. Although numerous studies have focused on deep transfer learning [28,42,64,128], most of these works utilized spatiotemporal transfer learning models due to their architectural design. However, to the best of our knowledge, deep temporal transfer learning remains unexplored. Deep temporal transfer learning can be achieved by designing the representation learning stage of the model, such as CNN layers, to focus solely on temporal features (for instance configuration in Fig. 9). By isolating the transfer of knowledge to temporal information alone, deep temporal transfer learning can address spatial mismatches between source and target datasets, including discrepancies in the number of input electrodes.

#### 6.2.5. Light-weight inference models for deployment on wearable hardware

Deep learning models are typically trained on high-performance computers equipped with abundant computational resources. In contrast, inference is usually performed on more resource-constrained hardware, where size, power efficiency, and real-time performance are critical, especially in applications like prosthetic hands and rehabilitation devices [3, 19]. This mismatch presents a challenge: many deep learning models are too computationally intensive to run efficiently on lightweight hardware while meeting the low-latency requirements of real-time prosthetic control [132,150]. To bridge this gap, researchers have developed various machine learning techniques to reduce the computational footprint and enable deployment on portable or embedded devices [159–162]. Despite

its practical importance, model size reduction remains a relatively underexplored area of research in the field of myoelectric hand gesture recognition. Overall, there are two main approaches for developing low-footprint inference models: designing lightweight deep learning architectures and applying model compression techniques.

- **Lightweight deep learning architectures** are a class of models specifically designed to minimize computational requirements while maintaining strong discriminative performance. MobileNet [160] and MobileNetV2 [163] are two notable examples, developed to reduce the computational footprint of convolution operations without significantly compromising accuracy. MobileNet [160] is a lightweight convolutional neural network architecture designed for efficient performance on mobile and embedded devices. Its core innovation is the use of depthwise separable convolutions, which break a standard convolution into two simpler operations: a depthwise convolution, which applies a single filter to each input channel, and a pointwise convolution, which uses  $1 \times 1$  filters to combine the outputs across channels (Fig. 20b). MobileNetV2 [163] extends MobileNet by introducing inverted residual blocks and linear bottlenecks, which improve computation efficiency and recognition accuracy. Depthwise separable convolutions, inverted residual blocks and linear bottlenecks has been featured in various myoelectric hand gesture recognition architecture [92,93].
- **Model compression techniques** are used to reduce the size, memory footprint, and computational cost of deep learning models while preserving their performance. **Quantization**, **pruning**, and **knowledge distillation** are techniques commonly used model compression techniques. **Quantization** reduces the precision of model weights and activations, for example, converting 32-bit floating-point to 16-bit floating-point, to enable faster inference and lower memory usage [162]. **Pruning** is a technique that removes redundant or less important parameters, such as small-weight connections or entire neurons, leading to sparser and more efficient models [164]. **Knowledge distillation** is a technique in which a smaller “student” model learns to replicate the behavior of a larger, more accurate “teacher” model by training on the “teacher” model’s soft output predictions or latent representations [161].

## 7. Conclusion

Myoelectric hand gesture recognition is challenging due to multiple sources of unpredictable disturbances such as *covariate*

shifts caused by the non-stationary nature of sEMG signals.... To enhance the accuracy of hand gesture recognition in real-world conditions under these disturbances, this survey offers comprehensive insights into the field. We introduce a new taxonomy for categorizing recognition models based on feature extraction and decision boundaries. Additionally, we review advancements in traditional and deep learning, along with strategies to address non-stationarity. Through our intensive survey of relevant literature, we identified 7 key challenges of myoelectric hand gesture recognition which are (1) leveraging cross-session, cross-subject data, (2) increased user training burden, (3) negative knowledge transfer, (4) generalization to unseen conditions, (5) low interpretability, (6) minimizing inference time, (7) wearable inference hardware. Finally, we highlight key research directions, including the need for a realistic benchmarking dataset, improved interpretability, real-time delay assessment, supervised transfer learning of temporal features and deployable lightweight machine learning models.

### CRedit authorship contribution statement

**Hongquan Le:** Writing – review & editing, Conceptualization, Writing – original draft. **Marc in Het Panhuis:** Supervision, Writing – review & editing, Funding acquisition. **Gursel Alici:** Writing – review & editing, Project administration, Conceptualization, Supervision, Funding acquisition.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

The research was made possible through financial support from the ARC Centre of Excellence for Electromaterials Science (CE140100012), the ARC-Discovery Project (DP210102911), and the University of Wollongong.

### References

- [1] N.H.H. Mohamad Hanif, P. Chappell, N. White, A. Cranny, A psychophysical investigation on vibrotactile sensing for transradial prosthesis users, *Cogent Eng.* 5 (1) (2018) 1539943, <http://dx.doi.org/10.1080/23311916.2018.1539943>.
- [2] H. van Duinen, S.C. Gandevia, Constraints for control of the human hand, *J. Physiol.* 589 (23) (2011) 5583–5593, <http://dx.doi.org/10.1111/jphysiol.2011.217810>.
- [3] F. Cordella, et al., Literature review on needs of upper limb prosthesis users, *Front. Neurosci.* 10 (2016) 209, <http://dx.doi.org/10.3389/fnins.2016.00209>.
- [4] A. Sahu, R. Sagar, S. Sarkar, S. Sagar, Psychological effects of amputation: A review of studies from India, *Ind. Psychiatry J.* 25 (1) (2016) 4, <http://dx.doi.org/10.4103/0972-6748.196041>.
- [5] G. Buckingham, J. Parr, G. Wood, S. Vine, P. Dimitriou, S. Day, The impact of using an upper-limb prosthesis on the perception of real and illusory weight differences, *Psychon. Bull. Rev.* 25 (4) (2018) 1507–1516, <http://dx.doi.org/10.3758/s13423-017-1425-2>.
- [6] C. Antfolk, M. D'Alonzo, B. Rosen, G. Lundborg, F. Sebelius, C. Cipriani, Sensory feedback in upper limb prosthetics, *Expert. Rev. Med. Devices* 10 (1) (2013) 45–54, <http://dx.doi.org/10.1586/erd.12.68>.
- [7] S.L. Carey, D.J. Lura, M.J. Highsmith, Differences in myoelectric and body-powered upper-limb prostheses: Systematic literature review, *J. Rehabil. Res. Dev.* 52 (3) (2015) <http://dx.doi.org/10.1097/JPO.000000000000159>.
- [8] M. Rooks, R. Vogel, L. Fleming, Myoelectric prostheses. A long-term follow-up and a study of the use of alternate prostheses, *J. Bone Jt. Surg. Am. Vol. 75* (12) (1993) 1781–1789.
- [9] A. Cloutier, J. Yang, Design, control, and sensory feedback of externally powered hand prostheses: a literature review, *Crit. Rev.™ Biomed. Eng.*

- [10] P. Weiner, J. Starke, S. Rader, F. Hundhausen, T. Asfour, Designing prosthetic hands with embodied intelligence: The KIT prosthetic hands, *Front. Neurobot.* 16 (2022) 815716, <http://dx.doi.org/10.3389/fnbot.2022.815716>.
- [11] Y. Huang, et al., Human-like dexterous manipulation for anthropomorphic five-fingered hands: A review, *Biomim. Intell. Robot.* 5 (1) (2025) 100212, <http://dx.doi.org/10.1016/j.birob.2025.100212>, 2025/03/01.
- [12] Y. Li, et al., SoftGrasp: Adaptive grasping for dexterous hand based on multimodal imitation learning, *Biomim. Intell. Robot.* 5 (2) (2025) 100217, <http://dx.doi.org/10.1016/j.birob.2025.100217>, 2025/06/01/.
- [13] X. Ma, J. Zhang, B. Wang, J. Huang, G. Bao, Continuous adaptive gaits manipulation for three-fingered robotic hands via bioinspired fingertip contact events, *Biomim. Intell. Robot.* 4 (1) (2024) 100144, <http://dx.doi.org/10.1016/j.birob.2024.100144>, 2024/03/01/.
- [14] H. Zhou, C. Tawk, G. Alici, A multipurpose human-machine interface via 3D-printed pressure-based force myography, *IEEE Trans. Ind. Informatics* 20 (6) (2024) 8838–8849, <http://dx.doi.org/10.1109/TII.2024.3375376>.
- [15] Z. Liao, et al., Human-robot interface based on sEMG envelope signal for the collaborative wearable robot, *Biomim. Intell. Robot.* 3 (1) (2023) 100079, <http://dx.doi.org/10.1016/j.birob.2022.100079>.
- [16] I. Kyranou, S. Vijayakumar, M.S. Erden, Causes of performance degradation in electromyographic pattern recognition in upper limb prostheses, *Front. Neurobot.* 12 (2018) 58, <http://dx.doi.org/10.3389/fnbot.2018.00058>.
- [17] B. Stephens-Fripp, M. Jean Walker, E. Goddard, G. Alici, A survey on what Australians with upper limb difference want in a prosthesis: justification for using soft robotics and additive manufacturing for customized prosthetic hands, *Disabil. Rehabil. Assist. Technol.* (2019) 1–8, <http://dx.doi.org/10.1080/17483107.2019.1580777>.
- [18] M.A. Powell, R.R. Kaliki, N.V. Thakor, User training for pattern recognition-based myoelectric prostheses: Improving phantom limb movement consistency and distinguishability, *IEEE Trans. Neural Syst. Rehabil. Eng.* 22 (3) (2013) 522–532, <http://dx.doi.org/10.1109/TNSRE.2013.2279737>.
- [19] D. Yang, Y. Gu, N.V. Thakor, H. Liu, Improving the functionality, robustness, and adaptability of myoelectric control for dexterous motion restoration, *Exp. Brain Res.* 237 (2) (2019) 291–311, <http://dx.doi.org/10.1007/s00221-018-5441-x>.
- [20] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, J. Yang, A framework for hand gesture recognition based on accelerometer and EMG sensors, *IEEE Trans. Syst. Man, Cybernetics- Part A: Syst. Humans* 41 (6) (2011) 1064–1076, <http://dx.doi.org/10.1109/TSMCA.2011.2116004>.
- [21] X. Zhang, X. Li, O.W. Samuel, Z. Huang, P. Fang, G. Li, Improving the robustness of electromyogram-pattern recognition for prosthetic control by a postprocessing strategy, *Front. Neurobot.* 11 (2017) 51, <http://dx.doi.org/10.3389/fnbot.2017.00051>.
- [22] G. Staude, W. Wolf, Objective motor response onset detection in surface myoelectric signals, *Med. Eng. Phys.* 21 (6–7) (1999) 449–467, [http://dx.doi.org/10.1016/S1350-4533\(99\)00067-3](http://dx.doi.org/10.1016/S1350-4533(99)00067-3).
- [23] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, W. Geng, A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition, *PLoS One* 13 (10) (2018) e0206049, <http://dx.doi.org/10.1371/journal.pone.0206049>.
- [24] W. Wei, Y. Wong, Y. Du, Y. Hu, M. Kankanhalli, W. Geng, A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface, *Pattern Recognit. Lett.* 119 (2019) 131–138, <http://dx.doi.org/10.1016/j.patrec.2017.12.005>.
- [25] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, Y. Laurillau, EMG feature evaluation for improving myoelectric pattern recognition robustness, *Expert Syst. Appl.* 40 (12) (2013) 4832–4840, <http://dx.doi.org/10.1016/j.eswa.2013.02.023>.
- [26] A.H. Al-Timemy, R.N. Khushaba, G. Bugmann, J. Escudero, Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees, *IEEE Trans. Neural Syst. Rehabil. Eng.* 24 (6) (2015) 650–661, <http://dx.doi.org/10.1109/TNSRE.2015.2445634>.
- [27] R.N. Khushaba, A. Al-Timemy, S. Kodagoda, K. Nazarpour, Combined influence of forearm orientation and muscular contraction on EMG pattern recognition, *Expert Syst. Appl.* 61 (2016) 154–161, <http://dx.doi.org/10.1016/j.eswa.2016.05.031>.
- [28] I. Ketykó, F. Kovács, K.Z. Varga, Domain adaptation for sEMG-based gesture recognition with recurrent neural networks, 2019, <http://dx.doi.org/10.1109/IJCNN.2019.8852018>, arXiv preprint arXiv:1901.06958.
- [29] A. Ameri, M.A. Akhaee, E. Scheme, K. Englehart, A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control, *IEEE Trans. Neural Syst. Rehabil. Eng.* (2019) <http://dx.doi.org/10.1109/TNSRE.2019.2962189>.
- [30] R.N. Khushaba, M. Takruri, J.V. Miro, S. Kodagoda, Towards limb position invariant myoelectric pattern recognition using time-dependent spectral features, *Neural Netw.* 55 (2014) 42–58, <http://dx.doi.org/10.1016/j.neunet.2014.03.010>.

- [31] A. Fougner, E. Scheme, A.D. Chan, K. Englehart, Ø. Stavdahl, Resolving the limb position effect in myoelectric pattern recognition, *IEEE Trans. Neural Syst. Rehabil. Eng.* 19 (6) (2011) 644–651, <http://dx.doi.org/10.1109/TNSRE.2011.2163529>.
- [32] M. Cifrek, S. Tonković, V. Medved, Measurement and analysis of surface myoelectric signals during fatigued cyclic dynamic contractions, *Measurement* 27 (2) (2000) 85–92, [http://dx.doi.org/10.1016/S0263-2241\(99\)00059-7](http://dx.doi.org/10.1016/S0263-2241(99)00059-7).
- [33] N.E. Bunderson, T.A. Kuiken, Quantification of feature space changes with experience during electromyogram pattern recognition control, *IEEE Trans. Neural Syst. Rehabil. Eng.* 20 (3) (2012) 239–246, <http://dx.doi.org/10.1109/TNSRE.2011.2182525>.
- [34] D. Farina, R. Merletti, D. Stegeman, Biophysics of the generation of EMG signals, *Electromyogr.: Physiol. Eng. Noninvasive Appl.* (2004) 81–105, <http://dx.doi.org/10.1002/9781119082934.ch02>.
- [35] R. Merletti, H. Hermens, Detection and conditioning of the surface EMG signal, *Electromyogr.: Physiol. Eng. Noninvasive Appl.* (2004) 107–131, <http://dx.doi.org/10.1002/9781119082934.ch03>.
- [36] G. Yao, T. Lei, J. Zhong, A review of convolutional-neural-network-based action recognition, *Pattern Recognit. Lett.* 118 (2019) 14–22, <http://dx.doi.org/10.1016/j.patrec.2018.05.018>.
- [37] K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos, in: *Advances in Neural Information Processing Systems*, 2014, pp. 568–576, <http://dx.doi.org/10.48550/arXiv.1406.2199>.
- [38] Z. Teng, G. Xu, R. Liang, M. Li, S. Zhang, Evaluation of synergy-based hand gesture recognition method against force variation for robust myoelectric control, *IEEE Trans. Neural Syst. Rehabil. Eng.* 29 (2021) 2345–2354, <http://dx.doi.org/10.1109/TNSRE.2021.3124744>.
- [39] A.J. Fuglevand, D.A. Winter, A.E. Patla, Models of recruitment and rate coding organization in motor-unit pools, *J. Neurophysiol.* 70 (6) (1993) 2470–2488, <http://dx.doi.org/10.1152/jn.1993.70.6.2470>.
- [40] S. Ni, M.A.A. Al-qaness, A. Hawbani, D. Al-Alimi, M. Abd Elaziz, A.A. Ewees, A survey on hand gesture recognition based on surface electromyography: Fundamentals, methods, applications, challenges and future trends, *Appl. Soft Comput.* 166 (2024) 112235, <http://dx.doi.org/10.1016/j.asoc.2024.112235>, 2024/11/01/.
- [41] M. Atzori, M. Cognolato, H. Müller, Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands, *Front. Neurobot.* 10 (2016) 9, <http://dx.doi.org/10.3389/fnbot.2016.00009>.
- [42] R. Soroushmojedhi, S. Javadzadeh, A. Pedrocchi, M. Gandolla, Transfer learning in hand movement intention detection based on surface electromyography signals, *Front. Neurosci.* 16 (2022) 977328, <http://dx.doi.org/10.3389/fnins.2022.977328>.
- [43] B. Ghoghj, M. Crowley, Linear and quadratic discriminant analysis: tutorial, 2019, <http://dx.doi.org/10.48550/arXiv.1906.02590>, arXiv preprint arXiv:1906.02590.
- [44] A. Phinyomark, S. Thongpanja, H. Hu, P. Phukpattaranont, C. Limsakul, The usefulness of mean and median frequencies in electromyography analysis, in: *Computational Intelligence in Electromyography Analysis—a Perspective on Current Applications and Future Challenges*, IntechOpen, 2012.
- [45] X. Zhang, L. Huang, H. Niu, Structural design and stiffness matching control of bionic variable stiffness joint for human–robot collaboration, *Biomim. Intell. Robot.* 3 (1) (2023) 100084, <http://dx.doi.org/10.1016/j.birob.2022.100084>.
- [46] O. Paiss, G.F. Inbar, Autoregressive modeling of surface EMG and its spectrum with application to fatigue, *IEEE Trans. Biomed. Eng.* BME-34 (10) (1987) 761–770, <http://dx.doi.org/10.1109/TBME.1987.325918>.
- [47] G. Li, et al., Multi-view fusion network-based gesture recognition using sEMG data, *IEEE J. Biomed. Heal. Informat.* (2023) <http://dx.doi.org/10.1109/JBHI.2023.3287979>.
- [48] J.L. Betthausen, et al., Limb position tolerant pattern recognition for myoelectric prosthesis control with adaptive sparse representations from extreme learning, *IEEE Trans. Biomed. Eng.* 65 (4) (2017) 770–778, <http://dx.doi.org/10.1109/TBME.2017.2719400>.
- [49] H. Huang, H.-B. Xie, J.-Y. Guo, H.-J. Chen, Ant colony optimization-based feature selection method for surface electromyography signals classification, *Comput. Biol. Med.* 42 (1) (2012) 30–38, <http://dx.doi.org/10.1016/j.compbiomed.2011.10.004>.
- [50] J. Kilby, H.G. Hosseini, Extracting effective features of SEMG using continuous wavelet transform, in: *2006 International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2006*, pp. 1704–1707, <http://dx.doi.org/10.1109/IEMBS.2006.260064>.
- [51] K. Kiatpanichagij, N. Afzulpurkar, Use of supervised discretization with PCA in wavelet packet transformation-based surface electromyogram classification, *Biomed. Signal Process. Control.* 4 (2) (2009) 127–138, <http://dx.doi.org/10.1016/j.bspc.2009.02.004>.
- [52] F. Duan, L. Dai, W. Chang, Z. Chen, C. Zhu, W. Li, sEMG-based identification of hand motion commands using wavelet neural network combined with discrete wavelet transform, *IEEE Trans. Ind. Electron.* 63 (3) (2015) 1923–1934, <http://dx.doi.org/10.1109/TIE.2015.2497212>.
- [53] R.N. Khushaba, A.H. Al-Timemy, A. Al-Ani, A. Al-Jumaily, A framework of temporal-spatial descriptors-based feature extraction for improved myoelectric pattern recognition, *IEEE Trans. Neural Syst. Rehabil. Eng.* 25 (10) (2017) 1821–1831, <http://dx.doi.org/10.1109/TNSRE.2017.2687520>.
- [54] Y. Zou, L. Cheng, Z. Li, A multimodal fusion model for estimating human hand force: Comparing surface electromyography and ultrasound signals, *IEEE Robot. Autom. Mag.* 29 (4) (2022) 10–24, <http://dx.doi.org/10.1109/MRA.2022.3177486>.
- [55] M.R. Canal, Comparison of wavelet and short time Fourier transform methods in the analysis of EMG signals, *J. Med. Syst.* 34 (1) (2010) 91–94, <http://dx.doi.org/10.1007/s10916-008-9219-8>.
- [56] A. Phinyomark, R.N. Khushaba, E. Ibáñez-Marcelo, A. Patania, E. Scheme, G. Petri, Navigating features: a topologically informed chart of electromyographic features space, *J. R. Soc. Interface* 14 (137) (2017) 20170734, <http://dx.doi.org/10.1098/rsif.2017.0734>.
- [57] K. Nazarpour, A.H. Al-Timemy, G. Bugmann, A. Jackson, A note on the probability distribution function of the surface electromyogram signal, *Brain Res. Bull.* 90 (2013) 88–91, <http://dx.doi.org/10.1016/j.brainresbull.2012.09.012>.
- [58] P.A. Kaplanis, C.S. Pattichis, D. Zazula, Multiscale entropy-based approach to automated surface EMG classification of neuromuscular disorders, *Med. Biol. Eng. Comput.* 48 (8) (2010) 773–781, <http://dx.doi.org/10.1007/s11517-010-0629-7>.
- [59] X. Zhang, X. Ren, X. Gao, X. Chen, P. Zhou, Complexity analysis of surface EMG for overcoming ECG interference toward proportional myoelectric control, *Entropy* 18 (4) (2016) 106, <http://dx.doi.org/10.3390/e18040106>.
- [60] E. Scheme, K. Englehart, On the robustness of EMG features for pattern recognition based myoelectric control; a multi-dataset comparison, in: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014*, pp. 650–653, <http://dx.doi.org/10.1109/EMBC.2014.6943675>.
- [61] M.A. Oskoei, H. Hu, J.Q. Gan, Manifestation of fatigue in myoelectric signals of dynamic contractions produced during playing PC games, in: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2008*, pp. 315–318, <http://dx.doi.org/10.1109/IEMBS.2008.4649153>.
- [62] T.W. Beck, et al., Comparison of Fourier and wavelet transform procedures for examining the mechanomyographic and electromyographic frequency domain responses during fatiguing isokinetic muscle actions of the biceps brachii, *J. Electromyography Kinesiol.* 15 (2) (2005) 190–199, <http://dx.doi.org/10.1016/j.jelekin.2004.08.007>.
- [63] K. Englehart, B. Hudgins, P.A. Parker, M. Stevenson, Classification of the myoelectric signal using time-frequency based representations, *Med. Eng. Phys.* 21 (6–7) (1999) 431–438, [http://dx.doi.org/10.1016/S1350-4533\(99\)00066-1](http://dx.doi.org/10.1016/S1350-4533(99)00066-1).
- [64] U. Cote-Allard, et al., Deep learning for electromyographic hand gesture signal classification using transfer learning, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (4) (2019) 760–771, <http://dx.doi.org/10.1109/TNSRE.2019.2896269>.
- [65] M.M.-C. Vidovic, H.-J. Hwang, S. Amsüss, J.M. Hahne, D. Farina, K.-R. Müller, Improving the robustness of myoelectric pattern recognition for upper limb prostheses by covariate shift adaptation, *IEEE Trans. Neural Syst. Rehabil. Eng.* 24 (9) (2015) 961–970, <http://dx.doi.org/10.1109/TNSRE.2015.2492619>.
- [66] T. Lorrain, N. Jiang, D. Farina, Influence of the training set on the accuracy of surface EMG classification in dynamic contractions for the control of multifunction prostheses, *J. Neuroeng. Rehabil.* 8 (1) (2011) 25, <http://dx.doi.org/10.1186/1743-0003-8-25>.
- [67] R.N. Khushaba, A. Al-Jumaily, Channel and feature selection in multifunction myoelectric control, in: *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2007*, pp. 5182–5185, <http://dx.doi.org/10.1109/IEMBS.2007.4353509>.
- [68] A. Phinyomark, P. Phukpattaranont, C. Limsakul, Feature reduction and selection for EMG signal classification, *Expert Syst. Appl.* 39 (8) (2012) 7420–7431, <http://dx.doi.org/10.1016/j.eswa.2012.01.102>.
- [69] K. Englehart, B. Hudgins, P. Parker, M. Stevenson, Time-frequency representation for classification of the transient myoelectric signal, in: *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 20 Biomedical Engineering Towards the Year 2000 and beyond (Cat. No. 98CH36286)*, vol. 5, IEEE, 1998, pp. 2627–2630, <http://dx.doi.org/10.1109/IEMBS.1998.745109>.
- [70] B. Ghoghj, F. Karray, M. Crowley, Fisher and kernel fisher discriminant analysis: tutorial, 2019, <http://dx.doi.org/10.48550/arXiv.1906.09436>, arXiv preprint arXiv:1906.09436.

- [71] S. Liang, et al., Identification of gesture based on combination of raw sEMG and sEMG envelope using supervised learning and univariate feature selection, *J. Bionic Eng.* 16 (4) (2019) 647–662, <http://dx.doi.org/10.1007/s42235-019-0052-1>.
- [72] A.S.M. Miah, J. Shin, M.A.M. Hasan, Effective features extraction and selection for hand gesture recognition using sEMG signal, *Multimedia Tools Appl.* (2024) 1–25, <http://dx.doi.org/10.1007/s11042-024-19468-2>.
- [73] H.M. Al-Angari, G. Kanitz, S. Tarantino, C. Cipriani, Distance and mutual information methods for EMG feature and channel subset selection for classification of hand movements, *Biomed. Signal Process. Control.* 27 (2016) 24–31, <http://dx.doi.org/10.1016/j.bspc.2016.01.011>.
- [74] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (8) (2005) 1226–1238, <http://dx.doi.org/10.1109/TPAMI.2005.159>.
- [75] J. Liu, X. Li, G. Li, P. Zhou, EMG feature assessment for myoelectric pattern recognition and channel selection: a study with incomplete spinal cord injury, *Med. Eng. Phys.* 36 (7) (2014) 975–980, <http://dx.doi.org/10.1016/j.medengphy.2014.04.003>.
- [76] I. Mesa, A. Rubio, I. Tubia, J. De No, J. Diaz, Channel and feature selection for a surface electromyographic pattern recognition task, *Expert Syst. Appl.* 41 (11) (2014) 5190–5200, <http://dx.doi.org/10.1016/j.eswa.2014.03.014>.
- [77] M. Atzori, et al., Electromyography data for non-invasive naturally-controlled robotic hand prostheses, *Nature* 1 (12/23) (2014) <http://dx.doi.org/10.1038/sdata.2014.53>.
- [78] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324, <http://dx.doi.org/10.1109/5.726791>.
- [79] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- [80] A.v.d. Oord, et al., Wavenet: A generative model for raw audio, 2016, <http://dx.doi.org/10.48550/arXiv.1609.03499>, arXiv preprint arXiv:1609.03499.
- [81] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, J. Li, Gesture recognition by instantaneous surface EMG images, *Sci. Rep.* 6 (1) (2016) 36571, <http://dx.doi.org/10.1038/srep36571>.
- [82] P. Koch, H. Phan, M. Maass, F. Katzberg, A. Mertins, Recurrent Neural Network Based Early Prediction of Future Hand Movements, *IEEE*, 2018, pp. 4710–4713, <http://dx.doi.org/10.1109/EMBC.2018.8513145>, 2018.
- [83] N.K. Karnam, S.R. Dubey, A.C. Turlapaty, B. Gokaraju, EMGHandNet: A hybrid CNN and Bi-LSTM architecture for hand activity classification using surface EMG signals, *Biocybern. Biomed. Eng.* 42 (1) (2022) 325–340, <http://dx.doi.org/10.1016/j.bbe.2022.02.005>.
- [84] Y. Wu, B. Zheng, Y. Zhao, Dynamic gesture recognition based on LSTM-CNN, in: 2018 Chinese Automation Congress, CAC, IEEE, 2018, pp. 2446–2450, <http://dx.doi.org/10.1109/CAC.2018.8623035>.
- [85] H. Le, G.M. Spinks, M. in Het Panhuis, G. Alici, Cross-day myoelectric gesture recognition with hybrid multistream CNN-bidirectional LSTM, in: 2025 IEEE International Conference on Mechatronics, ICM'25, Wollongong, Australia, 2025, <http://dx.doi.org/10.1109/ICM62621.2025.10934890>.
- [86] A. Vaswani, et al., Attention is all you need, *Adv. Neural Inf. Process. Syst.* (2017).
- [87] J.D.M.-W.C. Kenton, L.K. Toutanova, Bert: pre-training of deep bidirectional transformers for language understanding, in: *Proceedings of NaacL-HLT*, vol. 1, (2) Minneapolis, Minnesota, 2019, <http://dx.doi.org/10.18653/v1/N19-1423>.
- [88] J. Achiam, et al., Gpt-4 technical report, 2023, <http://dx.doi.org/10.48550/arXiv.2303.08774>, arXiv preprint arXiv:2303.08774.
- [89] A. Dosovitskiy, An image is worth 16x16 words: Transformers for image recognition at scale, 2020, <http://dx.doi.org/10.48550/arXiv.2010.11929>, arXiv preprint arXiv:2010.11929.
- [90] S. Zabihi, E. Rahimian, A. Asif, A. Mohammadi, Trahgr: transformer for hand gesture recognition via electromyography, *IEEE Trans. Neural Syst. Rehabil. Eng.* (2023) <http://dx.doi.org/10.1109/TNSRE.2023.3324252>.
- [91] S. Shen, X. Wang, F. Mao, L. Sun, M. Gu, Movements classification through sEMG with convolutional vision transformer and stacking ensemble learning, *IEEE Sensors J.* 22 (13) (2022) 13318–13325, <http://dx.doi.org/10.1109/JSEN.2022.3179535>.
- [92] S. Zabihi, E. Rahimian, A. Asif, A. Mohammadi, Light-weight CNN-attention based architecture for hand gesture recognition via electromyography, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2023, pp. 1–5, <http://dx.doi.org/10.1109/ICASSP49357.2023.10096292>.
- [93] Y. Liu, X. Li, L. Yang, G. Bian, H. Yu, A CNN-transformer hybrid recognition approach for sEMG-based dynamic gesture prediction, *IEEE Trans. Instrum. Meas.* 72 (2023) 1–16, <http://dx.doi.org/10.1109/TIM.2023.3273651>.
- [94] Z. Wang, J. Yao, M. Xu, M. Jiang, J. Su, Transformer-based network with temporal depthwise convolutions for sEMG recognition, *Pattern Recognit.* 145 (2024) 109967, <http://dx.doi.org/10.1016/j.patcog.2023.109967>.
- [95] S. Duan, L. Wu, A. Liu, X. Chen, A global-local fusion model exploring temporal-spatial dependence for multimodal hand gesture recognition, *IEEE Trans. Med. Robot. Bionics* (2025) <http://dx.doi.org/10.1109/TMRB.2025.3550646>.
- [96] L. Hargrove, K. Englehart, B. Hudgins, The effect of electrode displacements on pattern recognition based myoelectric control, in: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2006, pp. 2203–2206, <http://dx.doi.org/10.1109/IEMBS.2006.260681>.
- [97] L. Hargrove, K. Englehart, B. Hudgins, A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control, *Biomed. Signal Process. Control.* 3 (2) (2008) 175–180, <http://dx.doi.org/10.1016/j.bspc.2007.11.005>.
- [98] E. Scheme, A. Fougner, Ø. Stavdahl, A.D. Chan, K. Englehart, Examining the adverse effects of limb position on pattern recognition based myoelectric control, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010, pp. 6337–6340, <http://dx.doi.org/10.1109/IEMBS.2010.5627638>.
- [99] J. Liu, D. Zhang, X. Sheng, X. Zhu, Quantification and solutions of arm movements effect on sEMG pattern recognition, *Biomed. Signal Process. Control.* 13 (2014) 189–197, <http://dx.doi.org/10.1016/j.bspc.2014.05.001>.
- [100] A. Radmand, E. Scheme, K. Englehart, A characterization of the effect of limb position on EMG features to guide the development of effective prosthetic control schemes, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 662–667, <http://dx.doi.org/10.1109/EMBC.2014.6943678>.
- [101] X. Wang, H. Yu, S. Kold, O. Rahbek, S. Bai, Wearable sensors for activity monitoring and motion control: A review, *Biomim. Intell. Robot.* 3 (1) (2023) 100089, <http://dx.doi.org/10.1016/j.birob.2023.100089>.
- [102] R. Tchanchane, H. Zhou, S. Zhang, G. Alici, A review of hand gesture recognition systems based on noninvasive wearable sensors, *Adv. Intell. Syst.* 5 (10) (2023) 2300207, <http://dx.doi.org/10.1002/aisy.202300207>.
- [103] H.E. Williams, A.W. Shehata, M.R. Dawson, E. Scheme, J.S. Hebert, P.M. Pilarski, Recurrent convolutional neural networks as an approach to position-aware myoelectric prosthesis control, *IEEE Trans. Biomed. Eng.* 69 (7) (2022) 2243–2255, <http://dx.doi.org/10.1109/TBME.2022.3140269>.
- [104] O.W. Samuel, et al., Resolving the adverse impact of mobility on myoelectric pattern recognition in upper-limb multifunctional prostheses, *Comput. Biol. Med.* 90 (2017) 76–87, <http://dx.doi.org/10.1016/j.compbiomed.2017.09.013>.
- [105] W. Wei, L. Ren, M. Zhou, Z. Ma, K. Zhao, X. Xu, Improving unimodal sEMG-based pattern recognition through multimodal generative adversarial learning, *IEEE Trans. Instrum. Meas.* (2025) <http://dx.doi.org/10.1109/TIM.2025.3556214>.
- [106] S. Zhang, H. Zhou, R. Tchanchane, G. Alici, Hand gesture recognition across various limb positions using a multi-modal sensing system based on self-adaptive data-fusion and convolutional neural networks (CNNs), *IEEE Sensors J.* (2024) <http://dx.doi.org/10.1109/JSEN.2024.3389963>.
- [107] P.R. Young, et al., The effects of limb position and grasped load on hand gesture classification using electromyography, force myography, and their combination, *PLoS One* 20 (4) (2025) e0321319, <http://dx.doi.org/10.1371/journal.pone.0321319>.
- [108] T.T. Oyemakinde, et al., A novel sEMG-FMG combined sensor fusion approach based on an attention-driven CNN for dynamic hand gesture recognition, *IEEE Trans. Instrum. Meas.* (2025) <http://dx.doi.org/10.1109/TIM.2025.3552811>.
- [109] M. Rohr, et al., On the benefit of FMG and EMG sensor fusion for gesture recognition using cross-subject validation, *IEEE Trans. Neural Syst. Rehabil. Eng.* (2025) <http://dx.doi.org/10.1109/TNSRE.2025.3543649>.
- [110] J.L. Betthausser, L.E. Osborn, R.R. Kaliki, N.V. Thakor, Electrode-shift tolerant myoelectric movement-pattern classification using extreme learning for adaptive sparse representations, in: 2017 IEEE Biomedical Circuits and Systems Conference, BioCAS, IEEE, 2017, pp. 1–4, <http://dx.doi.org/10.1109/BIOCAS.2017.8325201>.
- [111] T. Farrell, R.F. Weir, Analysis window induced controller delay for multifunctional prostheses, in: *Myoelectric Symposium*, 2008.
- [112] R. Menon, G. Di Caterina, H. Lakany, L. Petropoulakis, B.A. Conway, J.J. Soraghan, Study on interaction between temporal and spatial information in classification of EMG signals for myoelectric prostheses, *IEEE Trans. Neural Syst. Rehabil. Eng.* 25 (10) (2017) 1832–1842, <http://dx.doi.org/10.1109/TNSRE.2017.2687761>.
- [113] X. Zhai, B. Jelfs, R.H. Chan, C. Tin, Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network, *Front. Neurosci.* 11 (2017) 379, <http://dx.doi.org/10.3389/fnins.2017.00379>.
- [114] L.I. Barona López, F.M. Ferri, J. Zea, Á.L. Valdivieso Caraguay, M.E. Benalcázar, CNN-LSTM and post-processing for EMG-based hand gesture recognition, *Intell. Syst. Appl.* 22 (2024) 200352, <http://dx.doi.org/10.1016/j.iswa.2024.200352>, 2024/06/01.

- [115] J.L. Betthauser, J.T. Krall, R.R. Kaliki, M.S. Fifer, N.V. Thakor, Stable electromyographic sequence prediction during movement transitions using temporal convolutional networks, in: 2019 9th International IEEE/EMBS Conference on Neural Engineering, NER, IEEE, 2019, pp. 1046–1049, <http://dx.doi.org/10.1109/NER.2019.8717169>.
- [116] J.L. Betthauser, et al., Stable responsive EMG sequence prediction and adaptive reinforcement with temporal convolutional networks, *IEEE Trans. Biomed. Eng.* 67 (6) (2019) 1707–1717, <http://dx.doi.org/10.1109/TBME.2019.2943309>.
- [117] S.A. Stuttaford, M. Dyson, K. Nazarpour, S.S. Dupan, Reducing motor variability enhances myoelectric control robustness across untrained limb positions, in: *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023, <http://dx.doi.org/10.1109/TNSRE.2023.3343621>.
- [118] J. He, D. Zhang, X. Sheng, S. Li, X. Zhu, Invariant surface EMG feature against varying contraction level for myoelectric control based on muscle coordination, *IEEE J. Biomed. Heal. Informat.* 19 (3) (2014) 874–882, <http://dx.doi.org/10.1109/JBHI.2014.2330356>.
- [119] F. Onay, A. Mert, Phasor represented EMG feature extraction against varying contraction level of prosthetic control, *Biomed. Signal Process. Control.* 59 (2020) 101881, <http://dx.doi.org/10.1016/j.bspc.2020.101881>.
- [120] F. Zhuang, et al., A comprehensive survey on transfer learning, *Proc. IEEE* 109 (1) (2020) 43–76, <http://dx.doi.org/10.1109/JPROC.2020.3004555>.
- [121] R.N. Khushaba, Correlation analysis of electromyogram signals for multi-user myoelectric interfaces, *IEEE Trans. Neural Syst. Rehabil. Eng.* 22 (4) (2014) 745–755, <http://dx.doi.org/10.1109/TNSRE.2014.2304470>.
- [122] F. Kulwa, et al., A robust feature adaptation approach against variation of muscle contraction forces for myoelectric pattern recognition-based gesture characterization, *Biomed. Signal Process. Control.* 95 (2024) 106446, <http://dx.doi.org/10.1016/j.bspc.2024.106446>.
- [123] T. Hoshino, S. Kanoga, M. Tsubaki, A. Aoyama, Comparing subject-to-subject transfer learning methods in surface electromyogram-based motion recognition with shallow and deep classifiers, *Neurocomputing* 489 (2022) 599–612, <http://dx.doi.org/10.1016/j.neucom.2021.12.081>.
- [124] J. Liu, X. Sheng, D. Zhang, J. He, X. Zhu, Reduced daily recalibration of myoelectric prosthesis classifiers based on domain adaptation, *IEEE J. Biomed. Heal. Informat.* 20 (1) (2014) 166–176, <http://dx.doi.org/10.1109/JBHI.2014.2380454>.
- [125] K.-T. Kim, C. Guan, S.-W. Lee, A subject-transfer framework based on single-trial EMG analysis using convolutional neural networks, *IEEE Trans. Neural Syst. Rehabil. Eng.* 28 (1) (2019) 94–103, <http://dx.doi.org/10.1109/TNSRE.2019.2946625>.
- [126] A.A. Rusu, et al., Progressive neural networks, 2016, <http://dx.doi.org/10.48550/arXiv.1606.04671>, arXiv preprint arXiv:1606.04671.
- [127] X. Chen, Y. Li, R. Hu, X. Zhang, X. Chen, Hand gesture recognition based on surface electromyography using convolutional neural network with transfer learning method, *IEEE J. Biomed. Heal. Informat.* 25 (4) (2020) 1292–1304, <http://dx.doi.org/10.1109/JBHI.2020.3009383>.
- [128] Y. Zou, L. Cheng, A transfer learning model for gesture recognition based on the deep features extracted by CNN, *IEEE Trans. Artif. Intell.* 2 (5) (2021) 447–458, <http://dx.doi.org/10.1109/TAI.2021.3098253>.
- [129] P. Shi, X. Zhang, W. Li, H. Yu, Improving the robustness and adaptability of sEMG-based pattern recognition using deep domain adaptation, *IEEE J. Biomed. Heal. Informat.* 26 (11) (2022) 5450–5460, <http://dx.doi.org/10.1109/JBHI.2022.3197831>.
- [130] J. Yang, D. Cha, D.-G. Lee, S. Ahn, Stcnet: Spatio-temporal cross network with subject-aware contrastive learning for hand gesture recognition in surface EMG, *Comput. Biol. Med.* 185 (2025) 109525, <http://dx.doi.org/10.1016/j.compbiomed.2024.109525>.
- [131] J. He, N. Jiang, Biometric from surface electromyogram (sEMG): Feasibility of user verification and identification based on gesture recognition, *Front. Bioeng. Biotechnol.* 8 (2020) 58.
- [132] M.H. Zafar, S.K.R. Moosavi, F. Sanfilippo, Federated learning-enhanced edge deep learning model for EMG-based gesture recognition in real-time human-robot interaction, *IEEE Sensors J.* (2025) <http://dx.doi.org/10.1109/JSEN.2025.3529841>.
- [133] Z. Zhang, Y. Ming, Y. Wang, A federated transfer learning approach for surface electromyographic hand gesture recognition with emphasis on privacy preservation, *Eng. Appl. Artif. Intell.* 136 (2024) 108952, <http://dx.doi.org/10.1016/j.engappai.2024.108952>.
- [134] H. Lee, M. Jiang, Q. Zhao, FedAssist: Federated learning in AI-powered prosthetics for sustainable and collaborative learning, in: 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, (EMBC) 15–19 2024, 2024, pp. 1–5, <http://dx.doi.org/10.1109/EMBC53108.2024.10781961>.
- [135] D.R. Hardoon, S. Szedmak, J. Shawe-Taylor, Canonical correlation analysis: An overview with application to learning methods, *Neural Comput.* 16 (12) (2004) 2639–2664, <http://dx.doi.org/10.1162/0899766042321814>.
- [136] B. Fernando, A. Habrard, M. Sebban, T. Tuytelaars, Unsupervised visual domain adaptation using subspace alignment, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2960–2967.
- [137] B. Gong, Y. Shi, F. Sha, K. Grauman, Geodesic flow kernel for unsupervised domain adaptation, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 2066–2073, <http://dx.doi.org/10.1109/CVPR.2012.6247911>.
- [138] B. Sun, J. Feng, K. Saenko, Correlation alignment for unsupervised domain adaptation, in: *Domain Adaptation in Computer Vision Applications*, Springer, 2017, pp. 153–171.
- [139] S.J. Pan, I.W. Tsang, J.T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, *IEEE Trans. Neural Netw.* 22 (2) (2010) 199–210, <http://dx.doi.org/10.1109/TNN.2010.2091281>.
- [140] H. Shi, X. Jiang, C. Dai, W. Chen, EMG-based multi-user hand gesture classification via unsupervised transfer learning using unknown calibration gestures, *IEEE Trans. Neural Syst. Rehabil. Eng.* 32 (2024) 1119–1131, <http://dx.doi.org/10.1109/TNSRE.2024.3372002>.
- [141] X. Zhang, X. Zhang, L. Wu, C. Li, X. Chen, X. Chen, Domain adaptation with self-guided adaptive sampling strategy: feature alignment for cross-user myoelectric pattern recognition, *IEEE Trans. Neural Syst. Rehabil. Eng.* 30 (2022) 1374–1383, <http://dx.doi.org/10.1109/TNSRE.2022.3173946>.
- [142] X. Chen, D. Zhang, X. Zhu, Application of a self-enhancing classification method to electromyography pattern recognition for multifunctional prosthesis control, *J. NeuroEng. Rehabil.* 10 (1) (2013) 44, <http://dx.doi.org/10.1186/1743-0003-10-44>, 2013/05/01.
- [143] S. Amsüss, P.M. Goebel, N. Jiang, B. Graimann, L. Paredes, D. Farina, Self-correcting pattern recognition system of surface EMG signals for upper limb prosthesis control, *IEEE Trans. Biomed. Eng.* 61 (4) (2013) 1167–1176, <http://dx.doi.org/10.1109/TBME.2013.2296274>.
- [144] Z. Wang, et al., Optimization of inter-subject sEMG-based hand gesture recognition tasks using unsupervised domain adaptation techniques, *Biomed. Signal Process. Control.* 92 (2024) 106086, <http://dx.doi.org/10.1016/j.bspc.2024.106086>.
- [145] H. Le, M. in Het Panhuis, G.M. Spinks, G. Alici, The effect of dataset size on EMG gesture recognition under diverse limb positions, in: 2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechanics (BioRob), IEEE, 2024, pp. 303–308, <http://dx.doi.org/10.1109/BioRob60516.2024.10719858>.
- [146] H. Le, M. in Het Panhuis, G.M. Spinks, G. Alici, Quantifying covariate shift and improving electromyography driven gesture recognition with calibration and sample selection, in: 2024 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, IEEE, 2024, pp. 1434–1440, <http://dx.doi.org/10.1109/AIM5361.2024.10636995>.
- [147] U. Côté-Allard, E. Campbell, A. Phinyomark, F. Lavolette, B. Gosselin, E. Scheme, Interpreting deep learning features for myoelectric control: A comparison with handcrafted features, *Front. Bioeng. Biotechnol.* 8 (2020) 158, <http://dx.doi.org/10.3389/fbioe.2020.00158>.
- [148] T.R. Farrell, Determining delay created by multifunctional prosthesis controllers, *J. Rehabil. Res. Dev.* 48 (6) (2011) xxi.
- [149] G. Akkad, A. Mansour, E. Inaty, Embedded deep learning accelerators: A survey on recent advances, *IEEE Trans. Artif. Intell.* 5 (5) (2023) 1954–1972, <http://dx.doi.org/10.1109/TAI.2023.3311776>.
- [150] A.T. Nguyen, et al., A portable, self-contained neuroprosthetic hand with deep learning-based finger control, *J. Neural Eng.* 18 (5) (2021) 056051, <http://dx.doi.org/10.1088/1741-2552/ac2a8d>.
- [151] T. Hastie, R. Tibshirani, J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2009.
- [152] B. Frénay, M. Verleysen, Classification in the presence of label noise: a survey, *IEEE Trans. Neural Networks Learn. Syst.* 25 (5) (2013) 845–869, <http://dx.doi.org/10.1109/TNNLS.2013.2292894>.
- [153] K. Nar, O. Ocal, S.S. Sastry, K. Ramchandran, Cross-entropy loss leads to poor margins, 2018.
- [154] R.J. Beaulieu, et al., Multi-position training improves robustness of pattern recognition and reduces limb-position effect in prosthetic control, *J. Prosthetics Orthot.*: JPO 29 (2) (2017) 54, <http://dx.doi.org/10.1097/JPO.000000000000121>.
- [155] N. Jiang, I. Vujaklija, H. Rehbaum, B. Graimann, D. Farina, Is accurate mapping of EMG signals on kinematics needed for precise online myoelectric control? *IEEE Trans. Neural Syst. Rehabil. Eng.* 22 (3) (2013) 549–558, <http://dx.doi.org/10.1109/TNSRE.2013.2287383>.
- [156] I.S. MacKenzie, C. Ware, Lag as a determinant of human performance in interactive systems, in: *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, ACM, 1993, pp. 488–493, <http://dx.doi.org/10.1145/169059.169431>.
- [157] B. Berberian, P. Le Blaye, C. Schulte, N. Kinani, P.R. Sim, Data transmission latency and sense of control, in: *International Conference on Engineering Psychology and Cognitive Ergonomics*, Springer, 2013, pp. 3–12, [http://dx.doi.org/10.1007/978-3-642-39360-0\\_1](http://dx.doi.org/10.1007/978-3-642-39360-0_1).
- [158] P.R. Cavanagh, P.V. Komi, Electromechanical delay in human skeletal muscle under concentric and eccentric contractions, *Eur. J. Appl. Physiol.* 42 (1979) 159–163, <http://dx.doi.org/10.1007/BF00431022>.

- [159] S. Mehta, M. Rastegari, Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer, 2021, <http://dx.doi.org/10.48550/arXiv.2110.02178>, arXiv preprint [arXiv:2110.02178](https://arxiv.org/abs/2110.02178).
- [160] A.G. Howard, Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017, <http://dx.doi.org/10.48550/arXiv.1704.04861>, arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861).
- [161] A. Moslemi, A. Briskina, Z. Dang, J. Li, A survey on knowledge distillation: Recent advancements, Mach. Learn. Appl. (2024) 100605, <http://dx.doi.org/10.1016/j.mlwa.2024.100605>.
- [162] L. Wei, Z. Ma, C. Yang, Q. Yao, Advances in the neural network quantization: A comprehensive review, Appl. Sci. 14 (17) (2024) 7445, <http://dx.doi.org/10.3390/app14177445>.
- [163] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.
- [164] H. Cheng, M. Zhang, J.Q. Shi, A survey on deep neural network pruning: Taxonomy, Comp. Anal. Recomm. IEEE Trans. Pattern Anal. Mach. Intell. (2024) <http://dx.doi.org/10.1109/TPAMI.2024.3447085>.