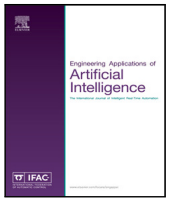
Engineering Applications of Artificial Intelligence 137 (2024) 109119

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Research Paper

Surface electromyography based explainable Artificial Intelligence fusion framework for feature selection of hand gesture recognition

Naveen Gehlot [a], Ashutosh Jena [a], Ankit Vijayvargiya [b,c,*], Rajesh Kumar [a]

[a] Department of Electrical Engineering, Malaviya National Institute of Technology, Jaipur, 302017, India
[b] Insight Science Foundation Ireland Research Centre for Data Analytics, School of Human and Health Performance, Dublin City University, Dublin, D09V209, Ireland
[c] Department of Electrical Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur, 302017, India

ARTICLE INFO

Keywords:
Hand gesture recognition
Surface electromyography
Feature selection
Fusion framework
Explainable Artificial Intelligence
Machine learning

ABSTRACT

Over the past decade, the utilization of machine learning (ML) models for recognizing hand gestures from surface electromyography (sEMG) signals has been in demand for the control of prosthetics. Such real-time control demands swift responses and efficient models. The model's efficiency heavily relies on selecting optimal handcrafted features. As the number of acquisition channels increases, the complexity of handcrafted features escalates the computational burden and necessitates the reduction of irrelevant or noisy features to enhance model efficiency. This study proposes an eXplainable Artificial Intelligence (XAI) fusion-based feature selection framework for sEMG-based hand gesture recognition (HGR). The proposed framework comprises two stages. Firstly, it combines three feature selection methods: filter, wrapper, and embedded. Secondly, it employs SHAP-based explainability for ML models to select relevant features. The first stage of feature selection retains ten relevant features, followed by a fine selection of 50% of features in the second stage. The performance of the proposed framework is compared with three baseline models. It shows that the proposed model performs better than other baseline models in terms of accuracy, precision, recall, f-score, and computational time. The proposed XAI fusion framework achieves an average classification accuracy of 80.75% with the Extra Tree classifier, with a computational time of 3.47 ms. Furthermore, to assess its robustness against baseline models, evaluation is conducted using a publicly available dataset, revealing superior performance compared to other baselines.

1. Introduction

Hand amputation is a rare yet devastating accident that profoundly affects the lifestyle of amputees, including employment, hobbies, and self-care (Fang et al., 2022). Hand amputations are more common in developing countries, frequently resulting from occupational injuries, traffic accidents, or combat-related injuries (Yang et al., 2005). Innovative solutions have been devised to assist individuals facing such challenges, with one notable approach being the development of prosthetic devices. These devices can be classified as either passive or active. A passive prosthesis is a cosmetic addition to the body that lacks functionality. The active device can assist the user in performing mechanical or bionic movements similar to those made by a natural hand. This device converts a signal into gestures by processing electromyography (EMG) signals to replicate the movements of a hand. The most commonly measured and collected EMG signals are needle and fine-wire EMG, or surface EMG (sEMG). Such devices are plentily available in the market in terms of price, accuracy, and number of channels.

Adopting assistive technology based on sEMG signals collected from untrustworthy devices can profoundly impact classification results. It may lead to the mechanical prosthesis's undesirable functioning. In Pradhan et al. (2016), Pradhan, et al. used INA128 to capture sEMG data for capturing hand gestures using the Texas ADS1294, a specific biomedical device. In Pancholi and Joshi (2019), Pancholi, et al. used the ADS1298 to acquire the sEMG signal, essentially equivalent to the ADS1294. In Vijayvargiya et al. (2022a), Ankit et al. conducted research using Myoware sensors to collect sEMG data for activity classification. Despite the market's abundance of signal conditioning tools and advanced EMG signal acquisition devices, it is worth noting that Biopac and BioNomadix are well-known for producing high-quality, research-grade EMG devices with a long history of accurate signal capturing (Ozdemir et al., 2022; Qiu et al., 2022; Kwon et al., 2023).

* Corresponding author at: Department of Electrical Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur, 302017, India.
E-mail address: ankitvijayvargiya29@gmail.com (A. Vijayvargiya).

https://doi.org/10.1016/j.engappai.2024.109119
Received 29 December 2023; Received in revised form 8 July 2024; Accepted 5 August 2024
Available online 12 August 2024

Table 1
Summary of attempted methodology for feature selection in hand gesture recognition.

Ref	Feature selection			Machine learning classifier					XAI
	Filter	Wrapper	Embedded	ET	k-NN	LR	NB	SVM	
He et al. (2023)	✗	✗	✓	✗	✓	✗	✗	✓	✗
Junior et al. (2020)	✗	✓	✗	✗	✗	✗	✗	✓	✗
Siddiqui and Chan (2020)	✓	✗	✗	✗	✗	✗	✗	✓	✗
Liu et al. (2020)	✗	✓	✗	✗	✓	✗	✗	✓	✗
Singha and Laskar (2017)	✓	✗	✗	✗	✓	✗	✓	✓	✗
Li et al. (2020)	✗	✗	✓	✗	✓	✗	✗	✓	✗
Proposed	✓	✓	✓	✓	✓	✓	✓	✓	✓

Post gathering real-time sEMG signal data, it undergoes classification to distinguish hand gestures. This can be achieved through machine learning (ML) or deep learning (DL) classifiers (Borlea et al., 2022; Jiang et al., 2022; Harrison et al., 2022). While DL excels in certain applications such as image classification, object detection, etc. Roy and Bhaduri (2023), Roy et al. (2022) and Arican and Aydin (2022), where large amounts of data and substantial computational resources are available, and interpretability is not a major concern. In the healthcare industry, interpretability is a necessity. Such models also require high-end processors and storage units during hardware implementation, which makes the device costlier and bulkier, which may be difficult for daily usage. For the above reasons, ML is preferred for this study. Researchers have explored various classification methods to assess their effectiveness, including Bayesian techniques, Support Vector Machines (SVM), Extra Trees (ET), k-Nearest Neighbors (k-NN), and Logistic Regression(LR) (Hassan et al., 2023; Hong et al., 2023). However, to achieve higher accuracy, ML models require an optimal selection of features. Some popular feature selection methods are based on statistics, iterative optimization, and interpretability. Therefore, ML models are trained with feature selection techniques to perform efficient hand gesture recognition in this study.

The features, like channel, power, amplitude, and frequency, can be derived directly from the sEMG signal, or they can be obtained after a brief preprocessing that computes parameters that may represent the nature of the signal, such as mean, RMS, and standard deviation. An sEMG signal, being lengthy and inherently random, poses challenges for interpretation. Training an ML classifier directly on such data without preprocessing can lead to slow training and increased misclassification. Therefore, numerical features are computed over segmented windows of the signal during preprocessing. This process reduces the number of training samples while meaningfully representing signal segments. This process is known as feature extraction. In Vijayvargiya et al. (2021), Phinyomark et al. (2012), Xue et al. (2023) and Narayan (2021), authors have extracted Willison Amplitude (WAMP), Variance (VAR), Mean Absolute Value (MAV), Waveform Length (WL), Zero Crossings (ZC), Root Mean Square (RMS), Average Amplitude Change (AAC), Difference Absolute Standard Deviation Value (DASDV), Integrated Electromyogram (iEMG), Myopulse (MYOP), Log Detector (Log), from the time domain signal, and Median Frequency (MDF), Mean Frequency (MNF), Frequency Ratio (FR) from the frequency domain of the sEMG signal.

The extraction of features is not only the essential part of enhancing the efficiency of an ML model, but the selection of appropriate features is crucial for the ML model to generate more accurate outputs and reduce the computational burden on the model (Jena et al., 2023; Jalilvand and Salim, 2017). All the features hold different meanings, so among the features that are extracted, many unrelated or redundant features may be present. Using all features as input to a classifier may lead to overfitting and increased computing costs. In the literature, two solutions have been addressed: feature selection and dimensionality reduction. The latter directly transforms features without relying on class information, which is mainly used in unsupervised learning. Examples of such techniques include linear (e.g., Principal Component analysis (PCA), Linear Discriminant Analysis (LDA)) and nonlinear methods (e.g., Kernel PCA). However, feature selection assesses the closeness between features and class labels to determine redundancy. This technique is predominantly used in supervised learning (Zhang et al., 2024).

Table 1 presents related work on feature selection for hand gesture classification. In He et al. (2023), the authors explore four different hand gestures using SVM and k-NN classifiers and enhance performance through feature selection using the embedded method, where the random forest is utilized to identify the most optimal features. In Junior et al. (2020), researchers improve the identification of six hand gestures by selecting optimal features using the wrapper method and employing SVM and LDA classifiers for gesture classification. In Siddiqui and Chan (2020), the identification of 36 gestures is achieved with improved accuracy by selecting influential features using the filter technique of Mutual Information, followed by classification using SVM and LDA classifiers. Similarly, Liu et al. (2020) introduces Sequential Forward Selection to identify influential features and enhance classification accuracy using SVM, k-NN, LDA, and Deep Neural Network classifiers for hand gesture identification. In Singha and Laskar (2017), the authors accurately identify 10 gestures by employing one of the filter techniques known as Anova F-value for optimal feature selection, followed by classification using SVM, k-NN, Naive Bayes, and Artificial Neural Network classifiers. Furthermore, authors of Li et al. (2020) achieve a more accurate classification of 13 gestures by selecting influential features using the embedded method of random forest and PCA, followed by classification with SVM, decision trees, k-NN, LDA, and Neural Network classifiers. Overall, the literature demonstrates the use of various feature selection techniques, including wrapper, filter, and embedded methods, each offering unique advantages and trade-offs. The wrapper method efficiently interacts with the classifier for feature selection. Conversely, the filter method does not interact with the classifier for feature selection, but it has a faster running time. Similarly, the embedded method is also faster compared to the wrapper. However, in the embedded method, identifying a small set of features can be problematic, while in the filter method, scaling high-dimensional data features is easier (Van Zyl et al., 2024; Biswas et al., 2016). After individually collecting influential features from these methods, they are consolidated into a comprehensive framework for feature selection, aimed at enhancing efficiency. This consolidated framework, known as a fusion framework, integrates the strengths of each method to produce superior results.

However, feature selection lacks interpretability in explaining features' importance using ML models in hand gesture classification. Lack of clarity in the interpretation of an ML model makes it challenging to trust. Due to this, it is regarded as a complex learning structure and often called a 'Black-Box'. Explainable Artificial Intelligence (XAI) is a rapidly expanding field, quickly emerging as one of the essential components of Artificial Intelligence. Interpretation from XAI models helps reveal feature importance and model design. Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are two commonly employed surrogate models that facilitate understanding complex models. LIME offers local explanations, elucidating predictions for specific instances (Gehlot et al., 2024; Vijayvargiya et al., 2023b). Conversely, SHAP generates explainers, furnishing both local and global explanations. Local explanations

dissect individual predictions using feature-based plots, while global explanations elucidate model predictions across features. SHAP's ability to make models interpretable and resolve the issue of feature impact on the model with explanations is valuable. Interpreting explanations in eXplainable Artificial Intelligence (XAI) requires careful consideration of potential dependencies between features. Hence, this study proposes integrating a fusion framework with an XAI technique to select features from sEMG signals for hand gesture recognition, termed the XAI fusion framework. This hybrid approach effectively reduces the number of features for ML models, thus decreasing computational time and improving model performance.

The motivation for this study arises from the need for real-time neural control of robotic arms, upper limb exoskeletons, or prostheses. It is observed that the diverse movements of the hand required for controlling these devices can be accurately identified using surface electromyography (sEMG) signals obtained from forearm muscles. So, the outcomes of our investigation will significantly contribute to future advancements in real-time neural-controlled robotic arms, upper limb prostheses, or exoskeletons capable of performing various hand gestures. Achieving real-time control primarily involves two steps: data acquisition and classification of different gestures. This study focuses on the acquisition of real-time data and subsequent signal classification. The main contributions of this study are:

1. Proposed an XAI fusion-based feature selection framework for sEMG-based hand gesture recognition (HGR), addressing the critical need for efficient prosthetic control.
2. An interpretability of the model is performed using SHAP, enhancing the interpretability of the proposed feature selection model.
3. A conventional feature selection method filter, wrapper, and embedded to select most 10 influencing features fuse together to develop a fusion framework.
4. A comparative analysis has been done in four different scenarios on two dataset cases: first on real-time acquired 10 subjects data and second on publicly existing data of 8 subjects with the performance metrics accuracy, precision, recall, F-score, and time.

2. System overview

The flow diagram of the proposed framework for hand gesture recognition is depicted in Fig. 1, consisting of 7 steps. In the initial step, sEMG signals were collected, with 10 subjects volunteering to perform six different gestures. The second step involves preprocessing the acquired data. Next, relevant features are extracted in the third step using a windowing technique (Gehlot et al., 2023). In the fourth step, a fusion framework is employed to rank and select influential features, considering only ten features while discarding the rest. Following this, in the fifth step, ML classifiers are trained. Specifically, five classifiers, Extra Tree (ET), Naive Bayes (NB), Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), and Logistic Regression (LR), are selected, and their outputs are analyzed to determine the most accurate classifier. Subsequently, in the sixth step, SHAP-based XAI is utilized to explain the high accuracy of this classifier. SHAP's explainability aids in organizing the features based on their impact on predicting a certain class for that classifier. Finally, 50% of these features are selected to retrain the most accurate model obtained at the end of the fifth step. The final step involves gesture recognition. The proposed model (B4) integrates all the aforementioned steps. For comparison, three baseline models (B1, B2, B3) are selected, each following the steps indicated by the colored arrows. These four scenarios, including the proposed model and the three baseline models, are further discussed in Section 3. The trained classifiers within these scenarios are evaluated based on performance metrics. The following sections provide detailed elaboration of these steps.

Table 2
Demographic information about subjects: Gender (F: Female M: Male), Age, Height, Weight, Forearm circumference (The circumference of the forearm next to the rubber ring), Forearm length (Forearm length: The measurement from the inner side of the elbow to the wrist line).

Participant	Gender	Age	Height (cm)	Weight (kg)	Forearm circumference (cm)	Forearm length (cm)
M1	M	28	170	79.0	26.0	27.5
M2	M	24	178	81.0	27.0	27.5
M3	M	19	171	90.0	29.8	27.4
M4	M	21	165	68.0	24.2	26.8
M5	M	20	175	55.0	24.0	27.5
M6	M	20	175	61.3	23.7	28.6
M7	M	20	170	54.0	22.5	27.0
F1	F	22	165	67.0	21.0	27.0
F2	F	26	163	62.0	22.0	28.0
F3	F	22	170	51.3	18.0	26.0

2.1. Datasets

To assess the proposed model's robustness against baseline models, this study encompassed two datasets. The first dataset was acquired from the RAMAN Lab at MNIT. The second dataset was considered from a publicly available dataset repository.

1. **Acquired Dataset**: sEMG signal data is acquired using the BIOPAC MP 150 acquisition device. For this study, data acquisition is conducted using two sEMG channels strategically placed on the surface of the forearm muscle. The acquired data from the muscles is then visualized on a monitor using AcqKnowledge software version 4.4. In the acquisition system, the amplifier gain is set at 1000, and the sampling frequency for this study is set at 2000 Hz. Ten healthy participants volunteered in this study during the data collection. Demographic data for these participants are presented in the Table 2. The samples of two-channel signals were acquired by placing electrodes on the participant's forearm muscles. Specifically, one set of electrodes was positioned on the posterior compartment muscles (extensor digitorum), and the other set was placed on the anterior compartment muscles (flexor pollicis longus). The data samples were acquired for six hand gesture activities (TE: Thumb Extension, ME: Middle Extension, FME: Fore Middle Extension, FMTE: Fore Middle Thumb Extension, FMR: Fore Middle Ring Extension, HC: Hand Close) are depicted in Fig. 2. The data is split 70% training and 30% testing in sequentially, as shown in Fig. 2.
2. **Publicly Dataset**: This dataset is sourced from the data repository of R. N. Khushaba et al. Khushaba and Kodagoda (2012). The dataset was collected using eight channels of DE2.x series EMG sensors placed around the forearm. The Delsys Bangoli-8 amplifier processed these acquired signals with a gain set to 1000. A 12-bit ADC converter was employed to sample the data at a frequency of 4000 Hz. The acquisition software used for the data was Delsys EMGworks. Subsequently, the EMG signals underwent band-pass filtering within the range of 20 Hz to 450 Hz, along with a 50 Hz notch filter to eliminate power line interference. The data acquisition involved the consent of 8 participants, comprising six males and two females lying in the range of age 20 to 35 years, performing 15 different hand gestures.

2.2. Preprocessing and feature extraction

During the acquisition phase, sEMG signals are susceptible to unwanted external influences, commonly called noise. This noise can manifest in various forms, including power line interference, electromagnetic noise, and interference from electronic devices such as phones

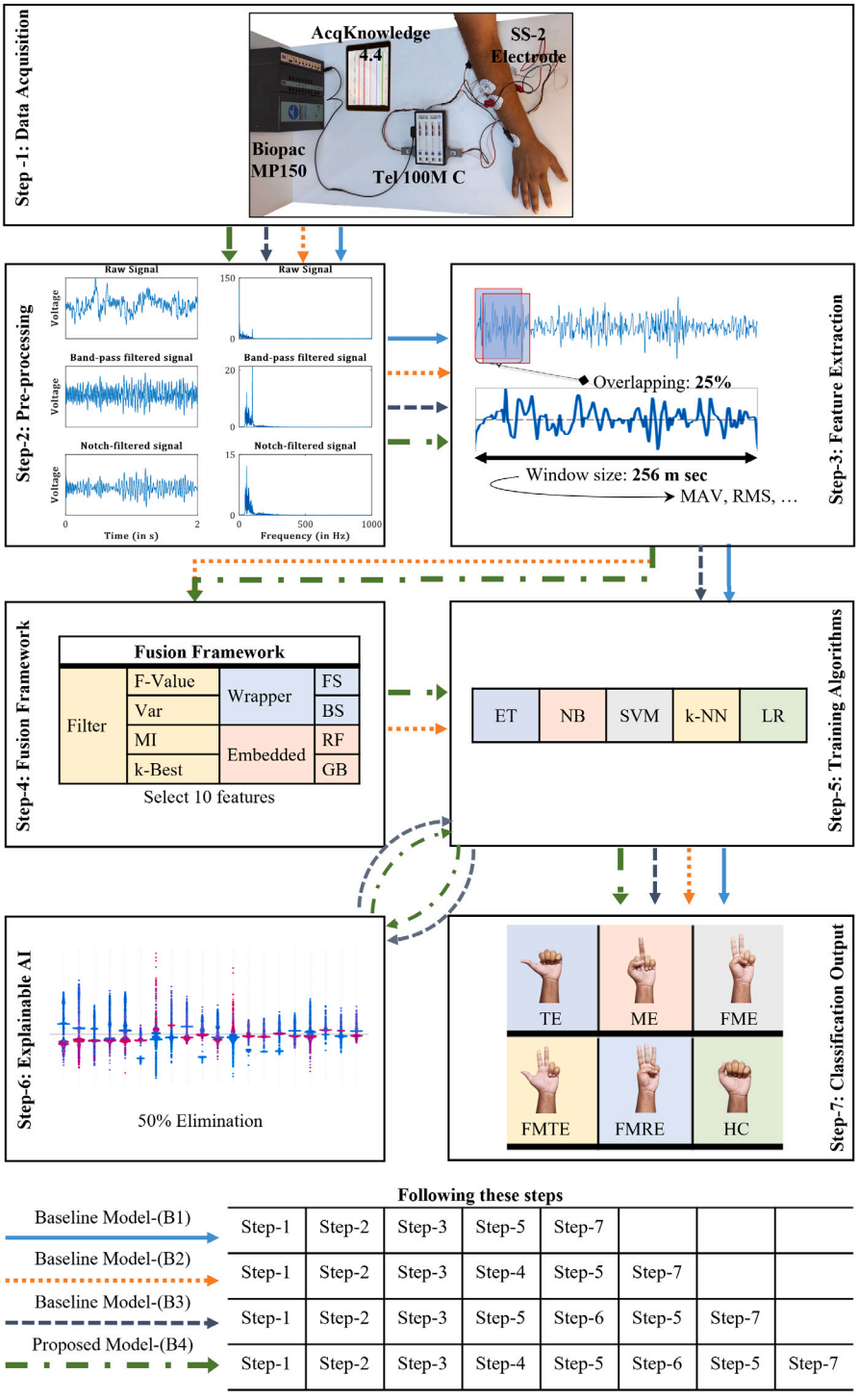


Fig. 1. Flow diagram of the proposed framework for optimal feature selection and classification of sEMG signals for hand gesture recognition.

and broadband equipment (Vijayvargiya et al., 2022b). To ensure data quality, filtering techniques are employed as a crucial step to reduce noise. Specifically, a notch filter is applied to eliminate the power line frequency, typically at 50 Hz. Additionally, a bandpass filter with a cutoff frequency ranging from 10 Hz to 500 Hz is employed. These filtering processes are applied using mathematical tools to reduce the impact of noise and enhance the original sEMG signals' suitability for further analysis.

After filtering, the signal is sequentially split into 70% training and 30% testing, as shown in Fig. 2. Further, both the splits are passed through feature extraction. Feature extraction involves segmenting the sEMG signal into 256 ms long windows with a 25% overlap between consecutive windows (Vijayvargiya et al., 2023a). This windowing strategy serves a dual purpose: it helps eliminate data redundancy and addresses the challenge of preserving instantaneous signal information. Within these segmented windows, six features are extracted from the frequency domain and eleven features from the time domain. These features are listed in Table 3 and play a pivotal role in further analysis of HGR.

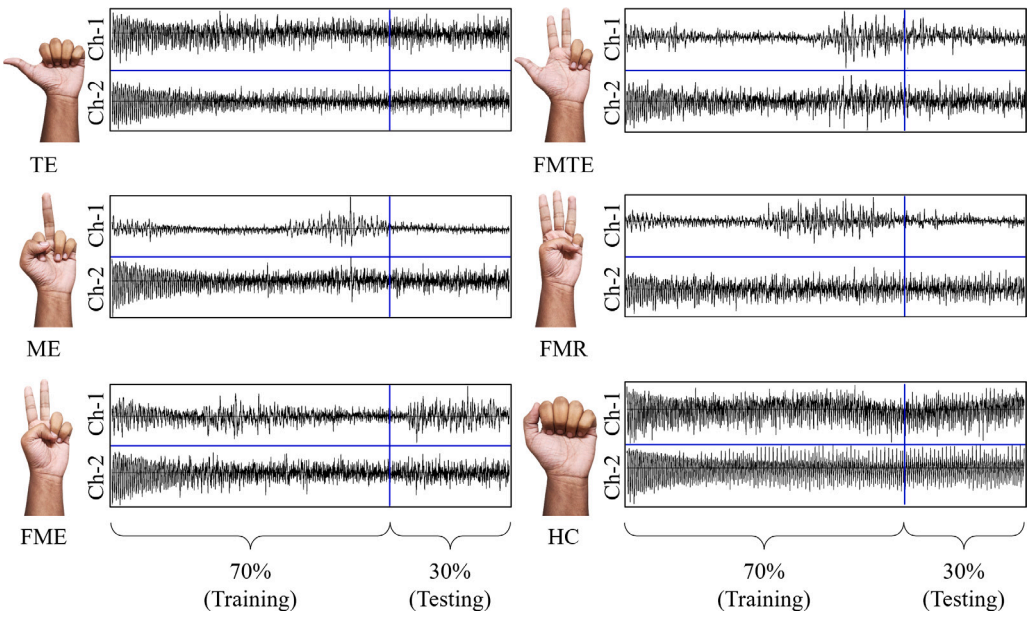


Fig. 2. Data Samples of sEMG signal classification.

Table 3
Features of the time–frequency domain.

Time domain features					
SN	Feature name	Mathematical expression	SN	Feature name	Mathematical expression
1	Mean Absolute Value (MAV)	$\frac{1}{N}\sum_{p=1}^{k}\left\|x_p\right\|$	2	Root Mean Square (RMS)	$\sqrt{\frac{1}{N}\sum_{p=1}^{k}\left\|x_p\right\|^2}$
3	Variance (VAR)	$\frac{1}{N-1}\sum_{i=1}^{p}x_p^2$	4	Average Amplitude Change (AAC)	$\frac{1}{N}\sum_{p=1}^{N-1}\left\|x_{p+1}-x_p\right\|$
5	Difference Absolute Standard Deviation Value (DASDV)	$\sqrt{\frac{1}{N-1}\sum_{p=1}^{N-1}(x_{p+1}-x_p)^2}$	6	Zero Crossing (ZC)	$\sum_{p=1}^{N-1} f(x_p)$ where $f(x_p)=\begin{cases}1 & \text{if } (x_p>0 \text{ and } x_{p+1}<0)\\ & \text{or } (x_p<0 \text{ and } x_{p+1}>0)\\ 0, & \text{otherwise}\end{cases}$
7	Waveform Length (WL)	$\frac{1}{N-1}\sum_{p=1}^{N}\left\|x_p\right\|^2$	8	Willson Amplitude (WAMP)	$\sum_{p=1}^{N-1} f(\left\|x_{p+1}-x_p\right\|)$ where $f(x_p)=\begin{cases}1 & \text{if } x\geq Threshold\\ 0, & \text{otherwise}\end{cases}$
9	Integrated Electromyogram (iEMG)	$\sum_{p=1}^{N}\left\|x_p\right\|$	10	Myopulse (MYOP)	$\frac{1}{N}\sum_{p=1}^{N} f(x_p)$ where $f(x_p)=\begin{cases}1 & \text{if } x\geq Threshold\\ 0, & \text{otherwise}\end{cases}$
11	Log Detector (Log)	$e^{\frac{1}{N}\sum_{p=1}^{N} log(\left\|x_p\right\|)}$			
Frequency domain features					
12	Total Power (TP)	$\sum_{k=1}^{M} P_k$	13	Peak Frequency (PKF)	$\frac{1}{2}\sum_{i=1}^{M} P_k$
14	Frequency Ratio (FR)	$\frac{\sum_{k-LLC}^{ULC} P_k}{\sum_{LHC}^{UHC} P_k}$	15	Mean Frequency (MNF)	$\frac{\sum_{k=1}^{M} f_k P_k}{\sum_{k=1}^{M} P_k}$
16	Median Frequency (MDF)	$\frac{1}{2}\sum_{k=1}^{M} P_k$	17	Mean Power (MNP)	$\sum_{k=1}^{M}\frac{P_k}{M}$

Here x_p is the p^{th} input sample of sEMG signal, N is the total no. of samples, P_k is the power at k^{th} frequency and M is the length of power spectrum density (Vijayvargiya et al., 2021; Phinyomark et al., 2012; Côté-Allard et al., 2019).

2.3. Feature selection

Feature selection plays a pivotal role in ML by identifying the most relevant features to enhance model performance while minimizing computational complexity. In this study, 17 features are extracted from each channel of the acquired signals, including 11 from the time domain and 6 from the frequency domain. Enhanced signal acquisition in the channel numbers leads to a higher number of input features for ML, thereby increasing computational complexity and potentially affecting classifier performance. Consequently, reducing the feature input alleviates the computational burden while enhancing model effectiveness. This study explores three selection techniques:

- Fusion based-Feature selection
- XAI based-Feature selection
- sEMG-based XAI Fusion Feature Selection (sXFS)

2.3.1. Fusion based-feature selection

In this study, feature selection is conducted using a fusion approach that integrates four filter methods, two wrapper methods, and two embedded methods. From each technique, the 10 most influential features are fused together to be utilized as input for the ML classifier (Bi et al., 2024; Sánchez-Maroño et al., 2007; Kaushik et al., 2023). The feature selection techniques employed in this study include:

1. **Filter Method:** In this, four techniques are used as described below:

   (i) Anova F-value: By using the analysis of variance (ANOVA), the F-value score is calculated as follows:

   $$F(\zeta) = \frac{V_b^2(\zeta)}{V_w^2(\zeta)} \tag{1}$$

   where, ζ denotes the set of instances of a particular feature, $V_b^2(\zeta)$ denotes the sample variance in between groups and $V_w^2(\zeta)$ sample of variance within groups (Qian et al., 2023; Lin and Ding, 2011). Based on this score, ranking is done in decreasing order.

   (ii) Variance Threshold: In statistics, variance denotes the spread of a set of data. This method takes into consideration the value of variance for each feature corresponding to different classes of data. A predefined threshold value is set, and when the spread of different classes within a feature is less than this threshold, that feature is eliminated (Fida et al., 2021).

   (iii) Mutual Information: It is the amount of uncertainty in A_x due to the knowledge of A_y represented as $MI(A_x, A_y)$. It is mathematically defined as follows:

   $$MI(A_x, A_y) = \sum_{A_x, A_y} p(A_x, A_y) log \frac{p(A_x, A_y)}{p(A_x)p(A_y)} \tag{2}$$

   where $p(A_x, A_y)$ is the joint probability distribution function (PDF) of A_x and A_y and $p(A_x)$ and $p(A_y)$ are the marginal PDF for A_x and A_y (Swingle, 2012).

   (iv) k-Best: This is an iterative method to select k-best (k = 10) features out of the total number of features fed to a wrapper classifier C (Effrosynidis and Arampatzis, 2021; Akman et al., 2023). $l_C(X', Y)$ is the loss function of that wrapper classifier, where $X' \subseteq X$ is having only k features. Iteratively, this method focuses on obtaining an optimal set of k features X^* such that it obeys the following Eq. (3).

   $$X^* = \arg \min_{X' \subseteq X} l_C(X', Y) \tag{3}$$

2. **Wrapper Method:** These set of methods are iterative and keep eliminating or adding features till the objective is satisfied. It includes the following two methods. In both methods, only the top 10 features are considered for further analysis (Bermejo et al., 2012).

   (i) Backward Search Elimination (BSE): BSE is a wrapper-based feature selection approach that eliminates features repeatedly to enhance prediction accuracy using leave-one-out cross-validation until no further increase in accuracy is found.

   (ii) Forward Sequential Selection (FSS): FSS is the inverse of BSE, progressively adding features that improve accuracy on the validation set until a drop is identified, making it computationally costly.

3. **Embedded Method:** These methods take the help of ML classifiers to identify the features with greater importance and keep ten of the best features for further analysis.

   (i) Random Forest: It has the benefit of assessing the significance score of each feature, allowing them to learn the influence of each feature on class prediction. The significance score is defined as follows:

   $$S_{score} = \frac{1}{n_t} \sum_{k \epsilon x_i} gain(x_i, \nu) \tag{4}$$

   Where n_t is the number of trees in RF, $k\ \epsilon\ x_i$ is the set of split nodes, and $gain(x_i, \nu)$ is the gain of the feature x_i at node ν (Deng and Runger, 2013).

   (ii) Gradient Boost: Unlike the other embedding methods for feature selection, this method relies on the feature importance score. This score is calculated based on the impact of a feature on making important decisions in a decision tree algorithm. The mathematical equations driving this technique are referred from (Xu et al., 2014) illustrated in the following set of equations:

   $$\min_{\theta} \sum_{\phi(x_i), y_i} l(\phi(x_i), y_i, \theta) + \lambda \|\theta\|_1 + \mu q_\epsilon(\theta) \tag{5}$$

   Eq. (5) represents the learning mechanism in a newly transformed space to finally converge, such that loss function $l(\phi(x_i), y_i, \theta)$ is reduced. θ denotes the learning parameter that changes every iteration as the output $\phi(x_i)$ converges toward target variable y_i.

   $$H(x) = \sum_{t=1}^{T} \theta_t h_t(x) \tag{6}$$

   Eq. (6) represents the hypothesis governing the regression tree.

2.3.2. XAI based feature selection

Using model interpretability techniques aids in selecting features. In this study, the XAI model SHAP is used to interpret the ability of selected features towards appropriate classification. SHAP is a renowned XAI system that offers local explainability for tabular and text datasets. It assigns values to features within a prediction, exposing their magnitude and direction (positive or negative) concerning the prediction. SHAP is rooted in the concept of Shapley values from cooperative game theory, interpreting any model by treating each feature as a player and the model's outcome as the payout (Shapley et al., 1953; Lundberg and Lee, 2017; Antwarg et al., 2023). SHAP's ability to make models interpretable and resolve the issue of feature impact on the model with explanations is valuable.

2.3.3. sEMG based XAI fusion feature selection (sXFS)

Combining Fusion feature selection with eXplainable Artificial Intelligence (XAI), we propose a sEMG-based XAI fusion feature selection technique termed sXFS. In this approach, features selected from the fusion feature selection undergo XAI model interpretation to ascertain their importance. Feature importance is evaluated using the additive feature attribution theory, which can be expressed as follows:

$$f(x') = \omega_0 + \sum_{i=1}^{N} \omega_i x_i' \quad \omega_i \epsilon R \tag{7}$$

where, $f(x')$ represents the interpretable model, N denotes the number of input features, $x' \epsilon \{0, 1\} N$, and ω_i signifies the weight function of the ith input feature, $\omega_i \epsilon R$, and ω_0 represents the mean weight function.

The impact of each feature on the model is determined as follows:

$$\omega_i(g, x) = \sum_{s' \subseteq x'} \frac{|s'|!(N - |s'| - 1)!}{N!} [f(s') - f(s'\pi)] \tag{8}$$

where $|s'|$ presents the nonzero entries in x', $|s'|$, and $[f(s') - f(s'\pi)]$ defines the difference between a model trained with the ith input feature and another model trained without it (Lundberg and Lee, 2017).

After determining the feature importance, the top 50% of the most influential features are selected, while the remaining features are discarded. The procedure for the proposed technique is elucidated through the pseudocode as shown in Algorithm 1.

Algorithm 1 Pseudocode for the sXFS approach

Require: sEMG signal acquired at $f_s = 2000\,Hz$ with labels.
Ensure: Filter, $f_c = 500Hz$ ▷ Preprocessing
Ensure: $pc_{train} = 70\%$ ▷ Train-Test Split
Ensure: Window Length = 256 ms & Overlap = 25%

Feature Extraction: $X_i = [x_{i,1,1}, x_{i,1,2}, x_{i,1,3}, ..., x_{i,2,17}]$ ▷ ∀ Windows
Labels: y_i
Rank Features: Using Fusion Framework.
Select Features : 10
Classifiers initialization (E_i).

$n \leftarrow$ No. of Evaluation Samples
$M \leftarrow$ No. of Classifiers
$N \leftarrow$ No. of Training Samples
while $M \neq 0$ **do** ▷ Training
  for $j = 1 : N$ **do**
    $E_M.fit(X_j, y_j)$
  end for
  $M \leftarrow M - 1$
end while
while $M \neq 0$ **do**
  k = 0;
  for $j = 1 : n$ **do** ▷ Evaluation
    $y_{pred_j} = E_M.predict(X_j, y_j)$
    $y_{pred} = append(y_{pred}, y_{pred_j})$
    $y_{act} = append(y_{act}, y_j)$
    if y_{act} is y_j **then**
      $k = k + 1$
    end if
  end for
  $Acc_{E_M} = k/n$
  $M \leftarrow M - 1$
end while

Select Classifier, $E_{best} = \underset{E_M}{argmax}(Acc_{E_M})$
Impact analysis of features, through Explainable-AI.
Eliminate 50% of features.

while $M \neq 0$ **do** ▷ Final Model Training
  for $j = 1 : N$ **do**
    $E_M.fit(X_j, y_j)$
  end for
  $M \leftarrow M - 1$
end while

2.4. Machine learning classifier

After selecting the input feature for the machine learning (ML) classifier, it is used to train or predict the output. In this study, five different types of ML classifiers are employed. These applied models are described as follows:

2.4.1. Extra tree (ET)

ET is an ensemble learning method closely related to Random Forests but with notable differences. It is a tree-based ML approach that employs multiple decision trees and added randomization to reduce variance and computational complexity (Geurts et al., 2006).

2.4.2. k-nearest neighbor (k-NN)

k-NN works based on the assumption that similar data points are nearby. It classifies new data by comparing them to the k-nearest neighbors in the training set. It assigns the majority class label among these neighbors as the predicted class (Dudani, 1976).

2.4.3. Logistic regression (LR)

Logistic Regression is a statistical ML approach primarily employed for classification tasks. It operates on labeled datasets to establish an optimal model that captures the relationship between dependent and independent variables (Indra et al., 2016).

Table 4
Parameters of machine learning classifiers.

Model	Parameter	Model	Parameter
ET	n_estimators = 100 criterion = gini min_samples_split = 2 min_samples_leaf = 1 min_weight_fraction_leaf = 0.0 max_features = sqrt min_impurity_decrease = 0.0 verbose = 0 ccp_alpha = 0.0	NB	alpha = 1.0 force_alpha = True fit_prior = True
k-NN	n_neighbors = 5 leaf_size = 30 p = 2 metric = minkowski	SVM	C = 1.0 kernel = rbf degree = 3 gamma = scale coef = 0.0 tol = 0.001 max_iter = −1 decision_function_shape = ovr cache_size = 200,
LR	penalty = l2 tol= 0.0001 C = 1.0 intercept_scaling = 1 solver = lbfgs max_iter = 100 verbose = 0		

2.4.4. Naive bayes (NB)

The NB classifier comprises fundamental probabilistic ML models widely employed for classification. It relies on Bayes' theorem with the assumptions of strong independence among features. The core concept behind Naive Bayes is that one feature's presence does not affect others' presence, making it fast and efficient (Leung et al., 2007).

2.4.5. Support vector machine (SVM)

SVM is a robust supervised learning technique suitable for classification and outlier detection. It identifies an optimal hyperplane within an N-dimensional space to separate data points into distinct classes in the feature space. The primary objective of the SVM algorithm is to maximize the margin between the nearest data points from different classes, as a more significant margin typically results in reduced generalization errors (Cristianini and Shawe-Taylor, 2000).

3. Results and discussion

This study is executed following four different scenarios as illustrated in Fig. 1 to evaluate the performance of the proposed method., To carry out this study, a 12^{th} generation Intel core i7 CPU with 16.0 GB of installed RAM is used. To run all the scripts and examine the findings, the Python 3.11.3 platform and Spyder IDE 5.4.3 are utilized. The machine learning models were implemented using the Scikit-learn library version 1.2.2. The performance metrics employed for hand gesture classification include accuracy, precision, recall, F-score, and time consumed by the classifier to predict (Vijayvargiya et al., 2020). These metrics are used to determine the model's capability to produce better results. Computational time analysis is done among the four scenarios to point out the scenarios' efficacy in practical applications. These metrics have been used to determine the best executable scenario out of the four presented in the study. To validate the robustness of the proposed scenario, it is also tested with a publicly available dataset, and a comparative analysis is provided for the same in this section.

This section begins by highlighting the model parameter settings under which all four scenarios are tested and evaluated, and it is illustrated in Table 4. Variations in these settings may offer biased results for the scenarios, therefore they are kept constant throughout the study. In the subsequent sections, the four scenarios are detailed, of which the first three scenarios depict the analyses of the baseline models, and the fourth one depicts the analysis of the proposed framework. Following the description of all scenarios, Table 5 highlights the performance metrics of different classifiers in all these scenarios. This

Table 5
Accuracy analysis of all scenarios with used classifiers.

Scenario	Classifier	M1	M2	M3	M4	M5	M6	M7	F1	F2	F3
B1	ET	85.75	85.75	70.96	70.43	75.53	76.61	76.34	86.82	72.84	88.70
	k-NN	72.31	78.49	67.20	53.22	64.78	72.03	71.50	85.75	55.10	86.29
	LR	60.75	68.01	63.44	57.52	62.09	62.63	68.81	79.83	56.72	83.06
	NB	59.94	54.56	56.18	49.73	40.05	58.06	56.45	84.13	61.02	82.25
	SVM	69.89	69.89	54.03	47.58	51.88	52.41	50.53	69.89	72.58	78.76
B2	ET	84.94	85.75	75.80	70.69	76.61	76.07	77.15	86.55	73.11	89.24
	k-NN	75.80	78.49	66.93	55.10	64.51	71.77	72.04	86.02	56.98	85.75
	LR	61.82	68.01	64.51	57.25	58.33	69.62	65.32	80.10	55.64	80.91
	NB	58.87	59.94	56.72	50.80	40.86	59.67	56.72	83.33	59.13	81.98
	SVM	67.20	69.62	53.22	49.73	51.88	52.85	56.72	74.73	64.24	78.49
B3	ET	83.60	81.98	75.26	71.23	74.46	75.53	78.76	87.09	73.11	89.24
	k-NN	67.47	78.22	66.39	57.79	66.39	69.35	72.58	86.55	63.97	85.21
	LR	69.62	78.49	63.44	59.13	58.60	67.20	65.05	86.82	66.12	80.37
	NB	56.18	59.67	53.49	45.16	47.04	55.64	55.91	85.48	62.36	72.04
	SVM	68.81	60.21	53.76	49.46	51.61	53.22	58.06	84.67	70.16	84.13
B4	ET	87.36	86.82	76.07	71.23	76.88	77.95	79.83	87.36	74.46	89.51
	k-NN	79.03	79.56	65.32	59.67	66.39	72.58	73.38	86.55	64.51	89.24
	LR	77.95	79.83	68.81	63.17	63.17	70.16	70.69	86.82	66.93	83.60
	NB	65.86	59.94	54.56	52.95	47.04	60.21	56.72	86.02	62.90	82.25
	SVM	79.03	71.77	54.03	49.46	66.93	53.22	58.33	84.94	73.92	84.13

Table 6
Performance analysis of highly accurate classifier Extra Tree.

Scenario	Measure	M1	M2	M3	M4	M5	M6	M7	F1	F2	F3
B1	Accuracy	85.75	85.75	70.96	70.43	75.53	76.61	76.34	86.82	72.84	88.70
	Precision	85.94	86.06	71.07	64.00	76.43	76.14	78.02	86.98	73.49	88.78
	Recall	85.75	85.75	70.96	70.43	75.53	76.61	76.34	86.82	72.84	88.70
	F-score	85.50	85.59	70.63	66.28	75.14	75.36	76.76	86.65	72.69	88.34
	Time	3.60	3.58	4.03	4.27	3.71	4.01	3.95	3.74	3.68	3.95
B2	Accuracy	84.94	85.75	75.80	70.69	76.61	76.07	77.15	86.55	73.11	89.24
	Precision	86.02	85.91	76.01	64.41	77.85	75.96	78.65	86.85	73.31	89.18
	Recall	84.94	85.75	75.80	70.69	76.61	76.07	77.15	86.55	73.11	89.24
	F-score	84.58	85.55	75.69	66.58	76.19	74.85	77.48	86.26	73.06	88.91
	Time	3.49	3.52	3.52	3.61	3.58	3.71	3.63	3.71	3.63	3.68
B3	Accuracy	83.60	81.98	75.26	71.23	74.46	75.53	78.76	87.09	73.11	89.24
	Precision	84.33	82.98	75.51	64.66	76.05	75.47	80.03	87.36	72.98	89.55
	Recall	83.60	81.98	75.26	71.23	74.46	75.53	78.76	87.09	73.11	89.24
	F-score	83.22	81.55	75.24	67.06	73.55	74.34	78.94	86.87	72.88	88.81
	Time	3.45	3.44	3.55	3.58	3.55	3.60	3.60	3.55	3.60	3.66
B4	Accuracy	87.36	86.82	76.07	71.23	76.88	77.95	79.83	87.36	74.46	89.51
	Precision	87.60	87.25	76.71	64.60	77.87	77.63	80.76	87.78	74.59	89.69
	Recall	87.36	86.82	76.07	71.23	76.88	77.95	79.83	87.36	74.46	89.51
	F-score	87.12	86.72	75.96	67.04	76.13	76.58	80.02	87.06	74.21	89.29
	Time	3.38	3.41	3.41	3.47	3.49	3.52	3.58	3.39	3.58	3.47

table is an indication of the best-performing ML classifier in terms of accuracy parameters. This is followed by Table 6, which exploits the best ML classifier in all four scenarios from Table 5 in terms of the other performance metrics. This table provides a proper justification for the better performance of one scenario over the others. After that, for observing the reliability of scenarios a bar plot is shown in Fig. 3 highlighting the mean accuracy and time consumption of the most accurate classification model obtained across different subjects in Table 6. The bar plot aims to elucidate the trade-off between accuracy and processing time across all scenarios.

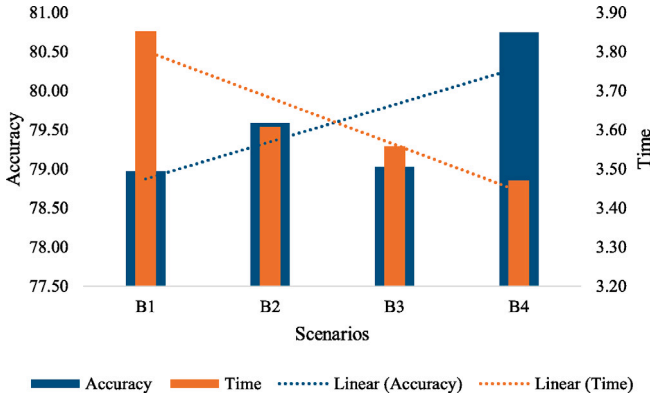


Fig. 3. Comparative analysis of baseline (B1, B2, B3) to proposed framework (B4).

3.1. Baseline model-B1: Performance analysis of machine learning classifiers

This scenario considers direct classification of all the features without any feature selection, this is interpreted as a baseline model for comparison in Fig. 3. The machine learning (ML) models include the five ML classifiers, ET, k-NN, LR, SVM, and NB, without the integration of any feature selection method. It directly takes all the features extracted during pre-processing as input. These models are trained with a training set of features and tested with the remaining set of features. The performance metrics are evaluated on the test results. The first row of Table 5 provides the accuracy of the five different ML classifiers on the testing dataset of 10 subjects corresponding to this scenario. In this scenario, the accuracies of subjects M1 to M7 and subjects F1 to F3 for the highly accurate classifier ET are 85.75%, 85.75%, 70.96%, 70.43%,

75.53%, 76.61%, 76.34%, 86.82%, 72.84%, and 88.70%, respectively. The robustness of this classifier can be verified from Table 6 by examining the other performance measures of ET for testing datasets from subjects M1 to M7 and F1 to F3. The mean accuracy for this scenario is 78.97%, and the mean time attained by this scenario is 3.85 ms as illustrated in Fig. 3.

3.2. Baseline model-B2: Performance analysis of fusion-based feature selection

This scenario is also considered as a baseline model which considers a fusion-based feature selection. The fusion framework combines three feature selection methods: filter, wrapper, and embedded. This fusion framework is applied to rank features based on relevance. Using the selected features, we analyze the impact of this scenario on the classification of hand gestures. The second row of Table 5 displays the accuracy of various ML classifiers employing this scenario. In this scenario, the accuracy of subjects M1 to M7 and subjects F1 to F3 for the highly accurate classifier ET are 84.94%, 85.75%, 75.80%, 70.69%, 76.61%, 76.07%, 77.15%, 86.55%, 73.11%, and 89.24%, respectively. The robustness of this classifier can be verified from Table 6 by examining the other performance measures of ET for testing datasets from subjects M1 to M7 and F1 to F3. The mean accuracy for this scenario is 79.59%, with a mean time of 3.61 ms, as illustrated in Fig. 3. The improvement observed between baseline models B2 and B1 in terms of mean accuracy and meantime is 0.62% higher and 0.24 ms reduction in time, respectively, as illustrated in Fig. 3.

3.3. Baseline model-B3: Performance analysis of XAI-based feature selection

In this scenario, initially, all the features are taken as input to the classifier, and then the output of the classifier is analyzed using SHAP. SHAP outputs the magnitude and direction of the impact of the feature on a particular class. The output of SHAP may vary from classifier to classifier. Since our objective is to maximize the performance of the classifier, it is decided that the SHAP be applied to the model with maximum accuracy. After Scenario-S1, we concluded that among the considered classifiers, ET is the most accurate. These findings are used here to interpret the impact of features to classify different classes while using the ET model. Viewing the impact of feature elimination on time consumption, it is decided that half of the features be eliminated for reduced time consumption. The results obtained in this scenario are displayed in the third row of Table 5. In this scenario, the accuracies of subjects M1 to M7 and subjects F1 to F3 for the highly accurate classifier ET are 83.60%, 81.98%, 75.26%, 71.23%, 74.46%, 75.53%, 78.76%, 87.09%, 73.11%, and 89.24%, respectively. The robustness of this classifier can be verified from Table 6 by examining the other performance measures of ET for testing datasets from subjects M1 to M7 and F1 to F3. The mean accuracy for this scenario is 79.03%, with a mean time of 3.56 ms as illustrated in Fig. 3. Comparison among the baseline models B3, B1, and B2 in terms of mean accuracy and meantime difference is a 0.06% improvement in accuracy compared to B1 and a 0.56% decrease compared to B2. Additionally, the processing time is shorter compared to B1 by 0.29 ms and compared to B2 by 0.05 ms, as depicted in Fig. 3.

3.4. Proposed model-B4: Performance analysis of sXFS

This scenario follows the proposed framework for feature selection in sEMG-based hand gesture identification. The framework combines them to create a fusion framework that leverages advantages from both scenarios, which includes model interpretability as well as feature importance. The performance of this fusion framework is presented in the fourth row of Table 5. In this scenario, the accuracies of subjects M1 to M7 and subjects F1 to F3 for the highly accurate classifier ET are 87.36%, 86.82%, 76.07%, 71.23%, 76.88%, 77.95%, 79.83%, 87.36%, 74.46%, and 89.51%, respectively. The robustness of this classifier can be verified from Table 6 by examining other performance measures of ET for testing datasets from subjects M1 to M7 and F1 to F3. The mean accuracy for this scenario is 80.75%, with a mean time of 3.74 ms, as illustrated in Fig. 3. The comparison between the proposed model B4 and baseline models B1, B2, and B3 shows significant improvements in mean accuracy: 1.78% higher than B1, 1.16% higher than B2, and 1.72% higher than B3. Furthermore, the proposed model demonstrates reduced processing times compared to B1, B2, and B3 by 0.28 ms, 0.14 ms, and 0.09 ms, respectively, as illustrated in Fig. 3.

Table 7
Shapiro–Wilk test for normal distribution of result analysis.

Model	B1		B2		B3		B4	
	Stat.	Sig.	Stat.	Sig.	Stat.	Sig.	Stat.	Sig.
ET	0.871	0.102	0.904	0.240	0.931	0.457	0.903	0.238
k-NN	0.942	0.573	0.950	0.673	0.927	0.416	0.951	0.676
LR	0.873	0.109	0.908	0.267	0.913	0.305	0.921	0.369
NB	0.875	0.114	0.851	0.060	0.894	0.186	0.890	0.169
SVM	0.859	0.074	0.909	0.276	0.871	0.104	0.920	0.360

Table 8
Paired t-test analysis for the proposed framework results.

Statistics	ML models				Baseline models		
	ET&k-NN	ET&LR	ET&NB	ET&SVM	B4-B1	B4-B2	B4-B1
Stat.	0.965	0.936	0.735	0.791	3.925	4.047	3.373
Sign.	0.001	0.001	0.015	0.006	0.003	0.003	0.008

3.5. Statistical significance analysis of result

The statistical analysis carried out to see whether the proposed model's acquired results differed significantly from those of other applied models. Therefore, the testing of the result's significant difference primary concern is to check the normal distribution of the data and, after the normal distribution, move on to the test, which gives clarity regarding the significant difference in the proposed model result to other applied models.

For the normality check, the most suited method is the Shapiro–Wilk test to check the normality of the distribution. For the checking of normality, the null hypothesis stated that data are normally distributed when the p value is greater than 0.05. Table 7 illustrates that the p value in all the cases is more than 0.05, so the null hypothesis is accepted, and the data is normally distributed.

After confirming normality, a paired t-test was conducted to determine the significant differences between the ET and the applied machine learning models (k-NN, LR, NB, and SVM). Tests were also conducted for the proposed model framework B4 and the other baseline models (B1, B2, and B3) to assess significant differences. If there is no significant difference between the proposed model and the other applied models, the null hypothesis (that the p-value is greater than 0.05) is accepted.

Table 8 shows that the most accurate ML classifier is ET, compared with other models (k-NN, LR, NB, and SVM), which indicates a significant difference analysis where the p-value is less than 0.05 in all cases. This leads to the rejection of the null hypothesis, meaning there is a significant difference between the ET model and other classifier models (k-NN, LR, NB, and SVM). Additionally, Table 8 illustrates that the p-value for the pair between the proposed B4 model and other baseline models (B1, B2, and B3) is less than 0.05. Therefore, the null hypothesis is rejected, and the alternative hypothesis (that there is a significant difference between the proposed B4 model and the other baseline models B1, B2, and B3) is accepted.

Table 9
Detailed comparison with other studies.

References	# Channels	# Subjects	# Gestures	# Features	Feature selection			Feature reduction		XAI	Classifier	Accuracy
					Filter	Wrapper	Embedded	PCA	LDA			
Zhang et al. (2014)	4	1	9	6	✗	✗	✗	✓	✓	✗	MDC	97.5%
Geng et al. (2014)	56	8	22	4	✓	✗	✗	✓	✗	✗	LDA	94.3%
Abbaspour et al. (2020)	4	20	11	46	✗	✗	✗	✓	✗	✗	LDA,k-NN,MLE,SVM,MLP	above 97%
La Banca Freitas et al. (2019)	6	3	6	4	✗	✗	✗	✗	✗	✗	NB,LDA,DT,QDA,NN, MLP	90%
Jena et al. (2023)	8	6	6	9	✓	✗	✗	✗	✗	✗	DT,NB,SVM,k-NN	89.75%
Gehlot et al. (2023)	2	1	6	17	✗	✗	✗	✗	✗	✗	ADB,RF,DT,LR,NB	78.49%
Proposed	2	10	6	17	✓	✓	✓	✗	✗	✓	ET,k-NN,LR,NB,SVM	mean(80.75%)

= Number of, PCA = Principal Component Analysis, LDA = Linear discriminant analysis, XAI = eXplainable Artificial Intelligence, MDC = Minimum distance classifier, QDA = Quadratic Discriminant Analysis, MLP = Multi-Layer Perceptron, k-NN = k-Nearest Neighbors, MLE = Maximum Likelihood Estimation, SVM = Support Vector Machine, NB = Naive Bayes , DT = Decision Tree, NN = Neural Network, ADB = AdaBoost, RF = Random Forest, LR = Logistic Regression, ET = Extra Tree.

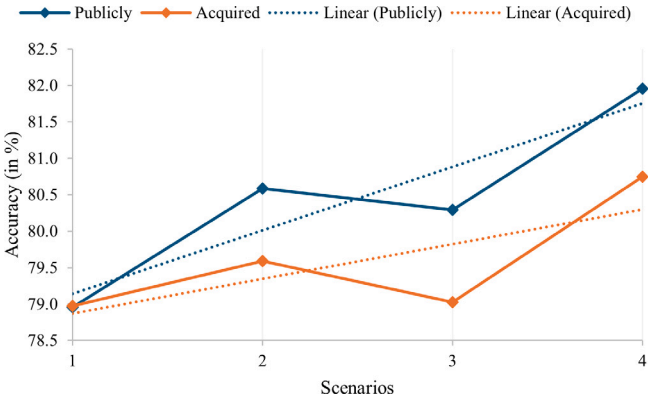


Fig. 4. Comparative analysis of publicly available dataset with collected dataset on baseline (B1, B2, B3) to proposed framework (B4).

3.6. Evaluation with publicly dataset

The proposed model was evaluated using the existing data repository from R. N. Khushaba et al. Khushaba and Kodagoda (2012), which consists of 15 different hand gestures similar to those in our acquired dataset. We followed the exact approach for the training and testing split, as well as maintaining a consistent environment across all baseline models. After ensuring uniformity in the environment, we observed the accuracy achieved in each baseline model and identified that our proposed model (B4) attained the highest accuracy among all baseline models. The mean accuracy curve of all subjects across all baselines is illustrated in Fig. 4. The figure clearly visualizes the accuracy at the baseline B1 for acquired data and publicly data is 78.97%, 78.95%, Similarly, for the proposed scenarios, the mean accuracy of the acquired is 80.75%, 81.95%. The figure demonstrates a consistent growth pattern across all scenarios when the environment remains consistent for both the acquired and existing data. The trend of the line from the baseline model B1 to the proposed model B4 indicates linear growth in classifier accuracy.

4. Discussion

4.1. Ablation study

This study encompasses four distinct scenarios, which include three baseline models (B1, B2, B3), (B1) comprises of classification without feature selection, (B2) comprises of classification with fusion framework for feature selection, (B3) comprises of classification with SHAP for feature selection, and (B4) comprises of classification with the proposed framework: a combination of XAI (SHAP) and fusion framework for feature selection. The primary objective of the experiment is to finalize a reduced set of optimal features that enhance the accuracy of the system while reducing time consumption. Five baseline models, ET, k-NN, LR, NB, and SVM, were chosen to measure the enhancement in performance. Performance metrics such as accuracy, precision, recall, and F-score, time are used to evaluate the predictability of these models. Time consumption is used to evaluate the four scenarios. It is observed that the proposed framework outperformed all baseline models by a significant difference in accuracy, precision, recall, F-score, and time consumption. Which justifies the viability of the proposed framework for feature selection. Additionally, the best-performing ML classifier is selected to enhance the accuracy of the final analysis. The combination of ET and the proposed framework provides very high accuracy and reduced time consumption. The robustness of this is also validated by testing the proposed framework on a public dataset.

4.2. Comparison with previous studies

This study introduces a peer-to-peer HGR model based on ML classifiers and a feature selection approach. The superior performance of the proposed model has been validated using both a collected dataset and an existing dataset. In the existing literature, the focus primarily revolves around enhancing gesture accuracy through conventional feature selection techniques combined with ML classifiers. However, researchers often overlook the interpretability of features in conjunction with the classifier. To address this gap, we propose a sXFS for HGR.

Table 9 summarizes state-of-the-art studies utilizing feature selection, highlighting their best performances for comparison with our study in terms of key parameters and accuracy. According to the literature, while the performance of models tends to improve, the aspect of feature selection with interpretability remains overlooked. Therefore, in this study, we provide a detailed comparison of two such studies.

4.3. Limitations and future works

The study also reveals four limitations of the proposed sXFS framework. First, the model requires a significant amount of time. This time factor encompasses feature extraction, feature selection through fusion techniques, interpreting feature importance using XAI, and feeding back the final feature importance. While this extended processing time enhances the overall model's performance, it may not be suitable for scenarios where time is a critical factor in the initial stages of analysis. Second, the study exclusively employs the SHAP XAI model for interpretability. Introducing additional XAI models could provide a more comprehensive understanding of classifier model behavior. Third, the default model parameters are used, and optimizing these parameters could potentially enhance the model's accuracy. Lastly, the study is limited by the number of subjects involved in the research. To strengthen the validity and reliability of the proposed approach, it should be validated with a larger and more diverse pool of subjects.

5. Conclusion

In this research, the article introduces the proposed model sXFS for optimal feature selection based on the eXplainable Artificial Intelligence (XAI) fusion framework, which incorporates the fusion of filter, wrapper, and embedded methods using SHAP (Shapley Additive exPlanations). To assess the robustness and reliability of the proposed model (B4), it was tested against three baseline models (B1, B2, B3) using two datasets: one acquired dataset and another publicly available dataset. The results obtained from the acquired dataset in terms of performance metrics, including accuracy, precision, recall, F-score, and processing time, were compared and analyzed. The proposed model sXFS, in conjunction with the classifier ET, achieved the highest accuracy among both datasets. Additionally, the computational time of the proposed model was found to be faster compared to all baseline models. This advantage holds significant potential for applications in real-time robotics and precise movement control, ensuring exceptionally high levels of accuracy and enhancing the generalization of myoelectric control systems.

Funding

No funding.

CRediT authorship contribution statement

Naveen Gehlot: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ashutosh Jena:** Writing – review & editing, Data curation. **Ankit Vijayvargiya:** Writing – review & editing, Visualization, Validation, Supervision. **Rajesh Kumar:** Writing – review & editing, Visualization, Validation, Supervision, Methodology.

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.

Data availability

Data will be made available on request.

References

Abbaspour, S., Lindén, M., Gholamhosseini, H., Naber, A., Ortiz-Catalan, M., 2020. Evaluation of surface EMG-based recognition algorithms for decoding hand movements. Med. Biol. Eng. Comput. 58, 83–100.

Akman, D.V., Malekipirbazari, M., Yenice, Z.D., Yeo, A., Adhikari, N., Wong, Y.K., Abbasi, B., Gumus, A.T., 2023. K-best feature selection and ranking via stochastic approximation. Expert Syst. Appl. 213, 118864.

Antwarg, L., Galed, C., Shimoni, N., Rokach, L., Shapira, B., 2023. Shapley-based feature augmentation. Inf. Fusion 96, 92–102.

Arican, E., Aydin, T., 2022. An RGB-D descriptor for object classification. Romanian J. Inf. Sci. Technol. (ROMJIST) 25 (3–4), 338–349.

Bermejo, P., de la Ossa, L., Gámez, J.A., Puerta, J.M., 2012. Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking. Knowl.-Based Syst. 25 (1), 35–44.

Bi, J., Wang, H., Hua, M., Yan, K., 2024. An interpretable feature selection method integrating ensemble models for chiller fault diagnosis. J. Build. Eng. 109029.

Biswas, S., Bordoloi, M., Purkayastha, B., 2016. Review on feature selection and classification using neuro-fuzzy approaches. Int. J. Appl. Evol. Comput. (IJAEC) 7 (4), 28–44.

Borlea, I.-D., Precup, R.-E., Borlea, A.-B., 2022. Improvement of K-means cluster quality by post processing resulted clusters. Procedia Comput. Sci. 199, 63–70.

Côté-Allard, U., Fall, C.L., Drouin, A., Campeau-Lecours, A., Gosselin, C., Glette, K., Laviolette, F., Gosselin, B., 2019. Deep learning for electromyographic hand gesture signal classification using transfer learning. IEEE Trans. Neural Syst. Rehabil. Eng. 27 (4), 760–771.

Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and other Kernel-Based Learning Methods. Cambridge University Press.

Deng, H., Runger, G., 2013. Gene selection with guided regularized random forest. Pattern Recognit. 46 (12), 3483–3489.

Dudani, S.A., 1976. The distance-weighted k-nearest-neighbor rule. IEEE Trans. Syst. Man Cybern. (4), 325–327.

Effrosynidis, D., Arampatzis, A., 2021. An evaluation of feature selection methods for environmental data. Ecol. Inform. 61, 101224.

Fang, B., Wang, C., Sun, F., Chen, Z., Shan, J., Liu, H., Ding, W., Liang, W., 2022. Simultaneous sEMG recognition of gestures and force levels for interaction with prosthetic hand. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 2426–2436. http://dx.doi.org/10.1109/TNSRE.2022.3199809.

Fida, M.A.F.A., Ahmad, T., Ntahobari, M., 2021. Variance threshold as early screening to boruta feature selection for intrusion detection system. In: 2021 13th International Conference on Information & Communication Technology and System. ICTS, IEEE, pp. 46–50.

Gehlot, N., Jena, A., Vijayvargiya, A., Kumar, R., 2023. sEMG-based classification of finger movement with machine learning. In: 2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications. IC2E3, IEEE, pp. 1–6.

Gehlot, N., Malik, S., Jena, A., Vijayvargiya, A., Kumar, R., 2024. XAI-driven sEMG feature analysis for hand gestures. In: 2024 Third International Conference on Power, Control and Computing Technologies. ICPC2T, IEEE, pp. 19–24.

Geng, Y., Kuang, X., Zhu, M., Zhang, Y., Li, G., Zhang, Y.-T., 2014. Exploration of data dimensionality reduction methods for improving classification performance of voluntary movements. In: The International Conference on Health Informatics: ICHI 2013, Vilamoura, Portugal on 7-9 November, 2013. Springer, pp. 126–129.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63, 3–42.

Harrison, K.R., Elsayed, S.M., Weir, T., Garanovich, I.L., Boswell, S.G., Sarker, R.A., 2022. Solving a novel multi-divisional project portfolio selection and scheduling problem. Eng. Appl. Artif. Intell. 112, 104771.

Hassan, F., Hussain, S.F., Qaisar, S.M., 2023. Fusion of multivariate EEG signals for schizophrenia detection using CNN and machine learning techniques. Inf. Fusion 92, 466–478.

He, S., Huang, S., Huang, L., Xie, F., Xie, L., 2023. LiDAR-based hand contralateral controlled functional electrical stimulation system. IEEE Trans. Neural Syst. Rehabil. Eng. 31, 1776–1785.

Hong, C., Park, S., Kim, K., 2023. sEMG-based gesture recognition using temporal history. IEEE Trans. Biomed. Eng. 70 (9), 2655–2666. http://dx.doi.org/10.1109/TBME.2023.3261336.

Indra, S., Wikarsa, L., Turang, R., 2016. Using logistic regression method to classify tweets into the selected topics. In: 2016 International Conference on Advanced Computer Science and Information Systems. Icacsis, IEEE, pp. 385–390.

Jalilvand, A., Salim, N., 2017. Feature unionization: a novel approach for dimension reduction. Appl. Soft Comput. 52, 1253–1261.

Jena, A., Baberwal, K., Gehlot, N., Kumar, R., 2023. Impact of feature selection on sEMG signal classification. In: 2023 14th International Conference on Computing Communication and Networking Technologies. ICCCNT, IEEE, pp. 1–6.

Jiang, B., Chen, S., Wang, B., Luo, B., 2022. MGLNN: Semi-supervised learning via multiple graph cooperative learning neural networks. Neural Netw. 153, 204–214.

Junior, J.J.A.M., Freitas, M., Siqueira, H., Lazzaretti, A.E., Stevan, S., Pichorim, S.F., 2020. Comparative analysis among feature selection of sEMG signal for hand gesture classification by armband. IEEE Latin Am. Trans. 18 (06), 1135–1143.

Kaushik, A., Gurucharan, K., Padmavathi, S., 2023. Enhancing human activity recognition: An exploration of machine learning models and explainable AI approaches for feature contribution analysis. In: 2023 International Conference on Energy, Materials and Communication Engineering. ICEMCE, IEEE, pp. 1–6.

Khushaba, R.N., Kodagoda, S., 2012. Electromyogram (EMG) feature reduction using mutual components analysis for multifunction prosthetic fingers control. In: 2012 12th International Conference on Control Automation Robotics & Vision. ICARCV, IEEE, pp. 1534–1539.

Kwon, J., Kwon, O., Oh, K.T., Kim, J., Yoo, S.K., 2023. Breathing-associated facial region segmentation for thermal camera-based indirect breathing monitoring. IEEE J. Transl. Eng. Health Med. 11, 505–514. http://dx.doi.org/10.1109/JTEHM.2023.3295775.

La Banca Freitas, M., Mendes, J.J.A., Campos, D.P., Stevan, S.L., 2019. Hand gestures classification using multichannel sEMG armband. In: XXVI Brazilian Congress on Biomedical Engineering: CBEB 2018, Armação de Buzios, RJ, Brazil, 21-25 October 2018 (Vol. 2). Springer, pp. 239–246.

Leung, K.M., et al., 2007. Naive Bayesian Classifier. Vol. 2007, Polytechnic University Department of Computer Science/Finance and Risk Engineering, pp. 123–156.

Li, Z., Li, K., Li, J., Wei, N., 2020. Optimization of semg classification model based on correlation analysis and feature selection. In: 2020 5th International Conference on Advanced Robotics and Mechatronics. ICARM, IEEE, pp. 402–407.

Lin, H., Ding, H., 2011. Predicting ion channels and their types by the dipeptide mode of pseudo amino acid composition. J. Theoret. Biol. 269 (1), 64–69.

Liu, M.-K., Lin, Y.-T., Qiu, Z.-W., Kuo, C.-K., Wu, C.-K., 2020. Hand gesture recognition by a MMG-based wearable device. IEEE Sens. J. 20 (24), 14703–14712.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.

Narayan, Y., 2021. SEMG signal classification using KNN classifier with FD and TFD features. Mater. Today: Proc. 37, 3219–3225.

Ozdemir, M.A., Kisa, D.H., Guren, O., Akan, A., 2022. Dataset for multi-channel surface electromyography (sEMG) signals of hand gestures. Data Brief 41, 107921.

Pancholi, S., Joshi, A.M., 2019. Electromyography-based hand gesture recognition system for upper limb amputees. IEEE Sens. Lett. 3 (3), 1–4. http://dx.doi.org/10.1109/LSENS.2019.2898257.

Phinyomark, A., Phukpattaranont, P., Limsakul, C., 2012. Feature reduction and selection for EMG signal classification. Expert Syst. Appl. 39 (8), 7420–7431.

Pradhan, A., Nayak, S.K., Pande, K., Ray, S.S., Pal, K., Champaty, B., Anis, A., Tibarewala, D., 2016. Acquisition and classification of EMG using a dual-channel EMG biopotential amplifier for controlling assistive devices. In: 2016 IEEE Annual India Conference. INDICON, IEEE, pp. 1–5.

Qian, W., Huang, J., Xu, F., Shu, W., Ding, W., 2023. A survey on multi-label feature selection from perspectives of label fusion. Inf. Fusion 100, 101948.

Qiu, S., Zhao, H., Jiang, N., Wang, Z., Liu, L., An, Y., Zhao, H., Miao, X., Liu, R., Fortino, G., 2022. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. Inf. Fusion 80, 241–265.

Roy, A.M., Bhaduri, J., 2023. DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and swin-transformer prediction head-enabled YOLOv5 with attention mechanism. Adv. Eng. Inform. 56, 102007.

Roy, A.M., Bose, R., Bhaduri, J., 2022. A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. Neural Comput. Appl. 34 (5), 3895–3921.

Sánchez-Maroño, N., Alonso-Betanzos, A., Tombilla-Sanromán, M., 2007. Filter methods for feature selection–a comparative study. In: International Conference on Intelligent Data Engineering and Automated Learning. Springer, pp. 178–187.

Shapley, L.S., et al., 1953. A Value for N-Person Games. Princeton University Press Princeton.

Siddiqui, N., Chan, R.H., 2020. Hand gesture recognition using multiple acoustic measurements at wrist. IEEE Trans. Hum.-Mach. Syst. 51 (1), 56–62.

Singha, J., Laskar, R.H., 2017. Hand gesture recognition using two-level speed normalization, feature selection and classifier fusion. Multimedia Syst. 23 (4), 499–514.

Swingle, B., 2012. Rényi entropy, mutual information, and fluctuation properties of Fermi liquids. Phys. Rev. B 86 (4), 045109.

Van Zyl, C., Ye, X., Naidoo, R., 2024. Harnessing explainable artificial intelligence for feature selection in time series energy forecasting: A comparative analysis of grad-CAM and SHAP. Appl. Energy 353, 122079.

Vijayvargiya, A., Gupta, V., Kumar, R., Dey, N., Tavares, J.M.R., 2021. A hybrid WD-EEMD sEMG feature extraction technique for lower limb activity recognition. IEEE Sens. J. 21 (18), 20431–20439.

Vijayvargiya, A., Kumar, R., Dey, N., Tavares, J.M.R., 2020. Comparative analysis of machine learning techniques for the classification of knee abnormality. In: 2020 IEEE 5th International Conference on Computing Communication and Automation. ICCCA, IEEE, pp. 1–6.

Vijayvargiya, A., Kumar, R., Sharma, P., 2023a. PC-GNN: Pearson correlation-based graph neural network for recognition of human lower limb activity using sEMG signal. IEEE Trans. Hum.-Mach. Syst..

Vijayvargiya, A., Raghav, A., Bhardwaj, A., Gehlot, N., Kumar, R., 2023b. A LIME-based explainable machine learning technique for the risk prediction of chronic kidney disease. In: 2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications. IC2E3, IEEE, pp. 1–6.

Vijayvargiya, A., Singh, P., Kumar, R., Dey, N., 2022a. Hardware implementation for lower limb surface EMG measurement and analysis using explainable AI for activity recognition. IEEE Trans. Instrum. Meas. 71, 1–9.

Vijayvargiya, A., Singh, B., Kumar, R., Tavares, J.M.R., 2022b. Human lower limb activity recognition techniques, databases, challenges and its applications using sEMG signal: an overview. Biomed. Eng. Lett. 12 (4), 343–358.

Xu, Z., Huang, G., Weinberger, K.Q., Zheng, A.X., 2014. Gradient boosted feature selection. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 522–531.

Xue, J., Sun, Z., Duan, F., Caiafa, C.F., Solé-Casals, J., 2023. Underwater sEMG-based recognition of hand gestures using tensor decomposition. Pattern Recognit. Lett. 165, 39–46.

Yang, J., Jin, D., Wang, R., Zhang, J., Ji, L., Fang, X., Zhou, D., 2005. The investigation on sEMG of lower extremity when a slip occurs in level walking. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. pp. 5934–5937. http://dx.doi.org/10.1109/IEMBS.2005.1615842.

Zhang, Z., Zhang, F., Mao, L., Chen, C., Ning, H., 2024. DFS-WR: A novel dual feature selection and weighting representation framework for classification. Inf. Fusion 104, 102191.

Zhang, D., Zhao, X., Han, J., Zhao, Y., 2014. A comparative study on PCA and LDA based EMG pattern recognition for anthropomorphic robotic hand. In: 2014 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 4850–4855.