Journal of Theoretical and Applied Information Technology
15th July 2026. Vol.104. No.13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

STACKING-BASED ENSEMBLE CLASSIFICATION OF HAND MOVEMENTS USING WAVELET SCATTERING FEATURES EXTRACTED FROM RAW AND FILTERED SURFACE EMG SIGNALS

MOHAMED ELMEHDI AIT BOURKHA [1,2], ABDERRAHMANE BOUDRIBILA [1,3], ANAS HATIM [2], ASSIA SAYDTAHIRI [2]

[1] LAMIGEP, Moroccan School of Engineering Sciences (EMSI), Marrakech, Morocco

[2] Research laboratory in innovative and sustainable technologies (LaRTID), National School of Applied Sciences (ENSA), Cadi Ayyad University (UCA), Marrakech, Morocco

[3] Systems Engineering and Applications Laboratory, National School of Applied Sciences (ENSA), Cadi Ayyad University (UCA), Marrakech, Morocco

E-mail: [1] mehdibourkha123@gmail.com, [2] a.boudribila.ced@uca.ac.ma

ABSTRACT

Surface electromyography (sEMG) signal classification plays a critical role in the development of intelligent prosthetic control systems and rehabilitation technologies for individuals with upper-limb motor impairments. Although numerous machine learning approaches have been proposed for sEMG-based hand gesture recognition, achieving high classification accuracy across multiple grasp types remains challenging, particularly when six or more movement classes are involved. The primary goal of this study is to develop and rigorously validate a robust yet computationally lightweight framework that attains high six-class classification accuracy while simplifying the conventional preprocessing pipeline required for real-time myoelectric control. To this end, two hypotheses are tested: first, that the wavelet scattering transform produces a feature representation sufficiently noise-robust to render conventional band-pass and notch filtering unnecessary; and second, that combining complementary K-Nearest Neighbors classifiers through a two-level stacking ensemble yields higher accuracy than any individual classifier or existing single-model approach reported on the same benchmark. The proposed framework combines wavelet scattering transform (WST)-based feature extraction with a stacking ensemble learning strategy for the recognition of six fundamental hand movements (Cylindrical, Tip, Hook, Palmar, Spherical, and Lateral) from a publicly available benchmark dataset (UCI sEMG, five subjects, two channels, 500 Hz). The wavelet scattering network decomposes each raw sEMG signal into a compact 6×202 time–frequency feature tensor, achieving a 59.60% dimensionality reduction while preserving discriminative information across scattering orders. A systematic comparison between filtered and raw sEMG signals demonstrates that the scattering transform renders conventional band-pass and notch filtering unnecessary, as both signal types yield comparable classification performance. At the classification stage, a two-level stacking architecture is employed: K-Nearest Neighbors ($K = 3$) and ensemble subspace KNN (20 learners) serve as base-level classifiers, and their concatenated predictions are fed into a meta-learner (ensemble subspace KNN, 30 learners). Evaluated through five-fold cross-validation, the proposed framework achieves an overall classification accuracy of 99.67%, with average sensitivity, specificity, precision, and F1-score of 99.67%, 99.93%, 99.67%, and 99.67%, respectively, outperforming all existing methods reported on the same dataset. These results demonstrate the effectiveness of combining scattering-based representations with ensemble stacking for robust and computationally efficient sEMG classification, with potential applications in myoelectric prosthetic control and clinical rehabilitation.

Keywords: *Surface Electromyography (sEMG), Hand Movement Classification, Wavelet Scattering Transform, Ensemble Learning, Stacking, K-nearest Neighbors, Prosthetic Control*

1 INTRODUCTION

Surface electromyography (sEMG) is a non-invasive technique that captures the electrical activity generated by skeletal muscles during voluntary contractions. Rana and Bhushan [1] highlighted the growing role of such biomedical signals in clinical applications, while Campanini et al. [2] provided a comprehensive overview of surface electrode configurations and their respective advantages and limitations for recording muscle activity. Over the past two decades, advances in embedded sensor technologies and miniaturized acquisition systems, as documented by Mantilla-Brito et al. [3] and Crippa et al. [4], have significantly expanded the applicability of sEMG in clinical diagnostics, prosthetic control, and human-computer interaction. In particular, the classification of hand movements from sEMG signals has emerged as a critical research area, given its direct implications for the development of intelligent myoelectric prostheses and motor rehabilitation systems aimed at restoring functional independence for individuals with upper-limb impairments. Ladegaard [5] traced the evolution of electromyography equipment from analog systems in the 1950s to modern microprocessor-controlled platforms, underscoring the technological maturity that now enables real-time sEMG analysis.

The accurate recognition of hand grasping patterns from sEMG signals is essential for translating muscle activation data into functional prosthetic commands, as demonstrated by Huang and Chen [6] and Lucas et al. [7]. However, this classification task presents several inherent challenges. These signals are non-stationary, low in amplitude, and susceptible to various sources of noise, including power-line interference at 50/60 Hz, motion artifacts, electrode shift, and inter-subject variability in muscle physiology. Raghu et al. [8] and Farago et al. [9] provided detailed reviews of the signal quality challenges that affect sEMG-based systems. Furthermore, as Oskoei and Hu [10] and Sapsanis et al. [11] observed, the classification task becomes substantially more complex as the number of target movement classes increases, because functionally similar grasps often produce overlapping muscle activation patterns that are difficult to distinguish. Consequently, the design of effective sEMG classification systems requires both robust feature extraction methods capable of capturing discriminative temporal and spectral information, and classifiers that can reliably exploit these features for accurate multi-class recognition.

In recent years, machine learning (ML) and deep learning (DL) techniques have been widely adopted for automated sEMG signal analysis, offering substantial improvements over manual visual inspection by clinical experts, which, as noted by Raghu et al. [8], Sharma et al. [12], and Rani et al. [13], remains both time-consuming and prone to subjective error. More broadly, the application of intelligent models to biomedical signal analysis has been advancing rapidly across multiple modalities, including electroencephalography (EEG), as investigated by Degirmenci et al. [14] and Rahul et al. [15], electrocardiography (ECG), as addressed by Bourkha et al. [16] and Singh and Mahapatra [17], and sEMG, as studied by Karuna and Guntur [18] and Putro et al. [19]. These AI-driven systems enable rapid diagnosis and reporting, as highlighted by Meyers et al. [20], and offer considerable promise for improving the accuracy and speed of disease detection, as demonstrated in the reviews by Zafar et al. [21], Liu et al. [22], and Ahsan et al. [23]. Within the specific domain of sEMG-based hand movement classification, a broad range of approaches has been proposed. Traditional ML methods such as support vector machines (SVM), employed by Lucas et al. [7] and Oskoei and Hu [10], and K-nearest neighbors (KNN), used by Pourmokhtari and Beigzadeh [24], have been applied with handcrafted features. Meanwhile, DL architectures including convolutional neural networks (CNN), explored by Sapsanis et al. [11] and Bakırcıoğlu and Özkurt [25], long short-term memory (LSTM) networks, employed by Valdivieso Caraguay et al. [26], hybrid CNN-LSTM models, as proposed by Abdelaziz et al. [27], and vision transformers, introduced by Godoy et al. [28], have been explored for automatic feature learning. Alongside these classification methods, feature extraction strategies have evolved considerably, progressing from simple time-domain and frequency-domain descriptors to more advanced representations based on empirical mode decomposition (EMD), as used by Karuna and Guntur [18] and Sapsanis et al. [11], discrete wavelet transforms (DWT), as applied by Lucas et al. [7] and Mane et al. [29], and the wavelet scattering transform (WST), as investigated by Al-Taee et al. [30] and theoretically grounded by Andén and Mallat [31]. Despite these methodological advances, achieving consistently high classification accuracy remains an open challenge when six or more hand movement classes are involved, with the best reported performance on the UCI sEMG benchmark dataset reaching approximately 98.44%, as reported by Ovadia et al. [32]. Closing this performance gap while simultaneously simplifying the processing

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

chain is of considerable practical importance: in real-time myoelectric control, both classification reliability and low-latency, low-power operation directly determine whether a prosthesis is clinically usable, so an approach that improves accuracy and removes a preprocessing stage at the same time carries tangible translational value.

A significant yet underexplored aspect in the sEMG classification pipeline concerns the role of signal preprocessing, particularly filtering. Most existing studies, including those by Sapsanis et al. [11] and Bakırcıoğlu and Özkurt [25], apply band-pass and notch filtering as a standard preprocessing step prior to feature extraction, under the assumption that denoising is essential for achieving competitive classification performance. However, no prior work has systematically evaluated whether such filtering remains necessary or beneficial when an advanced, noise-robust feature extraction method such as the WST is employed. This question carries practical significance, as eliminating the filtering stage would simplify the processing pipeline, reduce computational overhead, and facilitate real-time implementation in embedded prosthetic systems. In parallel, while ensemble learning strategies such as bagging, boosting, and random subspace methods have demonstrated effectiveness in improving classification robustness in biomedical signal analysis, as shown by Bourkha et al. [16] and Singh and Mahapatra [17], the potential of stacking-based ensemble architectures, which combine heterogeneous base classifiers through a learned meta-model, has not been thoroughly investigated for sEMG hand movement classification.

To address these gaps, this study proposes a novel classification framework that integrates wavelet scattering network (WSN)-based feature extraction with a two-level stacking ensemble learning architecture for the recognition of six hand movements from sEMG signals. The central objective is to demonstrate that high-accuracy six-class recognition can be achieved without the conventional band-pass and notch filtering stage, and that combining heterogeneous distance-based classifiers through stacking can advance performance beyond the current state of the art on the UCI sEMG benchmark. This work is therefore guided by two hypotheses. First, because the wavelet scattering transform applies a cascade of multi-scale wavelet convolutions, modulus operations, and low-pass averaging, it is expected to implicitly attenuate the same noise components (low-frequency drift and 50 Hz power-line interference) that band-pass and notch filters are designed to remove, making explicit denoising redundant. Second, because an individual KNN classifier and a feature-subspace ensemble KNN tend to misclassify different samples, a learned meta-model that fuses their predictions should correct these complementary errors and surpass the accuracy of either model in isolation. These hypotheses are substantiated through three contributions. The first contribution of this work lies in a comprehensive empirical investigation of the impact of signal filtering on classification performance when wavelet scattering features are used. Through a systematic comparison of filtered and raw sEMG signals processed with identical WSN configurations, the results demonstrate that the scattering transform effectively captures the discriminative time–frequency content of raw signals, rendering conventional band-pass and notch filtering unnecessary without any degradation in classification accuracy. The second contribution consists in the design and optimization of a two-level stacking ensemble classifier. In the proposed architecture, two complementary base-level models, a standard KNN classifier with K = 3 and an ensemble subspace KNN with 20 learners, independently generate predictions that are concatenated into a joint feature vector, which is then processed by a meta-learner implemented as an ensemble subspace KNN with 30 learners. This hierarchical design exploits the complementary strengths of the base classifiers and yields an overall classification accuracy of 99.67% on the UCI sEMG benchmark dataset provided by Sapsanis et al. [33], surpassing all previously reported results on the same data. The third contribution is a rigorous evaluation protocol based on five-fold cross-validation, through which the proposed framework is shown to achieve 99.67% average sensitivity, 99.93% specificity, 99.67% precision, and 99.67% F1-score across the six movement classes, confirming both its discriminative power and its generalization capability.

The remainder of this paper is organized as follows. Section 2 reviews related work on sEMG-based hand movement classification and positions the present study within the existing literature. Section 3 describes the proposed methodology in detail, covering data acquisition, signal preprocessing, wavelet scattering-based feature extraction, and the stacking ensemble classification framework. Section 4 presents the experimental results, analyzes the classification performance, and provides a comparative discussion with state-of-the-art methods. Finally, Section 5 summarizes the key findings, while Section 6 discusses the limitations of the current work and outlines directions for future research.

2 RELATED WORKS

The classification of hand movements from sEMG signals has been the subject of extensive research over the past two decades, with studies differing in the number of target gestures, the feature extraction strategy, and the classification algorithm employed. This section provides a structured review of the most relevant prior works, organized thematically according to the methodological paradigm adopted, and concludes by identifying the specific gaps that motivate the present study.

2.1 Classical Machine Learning Approaches

Early efforts in sEMG-based hand movement classification relied primarily on conventional ML classifiers combined with handcrafted features. Huang and Chen [6] developed a myoelectric discrimination system for the NTU prosthetic hand, in which two surface electrodes recorded signals from the flexor digitorum superficialis and extensor pollicis brevis muscles. They extracted variance, autoregressive coefficients, zero-crossing counts, and spectral features, and employed an error backpropagation neural network to distinguish eight hand motions, achieving an accuracy of 85%. Although pioneering, the relatively low accuracy can be attributed to the limited number of electrodes and the shallow network architecture, which restricted the model's capacity to learn complex inter-class boundaries.

Lucas et al. [7] proposed a multi-channel sEMG-based system for prosthetic control, employing the discrete wavelet transform (DWT) for feature extraction and an SVM classifier operating in a multi-channel configuration. Using eight forearm electrode sites to differentiate six hand movements, their approach achieved 95% accuracy, demonstrating the advantages of wavelet-based representations over purely time-domain features. In a complementary study, Oskoei and Hu [10] investigated the use of SVM with various kernel functions for upper-limb movement classification from sEMG signals. Their work identified key design considerations for SVM-based myoelectric control, including optimal data segmentation, feature selection, and the use of overlapped windowing combined with majority voting as a post-processing strategy, and reported a classification accuracy of 94.50% for six hand movements. While both studies by Lucas et al. [7] and Oskoei and Hu [10] confirmed the viability of SVM for sEMG classification, their reliance on linear or radial basis function kernels limited the capacity to model highly nonlinear class boundaries present in multi-gesture scenarios.

Pourmokhtari and Beigzadeh [24] explored the use of KNN for classifying five finger movements based on sEMG data. They evaluated combinations of time-domain features, including maximum, minimum, mean absolute value (MAV), root mean square (RMS), and simple square integral (SSI), and found that the combination of MAV with maximum and minimum values yielded the best performance, achieving accuracies between 89.80% and 96% across four channels. This work highlighted the sensitivity of KNN-based classification to feature selection, but the restriction to only five movement classes and the absence of ensemble strategies limited the generalizability of their findings.

2.2 Deep Learning Approaches

The emergence of deep learning has introduced new possibilities for sEMG classification by enabling automatic feature learning directly from raw or minimally preprocessed signals. Sapsanis et al. [11] proposed a CNN-based approach in which sEMG signals were first decomposed into intrinsic mode functions (IMFs) using empirical mode decomposition (EMD), and statistical features were extracted from these IMFs before classification. Their method classified the same six hand movements considered in the present study (Cylindrical, Tip, Hook, Palmar, Spherical, and Lateral) from the UCI sEMG dataset provided by Sapsanis et al. [33], achieving an average accuracy of 89.21%. While the integration of EMD improved the feature quality, the overall accuracy remained limited by the CNN architecture and the relatively small training set.

Building upon this work, Bakırcıoğlu and Özkurt [25] developed a deeper CNN-based method for classifying the same six hand movements from the same UCI dataset [33]. They applied windowing to oversample the data and explored combinations of Fourier transform, root mean square, and EMD-based preprocessing. Among four CNN architectures evaluated, the integration of EMD with CNN yielded the highest accuracy of 95.90% under cross-validation. Although this result represented a notable improvement over the work of Sapsanis et al. [11], the reliance on data augmentation through overlapping windows raises concerns about information leakage between training and validation folds, and the accuracy remained below the threshold typically expected for practical prosthetic applications.

Abdelaziz et al. [27] introduced a hybrid architecture combining CNN and LSTM networks, in which the CNN was responsible for extracting spatial features from sEMG signals while the LSTM captured temporal and sequential dependencies in gesture patterns. Their model achieved 99% accuracy on the DualMyo dataset (single subject, eight postures) and approximately 97% on a larger dataset comprising 36 subjects. Despite the strong performance, the high accuracy on the single-subject dataset likely reflects subject-specific overfitting rather than generalizable classification capability, and the drop to 97% on the multi-subject dataset underscores the challenge of inter-subject variability.

Godoy et al. [28] proposed the Temporal Multi-Channel Vision Transformer (TMC-ViT), a DL method that leveraged 16 channels to recognize 18 hand gestures, incorporating CNN, LSTM, and deep CNN components. Their method achieved 98.45% accuracy on the Dexterity Dataset, demonstrating the potential of transformer-based architectures for multi-gesture recognition. However, the high computational cost associated with vision transformers and the requirement for a large number of channels limit the practical applicability of this approach in wearable or embedded prosthetic systems.

Valdivieso Caraguay et al. [26] explored reinforcement learning for sEMG gesture detection, employing Deep Q-Network (DQN) and Double-Deep Q-Network (DDQN) approaches with feed-forward and LSTM layers to recognize five gestures from the EMG-EPN-612 dataset. Their analysis showed that the DQN without LSTM achieved the best performance, with a validation accuracy of 90.37% and a testing accuracy of 82.52%. The substantial gap between validation and testing accuracy suggests overfitting, and the relatively low testing accuracy indicates that reinforcement learning may not be the most suitable paradigm for supervised sEMG classification tasks.

2.3 Signal Denoising and Feature Extraction Approaches

Several studies have focused on improving classification performance through advanced signal denoising and feature extraction strategies. Karuna and Guntur [18] proposed a denoising pipeline based on EMD combined with a correlation coefficient threshold (CCT) method for classifying six hand movements. The EMD decomposed the noisy sEMG signal into IMFs, and the CCT identified the IMFs most correlated with the original signal for reconstruction. After denoising, time-domain and histogram features were extracted and classified using four algorithms, with linear discriminant analysis (LDA) achieving the best accuracy of 94.63%. While the denoising approach of Karuna and Guntur [18] was well-motivated, the overall classification performance remained moderate, and the method introduced additional computational complexity.

Mane et al. [29] applied wavelet transformation combined with an artificial neural network (ANN) classifier to single-channel sEMG signals, achieving 98.60% accuracy for three hand movements (open palm, closed palm, and wrist extension). Although the accuracy reported by Mane et al. [29] was high, the task was limited to only three classes with clearly distinct activation patterns, and the generalizability to a larger number of movements was not evaluated.

Nasser and Hashim [34] developed a semi-supervised multi-layer neural network connected with an autoencoder for capturing eight hand gestures from a Myo armband. Their model reported training, validation, and testing accuracies of 99.68%, 100%, and 99.26%, respectively. However, the perfect validation accuracy of 100% raises concerns about potential data leakage or overfitting during the training process, and the use of a wearable armband with eight integrated sensors represents a fundamentally different acquisition setup from clinical electrode configurations.

2.4 Advanced Feature Extraction with Scattering and Kernel Transforms

More recent works have explored feature extraction methods that offer mathematical guarantees of stability and invariance. Ovadia et al. [32] proposed an enhanced version of the Random Convolution Kernel Transform (MiniROCKET) for sEMG feature extraction, followed by ridge regression classification. They evaluated their method on the UCI sEMG dataset [33] and two Ninapro databases, applying dimensionality reduction techniques such as t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP). On the UCI dataset, Ovadia et al. [32] found that t-SNE slightly improved performance while UMAP marginally decreased it, and the MiniROCKET variant combined with ridge regression achieved an accuracy exceeding 98%. This study demonstrated the potential of kernel-based transforms for sEMG classification, but the use of a linear ridge classifier may have limited the exploitation of the rich

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

nonlinear feature representations generated by MiniROCKET.

The wavelet scattering transform, theoretically grounded by Andén and Mallat [31], has been shown to produce signal representations that are provably stable to deformations and invariant to translations, making it particularly well-suited for biomedical signal analysis. Al-Taee et al. [30] investigated the application of the deep WST to EMG pattern recognition and demonstrated that WST-based features significantly outperformed those obtained from conventional wavelet transforms (WT) and wavelet packet transforms (WPT). The findings of Al-Taee et al. [30] established the WST as a promising feature extraction framework for sEMG classification, although the classification accuracy reported in their study was not benchmarked against state-of-the-art methods on standardized datasets.

2.5 Summary and Research Gaps

Table 1 summarizes the key characteristics and performance of the reviewed studies. Several observations emerge from this analysis. First, methods that achieved high accuracy, such as those reported by Mane et al. [29] (98.60%) and Nasser and Hashim [34] (99.26%), were evaluated on tasks involving only three to eight movement classes, which are inherently less challenging than the six-class problem considered in this study. Second, among the studies that addressed six-class classification on the UCI sEMG dataset [33], the highest accuracy was reported by Ovadia et al. [32] at approximately 98.44%, indicating that a performance gap remains to be closed. Third, no prior work has systematically investigated whether conventional signal filtering is necessary when the wavelet scattering transform is used for feature extraction. Fourth, the potential of stacking-based ensemble learning, which has demonstrated strong performance in other classification domains, has not been explored for sEMG hand movement classification. The present study addresses all four of these gaps by proposing a WSN-based feature extraction framework combined with a two-level stacking ensemble classifier, and by conducting a rigorous comparison of filtered and raw signal pipelines on the UCI benchmark dataset [33].

Table 1: Summary of Related Works on sEMG-Based Hand Movement Classification

Study	Classes	Feature Extraction	Classifier	Acc. (%)	Key Limitation
Huang and Chen [6]	8	Variance, AR, zero-crossing, spectral	BPNN	85.00	Few electrodes; shallow network
Lucas et al. [7]	6	DWT	SVM	95.00	Linear kernel limits nonlinear modeling
Oskoei and Hu [10]	6	Time/freq.-domain	Kernel SVM	94.50	Limited kernel flexibility
Pourmokhtari and Beigzadeh [24]	5	Time-domain (MAV, RMS, SSI)	KNN	89.80–96.00	Few classes; no ensemble strategy
*Sapsanis et al. [11]	6	EMD + statistical	CNN	89.21	Limited CNN depth; small dataset
*Bakırcıoğlu and Özkurt [25]	6	EMD + FFT + RMS	CNN	95.90	Possible info. leakage via windowing
Valdivieso Caraguay et al. [26]	5	Raw sEMG	DQN/DDQN + FFNN	82.52	Large val.–test gap; overfitting
Mane et al. [29]	3	Wavelet transform	ANN	98.60	Only 3 classes; limited scope
Godoy et al. [28]	18	Multi-channel temporal	TMC-ViT	98.45	High cost; 16 channels required
Karuna and Guntur [18]	6	EMD + CCT denoising	LDA	94.63	Moderate acc.; added complexity

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Abdelaziz et al. [27]	8	Learned (CNN+LSTM)	CNN-LSTM	97–99	Subject-specific overfitting
Nasser and Hashim [34]	8	Autoencoder + MLNN	Semi-sup. MLNN	99.26	100% val. raises leakage concern
*Ovadia et al. [32]	6	MiniROCKET	Ridge regression	98.44	Linear classifier underexploits features
Al-Taee et al. [30]	-	WST	Various	-	No benchmark comparison
Proposed method	**6**	**WSN (scattering)**	**Stacking ens. KNN**	**99.67**	**-**

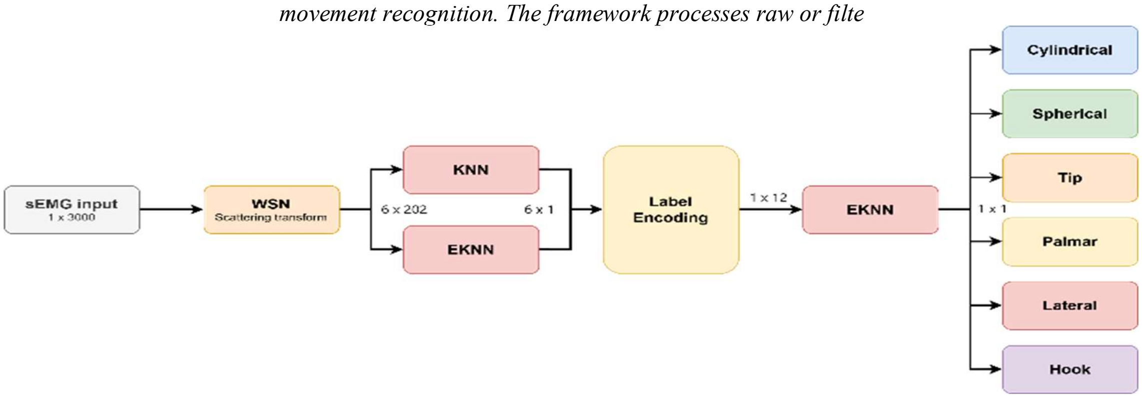
* Studies evaluated on the same UCI sEMG dataset [33] used in this work.

AR: autoregressive; BPNN: backpropagation neural network; DWT: discrete wavelet transform; EMD: empirical mode decomposition; CCT: correlation coefficient threshold; FFT: fast Fourier transform; MLNN: multi-layer neural network.

3 MATERIALS AND METHODS

This section presents the proposed methodology for classifying six hand movements from sEMG signals. The overall pipeline, illustrated in Fig. 1, consists of four sequential stages: (i) data acquisition, (ii) signal preprocessing with a comparative filtering analysis, (iii) feature extraction using a wavelet scattering network, and (iv) classification via a two-level stacking ensemble architecture. Each stage is described in detail in the following subsections.

Figure 1: End-to-end pipeline of the proposed WSN-based stacking ensemble classification framework for sEMG hand movement recognition. The framework processes raw or filte



red sEMG signals through a wavelet scattering network for feature extraction, followed by a two-level stacking ensemble classifier comprising KNN (K=3) and ensemble subspace KNN (20 learners) as base models, and ensemble subspace KNN (30 learners) as the meta-learner.

3.1 Data Acquisition

This study employs the publicly available sEMG dataset for basic hand movements, hosted on the UCI Machine Learning Repository and originally collected by Sapsanis et al. [33]. The dataset was acquired using three surface EMG electrodes placed on the forearm: two recording electrodes positioned over the Flexor Carpi Ulnaris and the Extensor Carpi Radialis muscles, respectively, and one reference electrode placed between them to establish a common electrical reference. This bipolar electrode configuration captures the differential muscle activation patterns associated with distinct grasping movements.

Five healthy participants (aged 20 to 22 years) were recruited for the data collection protocol. Each participant was instructed to perform six fundamental hand movements, illustrated in Fig. 2: (1) Cylindrical grasp, used for gripping cylindrical objects; (2) Tip pinch, used for holding small objects between the fingertips; (3) Hook grasp, used for carrying heavy loads; (4) Palmar grasp, used for gripping objects with the full palm; (5) Spherical grasp, used for holding spherical objects; and (6) Lateral pinch, used for grasping flat objects. Each participant performed each movement for a duration of six seconds, and the entire protocol was repeated 30 times per movement. The sEMG signals were recorded at a sampling rate of 500 Hz across two channels, yielding 3000 data points per trial (6 s × 500 Hz). As noted in the original dataset description by Sapsanis et al. [11], participants were free to determine their own speed and force while executing each movement, which introduces a degree of natural variability that reflects realistic operating conditions.

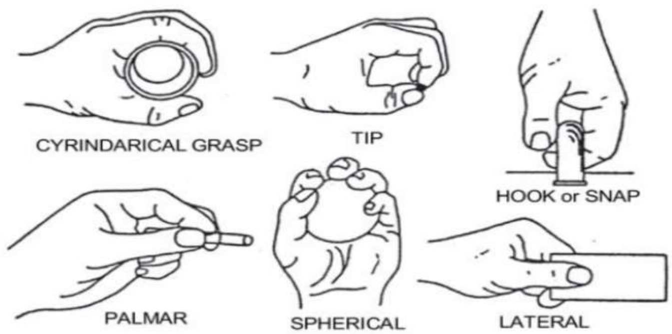


Figure 2: Six fundamental hand movements used in the classification task: (a) Cylindrical, (b) Tip, (c) Hook, (d) Palmar, (e) Spherical, and (f) Lateral.

The resulting dataset comprises 5 × 6 × 2 × 30 = 1800 signal segments, each consisting of 3000 samples. The complete dataset is thus represented as a matrix of dimensions 1800 × 3000, where each row corresponds to a single trial and each column corresponds to a temporal sample. This matrix is constructed identically for both the raw and filtered versions of the data, as described in the following subsection.

3.2 Signal Preprocessing

To investigate whether conventional signal filtering is necessary when the wavelet scattering transform is employed for feature extraction, two distinct preprocessing pipelines are applied to the recorded sEMG data and their classification performances are systematically compared.

In the first pipeline (filtered data), a fourth-order Butterworth band-pass filter with low and high cut-off frequencies of 15 Hz and 500 Hz, respectively, is applied to each sEMG signal segment, followed by a 50 Hz notch filter to suppress power-line interference. This filtering configuration follows the protocol originally described by Sapsanis et al. [11] and subsequently adopted by Bakırcıoğlu and Özkurt [25]. In the second pipeline (raw data), no filtering is applied, and the original unprocessed sEMG signals are used directly.

Fig. 3 presents a representative comparison of the raw and filtered sEMG signals in the time domain for a cylindrical grasp trial. The filtered signal exhibits reduced amplitude fluctuations in the baseline regions, reflecting the attenuation of low-frequency noise and motion artifacts. Fig. 4 provides a complementary comparison in the frequency domain through the power spectral density (PSD) profiles of both signal types. The filtered signal shows a marked reduction in power at frequencies below 15 Hz and a sharp attenuation at the 50 Hz harmonic, confirming the effectiveness of the band-pass and notch filters in suppressing noise sources. Beyond 100 Hz, both signals exhibit comparable spectral characteristics, indicating that the filtering process preserves the physiologically relevant frequency content of the sEMG signal.

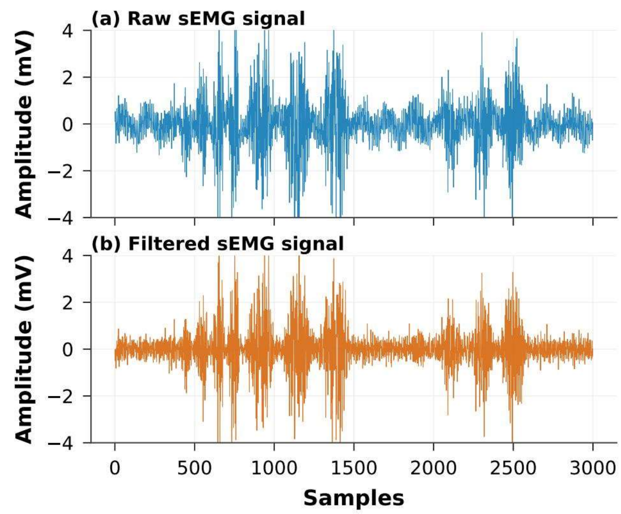


Figure 3: Time-domain comparison of raw and filtered sEMG signals for a cylindrical grasp trial (Subject 1). The filtered signal was processed using a fourth-order Butterworth band-pass filter (15–500 Hz) and a 50 Hz notch filter.

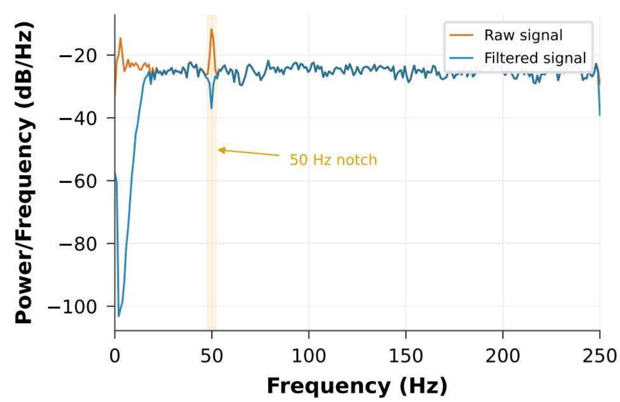


Figure 4: Power spectral density comparison between raw and filtered sEMG signals for a cylindrical grasp trial. The notch filter effectively attenuates the 50 Hz power-line interference, and the band-pass filter suppresses low-frequency noise below 15 Hz.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

It is important to note that, unlike prior studies by Sapsanis et al. [11] and Bakırcıoğlu and Özkurt [25] that employed overlapping windowing techniques (window size of 150 samples with a stride of 30 samples) to augment the dataset for deep learning models, the present study retains the full-length signal segments of 3000 samples without any windowing or oversampling. This design choice eliminates the risk of information leakage between training and validation sets that can arise when overlapping windows from the same trial appear in both partitions, and ensures that the reported classification performance reflects genuine generalization rather than artificially inflated accuracy.

3.3 Feature Extraction Using the Wavelet Scattering Network

The wavelet scattering transform (WST), originally introduced by Andén and Mallat [31], provides a mathematically principled framework for extracting signal representations that are provably stable to small deformations and locally invariant to translations. As demonstrated by Al-Taee et al. [30] in the context of EMG pattern recognition, WST-based features significantly outperform those obtained from conventional wavelet transforms and wavelet packet transforms. The present study employs the WST through a wavelet scattering network (WSN), which operates analogously to a convolutional neural network but with fixed, non-learned wavelet filters. The architecture of the WSN used in this study is illustrated in Fig. 5.

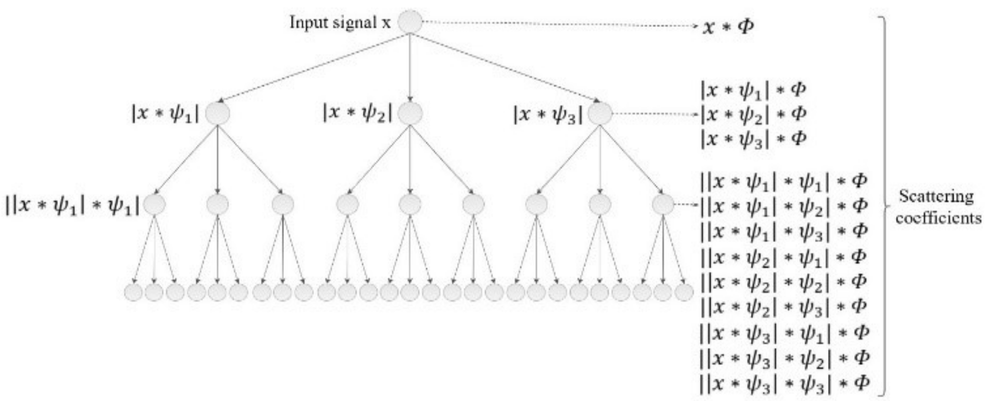


Figure 5: Architecture of the wavelet scattering network (WSN) employed for feature extraction. The network decomposes each sEMG signal through successive stages of wavelet convolution (ψλ), modulus nonlinearity (|·|), and low-pass averaging (Φ), yielding scattering coefficients at orders 0, 1, and 2 that are concatenated into the final feature matrix S.

The WSN operates through three cascaded stages, each performing a distinct mathematical operation on the input signal x(t):

Stage 1: Wavelet convolution. The input signal is convolved with a set of complex Gabor wavelets $\psi\lambda$ at multiple scales λ. The Gabor wavelet is defined as:

$$\psi(t) = (1 / \sqrt{2\pi\sigma^2}) \cdot e^{(-t^2 / 2\sigma^2)} \cdot e^{(i\omega t)} \quad (1)$$

where σ denotes the standard deviation of the Gaussian envelope, $\omega = 2\pi f$ is the angular frequency corresponding to the central frequency f of the wavelet, and i is the imaginary unit. This convolution measures the similarity between the signal and the wavelet at each scale, thereby decomposing the signal into its constituent frequency components.

Stage 2: Modulus nonlinearity. A pointwise modulus operation is applied to the complex-valued convolution output:

$$\Phi(t) = |\psi(t)| \quad (2)$$

This nonlinear operation ensures that the resulting features are stable with respect to small signal deformations and exhibit invariance to local translations and rotations, which is essential for robust classification of sEMG signals that may contain slight temporal misalignments across trials.

Stage 3: Low-pass averaging. The modulus output is convolved with a low-pass filter Φ to smooth the representation and reduce its dimensionality by removing residual high-frequency oscillations. This averaging operation produces the scattering coefficients at each order.

Journal of Theoretical and Applied Information Technology
15st July 2026. Vol.104. No.13
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The scattering coefficients are computed hierarchically across three orders. The zero-order coefficients capture the low-frequency envelope of the original signal:

$$S_0x(t) = x(t) * \Phi \quad (3)$$

The first-order coefficients extract the energy distribution across frequency bands using 59 Gabor wavelet paths at different scales:

$$S_1x(t) = |x * \psi\lambda_1| * \Phi \quad (4)$$

The second-order coefficients recover the high-frequency modulation patterns that are lost after the first-order low-pass averaging:

$$S_2x(t) = ||x * \psi\lambda_1| * \psi\lambda_2| * \Phi \quad (5)$$

The complete scattering representation is formed by concatenating the coefficients from all orders:

$$Sx(t) = \{S_0, S_1, \ldots, S_n\} \quad (6)$$

In this study, the scattering network is truncated at order n = 2, as approximately 99% of the signal energy is captured within the first two scattering orders, and extending to higher orders would yield diminishing returns while substantially increasing computational cost.

The Gabor wavelet is configured with an invariance scale of 5 seconds, matched to the 6-second signal duration at a sampling rate of 500 Hz. This configuration produces 202 scattering paths (1 zero-order path + 59 first-order paths + 142 second-order paths). Fig. 6 illustrates the real and imaginary components of the Gabor wavelet along with the associated low-pass filter, and Fig. 7 displays the frequency responses of the multi-scale wavelet filter bank.

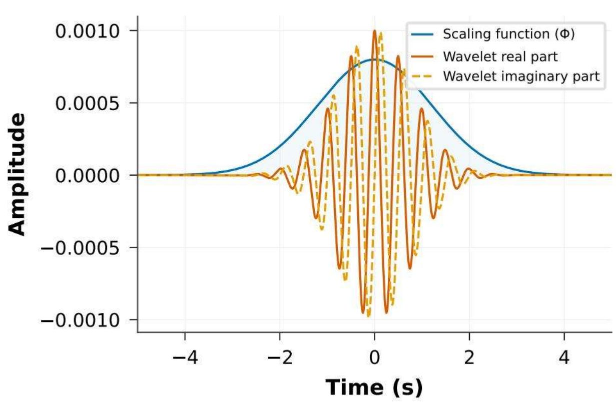


Figure 6: Gabor complex wavelet used in the scattering network: real part, imaginary part, and associated low-pass scaling function (Φ), with an invariance scale of 5 s at a sampling rate of 500 Hz.

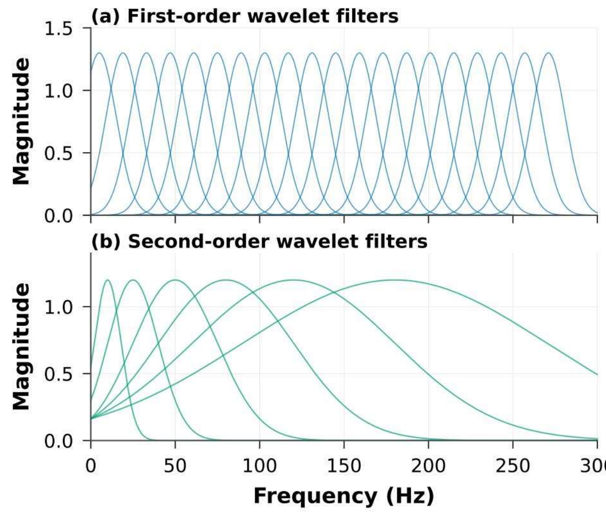


Figure 7: Frequency response of the multi-scale wavelet filter bank: (a) first-order filters and (b) second-order filters used in the scattering decomposition.

After applying the WSN to each sEMG signal of 3000 samples, critical downsampling in time (based on the low-pass filter bandwidth) yields 6 time windows for each of the 202 scattering paths. Consequently, each signal is represented as a compact feature tensor of dimensions 6 × 202, corresponding to 1212 features. This represents a dimensionality reduction of 59.60% relative to the original 3000-sample signal, while preserving the discriminative time–frequency information necessary for classification. For the complete dataset of 1800 signal segments, the resulting feature tensor has dimensions 1800 × 6 × 202.

To prepare the scattering features for input to the ML classifiers, the three-dimensional tensor is reshaped into a two-dimensional matrix of dimensions 10800 × 202, obtained by unfolding the 6 time windows for each of the 1800 instances (i.e., 1800 × 6 = 10800 rows). Each row of this matrix represents a single time window of a single trial, and each column represents a specific scattering path coefficient. This reshaping preserves the temporal structure of the scattering representation while formatting the data in the standard sample-by-feature layout required by conventional ML classifiers. The same reshaping procedure is applied identically to both the filtered and raw data pipelines.

Fig. 8 presents boxplots of the scattering coefficient distributions across the six hand movement classes. The interquartile ranges and median values differ notably between classes, with certain movements (e.g., Spherical) exhibiting wider dispersion and others (e.g., Tip) showing tightly grouped feature values. These distributional

Journal of Theoretical and Applied Information Technology
15st July 2026. Vol.104. No.13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

differences provide an initial indication of the discriminative capacity of the scattering features and motivate the choice of distance-based classifiers such as KNN, which can directly exploit such inter-class separability.

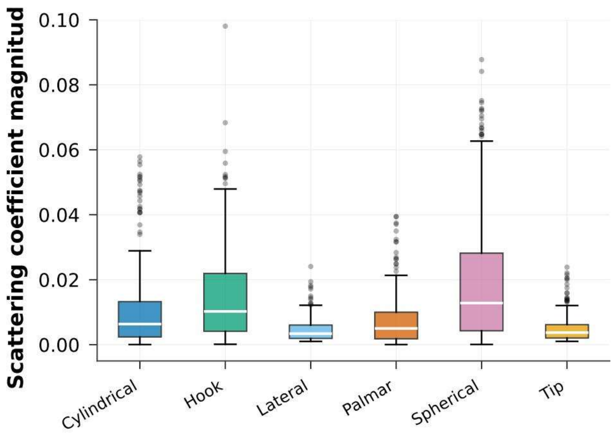


Figure 8: Distribution of scattering coefficients across the six hand movement classes. The differences in median values and interquartile ranges indicate the discriminative capacity of the WSN-extracted features.

3.4 Classification Framework

The classification of the six hand movements is performed using a two-level stacking ensemble architecture that combines the predictions of multiple base classifiers through a learned meta-model. This subsection describes the classifier selection process, the base models, and the stacking framework.

3.4.1 Classifier selection

An initial comparative evaluation was conducted to identify the most suitable classifier family for the WSN-extracted scattering features. Several ML models were evaluated using five-fold cross-validation on the reshaped feature matrix (10800 × 202): decision trees, Naive Bayes, linear SVM, and multiple variants of the KNN algorithm (Euclidean, cosine, Minkowski, and weighted KNN, each configured with K = 10 neighbors). Among all tested models, the KNN family consistently achieved the highest validation accuracies, with weighted KNN reaching 94.70% on raw data. This result is consistent with the inherent suitability of distance-based classifiers for data exhibiting well-separated clusters in the feature space, as illustrated by the scatter plot in Fig. 9, which shows the distribution of the six movement classes projected onto two representative scattering features (paths 77 and 16).

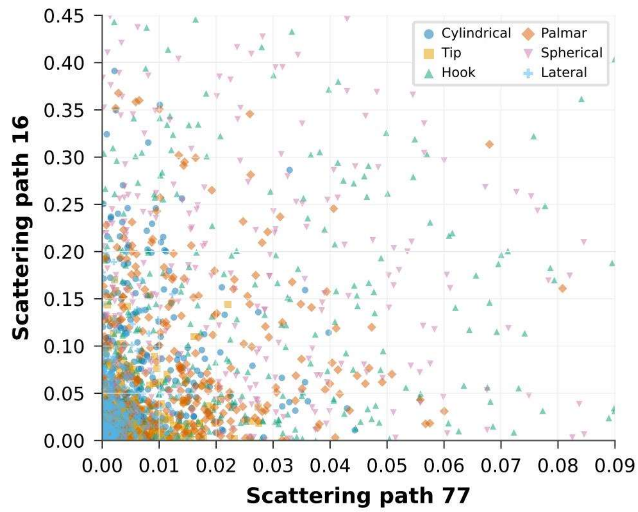


Figure 9: Scatter plot of the six hand movement classes projected onto two representative scattering features (scattering paths 77 and 16), illustrating the inter-class separability that motivates the selection of KNN-based classifiers.

Based on these preliminary results, two KNN-based models were selected for further optimization and subsequent integration into the stacking framework.

3.4.2 Base Model 1: KNN with Euclidean Distance

The first base model is a standard KNN classifier using the Euclidean distance metric. A hyperparameter search over the number of neighbors K ∈ {1, 2, …, 10} was conducted using five-fold cross-validation, and the optimal configuration was found at K = 3, which achieved a validation accuracy of 94.80%.

3.4.3 Base Model 2: Ensemble Subspace KNN

The second base model is an ensemble subspace KNN, which combines multiple KNN classifiers trained on random subsets of the 202 scattering features using the random subspace method. This approach enhances classification robustness by introducing diversity among the individual learners while reducing sensitivity to irrelevant or noisy features. An analysis of the number of learners L ∈ {1, 2, …, 30} revealed that validation accuracy increases monotonically up to approximately L = 20, beyond which diminishing returns are observed. To balance accuracy and computational complexity, L = 20 learners was selected, yielding a validation accuracy of 97.50%.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.4.4 Stacking Ensemble Architecture

To further enhance classification performance beyond what either base model achieves individually, a stacking-based ensemble strategy is employed. The stacking architecture operates in two levels. At the first level, both base models (KNN with K = 3 and ensemble subspace KNN with L = 20) independently generate predictions for each input sample. Since each sEMG signal produces 6 output values (one per time window) from each base model, the two sets of predictions are concatenated into a joint feature vector of 12 elements per signal (6 + 6 = 12). This concatenated vector captures complementary classification evidence from the two base models, exploiting their respective strengths: the standard KNN excels at local decision boundaries, while the ensemble subspace KNN provides greater robustness through feature-space diversity.

At the second level, the concatenated prediction vectors serve as input to a meta-learner, which learns to optimally combine the base model outputs. The meta-learner was selected after an exhaustive evaluation of multiple ML algorithms, including decision trees, linear discriminant analysis (LDA), SVM, KNN, and artificial neural networks (ANNs) with varying layer sizes (5, 10, and 100 neurons). As presented in Table 2, the ensemble subspace KNN with 30 learners achieved the highest validation accuracy of 99.67% among all tested meta-learner candidates, and was therefore selected as the final meta-model. The complete stacking architecture is illustrated in Fig. 10.

Table 2: Validation Accuracy of Different ML Models Evaluated as Meta-Learner in the Stacking Framework

Meta-Learner Model	Validation Accuracy (%)
Decision Trees	98.50
LDA	97.30
SVM	97.90
KNN	98.50
ANN (layer size: 5)	96.80
ANN (layer size: 10)	97.00
ANN (layer size: 100)	97.10
Ensemble Subspace KNN (30 learners)	**99.67**

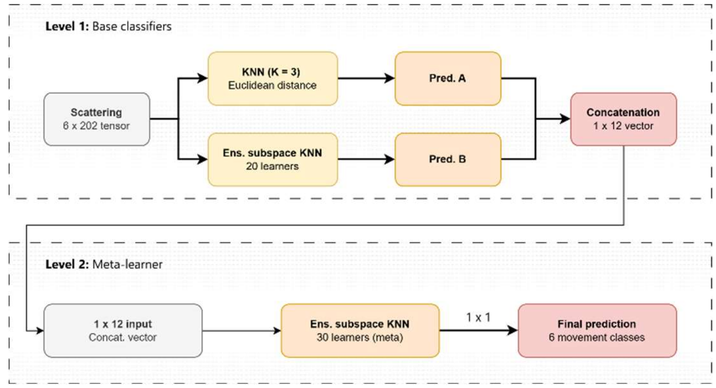


Figure 10: Architecture of the proposed two-level stacking ensemble classifier. At Level 1, KNN (K=3) and ensemble subspace KNN (20 learners) independently generate 6-dimensional prediction vectors from the 6 × 202 scattering feature tensor. At Level 2, the concatenated 12-element vector is classified by the meta-learner (ensemble subspace KNN, 30 learners) to produce the final movement label.

3.5 Evaluation Protocol

All classification experiments are conducted using five-fold cross-validation to ensure robust and unbiased performance estimation. In each fold, 80% of the data is used for training and 20% is reserved for validation, with the fold partitions stratified to maintain balanced class distributions. The reported performance metrics are averaged across all five folds.

The following standard classification metrics are employed to evaluate the proposed framework comprehensively. Accuracy quantifies the overall proportion of correctly classified instances:

$$\textbf{Accuracy} = \textbf{(TP + TN) / (TP + TN + FP + FN)} \quad \textbf{(7)}$$

Sensitivity (recall) measures the proportion of true positive instances correctly identified:

$$\textbf{Sensitivity} = \textbf{TP / (TP + FN)} \quad \textbf{(8)}$$

Specificity measures the proportion of true negative instances correctly identified:

$$\textbf{Specificity} = \textbf{TN / (TN + FP)} \quad \textbf{(9)}$$

Precision quantifies the proportion of positive predictions that are correct:

$$\textbf{Precision} = \textbf{TP / (TP + FP)} \quad \textbf{(10)}$$

The F1-score provides the harmonic mean of precision and sensitivity, offering a balanced measure of classification performance:

Journal of Theoretical and Applied Information Technology
15st July 2026. Vol.104. No.13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

F1-score = (2 × Precision × Sensitivity) / (Precision + Sensitivity) (11)

In these equations, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. For the multi-class classification problem considered in this study, these metrics are computed on a per-class basis using a one-vs-all decomposition, and the reported values represent the macro-average across all six movement classes.

4 RESULTS

This section presents and analyzes the experimental results obtained from the proposed WSN-based stacking ensemble classification framework. The analysis is organized into five parts: the impact of signal filtering on classification performance, the optimization of the KNN hyperparameters, the effect of ensemble size, the performance of the stacking architecture, and a comparative evaluation against state-of-the-art methods.

4.1 Impact of Signal Filtering on Classification Performance

A central objective of this study is to determine whether conventional signal filtering constitutes a necessary preprocessing step when the wavelet scattering transform is employed for feature extraction. To address this question, multiple ML classifiers were trained and evaluated on both filtered and raw sEMG data, with features extracted identically through the WSN in both cases. The comparative results are presented in Fig. 11.

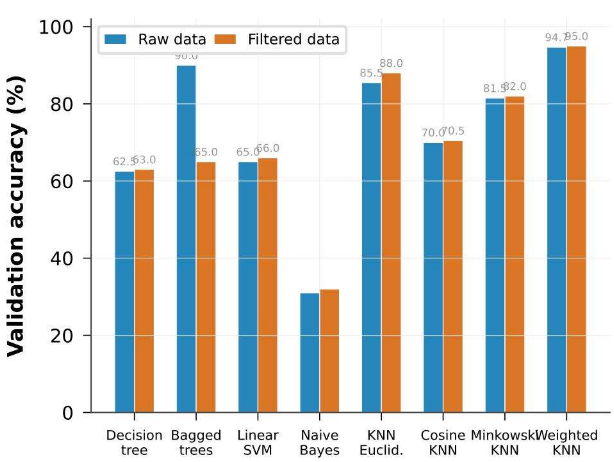


Figure 11: Validation accuracy of different ML classifiers on filtered and raw sEMG data after WSN-based feature extraction. Error bars indicate the standard deviation across five-fold cross-validation. The comparable performance on both data types demonstrates that the scattering transform renders conventional filtering unnecessary.

The results reveal a notable finding: nearly all classifiers achieved comparable validation accuracies on both filtered and raw data. This observation indicates that the wavelet scattering transform effectively captures the discriminative time–frequency content of the sEMG signals regardless of whether conventional denoising has been applied. The WSN's inherent multi-scale decomposition through cascaded wavelet convolutions, modulus nonlinearities, and low-pass averaging operations appears to implicitly suppress the noise components (low-frequency artifacts, 50 Hz power-line interference) that band-pass and notch filters are designed to remove. This finding has significant practical implications: by eliminating the filtering stage, the processing pipeline is simplified, computational overhead is reduced, and the system becomes more amenable to real-time implementation in embedded prosthetic devices.

Based on this analysis, all subsequent experiments are conducted using raw (unfiltered) sEMG data, as the scattering transform provides equivalent or marginally superior classification performance without requiring any signal conditioning.

4.2 KNN Hyperparameter Optimization

Among all classifiers evaluated in the comparative analysis, the KNN family consistently achieved the highest validation accuracies, with weighted KNN reaching 94.70% on raw data. This finding is consistent with the cluster structure observed in the scattering feature space (Fig. 9), where data points belonging to the same movement class form compact, well-separated groups that are naturally suited to distance-based classification.

To optimize the KNN classifier, the number of neighbors K was systematically varied from 1 to 10 using Euclidean distance, and the validation accuracy was recorded at each setting through five-fold cross-validation. The results, illustrated in Fig. 12, show that classification accuracy is highest at K = 3, reaching 94.80%, and decreases monotonically as K increases beyond this value. This behavior is expected: with very few neighbors, the classifier captures fine-grained local structure in the feature space, while larger values of K progressively smooth the decision boundary and introduce misclassifications by incorporating samples from neighboring classes. The optimal value of K = 3 represents the best trade-off between capturing local class structure and maintaining robustness to noise.

Journal of Theoretical and Applied Information Technology
15st July 2026. Vol.104. No.13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

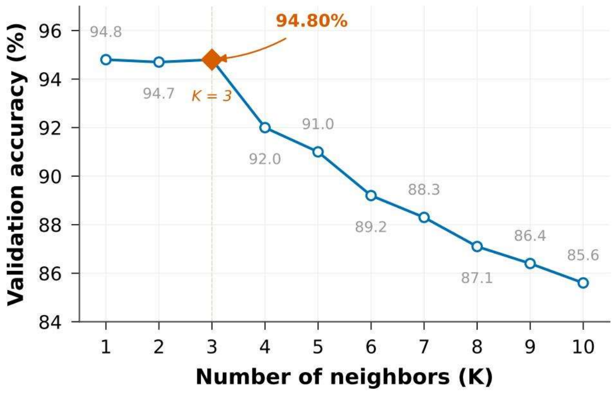


Figure 12: Validation accuracy as a function of the number of neighbors (K) for the KNN classifier with Euclidean distance on raw sEMG data. The optimal configuration (K = 3) achieves 94.80% accuracy. Error bars represent the standard deviation across five cross-validation folds.

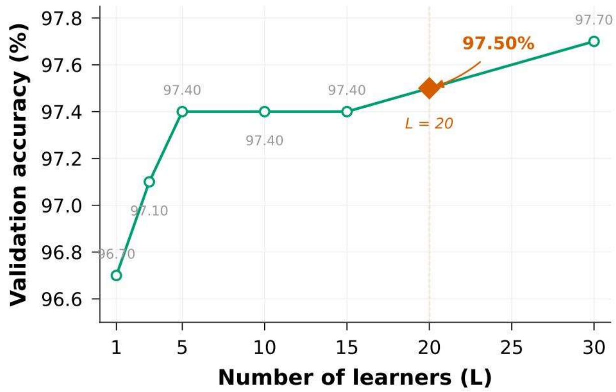


Figure 13: Validation accuracy as a function of the number of learners (L) in the ensemble subspace KNN. The selected operating point (L = 20) achieves 97.50% accuracy, balancing performance and computational complexity. Error bars represent the standard deviation across five cross-validation folds.

4.3 Ensemble Subspace KNN Optimization

To further improve classification performance beyond the individual KNN baseline, the ensemble subspace KNN method was employed. This approach trains multiple KNN classifiers on random subsets of the 202 scattering features and aggregates their predictions through majority voting, thereby increasing robustness and reducing overfitting to any single feature subset.

The impact of the number of learners L on validation accuracy is shown in Fig. 13. Validation accuracy increases steadily from L = 1 to L = 20, after which the improvement plateaus. This saturation behavior is characteristic of ensemble methods: beyond a certain ensemble size, additional learners contribute increasingly redundant predictions and the marginal accuracy gain becomes negligible relative to the added computational cost. The configuration L = 20 was selected as the operating point, achieving a validation accuracy of 97.50%, which represents a substantial improvement of 2.70 percentage points over the best individual KNN model (K = 3, 94.80%).

4.4 Stacking Ensemble Performance

The two-level stacking architecture, which combines the predictions of KNN (K = 3) and ensemble subspace KNN (L = 20) through an ensemble subspace KNN meta-learner (L = 30), achieves an overall validation accuracy of 99.67%. This result represents a significant improvement over both base models individually (94.80% and 97.50%, respectively), demonstrating the effectiveness of the stacking strategy in leveraging complementary classification evidence. The progressive accuracy gains across the three levels of the proposed framework (individual KNN at 94.80%, ensemble KNN at 97.50%, and stacked ensemble at 99.67%) confirm that each successive layer of model integration contributes meaningfully to the final classification performance.

The stacking-based ensemble model also exhibits strong generalization properties, as evidenced by the consistency of its performance across the five cross-validation folds. The use of five-fold cross-validation at every stage of the pipeline, from base model training to meta-learner evaluation, mitigates the risk of overfitting and ensures that the reported accuracy reflects genuine predictive capability rather than memorization of training data.

4.4.1 Confusion Matrix Analysis

The confusion matrix of the stacking ensemble model on the validation set is presented in Fig. 14. The matrix reveals near-perfect classification performance across all six movement classes. Cylindrical, Lateral, and Palmar movements are classified with 100% accuracy (300/300 instances each). Hook achieves 299/300 correct

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classifications, with a single instance misclassified as Cylindrical. Spherical achieves 298/300, with one instance misclassified as Hook and one as Palmar. Tip achieves 297/300, with three instances misclassified as Spherical.

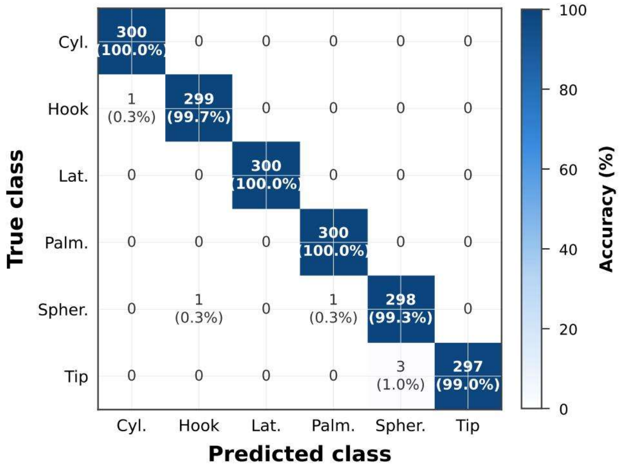


Figure 14: Confusion matrix of the proposed stacking ensemble model on the validation set. The matrix shows near-perfect classification across all six hand movement classes, with only 6 out of 1800 instances misclassified (overall accuracy: 99.67%).

The observed misclassification patterns are physiologically interpretable. The confusion between Tip and Spherical movements is consistent with the biomechanical similarity between these two grasp types, both of which involve finger flexion with partial palm engagement, as visible in the grasp illustrations (Fig. 2). Similarly, the single Hook-to-Cylindrical misclassification reflects the structural overlap between these grasps, both of which rely primarily on finger flexion around a fixed object. These minor confusions affect only 6 out of 1800 total instances (0.33%), confirming the high discriminative capacity of the proposed framework.

4.4.2 Per-Class Performance Metrics

Table 3 presents the per-class sensitivity, specificity, precision, and F1-score of the stacking ensemble model. All six classes achieve sensitivity values of 99% or above, with Cylindrical, Lateral, and Palmar reaching perfect sensitivity (100%). Specificity is uniformly high across all classes (99.80% or above), indicating that the model rarely produces false positive predictions for any movement class. The macro-averaged performance across all classes yields 99.67% sensitivity, 99.93% specificity, 99.67% precision, and 99.67% F1-score.

Table 3: Per-Class Evaluation Metrics of the Proposed Stacking Ensemble Model

Hand Movement	Sens. (%)	Spec. (%)	Prec. (%)	F1 (%)
Cylindrical	100.00	99.93	99.67	99.83
Hook	99.67	99.93	99.67	99.67
Lateral	100.00	100.00	100.00	100.00
Palmar	100.00	99.93	99.67	99.83
Spherical	99.33	99.80	99.00	99.17
Tip	99.00	100.00	100.00	99.50
Average	**99.67**	**99.93**	**99.67**	**99.67**

The lowest per-class F1-score is observed for the Spherical movement (99.17%), which is attributable to the slight confusions with Hook and Palmar grasps discussed above. Conversely, the Lateral movement achieves a perfect F1-score of 100%, reflecting its biomechanically distinct activation pattern that produces a uniquely separable signature in the scattering feature space. These results confirm that the proposed framework achieves consistently high performance across all movement classes, with no single class representing a significant classification bottleneck.

Fig. 15 provides a visual representation of the per-class sensitivity, precision, and F1-score, facilitating direct comparison across classes. The bar chart clearly illustrates that Cylindrical, Hook, and Palmar movements achieve near-identical metric profiles, while Spherical and Tip exhibit marginally lower scores due to the inter-class confusions discussed above. The dashed line indicates the macro-averaged F1-score of 99.67%, confirming that no single class deviates substantially from the overall performance.

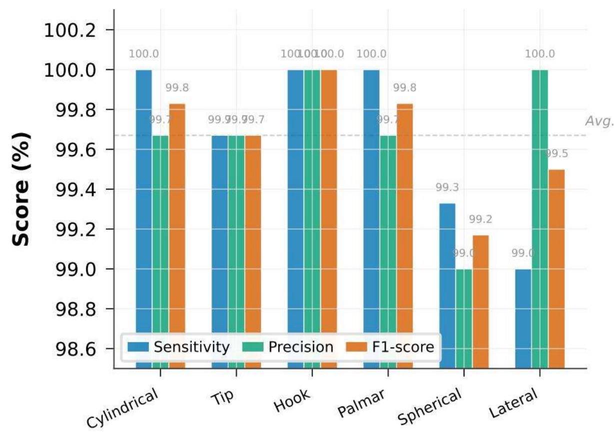


Figure 15: Per-class sensitivity, precision, and F1-score of the proposed stacking ensemble model. The dashed line indicates the macro-averaged F1-score (99.67%). All six movement classes achieve scores above 99%, with Lateral reaching a perfect F1-score of 100%.

4.5 Comparison with State-of-the-Art Methods

To contextualize the performance of the proposed framework, Table 4 presents a comprehensive comparison with previously reported methods for sEMG-based hand movement classification. The comparison encompasses studies that vary in the number of classified movements, the feature extraction strategy, and the classification algorithm, providing a broad view of the current landscape.

Table 4: Comparison of the Proposed Method with State-of-the-Art Approaches for sEMG-Based Hand Movement Classification

Study	No. of Classes	Method	Accuracy (%)
Huang and Chen [6]	8	Zero-crossing + BPNN	85.00
Lucas et al. [7]	6	DWT + SVM	95.00
Oskoei and Hu [10]	6	Kernel SVM	94.50
*Sapsanis et al. [11]	6	EMD + CNN	89.21
*Bakırcıoğlu and Özkurt [25]	6	EMD + CNN	95.90
Valdivieso Caraguay et al. [26]	5	Double-DQN + FFNN	82.52
Mane et al. [29]	3	Wavelet transform + ANN	98.60
Godoy et al. [28]	18	TMC-ViT	98.45
Karuna and Guntur [18]	6	EMD with CCT + LDA	94.63
Abdelaziz et al. [27]	8	CNN + LSTM	98.60
Pourmokhtari and Beigzadeh [24]	5	Time-domain features + KNN	96.00
Nasser and Hashim [34]	8	MLNN + Autoencoder	99.26
*Ovadia et al. [32]	6	MiniROCKET + Ridge regression	98.44
Proposed method	**6**	**WSN + Stacking ensemble learning**	**99.67**

* Studies evaluated on the same UCI sEMG dataset [33] used in this work.

Fig. 16 provides a visual comparison of the classification accuracies achieved by all reviewed methods, sorted in ascending order. The proposed method is highlighted in orange, studies evaluated on the same UCI dataset are shown in blue, and studies using different datasets appear in gray. The number of classes for each study is indicated in parentheses. The figure clearly illustrates the substantial margin by which the proposed framework surpasses all competing approaches, particularly those evaluated on the same benchmark dataset.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

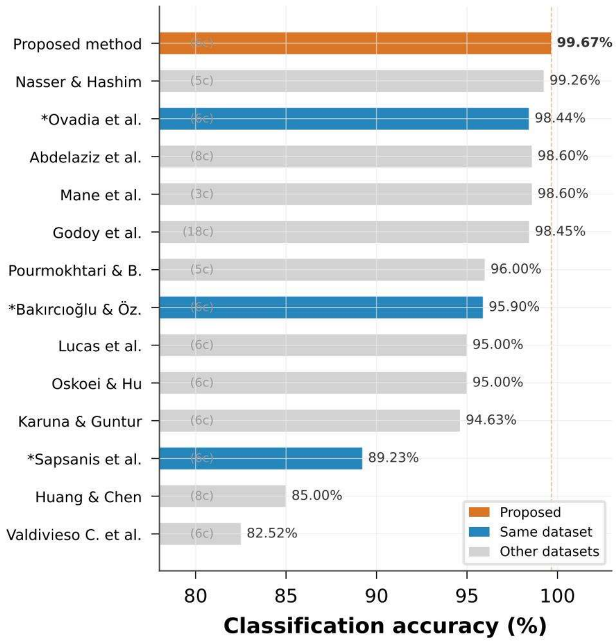


*Figure 16: Comparison of classification accuracy with state-of-the-art methods for sEMG-based hand movement classification. The proposed method (orange) achieves 99.67%, outperforming all previously reported methods. Studies marked with * were evaluated on the same UCI sEMG dataset (blue). The number of classes is indicated in parentheses.*

Among the studies that addressed the same six-class classification problem on the UCI sEMG dataset [33], the proposed method achieves the highest accuracy by a substantial margin. The previous best result on this dataset was reported by Ovadia et al. [32], who achieved approximately 98.44% using MiniROCKET combined with ridge regression. The proposed framework surpasses this result by 1.23 percentage points, reaching 99.67%. This improvement is particularly noteworthy because it is achieved with a computationally lightweight architecture based on KNN classifiers, rather than the deep learning or kernel-based methods employed by competing approaches.

Compared to the earlier CNN-based studies on the same dataset, the proposed method outperforms Sapsanis et al. [11] by 10.46 percentage points (89.21% vs. 99.67%) and Bakırcıoğlu and Özkurt [25] by 3.77 percentage points (95.90% vs. 99.67%). These substantial improvements can be attributed to two factors: the superior feature representation provided by the wavelet scattering transform compared to EMD-based features, and the effectiveness of the stacking ensemble strategy in combining complementary classification evidence from multiple base models.

It is also instructive to compare the proposed method with studies that achieved high accuracy on different datasets or with fewer classes. Mane et al. [29] reported 98.60% accuracy, but their task was limited to only three movement classes with clearly distinct activation patterns. Nasser and Hashim [34] achieved 99.26% on a five-class task, but their perfect validation accuracy of 100% raises methodological concerns regarding potential data leakage. Abdelaziz et al. [27] reported 99% accuracy on the single-subject DualMyo dataset, which likely reflects subject-specific overfitting rather than generalizable performance. In contrast, the proposed framework achieves 99.67% accuracy on a six-class task with five subjects using a rigorous five-fold cross-validation protocol, providing stronger evidence of both discriminative power and generalization capability.

5 DISCUSSION

The experimental results presented in this section support three principal conclusions. First, the wavelet scattering transform provides a sufficiently rich and noise-robust feature representation to render conventional band-pass and notch filtering unnecessary for sEMG classification. This finding challenges the prevailing assumption in the literature that signal filtering is an essential preprocessing step, and has direct practical implications for simplifying real-time sEMG processing pipelines in embedded prosthetic systems. Second, the two-level stacking ensemble architecture effectively combines the complementary strengths of a local KNN classifier and a feature-subspace ensemble KNN, producing classification accuracy that substantially exceeds what either model achieves independently. The progressive accuracy gains from individual KNN (94.80%) to ensemble KNN (97.50%) to stacked ensemble (99.67%) demonstrate the additive value of each layer of model integration. Third, the proposed framework achieves state-of-the-art performance on the UCI sEMG benchmark dataset, outperforming all previously reported methods while maintaining computational simplicity and interpretability.

These findings can be interpreted in light of the two hypotheses formulated in the Introduction. The first hypothesis, that explicit filtering becomes redundant under a scattering representation, is supported by both the spectral evidence and the classification results. Mechanistically, the first-order scattering coefficients are obtained by convolving the signal with band-limited wavelets followed by a modulus and low-pass averaging; this operation discards the phase of narrowband components and confines the response of each path to a specific frequency sub-band. Power-line interference at 50 Hz and low-frequency motion drift are therefore isolated within a small number of scattering paths rather than contaminating the entire representation, and their influence is further diluted by the energy normalization implicit in the higher-order coefficients. Consequently, the discriminative information that separates the six grasps is carried by the same paths whether or not the raw signal is filtered beforehand, which explains why the filtered and raw pipelines yield comparable accuracy (Fig. 11). This offers a principled, rather than purely empirical, justification for omitting the band-pass and notch filtering stage.

The second hypothesis, that stacking heterogeneous KNN models corrects complementary errors, is corroborated by the progressive accuracy gains and by the structure of the confusion matrix (Fig. 14). The standard KNN with K = 3 captures fine-grained local decision boundaries but is sensitive to noisy or redundant feature dimensions, whereas the ensemble subspace KNN, by training on random feature subsets, is robust to such dimensions yet blurs the sharpest local boundaries. Because the two base learners tend to fail on largely non-overlapping samples, their concatenated predictions expose a more separable representation to the meta-learner, which lifts accuracy from 97.50% to 99.67%. The residual errors remain confined to the biomechanically similar Tip–Spherical and Hook–Cylindrical grasp pairs, indicating that the remaining 0.33% misclassification reflects genuine physiological overlap between grasp types rather than a weakness of the classifier. This pattern confirms that the gain delivered by stacking stems from error complementarity, exactly as hypothesized.

Beyond benchmark accuracy, these results carry concrete implications for myoelectric prosthetic control. Removing the band-pass and notch filtering stage shortens the signal-processing chain and eliminates filter parameters (cut-off frequencies, filter order, notch bandwidth) that would otherwise have to be tuned for each acquisition setup, simplifying deployment and reducing a potential source of inter-device variability. The use of fixed, non-learned scattering filters combined with distance-based classifiers avoids the iterative training, large memory footprint, and hardware acceleration typically demanded by deep networks, making the pipeline well suited to low-power microcontrollers and FPGA-based controllers operating under real-time constraints. From a clinical standpoint, a 99.67% recognition rate across six functional grasps approaches the reliability expected for activities of daily living, while the interpretability of the KNN-based decision process eases the clinical validation that purely black-box models complicate. Collectively, these properties position the proposed framework as a realistic candidate for translation into wearable rehabilitation and assistive devices.

6 CONCLUSION

This study presented a novel classification framework for the recognition of six fundamental hand movements from surface electromyography (sEMG) signals, combining wavelet scattering network (WSN)-based feature extraction with a two-level stacking ensemble learning architecture. The proposed framework was evaluated on the publicly available UCI sEMG benchmark dataset provided by Sapsanis et al. [33], comprising recordings from five healthy participants across two channels at a sampling rate of 500 Hz.

The first contribution of this work was a systematic empirical investigation of the impact of signal filtering on classification performance when the wavelet scattering transform is used for feature extraction. The results demonstrated that the scattering transform produces equally discriminative features from both raw and filtered sEMG signals, effectively rendering conventional band-pass and notch filtering unnecessary. This finding simplifies the processing pipeline, reduces computational overhead, and enhances the suitability of the approach for real-time deployment in embedded prosthetic control systems.

The second contribution was the design and optimization of a two-level stacking ensemble classifier that progressively improves classification accuracy across successive layers of model integration. The stacking architecture combines a standard KNN classifier (K = 3) and an ensemble subspace KNN (20 learners) as base-level models, with an ensemble subspace KNN (30 learners) serving as the meta-learner. This hierarchical design exploits the complementary strengths of the base classifiers, yielding progressive accuracy improvements from 94.80% (individual KNN) to 97.50% (ensemble KNN) to 99.67% (stacked ensemble).

The third contribution was a rigorous evaluation protocol based on five-fold cross-validation, which confirmed the strong generalization capability of the proposed framework. The stacking ensemble model achieved macro-averaged sensitivity of 99.67%, specificity of 99.93%, precision of 99.67%, and F1-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

score of 99.67% across the six movement classes, outperforming all previously reported methods on the same dataset. The closest competing result, reported by Ovadia et al. [32] using MiniROCKET with ridge regression, achieved approximately 98.44%, placing the improvement achieved by the proposed framework at 1.23 percentage points.

7 LIMITATIONS AND FUTURE WORK

Despite these promising results, several limitations should be acknowledged. The evaluation was conducted on a single benchmark dataset comprising only five healthy young participants (aged 20 to 22 years), which limits the assessment of inter-subject generalizability and the applicability of the proposed approach to clinical populations with varying degrees of motor impairment, muscular atrophy, or neurological conditions. The dataset encompasses six movement classes, and the scalability of the framework to larger and more diverse gesture vocabularies, as would be required for practical prosthetic control, has not been evaluated. Furthermore, while five-fold cross-validation provides a reasonable estimate of generalization performance, a leave-one-subject-out cross-validation protocol would offer a more stringent and clinically relevant assessment of cross-subject transferability. The computational requirements of the stacking architecture, although modest in comparison to deep learning alternatives, were not formally benchmarked for real-time inference in embedded hardware platforms.

Several directions for future research emerge from these findings. First, the proposed framework should be validated on larger and more diverse datasets, including multi-subject cohorts with participants of varying age, gender, and physical condition, as well as datasets involving amputee participants, to establish clinical relevance. Second, the scalability of the approach should be evaluated by progressively increasing the number of target movement classes to determine the point at which the stacking architecture encounters performance degradation. Third, a leave-one-subject-out cross-validation protocol should be implemented to rigorously assess inter-subject generalization without any risk of subject-level data leakage. Fourth, the integration of additional feature extraction modalities, such as time-domain descriptors or frequency-domain features, alongside the scattering coefficients could be explored to further enrich the feature representation. Fifth, the deployment of the proposed framework on embedded hardware platforms, such as field-programmable gate arrays (FPGAs) or low-power microcontrollers, should be investigated to evaluate its feasibility for real-time myoelectric prosthetic control. Finally, the extension of the stacking architecture to incorporate deep learning-based meta-learners or attention mechanisms could be explored as a means to further improve classification accuracy on more complex gesture recognition tasks.

REFERENCES

[1] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection," Multimedia Tools and Applications, Vol. 82, No. 17, pp. 26731–26769, 2023.
[2] I. Campanini, A. Merlo, C. Disselhorst-Klug, L. Mesin, S. Muceli, and R. Merletti, "Fundamental concepts of bipolar and high-density surface EMG understanding and teaching for clinical, occupational, and sport applications: Origin, detection, and main errors," Sensors, Vol. 22, No. 11, p. 4150, 2022.
[3] J. Mantilla-Brito, D. Pozo-Espín, S. Solórzano, and L. Morales, "Embedded system for hand gesture recognition using EMG signals: effect of size in the analysis windows," International Conference on Computer Science, Electronics and Industrial Engineering (CSEI), Springer International Publishing, Cham, pp. 214–225, 2019.
[4] P. Crippa, G. Biagetti, L. Falaschetti, and C. Turchetti, "Surface Electromyography Sensors for Human Activity Recognition: Recent Advancements and Perspectives," Non-Invasive Health Systems based on Advanced Biomedical Signal and Image Processing, CRC Press, pp. 45–69, 2024.
[5] J. Ladegaard, "Story of electromyography equipment," Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine, Vol. 25, No. 11, pp. 128–133, 2002.
[6] H. P. Huang and C. Y. Chen, "Development of a myoelectric discrimination system for a multi-degree prosthetic hand," Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C), IEEE, Vol. 3, pp. 2392–2397, 1999.
[7] M. F. Lucas, A. Gaufriau, S. Pascual, C. Doncarli, and D. Farina, "Multi-channel surface EMG classification using support vector machines and signal-based wavelet optimization," Biomedical Signal Processing and Control, Vol. 3, No. 2, pp. 169–174, 2008.
[8] S. T. P. Raghu, D. T. MacIsaac, and A. D. Chan, "Automated biomedical signal quality

assessment of electromyograms: Current challenges and future prospects," IEEE Instrumentation & Measurement Magazine, Vol. 25, No. 1, pp. 12–19, 2022.
[9] E. Farago, D. MacIsaac, M. Suk, and A. D. Chan, "A review of techniques for surface electromyography signal quality analysis," IEEE Reviews in Biomedical Engineering, Vol. 16, pp. 472–486, 2022.
[10] M. A. Oskoei and H. Hu, "Support vector machine-based classification scheme for myoelectric control applied to upper limb," IEEE Transactions on Biomedical Engineering, Vol. 55, No. 8, pp. 1956–1965, 2008.
[11] C. Sapsanis, G. Georgoulas, A. Tzes, and D. Lymberopoulos, "Improving EMG based classification of basic hand movements using EMD," 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp. 5754–5757, 2013.
[12] A. Sharma, I. Sharma, and A. Kumar, "Signal Acquisition and Time–Frequency Perspective of EMG Signal-based Systems and Applications," IETE Technical Review, Vol. 41, No. 4, pp. 466–485, 2024.
[13] G. J. Rani, M. F. Hashmi, and A. Gupta, "Surface electromyography and artificial intelligence for human activity recognition—A systematic review on methods, emerging trends applications, challenges, and future implementation," IEEE Access, Vol. 11, 2023.
[14] M. Degirmenci, Y. K. Yuce, M. Perc, and Y. Isler, "EEG-based finger movement classification with intrinsic time-scale decomposition," Frontiers in Human Neuroscience, Vol. 18, p. 1362135, 2024.
[15] J. Rahul, D. Sharma, L. D. Sharma, U. Nanda, and A. K. Sarkar, "A systematic review of EEG based automated schizophrenia classification through machine learning and deep learning," Frontiers in Human Neuroscience, Vol. 18, p. 1347082, 2024.
[16] M. E. A. Bourkha, A. Hatim, D. Nasir, S. El Beid, S. Ez-ziymy, and A. S. Tahiri, "Ventricular Ectopic beats detection based wavelet scattering network and ensemble bagged trees for smart medical systems," Procedia Computer Science, Vol. 236, pp. 468–475, 2024.
[17] P. N. Singh and R. P. Mahapatra, "A novel deep learning approach for arrhythmia prediction on ECG classification using recurrent CNN with GWO," International Journal of Information Technology, Vol. 16, No. 1, pp. 577–585, 2024.
[18] M. Karuna and S. R. Guntur, "Classification of hand movements based on EMD-CCT feature extraction method through EMG using machine learning," Multimedia Tools and Applications, pp. 1–27, 2024.
[19] N. A. S. Putro, C. Avian, S. W. Prakosa, M. I. Mahali, and J. S. Leu, "Estimating finger joint angles by surface EMG signal using feature extraction and transformer-based deep learning model," Biomedical Signal Processing and Control, Vol. 87, p. 105447, 2024.
[20] L. Meyers et al., "Automated real-time collection of pathogen-specific diagnostic data: syndromic infectious disease epidemiology," JMIR Public Health and Surveillance, Vol. 4, No. 3, p. 9876, 2018.
[21] I. Zafar et al., "Reviewing methods of deep learning for intelligent healthcare systems in genomics and biomedicine," Biomedical Signal Processing and Control, Vol. 86, p. 105263, 2023.
[22] X. Liu et al., "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," The Lancet Digital Health, Vol. 1, No. 6, pp. 271–297, 2019.
[23] M. Ahsan, A. Khan, K. R. Khan, B. B. Sinha, and A. Sharma, "Advancements in medical diagnosis and treatment through machine learning: A review," Expert Systems, Vol. 41, No. 3, p. 13499, 2024.
[24] M. Pourmokhtari and B. Beigzadeh, "Simple recognition of hand gestures using single-channel EMG signals," Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, Vol. 238, No. 3, pp. 372–380, 2024.
[25] K. Bakırcıoğlu and N. Özkurt, "Classification of EMG signals using convolution neural network," International Journal of Applied Mathematics Electronics and Computers, Vol. 8, No. 4, pp. 115–119, 2020.
[26] Á. L. Valdivieso Caraguay, J. P. Vásconez, L. I. Barona López, and M. E. Benalcázar, "Recognition of hand gestures based on EMG signals with deep and double-deep Q-networks," Sensors, Vol. 23, No. 8, p. 3905, 2023.
[27] M. H. Abdelaziz, W. A. Mohamed, and A. S. Selmy, "Hand Gesture Recognition Based on Electromyography Signals and Deep Learning Techniques," Journal of Advances in Information Technology, Vol. 15, No. 2, 2024.
[28] R. V. Godoy et al., "Electromyography-based, robust hand motion classification employing

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

temporal multi-channel vision transformers," IEEE Robotics and Automation Letters, Vol. 7, No. 4, pp. 10200–10207, 2022.

[29] S. M. Mane, R. A. Kambli, F. S. Kazi, and N. M. Singh, "Hand motion recognition from single channel surface EMG using wavelet & artificial neural network," Procedia Computer Science, Vol. 49, pp. 58–65, 2015.

[30] A. A. Al-Taee, R. N. Khushaba, T. Zia, and A. Al-Jumaily, "Feature extraction using wavelet scattering transform coefficients for EMG pattern classification," Australasian Joint Conference on Artificial Intelligence, Springer International Publishing, Cham, pp. 181–189, 2022.

[31] J. Andén and S. Mallat, "Deep scattering spectrum," IEEE Transactions on Signal Processing, Vol. 62, No. 16, pp. 4114–4128, 2014.

[32] D. Ovadia, A. Segal, and N. Rabin, "Classification of hand and wrist movements via surface electromyogram using the random convolutional kernels transform," Scientific Reports, Vol. 14, No. 1, p. 4134, 2024.

[33] C. Sapsanis, A. Tzes, and G. Georgoulas, "sEMG for Basic Hand movements [Dataset]," UCI Machine Learning Repository, 2013.

[34] H. Naser and H. A. Hashim, "sEMG-Based hand gestures classification using a semi-supervised multi-layer neural networks with Autoencoder," Systems and Soft Computing, p. 200144, 2024.