Tuan Rohisham T. N., Abas N., Nordin N., Ali S.K., Abas M.A., Tokhi M. O.
Jurnal Mekanikal, June 2026, 49: 238-257

HYBRID MULTIMODAL FUSION FRAMEWORK INTEGRATING EMG AND FORCE SIGNALS FOR ENHANCED HAND MOVEMENT PREDICTION

Tuan Nursabrina Tuan Rohisham[a], Norafizah Abas[a]*, Nurdiana Nordin[a], Siti Khadijah Ali[b], Mohd Azman Abas[c], Mohammad Osman Tokhi[d]

[a]Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia
[b]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia.
[c]Pusat Kesihatan Universiti, Universiti Kebangsaan Malaysia
[d]School of Engineering, London South Bank University, London, UK

*Corresponding email: norafizahabas@utem.edu.my

Article history
Received
15th January 2026
Revised
5th April 2026
Accepted
16th April 2026
Published
3rd June 2026

ABSTRACT

Accurate prediction of hand movements is important to improve human–machine interaction, especially in rehabilitation and assistive applications. However, the nonlinearity of the electromyography (EMG) signal often limits the reliability of motion classification. It causes the sensor fusion to be unstable and not intelligent enough to continuously predict the user hand movements. To address this, we propose a hybrid multimodal fusion framework that integrates EMG and force signals to improve prediction accuracy and robustness. The framework investigates the relationship between forearm EMG signals, various grasping tasks, and finger/wrist joint angles. It goes beyond discrete classification to allow continuous motion intention prediction of wearable hand control. The proposed hybrid multimodal fusion framework has two levels: feature-level fusion and decision-level fusion. Canonical Correlation Analysis (CCA) is used to extract highly correlated features across modalities at the feature level. Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) are used to classify these features to identify six different hand movements. To strengthening the decision consistency, majority voting is used at the classifier level. The performance of the system is evaluated based on the confusion matrices, accuracy, and F1-scores. Results show the proposed framework is significantly better than unimodal approaches, with the highest accuracy of 97.86% being achieved by the Waveform Length. Through experiments using data from 10 healthy subjects, it was established that multimodal fusion is effective in addressing the nonlinearity of the EMG signal, which results in more accurate hand gesture recognition. The findings assist in developing an efficient control scheme for wearable hand devices that provide smooth, user-intent-driven motions. By enhancing accuracy and responsiveness, the proposed approach improves support for activities of daily living (ADL), reducing the possibility of user dissatisfaction and the discontinuation of the device.

Keywords: *Hybrid multimodal fusion framework, Electromyography (EMG), Hand movement prediction, Gesture recognition, Canonical correlation analysis.*

10.11113/jm.v49.791

1.0 INTRODUCTION

Traditional gesture recognition systems use unimodal electromyography (EMG) signals to detect muscle activation. Although electromyography (EMG) provides useful information about neural intent, its nonlinear characteristics and sensitivity to inter-subject variability, muscle fatigue, and electrode displacement make it less reliable in real-world applications [1]. As an alternative, stable measurements for pressure and hand grasping force have been obtained using force sensors. However, they are unable to capture the underlying neural intention because they only reflect mechanical output [2]. Therefore, relying solely on one modality may result in inconsistent control of assistive or rehabilitation devices.

Recent research has explored multimodal sensor fusion methods to overcome these limitations. Combined EMG and kinetic signals improve robustness and classification accuracy by capturing neural intent and mechanical response simultaneously. Integration of EMG with kinematics or kinetics has consistently improved recognition performance [3,4]. Studies report that surface EMG (sEMG) combined with force myography or inertial measurement units (IMUs) outperforms single-modality approaches, particularly for stroke rehabilitation [5,6]. However, a framework that incorporates both feature-level correlation and decision-level error mitigation across EMG and force signals remains unexplored.

The three types of gesture recognition fusion strategies are feature-level, decision-level, and hybrid. Before classification, feature-level fusion combines raw or pre-processed signals, often employing methods such as Canonical Correlation Analysis (CCA) to identify the components that are most strongly correlated [7]. Decision-level fusion, on the other hand, improves robustness to noise and sensor failures by combining the output of multiple classifiers, usually through majority voting [7]. The benefits of a hybrid multimodal fusion approach are still largely unexplored, though, as most recent research concentrates on these methods separately [8].

To control multimodal variability and improve generalisation, a hybrid multimodal fusion framework that integrates the two levels has shown promise. Due to their effectiveness, machine learning models like Support Vector Machines (SVM) and Linear Discriminant Analysis (LDA) are still widely used in these types of systems. Meanwhile, because of their ability to learn complex data and discover nonlinear relationships, Artificial Neural Networks (ANNs) have become a growing trend [1,9]. Confusion matrices are commonly used to assess performance, and metrics like accuracy and F1-score are employed to ascertain system performance [4].

This study proposes a novel hybrid multimodal fusion framework to fulfil the research gap. The framework classifies EMG and force signals using LDA, SVM, and ANN models, and then combines them through CCA at the feature-level fusion. To improve prediction consistency, a majority voting system is used at the decision level. The results of the experiment show that the proposed hybrid multimodal fusion framework performs better than unimodal approaches, which suggests that it could improve real-time gesture recognition in assistive and rehabilitation technologies.

2.0 RESEARCH METHODOLOGY

This section explains the methodology of this study, which gives an overview of the hybrid multimodal fusion framework, starting with the experimental setup, data collection, signal acquisition, and feature extraction processes. Then, the baseline fusion approach and the proposed hybrid multimodal fusion framework are described, followed by the machine learning classifiers used and the procedures of out-of-sample model validation.

2.1 Experimental Setup and Data Collection

This research is conducted with the approval of the Centre of Research and Innovation Management (CRIM) at Universiti Teknikal Malaysia Melaka (UTeM) under reference number UTeM.11.02/500-25/1/4 JILID 2 (57). The experiments were carried out in the Faculty of Electrical Technology and Engineering (FTKE), Advanced Digital Signal Processing (ADSP) Laboratory, UTeM. All experiments adhered to ethical research standards to ensure participant safety and data integrity. Participants were recruited through direct invitations and email announcements.

A total of ten healthy male subjects, aged between 20 and 40 years, participated in the experiment. To ensure data consistency across the cohort, all participants were right-hand dominant and had no prior experience with EMG data collection. Only individuals with no prior history of stroke, cerebrovascular diseases, muscle disorders, or mental illnesses (e.g., anxiety, mood, or psychotic disorders) were considered. These exclusion criteria are applied to minimize factors that could affect movement performance and introduce variability in the dataset. All subjects provided written informed consent and were briefed on experimental procedures.

The experimental setup is designed to ensure accurate and reliable data collection of EMG and force signals during various hand movements. During the recording sessions, subjects are seated in a neutral position with the dominant forearm supported at a 90-degree angle. Figure 1 shows the block diagram of the system, which incorporates two primary inputs: EMG signals (Input I) and force signals (Input II). EMG signals are recorded using a LabQuest Mini Data Acquisition system (DAQ), while force signals are measured with a hand dynamometer. Both inputs were interfaced with a laptop running Logger Lite software for real-time acquisition and visualization. Before each session, the DAQ and dynamometer are calibrated to ensure the measurement is accurate. The recorded signals were subsequently processed and used as inputs to machine learning classifiers.

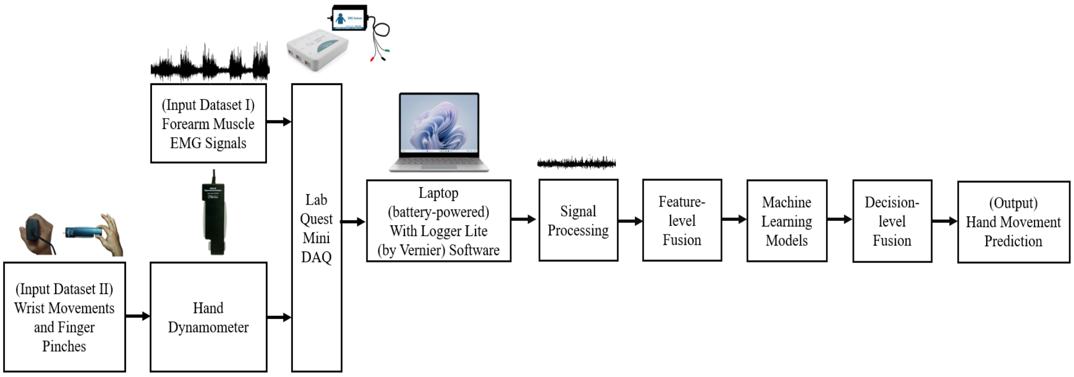


Figure 1: Block diagram of the experimental setup for hand movement prediction using EMG signals and a hand dynamometer.

The muscle selection procedure began by identifying the forearm muscles on the dominant hand of the subject that are responsible for various hand movements. Electrodes are strategically placed on the dominant forearm muscles, Flexor Digitorum Superficialis (FDS) and Flexor Carpi Radialis (FCR) on the anterior compartment, and Extensor Digitorum Communis (EDC) and Extensor Carpi Radialis Longus (ECRL) on the posterior. Figure 2 illustrates these muscle locations in both (a) anterior and (b) posterior forearm compartments.

Electrode placement followed SENIAM recommendations for a consistent signal [10]. Muscle contraction signals are detected by positive and negative electrodes placed 24 mm apart on the same muscle, with a reference electrode on a bony area to reduce interference. Skin preparation with alcohol enhanced signal quality. Table 1 presents the specific muscle identification tests and electrode placement locations for each of the forearm muscles, crucial for ensuring proper signal acquisition.

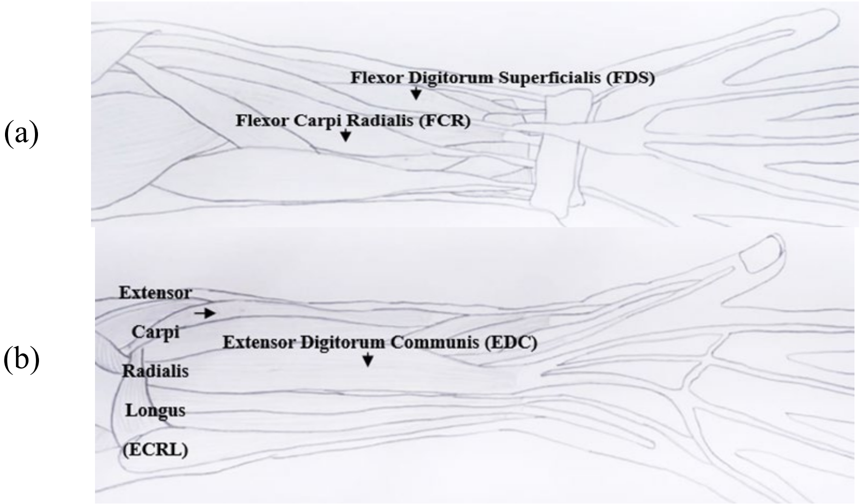


Figure 2: Electrode placement on (a) anterior compartment, (b) posterior compartment.

EMG signals were recorded using a Vernier multichannel sensor at 2000 Hz [10]. Electrodes were connected to the DAQ system via three-channel wires. Data were collected across seven consecutive days, with one session per subject per day. Each session included five repetitions of each movement, with five-minute breaks between repetitions to minimize fatigue. Simultaneously, force signals were measured using a Vernier hand dynamometer to capture grip strength and pressure changes during movements.

Table 1: Muscle identification tests and electrode placement [11]

Muscle	Test	Location
FCR	The wrist is flexed against resistance.	Three or four finger breadths away from the midpoint of a line connecting the medial epicondyle and biceps tendon.
FDS	Forearm in supination, one finger is flexed at the PIP joint, while the DIP is kept extended against resistance, and the other three fingers are held extended to inactive the FDP.	Middle of the forearm, index fingers flexed towards the biceps tendon, just medial to the finger.
EDC	Forearm in pronation, the fingers are extended at the MCP joint while pressure is exerted at the PIP joints by attempting to flex them.	Upper 1/3 of the forearm between the radius and ulna.
ECRL	The wrist is extended and abducted with the forearm pronated.	Two finger breadths away from the lateral epicondyle.

Participants performed eight distinct hand and finger movements: hand open (HO), hand grip at neutral wrist position (HGN), hand grip at wrist extension (HGE), hand grip at wrist flexion (HGF), index to thumb finger pinches (FP1), middle to thumb finger pinches (FP2), ring to thumb finger pinches (FP3), and little to thumb finger pinches (FP4). Posture was standardized, and all movements were supervised to ensure correct execution. These movements are visually represented in Figure 3.

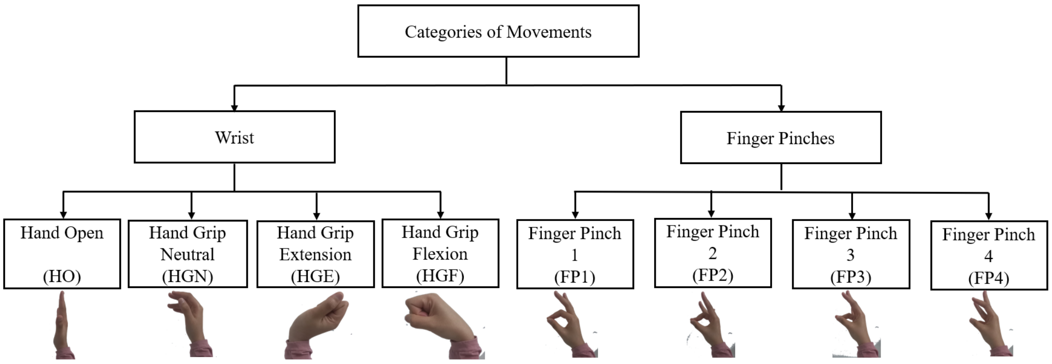


Figure 3: Categorization of wrist movements and a thumb-index middle finger pinch.

Each recording session produced approximately 90 seconds of data per subject. The data was split using a hold-out validation approach; the first 60 seconds were utilized for training, and the remaining 30 seconds were used for testing. Specifically, the total dataset of 1405 rows were divided into a 70/30 split ratio: 937 rows (derived from 2 Subjects) were used to train the model, while 468 rows (derived from 1 Subject) were held out for testing. This ensured a fair assessment of model performance on unseen data and verified the framework's ability to generalize to new subjects.

2.2 Signal Acquisition and Feature Extraction

Raw EMG and force signals were first corrected for DC offset to remove baseline shifts and prevent distortion [12]. A first-order Butterworth bandpass filter is then applied to the data, with a frequency range of 20 to 450Hz. This range attenuates low-frequency motion artifacts and high-frequency noise, preserving essential muscle activation characteristics. It enclosed the typical EMG spectrum while filtering unwanted artifacts.

After filtering, the EMG and force signals are segmented into overlapping time windows. Each window contains 256 samples with 50% overlap (an increment of 128 samples). This provides smooth transitions and robust feature extraction, with the chosen window size capturing sufficient muscle activity and preventing data loss. Signals are then normalized to a 0-1 range to minimize inter-subject variability and enhance classification performance, forming the basis for feature extraction.

Time-domain features were chosen for feature extraction due to their ease of use, low computational cost, and demonstrated efficacy in real-time EMG applications [13]. Based on a prior study, three widely used features, Waveform Length (WL), Mean Absolute Value (MAV), and Root Mean Square (RMS), are selected [14]. These features are chosen due to their demonstrated ability to capture various aspects of muscle activation while maintaining computational efficiency for real-time applications. RMS measures signal power (muscle effort), WL depicts the complex and dynamic nature of muscle contraction patterns, and MAV provides a reliable measure of signal amplitude. This combination results in a computationally light signal representation.

The RMS, which measures the EMG signal's energy level, is defined by Equation (1) [15]:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)^2} \tag{1}$$

where N is the number of samples, while x_i denotes the amplitude of the i^{th} EMG signal sample. This feature helps measure muscular effort because stronger muscle contractions produce higher RMS values.

The MAV determines the absolute mean values of the EMG signal, as shown in Equation (2) [15]:

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|x_i| \tag{2}$$

where N is the number of samples and $|x_i|$ indicates the absolute value of the i^{th} sample amplitude. MAV provides a simple yet reliable indicator of muscle activity, making it useful for applications requiring minimal computational overhead.

The WL, which measures the total absolute difference between consecutive data points, is defined by Equation (3) [15]:

$$WL = \sum_{i=1}^{N-1}|x_{i+1} - x_i| \tag{3}$$

where $N - 1$ represents the number of consecutive intervals over which the sum is calculated, while x_{i+1} an x_i represent the amplitudes of two adjacent samples. This summation represents the total path length of the signal, providing a measure of both frequency and amplitude complexity. Higher WL values indicate dynamic or complex muscle activity, while lower values indicate more consistent movement patterns.

2.3 Baseline Fusion Approaches

Sensor fusion is necessary to improve the accuracy and reliability of the prediction of hand movements by integrating data from multiple sensor approaches. Three baseline approaches, EMG-only, force-only, and a simple multimodal fusion strategy, are implemented and analyzed to create standards for performance evaluation. These strategies serve as benchmarks for the proposed hybrid multimodal fusion framework and are representative of conventional approaches found in the literature.

The EMG-only model is applied solely to electromyography signals obtained from the forearm muscles. Since it reflects motor intention, this method, which records neuromuscular activity, has been extensively employed in gesture recognition studies. Even though neuromuscular activation is reflected by EMG, accuracy may be diminished by noise and variability. Following these limitations, classifications are frequently unstable, causing EMG-only systems to be insufficiently reliable for practical use.

The force-only model uses signals obtained from a hand dynamometer to measure the grip strength and mechanical force applied during hand movements. Compared to EMG, force signals are relatively stable and less affected by external noise, thereby providing consistent measurements of executed motion. However, this approach cannot capture the underlying neural intention, as force is only observable once physical movement is initiated. Consequently, while force-only systems provide reliable mechanical feedback, they lack predictive capability and are unsuitable for applications that require anticipatory control.

The multimodal baseline is an integration of the EMG and force features, and integrates them into single feature vectors that are to be classified. This strategy exploits complementary information, with EMG features representing motor intention and force features reflecting mechanical execution. Although simple multimodal fusion typically yields higher accuracy than unimodal methods, it treats all features equally and does not account for inter-modal correlations or redundancies. This may introduce dimensionality challenges and reduce classifier efficiency, particularly when dealing with complex, high-variance datasets.

For all baselines, the approach begins with preprocessing and feature extraction of both signal types before their fusion, which significantly strengthens the prediction performance, especially when unimodal EMG signals are less reliable. EMG and force signals from four forearm muscles are preprocessed, segmented, and three time-domain features (RMS, MAV, WL) are extracted. These features are then input into machine learning models to classify the eight hand movements as discussed earlier.

This study highlights the advantages of a hybrid multimodal fusion framework that combines feature-level correlation maximized through CCA and decision-level through majority voting by comparing the proposed framework to the three approaches. This dual-level design can predict hand movements more accurately and consistently while addressing the limitations of unimodal and simple multimodal approaches.

2.4 Proposed Hybrid Multimodal Fusion Framework

The proposed hybrid multimodal fusion framework aims to enhance the accuracy and robustness of hand movement prediction through the integration of electromyography (EMG) and force signals. This framework employs a two-stage fusion approach (feature-level and decision-level) to produce an efficient prediction system using complementary data.

CCA is employed at the feature level. The EMG and force feature datasets are transformed by CCA after preprocessing and time-domain feature extraction. To create a CCA-transformed EMG and force feature that improves data representation and class discriminability, it identifies and merges the most correlated components.

These features of CCA-transformed EMG and force are then trained on three machine learning classifiers, LDA, SVM, and ANN. Their outputs are then passed to the decision-level fusion stage, where a majority voting strategy is used. This method assigns the final class label based on the most frequently predicted outcome, reducing individual misclassifications and enhancing prediction consistency. This hybrid multimodal fusion framework, as shown in Figure 4, enables precise and dependable real-time predictions and, therefore, applies to rehabilitation and assistive technologies.

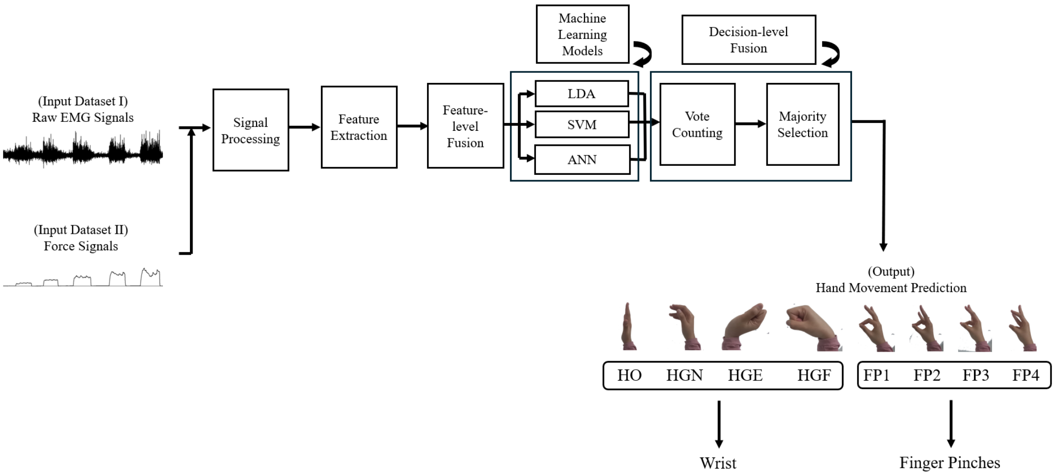


Figure 4: Block diagram for hand movement prediction using a hybrid multimodal fusion framework

Overall, the proposed hybrid multimodal fusion framework, which integrates EMG and force sensor at the feature-level, CCA, and at the decision-level, through majority voting, to predict the hand movements more accurately and reliably. The flowchart of the proposed framework is depicted in Figure 5. The results provide a quantitative evaluation and discussion of the effect of unimodal, multimodal, and hybrid multimodal fusion frameworks on prediction performance.

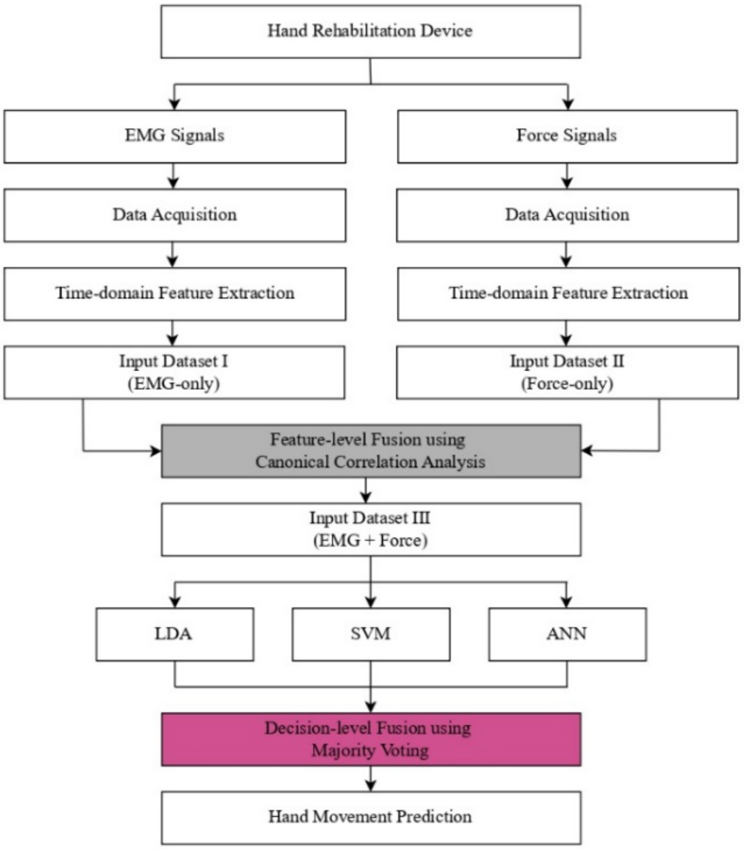


Figure 5: Flowchart illustrating the proposed hybrid multimodal fusion framework for hand movement prediction based on EMG and force signals.

2.5 Machine Learning Classifiers

This study implements a machine learning framework to classify eight distinct hand movements (HO, HGN, HGE, HGF, FP1, FP2, FP3, and FP4) using features extracted from electromyography (EMG) and force signals. Three different classification models are employed: LDA, SVM, and ANN. The selection of these models is made due to complementary characteristics in the ability to deal with both nonlinear and linear data distributions, making them suitable for predicting EMG and force-based hand movements.

In EMG signal processing, the statistical classification technique known as LDA is used due to its ease of use and computational effectiveness [11]. Assuming normally distributed data, LDA is a high-to-low-dimensional feature learner that maximizes the ratio of between-class to within-class variance to build a linear decision boundary. Since LDA works best with linearly separable data, performance may be limited by nonlinear or noisy EMG signals, even though it is computationally efficient for real-time applications.

In contrast, the SVM classifier constructs optimal hyperplanes in the feature space to maximise the margin between different movement classes [16]. Non-linearly separable data can be transformed into a higher-dimensional space to achieve better separation with the help of kernel functions (such as the polynomial kernel) [17,18]. This approach performs well for EMG-based movement prediction due to its robustness in handling complex and high-dimensional signals. SVMs are also known for their strong generalization capabilities and robust resistance to overfitting, especially when working with noisy sensor data or small sample sizes.

A feedforward neural network architecture is used in this study's ANN model. The hand movement classes are represented by the input layer, one hidden layer with ten neurons, and an output layer with eight neurons [16,17]. In order to efficiently model data nonlinearities, hidden layer neurons use a nonlinear activation function (such as softmax), enabling the ANN to discover complex relationships within multimodal, complex signal patterns. Table 2 shows the hyperparameters for the LDA, SVM, and ANN models in order to ensure reproducibility and optimal performance. Crucially, all models rigorously apply a consistent subject-wise data split (2 subjects for training, 1 for testing) to assess generalisation abilities on unseen individuals.

Table 2: Hyperparameters used for the LDA, SVM, and ANN models.

Model	Hyperparameter	Final Configuration/Value	Justification/Purpose
LDA	Discriminant Type	Linear	Creates linear decision boundaries.
	Class Priors	Estimated from training data	Reflects class proportions.
	Data Split Strategy	2 Subjects for Training (60s; 937 rows), 1 Subject for Testing (30s; 468 rows)	Evaluates generalization to unseen subjects.
SVM	Kernel Function	Linear	Enables linear decision boundaries; efficient.
	Box Constraint (C)	Default by templateSVM	Balances margin maximization and error.
	Multi-class Strategy	ECOC (Error Correcting Output Codes)	Adapts binary SVMs for multi-class problems.
	Data Split Strategy	2 Subjects for Training (60s; 937 rows), 1 Subject for Testing (30s; 468 rows)	Evaluates generalization to unseen subjects.
ANN	Number of Hidden Layers	1	Balances simplicity and learning capacity.
	Neurons in Hidden Layer	10	Provides learning capacity, limits overfitting.
	Hidden Layer Activation	Tangent Sigmoid (tansig)	Introduces non-linearity for complex patterns.
	Output Layer Activation	Softmax (softmax)	Generates class probabilities for classification.
	Training Algorithm	Scaled Conjugate Gradient (trainscg)	Efficiently optimizes network weights.
	Performance Function	Cross-Entropy (crossentropy)	Minimizes error for classification accuracy.
	Data Split Strategy	2 Subjects for Training (60s; 937 rows), 1 Subject for Testing (30s; 468 rows)	Evaluates generalization to unseen subjects.

Table 2 summarizes the key hyperparameters used for each machine learning model, including solver types, regularization parameters, and neural network configurations, all set to optimize performance and ensure replicability [19-21].

The strategic inclusion of LDA alongside SVM and ANN enhances the robustness of the machine learning framework. LDA provides a fast baseline for linearly separable data, while SVM and ANN offer greater flexibility for nonlinear and noisy sensor input. Within the hybrid multimodal fusion framework, these classifiers are applied at the decision-level fusion stage. Each model independently predicts movement classes, and their predictions contribute to the final decision via majority voting. This integration utilizes the strengths of each model to yield more accurate and reliable movement prediction. Model performance is subsequently evaluated using the confusion matrix, accuracy, and F1-score to assess its effectiveness in accurately distinguishing between the eight distinct hand movements.

2.6 Out-of-Sample Model Validation

The practical evaluation of EMG and force-based classifiers depends significantly on selecting appropriate training and testing strategies, as these factors influence the generalizability and reliability of movement prediction systems. Using data from the same subject for both testing and training is necessary for in-sample evaluation. For in-sample evaluation, methods like k-fold cross-validation and hold-out validation are typically used. The hold-out validation divides the data into training and testing sets (e.g., 70/30 split) and provides a fast estimate of the model's performance. This approach may yield overly optimistic results if the split does not accurately represent the whole range of signal variability. K-fold cross-validation, on the other hand, typically uses 5 or 10 folds to rotate the training and testing data across multiple partitions. This reduces the impact of random sampling and provides a more balanced and accurate performance estimate.

Applications that aim for real-time performance must evaluate classifier performance on out-of-sample data, which is entirely separate from the training set. Out-of-sample testing may involve different users, recording sessions, or experimental conditions. This technique helps assess how robust the classifier is to changes in physiological conditions, sensor positions, or muscle activity patterns. Strong out-of-sample performance is essential for myoelectric control systems, according to a recent study. The availability of large, multi-user datasets has enabled user-independent cross-user classification, demonstrating the potential of data-driven approaches to address individual variability and achieve encouraging accuracy [19]. By evaluating the model on real, unseen data, out-of-sample testing provides an accurate representation of system performance in the real world.

3.0 RESULTS, ANALYSIS, AND DISCUSSION

This section presents the hand movement prediction results and compares them with EMG-only, Force-only, multimodal, and a proposed hybrid multimodal fusion framework (which combines majority voting-based decision-level fusion with CCA-based feature-level fusion). The confusion matrix, accuracy, and F1-scores are used to evaluate the classification performance of the ANN, SVM, and LDA classifiers.

3.1 Preliminary Evaluation and Justification of the Dataset

The dataset design had to be refined to generate reproducible and reliable results of the sensor fusion approach. The initial set of eight hand movements was chosen to represent both gross and fine motor functions, consistent with the diversity highlighted in prior studies. Raw data inspection and early feature extraction across all eight gestures revealed limitations that called for reducing the movement set before applying the hybrid multimodal fusion framework.

As illustrated in Figures 6 and 7, preliminary trials using the full 8-movement set confirmed the inclusion of FP3 (thumb–ring pinch) and FP4 (thumb–little pinch), which introduced significant inter-class confusion. The scatter plots for both wrist-related muscles

(FCR and ECRL) and four medial finger muscles (FDS and EDC) show that the feature distributions for FP3 and FP4 are almost entirely overlapping.

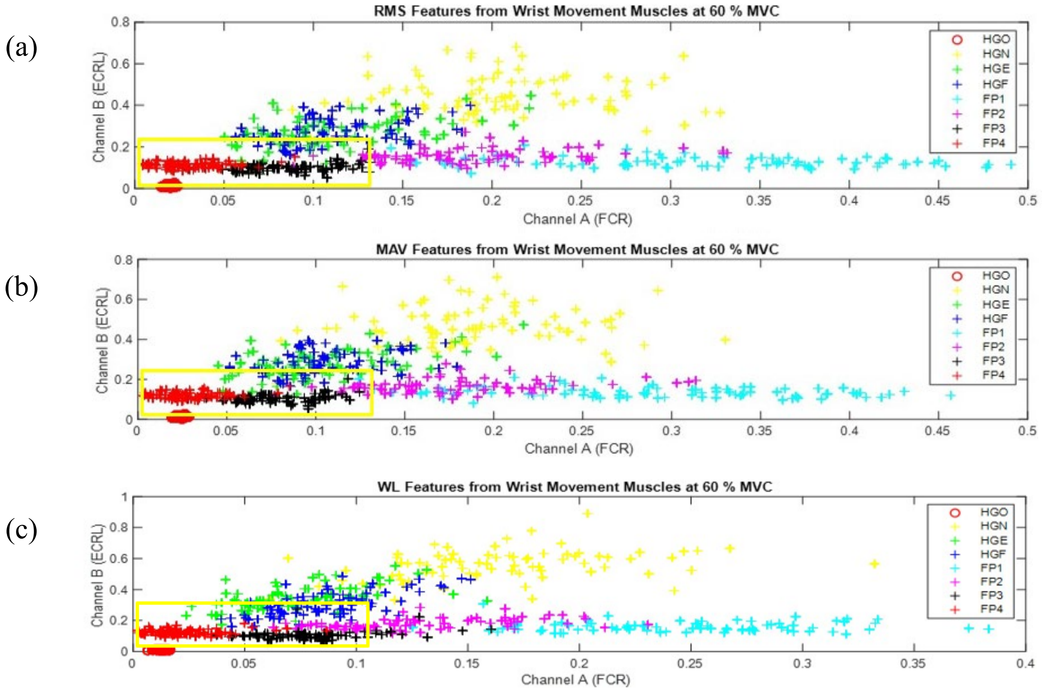
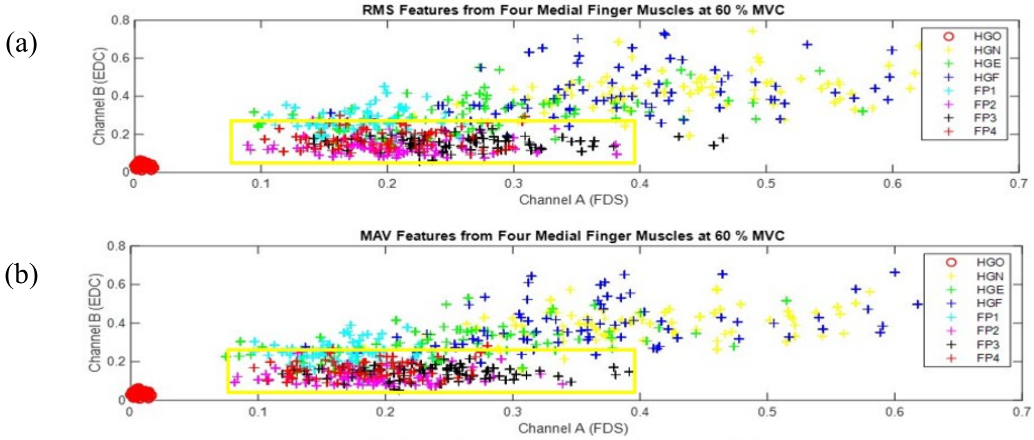


Figure 6: Scatter plot showing (a) RMS, (b) MAV, and (c) WL features from wrist movement muscles at 60% of MVC for Channel A (FCR) – Channel B (ECRL).

While Figure 6 illustrates the overlap in muscles contributing to wrist movement, a similar pattern of redundancy is observed in Figure 7 for the EMG signals extracted from muscles primarily responsible for four medial finger pinches.



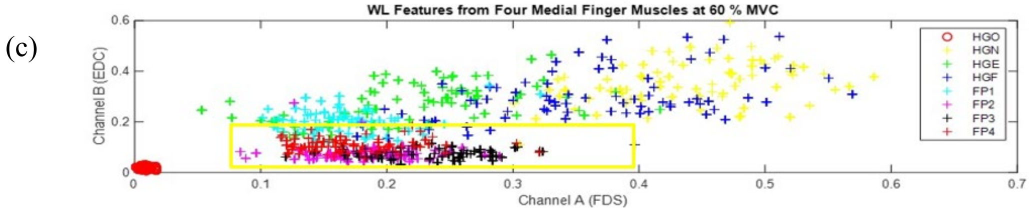


Figure 7: Scatter plot showing (a) RMS, (b) MAV, and (c) WL features from four medial finger muscles at 60% MVC for Channel A (FDS) – Channel B (EDC).

The exclusion of FP3 and FP4 was necessitated by subject-dependent physical limitations identified during the experimental phase. Unlike the more robust signals generated by FP1 and FP2, most subjects demonstrated insufficient motor strength and lacked the consistent coordination required to perform these specific pinches. Consequently, these subjects could not produce a strong or distinct physical output for FP3 and FP4, leading to weak and highly inconsistent signal patterns.

Furthermore, the hand dynamometer confirmed that the kinetic data were unable to distinguish between these gestures due to the lack of clear mechanical separation. Crucially, the addition of further force-sensing hardware or orientation sensors would not resolve this issue, because the subjects were unable to produce distinct pressure patterns or postural changes; additional equipment would only capture the same redundant, indistinct data. To maintain the robustness and reliability of the predictive system, these inconsistent classes were removed to focus on the six strongest and most achievable movements (HO, HGN, HGE, HGF, FP1, and FP2), ensuring high-accuracy performance across all users.

Figure 8 illustrates the raw signals used in this analysis. Across a 90-second session (60 seconds of training, 30 seconds testing), EMG data were recorded from four forearm muscles (FDS, EDC, FCR, ECRL), normalized to % MVC, alongside synchronized force and wrist angle signals. The recordings highlight distinct muscle activations: EDC/ECRL during extension tasks (HO, HGE) and FDS/FCR during grasping actions (HGN, HGF). Force traces rise consistently with pinching and grasping, while wrist angle transitions correspond with flexion and extension states. Together, these patterns confirm that the dataset effectively captures the coordinated interaction between neuromuscular activity, applied force, and joint motion [22,23].

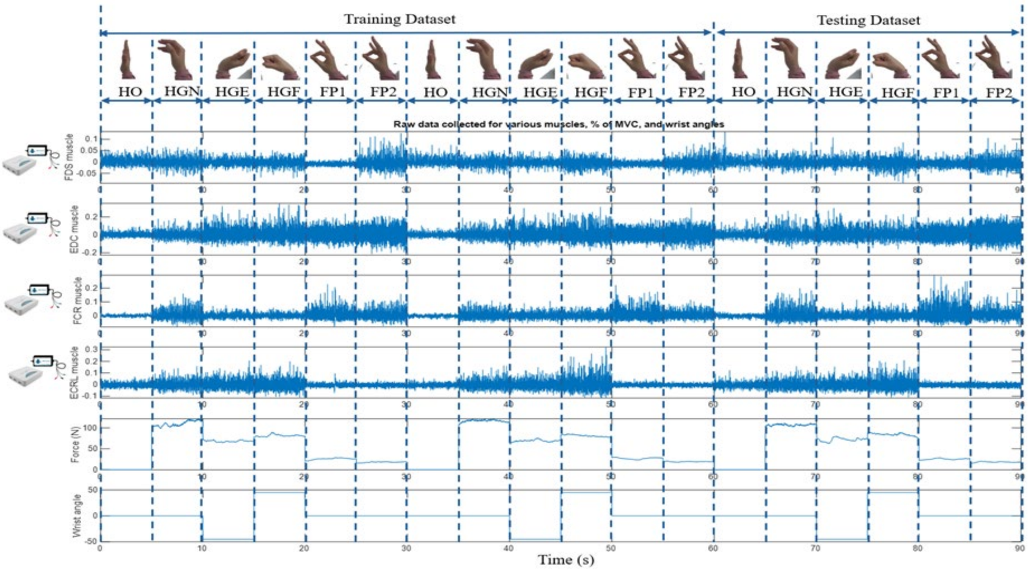


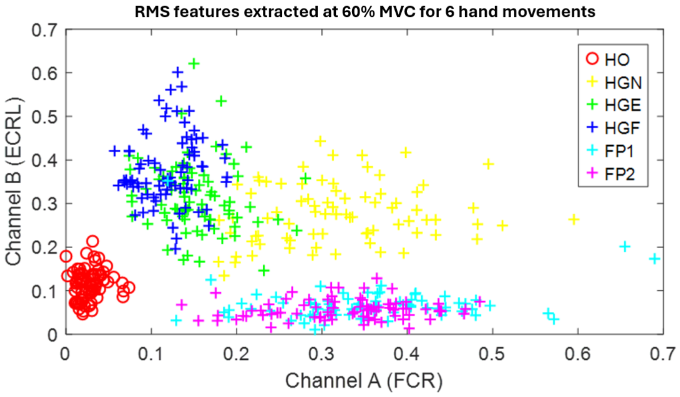
Figure 8: Raw EMG, force, and wrist angle signals over 90 seconds, separated into training and testing phases.

3.2 Time-Domain Feature Extraction and Visualization

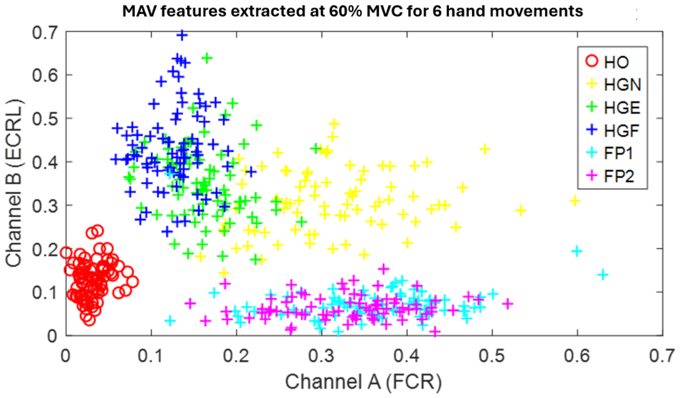
The time-domain features extracted from the six redefine dataset were plotted and analyzed with respect to the muscles that contributed to wrist movement (the FCR and ECRL muscles) and muscles that contributed to finger pinches (the FDS and EDC muscles), respectively, as shown in Figures 9 and10. Based on Figure 9, feature distribution plots for WL, RMS, and MAV at 60% MVC illustrate distinct patterns of class separability across the FCR (x-axis) and ECRL (y-axis) muscles. Both RMS and MAV exhibited highly similar patterns, as expected from their amplitude-based nature. In these two features, HO was consistently positioned near the origin with a compact distribution, making it reliably separable. For the finger pinches (FP1 and FP2), RMS and MAV placed the clusters in the low ECRL region but introduced wider horizontal spread, especially for FP1. This spread, combined with the evident overlap between FP1 and FP2, reduces their discriminability in amplitude-based features. Similarly, the grasping movements (HGF, HGN, HGE) showed broader and more diffuse clusters in RMS and MAV, leading to significant inter-class overlap.

In contrast, WL produced more compact and stable clusters with lower intra-class variance. HO remained tightly grouped, while FP1 and FP2 were more localized compared to RMS and MAV. Importantly, WL offered clearer separation among the grasping classes, with HGF forming the highest ECRL cluster and HGN/HGE occupying distinct adjacent regions. This indicates that WL captures both amplitude and time-based variation more effectively, yielding greater inter-class separability.

(a)



(b)



(c)

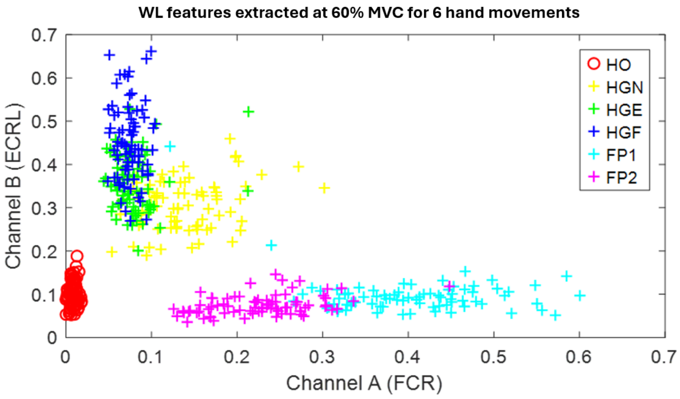
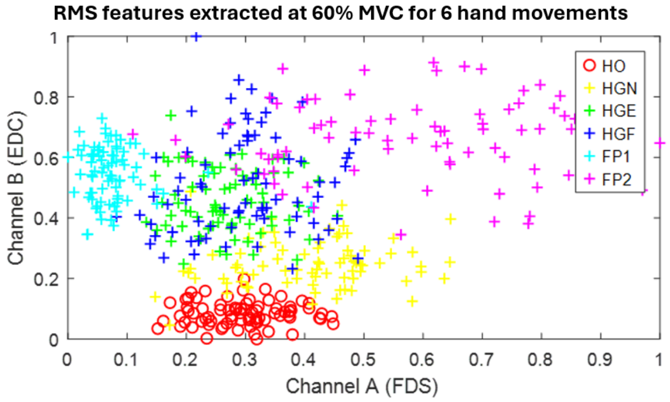


Figure 9: Time-domain features extracted at 60% MVC for 6 hand movements based on muscles (ECRL and FCR) contributing to wrist movements: (a) RMS feature, (b) MAV feature, (c) WL feature.

A similar pattern is observed for EMG extracted from muscles that contribute to finger pinches, where the WL feature exhibits a good separability between movement classes. Thus, suggesting the best feature to be used for the classification, further testing, and analysis.

Figure 10 shows the HO remains tightly clustered near the origin, confirming its consistent separability across muscle groups. More importantly, the finger pinches (FP1 and FP2), which typically present a challenge due to their similar motor activation, form relatively compact and localized distributions under WL. Although some overlap remains, the clusters show reduced spread and clearer positioning compared to RMS and MAV features. The grasping movements (HGF, HGN, HGE) also demonstrate improved organization, with WL yielding tighter boundaries along the FDS axis. It allows for more distinct differentiation among the three classes.

(a)



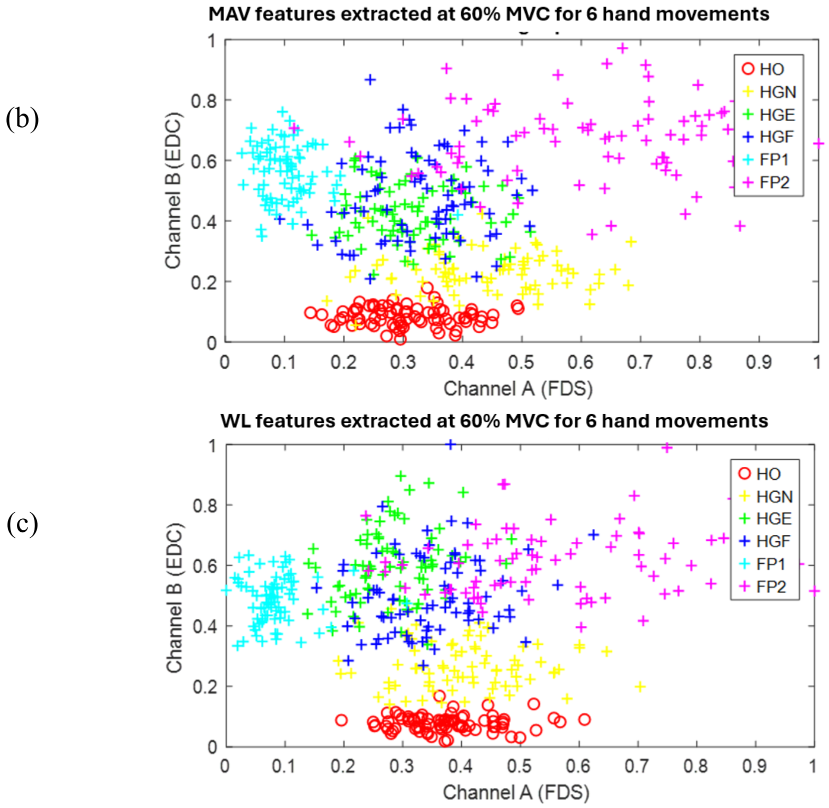


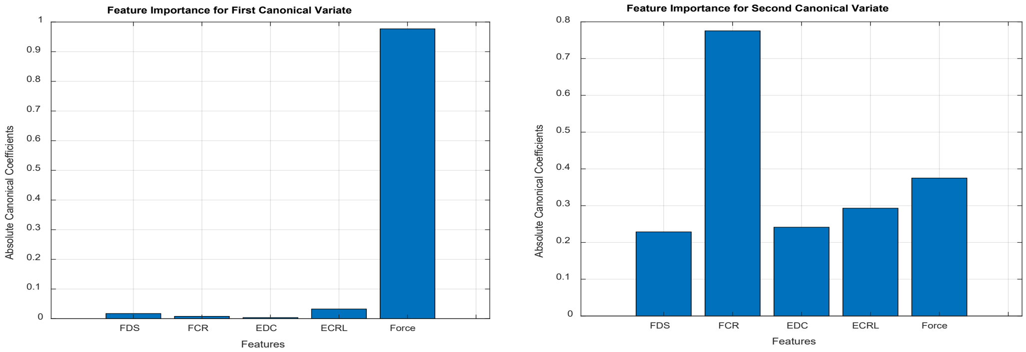
Figure 10: Time-domain features extracted at 60% MVC for 6 hand movements based on muscles (EDC and FDS) contributing to finger movements: (a) RMS feature, (b) MAV feature, (c) WL feature.

Overall, while RMS and MAV provide useful amplitude information, their similarity and higher variability reduce their effectiveness for distinguishing closely related movements. WL, on the other hand, demonstrates superior discriminative power by combining amplitude and complexity measures. This makes WL the most robust standalone feature in this dataset. Therefore, WL should be prioritized for feature-level fusion using CCA methods, while RMS and MAV may serve as complementary features in a multimodal feature set. Classifiers trained with WL are also expected to achieve higher accuracy and lower misclassification rates than those relying solely on RMS or MAV.

3.3 Feature-level Fusion Analysis using CCA Methods

CCA is employed for feature-level fusion, transforming EMG and force features into canonical variates that maximize their correlation. This analysis is crucial for understanding the contributions of original features to the fused representation and demonstrating the effectiveness of combining complementary information.

CCA was selected for its distinct advantages over other fusion methods. It provides optimal linear correlation mapping by identifying latent, shared structures between EMG and force, resulting in a fused representation that captures mutual information and leads to better class separability. Furthermore, CCA ensures dimensionality reduction with relevance. Unlike Principal Component Analysis (PCA), which is an unsupervised method, CCA ensures that the retained components are maximally relevant across both modalities, thereby effectively eliminating noise and irrelevant variance. Ultimately, CCA was chosen because it enables the projection of modality-specific features into a shared subspace of maximum correlation, resulting in a more informative and compact feature set. Figure 11 illustrates the importance of original features to the first and second canonical variates, respectively, using the WL feature, as demonstrated in the previous section.



(a) (b)

Figure 11: The importance of WL features for (a) first canonical variate, (b) second canonical variate using CCA.

Based on Figure 11(a), the force feature shows the highest contribution (≈1.0) to the first canonical variate, while all EMG channels (FDS, FCR, EDC, ECRL) show the minimum contribution. This indicates that force data is the dominant source of information for capturing the general patterns of hand movement and is primarily responsible for the most significant correlation between the EMG and force signals. This aligns with its strong performance in the unimodal analysis, where it effectively captured the main kinetic output of the movements.

In contrast, as depicted in Figure 11(b), the FCR channel stands out with the highest EMG contribution (≈0.8) to the second canonical variate, followed by ECRL, EDC, and then force. This suggests that specific EMG channels, particularly FCR, ECRL, and EDC, provide valuable complementary information, especially for detecting more refined or specific muscle activations that are not fully represented by the first canonical variate. These findings highlight that while force features are effective for capturing major movements, EMG features significantly enhance the model's sensitivity to precise motor activity and refined characteristics of hand movements.

As a result, the multimodal (EMG and force) effectively integrates both types of features at the feature level through CCA. It combines the dominant force signals with the most informative EMG signals to create a more robust input for classification [24]. This fusion, utilizing the strengths of both modalities, significantly improves the ability to accurately predict diverse hand movements by capturing both the kinetic patterns and the neuromuscular activations.

3.4 Comparative Evaluation of Hybrid Multimodal Fusion Framework

Table 3 presents a comparative analysis of unimodal, simple multimodal concatenation, and the proposed hybrid multimodal fusion framework across three features: RMS, MAV, and WL features. The unimodal EMG results demonstrate the lowest performance among all modalities, with accuracies between 79.06% and 82.69% and F1-scores in the range of 0.77–0.83. These values indicate that EMG signals alone do not provide sufficient accuracy for robust classification. In contrast, force-only features achieved significantly higher performance. The ANN classifier consistently outperformed others, with accuracies reaching 96.15% (RMS), 96.37% (MAV), and 96.58% (WL), and corresponding F1-scores around 0.96. These results highlight the strength of force modality as a reliable predictor of hand movement.

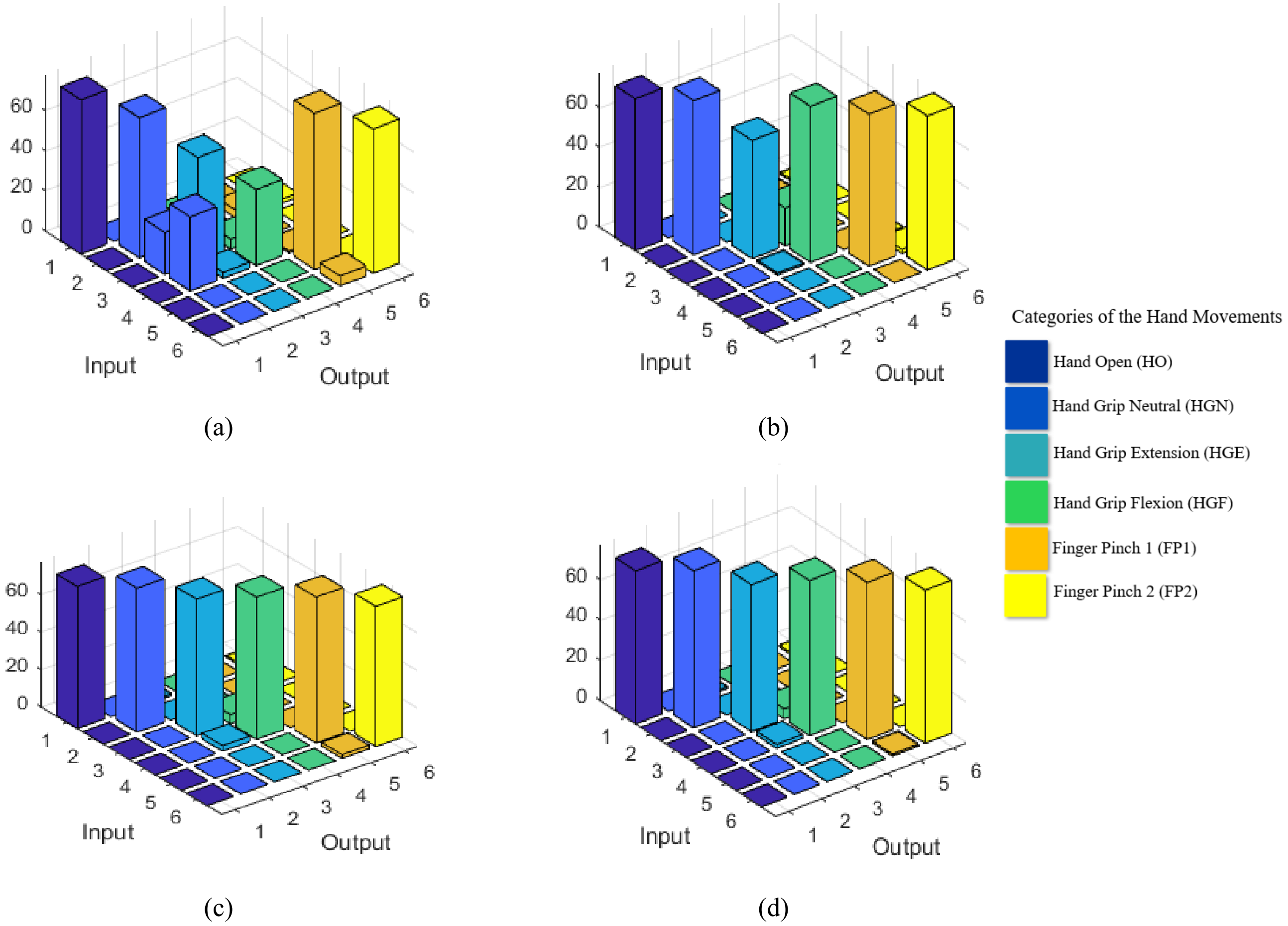
The concatenation of EMG and force features improved performance relative to EMG-only across all features. However, the level of improvement was feature-dependent. For RMS and MAV, accuracies ranged from 83.60% to 91.35%, depending on the classifier, and remained slightly lower than the best force-only baseline. WL features produced the strongest concatenation results, with SVM and ANN achieving 96.31% and 95.51%, respectively, which are comparable to force-only. Thus, concatenation helped balance both modalities but did not guarantee consistent gains over force-only features.

The proposed hybrid multimodal fusion framework, combining CCA at the feature level and majority voting at the decision level, produced the most consistent high performance [25]. Hybrid multimodal fusion framework accuracies reached 94.87% (RMS), 95.51% (MAV), and 97.86% (WL), with F1-scores between 0.9486 and 0.9787. Compared with simple concatenation, the hybrid multimodal fusion framework yielded clear improvements: +3.78% for RMS, +4.16% for MAV, and +1.55% for WL. These gains show that aligning modalities through CCA before classification enables more effective information fusion than direct concatenation. This indicates that the hybrid multimodal fusion framework is particularly advantageous when both EMG and force contribute complementary information.

Table 3: Comparative Classification Performance of Individual Classifiers and Hybrid Multimodal Fusion Framework Across Sensor Modalities Features

Sensor Fusion Modality	Features	LDA Classifier		SVM Classifier		ANN Classifier		Hybrid Multimodal Fusion Framework	
		Accuracy (%)	F1-Score	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
EMG-Only	RMS	79.06	0.79	73.29	0.7318	77.14	0.7712	-	-
	MAV	80.77	0.81	73.08	0.7315	78.63	0.7817	-	-
	WL	82.69	0.83	77.35	0.7823	82.26	0.8166	-	-
Force-Only	RMS	94.66	0.95	89.10	0.8885	96.15	0.9617	-	-
	MAV	94.66	0.95	93.59	0.9358	96.37	0.9637	-	-
	WL	94.66	0.95	89.10	0.8885	96.58	0.9660	-	-
Multimodal (simple concatenating method)	RMS	84.24	0.83	91.09	0.91	90.38	0.90	-	-
	MAV	83.60	0.82	87.14	0.87	91.35	0.91	-	-
	WL	88.39	0.88	96.31	0.96	95.51	0.96	-	-
Hybrid Multimodal Fusion Framework	RMS	-	-	-	-	-	-	94.87	0.9486
	MAV	-	-	-	-	-	-	95.51	0.9551
	WL	-	-	-	-	-	-	97.86	0.9787

The analysis confirms three important findings. First, EMG alone is insufficient for robust classification, while force signals are highly reliable. Second, multimodal concatenation improves over EMG-only but cannot consistently surpass the force-only modality. Third, the hybrid multimodal fusion framework offers stable performance across features and delivers the best results for WL. Figure 12 illustrates the confusion matrices for the hybrid multimodal fusion framework and the baselines, using out-of-sample validation.



(a) (b)

(c) (d)

Figure 12: Confusion matrices for different classification approaches using WL feature: (a) LDA classifier for unimodal EMG-only, (b) LDA classifier for unimodal Force-only, (c) LDA classifier for multimodal (simple concatenating method), and (d) hybrid multimodal fusion framework.

The slight reduction in RMS and MAV performance under the hybrid multimodal fusion framework may be due to two factors: (1) CCA focuses on shared variance across modalities, which can suppress force-specific discriminative cues, and (2) simple majority voting does not account for the relative strengths of individual classifiers, allowing weaker classifiers to dilute the influence of stronger ones. Future work could address these limitations by using weighted voting or stacked generalization with probability averaging, which would allow stronger classifiers, such as ANN, to contribute more influence in the final decision.

Overall, the results validate the effectiveness of the proposed hybrid multimodal fusion framework, particularly for WL features where both EMG and force contribute unique discriminative power. This highlights the potential of multimodal fusion strategies in improving classification robustness for human movement prediction tasks.

4.0 CONCLUSION AND FUTURE WORK

This study shows the significance of the implementation of the hybrid multimodal fusion framework as compared to three baselines for enhanced prediction of hand movements. By combining EMG and force signals through CCA at the feature-level, followed by decision-level majority voting, a significant boost in classification performance is achieved. The hybrid multimodal fusion framework yielded an impressive overall accuracy of 97.86%. This enhancement underscores the effectiveness of CCA in improving feature discriminability and the role of majority voting in refining predictions. By integrating muscle electrical activity with physical force, this method enables the development of more accurate and reliable movement recognition systems, essential for applications in prosthetic control, human-machine interfaces, and rehabilitation technologies.

Despite these promising results, this study has several limitations. While ten healthy subjects were initially recruited for data collection, the rigorous data-quality screening necessary for signal integrity resulted in a refined cohort for the final analysis. A larger, more diverse sample size is required to further increase statistical power and account for the high degree of inter-subject biological variability inherent in EMG and kinetic signal patterns. Furthermore, the current cohort lacks demographic and clinical diversity, specifically regarding the inclusion of female participants. The study also did not include physically impaired individuals or amputees, meaning the findings have limited generalizability to broader clinical populations. Additionally, while the proposed hybrid multimodal fusion framework achieved a high accuracy, it utilized classical machine learning models; future research could explore deep learning architectures to further enhance the system's adaptability to a wider range of users.

Future work will focus on three key areas to advance this research. Firstly, the integration of advanced Artificial Intelligence (AI) models, specifically deep learning architectures such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), will be explored for multimodal sensor data. These AI-driven approaches aim to enhance model sophistication, capture complex temporal signal patterns, and improve overall system robustness. Secondly, to improve generalizability and clinical relevance, the sample size will be expanded (targeting a minimum of 20 subjects) to include a more diverse demographic and clinical participants, and cross-subject validation will be performed. Lastly, additional sensor combinations, such as EMG with IMUs, will be investigated to broaden the scope and applicability of the movement prediction system. These steps aim to provide novel methodological contributions and further advance the state of the art.

ACKNOWLEDGEMENT

The authors acknowledge funding support from the Ministry of Higher Education (FRGS/1/2022/TK07/UTEM/03/13) for the financial aid under the Fundamental Research Grant Scheme, Universiti Teknikal Malaysia Melaka, for the research support provided, and the Rehabilitation Engineering and Assistive Technology (REAT) research group for the facilities provided.

REFERENCES

1. Shin J, Miah ASM, Konnai S, Takahashi I, Hirooka K. Hand gesture recognition using sEMG signals with a multi-stream time-varying feature enhancement approach. Sci Rep [Internet]. 2024 Dec 1 [cited 2025 May 14];14(1):22061. Available from: https://www.nature.com/articles/s41598-024-72996-7.
2. Fang C, He B, Wang Y, Cao J, Gao S. EMG-Centered Multisensory Based Technologies for Pattern Recognition in Rehabilitation: State of the Art and Challenges. Biosensors 2020, Vol 10, Page 85 [Internet]. 2020 Jul 26 [cited 2023 Aug 25];10(8):85. Available from: https://www.mdpi.com/2079-6374/10/8/85/htm.
3. Zhang X, Chen X, Li Y, Lantz V, Wang K, Yang J. A Framework for Hand Gesture Recognition based on Accelerometer and EMG Sensors. IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans. 2011 Nov;41(6):1064–76.
4. Duan S, Wu L, Xue B, Liu A, Qian R, Chen X. A Hybrid Multimodal Fusion Framework for sEMG-ACC-Based Hand Gesture Recognition. IEEE Sens J. 2023 Feb 1;23(3):2773–82.

5. Chen P, Li Z, Togo S, Yokoi H, Jiang Y. A Layered sEMG-FMG Hybrid Sensor for Hand Motion Recognition from Forearm Muscle Activities. IEEE Trans Hum Mach Syst. 2023 Oct 1;53(5):935–44.
6. Song X, van de Ven SS, Chen S, Kang P, Gao Q, Jia J, et al. Proposal of a Wearable Multimodal Sensing-Based Serious Games Approach for Hand Movement Training After Stroke. Front Physiol. 2022 Jun 3;13.
7. Zhou H, Zhang Q, Zhang M, Shahnewaz S, Wei S, Ruan J, et al. Toward Hand Pattern Recognition in Assistive and Rehabilitation Robotics Using EMG and Kinematics. Front Neurorobot [Internet]. 2021 May 13 [cited 2025 Jan 10];15. Available from: https://pubmed.ncbi.nlm.nih.gov/34054455/.
8. Vortmann LM, Ceh S, Putze F. Multimodal EEG and Eye Tracking Feature Fusion Approaches for Attention Classification in Hybrid BCIs. Front Comput Sci [Internet]. 2022 Mar 21 [cited 2025 May 15];4:780580. Available from: www.frontiersin.org.
9. Suhaimi MM, Ghazali AS, Mohideen AJH, Hafizalshah MH, Sidek SN. Enhancing Prosthetic Control: Neural Network Classification of Thumb Muscle Contraction Using HD-sEMG Signals. IIUM Engineering Journal [Internet]. 2024 Jul 14 [cited 2025 May 20];25(2):338–49. Available from: https://journals.iium.edu.my/ejournal/index.php/iiumej/article/view/3029.
10. Gohel V, Mehendale N. Review on Electromyography Signal Acquisition and Processing. Biophys Rev [Internet]. 2020 Dec 1 [cited 2024 Dec 26];12(6):1361. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7755956/.
11. Abas N, Bukhari WM, Abas MA, Tokhi MO. Electromyography Assessment of Forearm Muscles: Towards the Control of Exoskeleton Hand. 2018 5th International Conference on Control, Decision and Information Technologies, CoDIT 2018. 2018 Jun 22;822–8.
12. Binti Abas N. Modelling and EMG based Control of Upper Limb Exoskeletons for Hand Impairments. University of Sheffield; 2019.
13. Manikanta DCS, Gowtham G, Gantasala K. Implementation of Feature Extraction of Neuro Muscular EMG Signal. In: 2nd IEEE International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering, ICATIECE 2022. Institute of Electrical and Electronics Engineers Inc.; 2022.
14. Arland F, Setiawan AW, Adzkia M. Multi-Feature Extraction EMG For Classification of Finger Movement. In: Proceedings of the International Conference on Electrical Engineering and Informatics. Institute of Electrical and Electronics Engineers Inc.; 2023.
15. Moctar SMS, Rida I, Boudaoud S. Time-domain Features for sEMG Signal Classification: A Brief Survey. 2023 Jun 1 [cited 2024 Dec 12]; Available from: https://hal.science/hal-04199535.
16. Prasad VVKD V., Nagasirisha B, Janitha JY, Venkatesh NR, Lalithadithya NSB, Ramya T. Feature extraction and classification of different hand movements from the emg signal using linear discriminant analysis classifier. i-manager's Journal on Electronics Engineering. 2024;14(2):19.
17. Samuri SM, Nova TV, Bahbibirahmatullah, Li WS, Al-Qaysi ZT. Classification Model for Breast Cancer Mammograms. IIUM Engineering Journal [Internet]. 2022 Jan 4 [cited 2025 May 20];23(1):187–99. Available from: https://journals.iium.edu.my/ejournal/index.php/iiumej/article/view/1825.
18. Hamim M, Moudden I El, Ouzir M, Moutachaouik H, Hain M. A Novel Dimensionality Reduction Approach to Improve Microarray Data Classification. IIUM Engineering Journal [Internet]. 2021 Jan 4 [cited 2025 May 20];22(1):1–22. Available from: https://journals.iium.edu.my/ejournal/index.php/iiumej/article/view/1447.
19. Eddy E, Campbell E, Bateman S, Scheme E. Big Data in Myoelectric Control: Large Multi-user Models Enable Robust Zero-shot EMG-based Discrete Gesture Recognition. 2024;.
20. Kabanov AA. Application of Support Vector Machines to the Multiclass Classification Electromyography Signal Patterns. Proceedings of the 2021 15th International Scientific-Technical Conference on Actual Problems of Electronic Instrument Engineering, APEIE 2021. 2021;92–5.
21. Ahmed J, Ahmed M. Online News Classification using Machine Learning Techniques. IIUM Engineering Journal [Internet]. 2021 Jul 4 [cited 2025 May 20];22(2):210–25. Available from: https://journals.iium.edu.my/ejournal/index.php/iiumej/article/view/1662.
22. Abas N, Ganasegaran K, Ghani NMA, Kassim AM, Abas MA. Enhancing Control Strategies for Hand Exoskeletons Through Modeling and Electromyogram-based Control. Proceedings of the International Joint Conference on Neural Networks [Internet]. 2024 [cited 2025 Dec 3]; Available from: https://ieeexplore.ieee.org/document/10651023.
23. Seng CH, Abas N, Kasdirin HA, Abas MA, Ghani NA, Hanafi AN. EMG-based Assessment Device for Hand Rehabilitation with Cloud Analysis. Proceedings - 12th IEEE International Conference on Control, Automation and Information Sciences, ICCAIS 2023 [Internet]. 2023 [cited 2025 Dec 3];763–8. Available from: https://ieeexplore.ieee.org/document/10382377.
24. Rohr M, Haidamous J, Schafer N, Schaumann S, Latsch B, Kupnik M, et al. On the Benefit of FMG and EMG Sensor Fusion for Gesture Recognition using Cross-subject Validation. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2025;.
25. Patil P, Kale G, Bivalkar N, Kolhatkar A. Comparative Analysis of Weighted Ensemble and Majority Voting Algorithms for Intrusion Detection in OpenStack Cloud Environments. International Journal of Advanced Computer Science and Applications [Internet]. 2023 Dec 29 [cited 2025 Sep 27];14(12):741–7. Available from: https://thesai.org/Publications/ViewPaper?Volume=14&Issue=12&Code=IJACSA&SerialNo=76.