





Article

Evaluation of Finger Movement Impairment Level Recognition Method Based on Fugl-Meyer Assessment Using Surface EMG

Adhe Rahmatullah Sugiharto Suwito P¹, Ayumi Ohnishi¹, Yudith Dian Prawitri², Riries Rulaningtyas³,
Tsutomu Terada^{1,*} and Masahiko Tsukamoto¹

¹ Graduate School of Engineering, Kobe University, 1-1 Rokkodaicho, Nada, Kobe 657-8501, Japan; ohnishi@eedept.kobe-u.ac.jp (A.O.); tuka@kobe-u.ac.jp (M.T.)

² Department of Physical Medicine and Rehabilitation, Faculty of Medicine, Airlangga University, Airlangga University Hospital, Surabaya 60115, Indonesia

³ Biomedical Engineering Study Program, Department of Physics, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia; riries-r@fst.unair.ac.id

* Correspondence: tsutomu@eedept.kobe-u.ac.jp

Abstract: Subjectivity has been an inherent issue in the conventional Fugl-Meyer assessment, which has been the focus of impairment-level recognition in several studies. This study continues our previous work on the use of EMG to recognize finger movement impairment levels. In contrast to our previous work, this study provided a better and more reliable recognition result with improved experimental settings, such as an increased sampling frequency, EMG channels, and extensive patient data. This study employed two data processing mechanisms, inter-subject cross-validation (ISCV) and data-scaled inter-subject cross-validation (DS-ISCV), resulting in two evaluation methods. The machine learning algorithms employed in this study were SVM, random forest (RF), and multi-layer perceptron (MLP). MLP_ISCV achieved the highest average recall score of 0.73 across impairment levels in the spherical grasp task. Subsequently, the highest average recall score of 0.72 among non-majority classes was achieved by SVM_DS-ISCV in the mass extension task. The cross-validation result shows that the proposed method effectively handled the imbalanced dataset without being biased toward the majority class. The proposed method demonstrated the potential to assist doctors in clarifying the subjective assessment of finger movement impairment levels.

Keywords: electromyography; finger movement; Fugl-Meyer assessment; imbalance data; impairment level; post-stroke patients; recognition



Citation: P, A.R.S.S.; Ohnishi, A.; Prawitri, Y.D.; Rulangingtyas, R.; Terada, T.; Tsukamoto, M. Evaluation of Finger Movement Impairment Level Recognition Method Based on Fugl-Meyer Assessment Using Surface EMG. *Appl. Sci.* **2024**, *14*, 10830. <https://doi.org/10.3390/app142310830>

Academic Editors: Federica Verdini, Alessandro Mengarelli, Andrea Tigrini, Francesca Lunardini and Jesús Tornero

Received: 24 September 2024
Revised: 16 November 2024
Accepted: 18 November 2024
Published: 22 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Stroke is one of the major global health issues, with over 13 million new cases annually and representing the second leading cause of mortality and disability worldwide. According to the Global Burden of Disease Study (GBD) in 2016, there was an increase in frequency among younger groups (under 50 years old), depicting the distribution of incidence across ages. Moreover, the incidence of stroke, along with stroke-related mortality and disability, increased nearly two-fold from 1990 to 2016 [1]. According to the consequences of disability, muscle dysfunction is the predominant form of impairment following a stroke. Muscle dysfunction significantly worsens the risk of arm paralysis after a stroke and is frequently associated with increased impairment, reduced work capacity, and diminished quality of life. The accumulation of inactive muscle fibers due to dysfunction leads to abnormal muscle activation patterns, such as spastic muscle contractions, which may contribute to further muscle-related disorders [2,3]. Consequently, individuals who have experienced a stroke affecting their arm may struggle to perform daily activities [4].

In order to restore the function of muscle fiber after a stroke, rehabilitation is necessary as a routine muscle recovery process [3,5]. Rehabilitation refers to the combined and coordinated use of medical, social, educational, and occupational measures to retrain

a person to the highest functional skills level [5]. The rehabilitation process includes assessing the patient's impairment condition, followed by further medication, training, and reassessment. The assessment is intended to observe and evaluate the patient's actual condition. Thus, a suitable training and medication program can be specifically arranged for each patient. In this regard, the assessment process is essential, as it determines the patient's actual condition at the beginning of the rehabilitation program and after several medication and training processes.

Several clinical methods to assess the impairment level of post-stroke patients are currently being employed worldwide. Regarding the assessment of motorfunction, the Fugl-Meyer assessment (FMA) is a notable tool, as it provides a detailed assessment protocol with a scoring system in most human extremities. The Fugl-Meyer scale is a groundbreaking quantitative evaluative tool to measure sensorimotor stroke recovery. The motor-domain function of this tool encompasses large and small parts of the upper-lower extremities, suggesting the most comprehensive measures of motor impairment following stroke. One of the assessment components the assessment a complex motor function, such as finger movement, that is composed of a combination of complex muscles [3]. In the case of finger movement, the Fugl-Meyer assessment of the hand's upper extremity (FMA-UE) is typically selected. As shown in Table 1, this questionnaire consists of seven finger movement tasks, namely Mass Extension (ME), Mass Flexion (MF), Hook Grasp (HG), Thumb Adduction (TA), Pincer Grasp (PG), Cylinder Grasp (CG), and Spherical Grasp (SG). Meanwhile, the impairment level consists of three conditions (i.e., full, partial, and none).

Table 1. Motor-domain function of Fugl-Meyer assessment for finger movement [3].

Movement	Full (2)	Partial (1)	None (0)
Mass Extension	Full active extension	Some but not active extension	No extension
Mass Flexion	Full active flexion	Some but not active flexion	No flexion occurs
Hook Grasp	Maintains position against resistance	Can hold position but weak	Cannot be performed
Thumb Adduction	Can hold paper against a tug	Can hold paper but not against a tug	Cannot be performed
Pincher Grasp	Can hold a pencil against a tug	Can hold pencil but not against a tug	Cannot be performed
Cylinder Grasp	Can hold a cylinder against a tug	Can hold a cylinder but not against a tug	Cannot be performed
Spherical Grasp	Can hold a ball against a tug	Can hold a ball but not against a tug	Cannot be performed

The doctor or physiotherapist has an inherent role in the assessment. Several finger movements in FMA-UE are object-dependent tasks that require a patient to hold an object while the doctor tugs it. This mechanism in the Fugl-Meyer assessment is the conventional method that requires the doctor to observe the impairment level manually through visual and tug inspection. Nevertheless, the inherent role of a doctor in the conventional method raises another issue, i.e., inherent subjectivity, which also promotes the subjectivity of the assessment result. Consequently, this issue influences the accuracy and repeatability of the assessment score, which is particularly essential for repeated assessments in the rehabilitation process.

Considering a potential modality to address the subjectivity issue in FMA, electromyography (EMG) is a golden standard for objectively assessing muscle and nervous system function. It is widely used in rehabilitation and gesture recognition studies [6,7]. However, few studies have examined recognition performance in EMG-related subjective levels.

Our earlier study examined the performance of the EMG-based impairment-level recognition method in post-stroke patients for finger movement [8]. We reported good recognition results on several FMA-based finger movement tasks, highlighting the promising performance of EMG. We also pointed out several crucial areas for improvement in the

experimental settings, as the number of patients was only four, with incomplete impairment levels collected, consisting of full and partial levels only. Additionally, there were only two EMG channels, and the employed sampling frequency was too small. The occurrence of imbalanced dataset conditions could be addressed with a standard resampling method with a SMOTE filter algorithm [9]. The employed machine learning methods were Support Vector Machine (SVM) and Random Forest (RF), with average accuracies of 67.1 and 64.6, respectively. However, the score could not be generalized due to critical issues in the experimental settings and accuracy bias due to the imbalanced dataset. Consequently, the machine learning models were at risk of recognizing the participant personally instead of the impairment levels. Another highlighted point is the tendency of imbalance of the collected impairment level target, where one patient can only exhibit one or two levels without certain information prior to assessment. Therefore, an improvement in the experimental settings is required.

This study focused on improving the recognition of finger movement impairment levels in the FMA, aiming to reduce the issue of subjectivity. This study also addressed the subjective assessment issue of FMA and the doctor's perspective according to the performance of the machine learning model. The final aim of this study is to assist doctors in deciding patients' actual impairment levels in finger movement. Therefore, this study also introduced the implementation of the constructed method in a desktop application that can recognize and display the impairment level of each finger movement and output an assessment video. Thus, a doctor or physiotherapist can double check and evaluate the impairment level separately.

2. Related Research

2.1. Automation of Fugl-Meyer Assessment

Subjectivity and inflexibility are the most crucial issues in the Fugl-Meyer assessment. Several studies have been conducted to overcome these inherent problems. In 2014, a study was conducted on 24 patients with various upper-extremity hemiplegia levels after stroke [10]. Support vector regression (SVR) was employed to predict the FMA score on shoulder–elbow tasks. Another study proposed an automatic acquisition system for one task, pincer grasp, from the FMA of the upper extremity of the hand [11]. Several studies proposed a recognition system that employs complete FMA hand/finger movement tasks of MF, ME, HG, TA, PG, CG, and SG [12,13]. Lee et al. constructed a grasping object for each hand-function task [13], while Formstone et al. employed a muscle-related modality [12]. The studies mentioned above utilized different approaches to address the inherent subjectivity of the FMA, especially in hand-function tasks due to the complex movement of the fingers and corresponding muscles. However, most of the studies mentioned above utilized supervised machine learning as the recognition method, and the learning target was the doctor's subjective assessment. Furthermore, the doctor's final decision was to perform further rehabilitation procedures following the assessment. Thus, the subjectivity of the FMA could not be eliminated due to the inherent role of a doctor.

The aforementioned studies imply that the subjectivity issue of the FMA can be minimized by conducting multiple assessments. Given the inherent subjectivity in individual patient assessments, repetitive assessments minimize uncertainty and reinforce the doctor's judgment. Consequently, the repetitive assessment of numerous patients enhances the reliability of assessment outcomes. Furthermore, integrating sensor technology to capture context information relating to impairment levels can increase the precision of a doctor's assessments. This highlights the possibility of achieving accurate subjective assessments by gathering extensive training data for a machine learning model.

2.2. Employable Sensors in Finger Movement Impairment-Level Recognition of the Fugl-Meyer Assessment

Several approaches have been applied to address inherent subjectivity in the FMA. Depending on the task of the FMA, appropriate modalities should be utilized. The assessment

of non-hand tasks is simple, as they only comprise a single and long extremity part, such as an arm. Thus, any sensor has the potential to record the FMA task activity from this part of the body. Several studies have introduced the idea that the movement activity of non-hand FMA tasks could be captured using proprioceptive and camera-based modalities such as accelerometer, Force Sensing Resisto (FSR), and Kinect sensors [10,12,13].

Nevertheless, employing the same sensors in the hand tasks of the FMA is less appropriate, as the corresponding tasks involve the fingers, which are small and short extremities. Therefore, a specific or customized sensor is needed to capture the movement activity of the FMA's hand tasks. For instance, two studies constructed grasping tools embedded with FSR sensors to capture the activity of the fingers, especially in pincher grasping [11,13]. The studies indicated that a finger requires a customized sensor to capture its complex movement mechanism due to its small size. However, a customized grasping object embedded with a sensor is limited in terms of the type of movement and the size of the finger.

Another utilizable sensor in the FMA hand tasks is EMG sensors. EMG enables the recording of the complex movement of fingers by capturing the electrical activity of muscles that control the fingers [8,12]. Furthermore, EMG can capture several channels, facilitating the recording of muscle activities of individual finger movements, regardless of the various sizes of fingers and the types of movements. Accordingly, EMG has prominent potential to capture finger activity of the FMA's hand task, as it addresses the limitation of the aforementioned customized grasping object embedded with sensors.

2.3. Data Imbalance Nature in an Actual Experiment Involving the Collection of Patient Data

An imbalanced dataset is a common problem, especially in the medicine and rehabilitation fields. This condition occurs when one or more groups have an extensive amount of data while another is limited [14–16]. Several studies have shown the inherent nature of the dataset imbalance problem in experiments with actual participants or patients. Vijayvargiya et al. encountered an imbalance between healthy and unhealthy participants in a dataset used for knee abnormality detection [14]. Choi et al. proposed an abnormal electrocardiography (ECG) classification method to address the dataset imbalance problem [15]. Hasni et al. developed a classification algorithm based on SVM to address the imbalance between normal and neuropathy–myopathy samples in an EMG dataset [16]. The studies mentioned above showed that imbalances are evident in datasets when experimenting with normal and abnormal participants.

The FMA is prone to produce imbalanced datasets. Ideally, a dataset will have balanced data for each impairment level. In actual situations, the number of data depends on the availability of patients in a rehabilitation facility. This situation highlights the tradeoff in terms of the number of patients across different impairment levels in data collection experiments, with some levels having an abundance of patients and others fewer. Additionally, a patient may exhibit multiple levels of impairment in repetitive tasks, which may pose a risk of reducing the number of data for minority classes. Furthermore, this condition limits further evaluation of patient characteristics if information such as age and gender is imbalanced.

This study addresses the subjectivity issue of the FMA by proposing an improved method for recognizing finger movement impairment levels. By fulfilling the identified research gap, the method utilized a higher sampling frequency of EMG data and incorporated more EMG channels to capture the various muscles involved in finger movement. An extensive number of patients was recruited to enhance the reliability of the machine learning model, resulting in improved recognition performance. Additionally, the study highlighted the challenges of an imbalanced dataset and suggested effective processing methods for the collected data.

3. Methods

In this section, we propose an improved recognition method to detect the impairment level of finger movement based on the Fugl-Meyer assessment in a non-ideal condition

of imbalanced datasets. The improved recognition method is intended to benefit the system's reliability and reproducibility in addressing the subjectivity issue of the FMA. By constructing the system under the non-ideal condition of an imbalanced dataset, the system adapts to the actual assessment condition in rehabilitation facilities, thereby enabling it to assist doctors and efficiently minimize the subjective nature of the assessment.

3.1. Participants

Prior to recruiting the participants, several criteria had to be strictly followed. All participants had to be partially impaired in one arm and be able to maintain a sitting position either on a chair or in a wheelchair by themselves. According to the assessment of a rehabilitation doctor, the participants had undergone at least three months of a rehabilitation program. The participants were all 18 years old or older on the day of the experiment. The participants were not congenital stroke patients and had no permanent finger injuries. The participants did not have any cognitive impairments and were capable of understanding the doctor's instructions. After fulfilling the inclusion and exclusion criteria, the participants were selected based on a random sampling process. The participants, who never underwent a Fugl-Meyer assessment, were requested to participate in the experiment on the day of their rehabilitation program in the Airlangga University Hospital.

In this study, the participants comprised 28 stroke patients (17 males and 11 females). All experiment procedures were explained to the participants and their relatives, and they voluntarily agreed to participate. All participants sat on a chair or wheelchair alone or with the help of relatives or clinicians. Prior to the experiment, each participant was explained all of the experimental protocols and learned the FMA-based finger movement, as shown in Table 1. The ethical committee of Airlangga University Hospital, Surabaya, Indonesia (No. 125/KEP/2023) approved all participant selection criteria and experimental procedures.

3.2. Instrumentation

This study employed four channels of the EMG sensors of two TSND151 and AMP151 sensors (ATR-Promotion Inc., Kyoto, Japan). The TSND151 is a compact, wireless, multi-function sensor embedded with inertial measurement unit sensors (IMUs) and an external terminal 16-bit analog–digital input. Regarding EMG measurement, the external terminal was connected to the AMP151, an extended amplifier specified for biological signals. The sampling frequency was set to a maximal setting of 1000 Hz with $1000\times$ signal amplification. The AMP151 has a common mode rejection ratio exceeding 90 dB, an input impedance of 200 G Ω , and a power supply rejection ratio above 105 dB. A rectangular surface electrode of 19×38 mm in size was utilized. The electrode featured a polymer gel and Al/AgCl electrode material. Other instruments were used in relation to the FMA task for finger movement, such as a pen, a piece of paper, a cylinder, and a tennis ball.

3.3. Electrode Attachment

In this study, we improved the data collection process by capturing EMG signals from three extrinsic and one intrinsic muscle activity. In contrast, the previous experiment only captured two extrinsic forearm muscle activities [8]. The attachment locations of the electrodes were the extensor digitorum muscle (Channel 1), flexor digitorum muscle (Channel 2), extensor pollicis brevis muscle (Channel 3), and flexor pollicis brevis muscle (Channel 4). Channels 1–3 corresponded to extrinsic muscles, and channel 4 corresponded to intrinsic muscle [17]. The utilization of intrinsic muscle must be considered due to its potential to provide a prominent EMG signal of a specific hand finger. The muscle of channel 4 was selected due to its high relative importance in providing an EMG signal of pinch-related movement [18]. The locations of muscle attachment are shown in Figure 1.

Prior to attaching the electrodes, an alcohol swab was applied to cleanse the skin. The electrodes were arranged so that the inter-electrode distance of an EMG channel was approximately 20 mm. Additionally, the electrodes were attached in a parallel orientation with respect to the muscle. These settings were utilized to obtain a good-quality EMG signal

and minimize the occurrence of cross-talk [19]. The positions of the employed muscles were not adjacent to each other, avoiding cross-talk interaction between the employed muscles, as shown in Figure 1. In this study, the doctor guided the attachment of the electrode to find the best position concerning the target muscles. Hence, a good-quality EMG signal could be obtained. The subtle cross-talk interaction of the adjacent muscles with the employed muscle might have still occurred due to a muscle coordination event. However, given the anatomical positioning and functional roles of the employed muscles, this condition may be deemed negligible, as there is unlikely a cross-talk interaction among them [17,18].

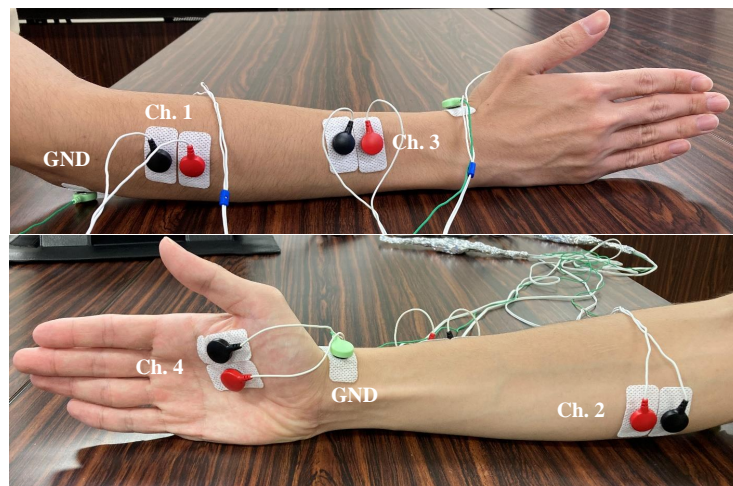


Figure 1. Electrode attachment positions.

3.4. Data Processing Flow

The data were processed to be prepared as machine learning input. The flow encompassed several steps: signal filtering, scaling, movement event exporting, and feature extraction. Following the feature extraction process, two mechanisms were employed to evaluate the performance of machine learning models. The first mechanism was inter-subject cross-validation, hereinafter referred to as ISCV, which consisted of a data resampling process to address imbalanced dataset conditions. Subsequently, the data were directly deployed in machine learning after the feature extraction step. Another mechanism that was employed is data-scaled inter-subject cross-validation, hereinafter referred to as DS-ISCV, which added data scaling before the classification process. The signal filtering process until feature extraction was performed individually for each participant’s data. Subsequently, the classification processes of both mechanisms were performed using inter-subject cross-validation. Finally, we observed and analyzed the classification outcomes of both mechanisms for each FMA-based movement task. The data process flow is shown in Figure 2.

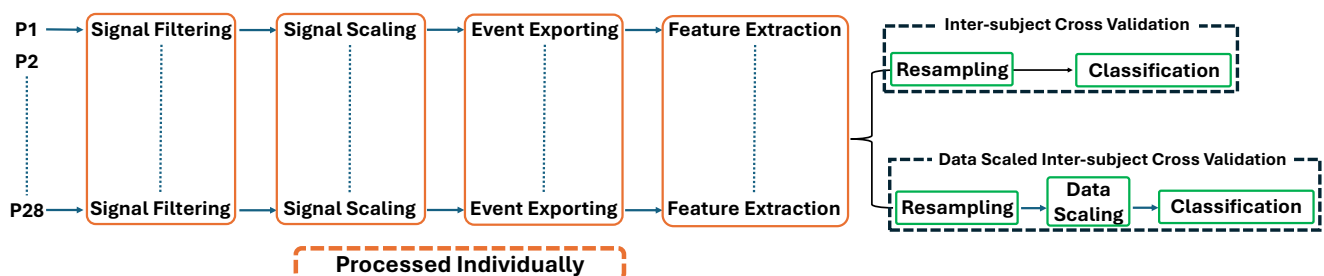


Figure 2. Diagram of data processing flow.

3.4.1. Signal Filtering

Motion artifacts, baseline, and power-line interference are primary noise sources in EMG signals [20,21]. The frequency of power-line interference is typically 50 Hz, depending

on the location. The range of baseline and motion disturbances was between 0 and 20 Hz. Power-line interference also tends to induce harmonic noises that yield high spikes at multiples of 50 Hz, which are observable in the frequency domain. A typical process used to address this is a signal-filtering technique. In the previous experiment, a Butterworth high-pass filter was able to generate a clean signal [8]. Nevertheless, conventional digital filters still leave undesirable noise within non-muscle-contracting signals such as a relaxed event.

One commonly used digital filtering method is wavelet denoising (WD). In contrast to Butterworth filters, which produce flat or smooth frequency responses, WD applies wavelet transform to decompose the signal into frequency components with matched resolutions [22]. Consequently, unlike Butterworth filters, WD offers superior time-frequency localization, ensuring the retention of precise signal features while minimizing noise. The process of wavelet denoising involves an initial decomposition of the signal through a wavelet transform (WT), followed by the application of appropriate thresholds to the detail coefficients. This step entails setting all coefficients below the associated thresholds to zero. Subsequently, the denoised signal is reconstructed based on the modified detail coefficients [23].

Parameter selection for wavelet denoising is essential. Preserving the muscle-contracting amplitude of the EMG signal is mandatory when eliminating undesirable noises. Additionally, addressing the non-stationary behavior of EMG signals remains a challenge. Thus, selecting incorrect parameters may result in poor noise removal or the elimination of essential amplitudes. A previous study investigated the optimum wavelet function to identify and denoise EMG signals [24]. According to the study findings, the Daubechies1 (db1 or haar) wavelet with a hard transformation at 0 dB signal-to-noise ratio (SNR) achieved the best denoising performance. In this study, we combined the Butterworth filter and wavelet denoising techniques. We employed high-pass and bandstop Butterworth filters to eliminate the specific frequencies of the noises. Wavelet filters with a db1 function, hard transformation, and level 1 decomposition were employed to eliminate the remaining noises within the relaxed event signal. Figure 3 illustrates the result after the signal filtering process.

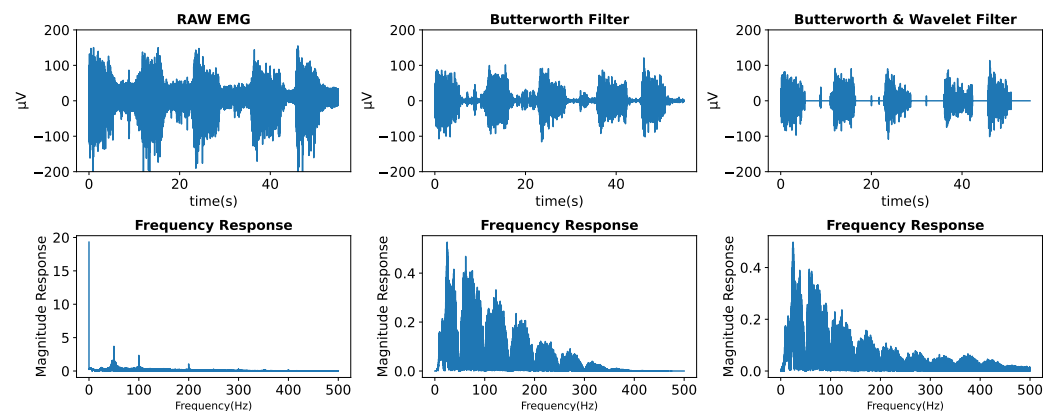


Figure 3. EMG signal-filtering process.

3.4.2. Signal Scaling

An EMG signal represents the combined motor unit action potentials during contraction, recorded at a specific electrode location. The surface EMG voltage potential is greatly influenced by various factors, which differ among individuals and may also change over time within the same individual. Consequently, the amplitude of an EMG signal is ineffective for group comparisons or monitoring over extended durations [25]. Subsequently, several studies have shown that a scaling method such as signal normalization or standardization can effectively reduce the differences between records within and across participants [25–27]. Several studies have used the z-score normalization method to improve the consistency of EMG signals [8,25,28–31]. This study employed the z-score

method for EMG signal scaling. The z-score method scales the signal instances by removing and scaling the mean feature to unit variance.

3.4.3. Event Exporting and Feature Extraction

Following the signal-filtering process, the EMG signal still consisted of all events of FMA movement tasks, including the relaxation period. Separating task-related amplitude from a relaxed state is essential to avoid undesirable data deployment in the feature extraction process and machine learning algorithms, such as the relax-related signal. In the following step, feature extraction was conducted on each exported event separately. As a result, only EMG signals corresponding to task-related events were extracted. An illustration of the event-exporting and feature extraction process is depicted in Figure 4.

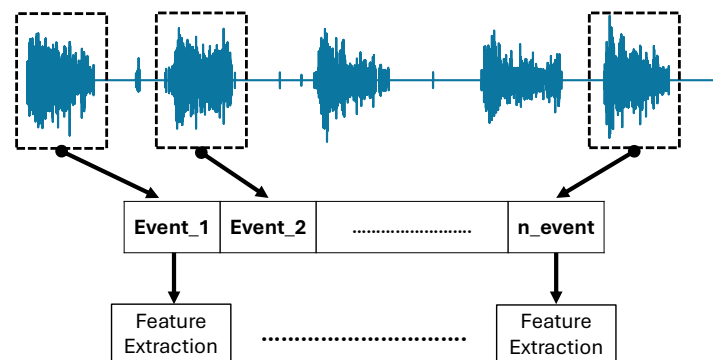


Figure 4. Event-exporting and feature extraction process.

Many features have been introduced for recognition purposes, with time- and frequency-domain features mainly employed due to their simplicity, low computational costs, and promising recognition results. In the previous experiment, seven time-domain features and one frequency-domain feature were employed [8]. The extracted features were mean absolute value (MAV), variance (VAR), root mean square (RMS), waveform length (WL), slope sign change (SSC), zero crossing (ZC), Willison amplitude (WAMP), and mean power frequency (MPF). Many studies have utilized these features to identify the amplitude, complexity, and frequency characteristics of EMG signals [32–35]. However, in the case of impairment-level recognition, the amplitude-related features can marginally differ in both between and within participants. Therefore, emphasizing the complexity-related features provides distinct information between impairment levels.

Several studies have introduced features that depict the complexity of EMG signals. Thongpanja et al. and Oo et al. utilized the skewness feature to provide complexity information from EMG signals [36,37]. Skewness measures the asymmetry of a variable within a distribution. Zero skewness signifies a symmetric distribution, while positive skewness indicates a right-skewed distribution and vice-versa [36]. Another promising feature that provides complexity information for EMG signal is Shannon entropy. This feature describes a signal's irregularity, complexity, and unpredictability characteristics [38].

Furthermore, the inherent non-stationary nature of EMG signals potentially provides vital feedback with respect to the entropy feature [39]. Typically, an increased value of Shannon entropy of an EMG signal represents good muscle condition, with many muscle motor units contributing to muscle contraction and a high-amplitude and randomized EMG signal. On the contrary, an impaired muscle makes it difficult to capture an EMG signal, resulting in a low value of Shannon entropy. In this regard, this study employed several features to focus more on the complex characteristics of EMG signals, including 5 time-domain features and 1 frequency-domain feature: waveform length (WL), zero crossing (ZC), mean absolute value (MAV), skewness, Shannon entropy, and mean power frequency (MPF).

3.4.4. Data Resampling and Scaling

Data resampling was performed to address the imbalance issue in the training datasets, as shown in Figure 5. The famous resampling approaches include oversampling, undersampling, and both combinations. Incorrect selection of the resampling method may lead to invalid classification results. In this study, the actual number and value of the minority classes were essential, as they correspond to the actual impairment level of the patient. Several studies have shown the promising performance of undersampling approaches to maintain the number of minority classes and produce good classification results [40–42]. This study utilized a random oversampling approach, employing Imblearn library in Python with a sampling strategy to resample all classes except the minority class.

Data scaling was performed for the second mechanism in the following step, as shown in Figure 5. Data scaling is a standard measure involving the treatment of data before inputting them to machine learning algorithms. This technique is instrumental in managing diverse data input scales and improving machine learning models' performance. One such data-scaling techniques is z-score scaling, in which z-score normalization or a standardization term is employed. The z score scales the data instances by removing the unit mean and dividing it by the unit variance. Several studies have proven the performance of the z-score method in enhancing classification accuracy and efficiency. Al-Faiz et al. showed that z-score scaling decreased the number of epochs required for the learning network [43]. Suma et al. achieved improved accuracy after z-score scaling [44]. Long et al. reported that classification performance was significantly improved after z-score scaling [45].

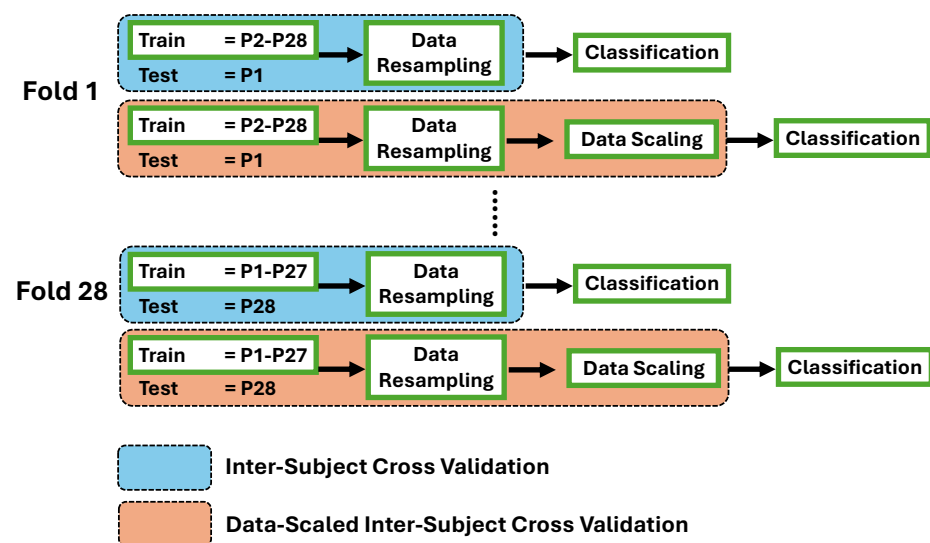


Figure 5. Inter-subject cross-validation data processing and classification.

3.5. Classification

In the previous experiment, we built machine learning models using the support vector machine (SVM) and random forest (RF) algorithms [8]. Despite being known as traditional machine learning methods, these algorithms are capable of recognizing complex patterns in the field of EMG-related classification [46–49]. The SVM and RF algorithms are based on support vector and ensemble-type architectures, respectively. In this study, we employed an additional algorithm, namely multi-layer perceptron (MLP), a neural network architecture. A machine learning model was built for each FMA movement task to ensure that the final output of each model was focused on the impairment level of finger movement. As shown in Figure 5, two mechanisms of data processing flow were applied, resulting in two classification results of the first (ISCV) and second (DS-ISCV) mechanisms

for each machine learning model. In the present study, the machine learning models of the second mechanism are denoted as SVM_scaled, RF_scaled, and MLP_scaled.

A Python library named scikit-learn was employed to implement the machine learning models. The parameters of each machine learning algorithm were pre-determined. SVM utilized a poly kernel of degree 3, a regularization parameter of 3, and a gamma parameter of 1 divided by the number of features. Regarding the random forest algorithm, the employed number of trees was 700 with a Gini criterion, and the maximum depth of the tree was 15. Lastly, the MLP parameters comprised a hidden-layer size of 100, a maximum number of iterations of 200, an alpha parameter of 1×10^{-4} , an initial learning rate of 1×10^{-4} with an adaptive method, and a stopping tolerance of 1×10^{-7} , with rectification as the activation function and a stochastic gradient-based optimizer.

3.6. Evaluation

This study evaluated the performance of the employed machine learning models through the inter-subject cross-validation method. The inter-subject cross-validation process was chosen to deal with cases in which a patient can exhibit only one or two impairment levels for a movement task. In this study, the inter-subject cross-validation procedure encompassed the classification process and the prior data preparation processes, as shown in Figure 2. The objective of this approach was to correctly classify inter-subject cross-validation, especially in train–test dataset selection, and evaluate the outcome of the machine learning models. An illustration of the inter-subject cross-validation approach is shown in Figure 5. The ISCV mechanism encompasses data resampling and classification, while DS-ISCV includes the data scaling process after resampling. The first cross-validation fold encompassed selecting the test data from participant number 1, with the remaining data used as a training dataset. Subsequently, the last fold utilized test data from participant number 28, while the remaining data were used as a training dataset.

3.7. Data Collection Experiment

During the experiment, four channels of EMG sensors were attached to the participant's forearm and palm on the impaired side, where a cleaning measure with an alcohol swab was performed prior to attachment in the skin area of the electrode attachment position. A doctor sat before the participant to assess and give finger movement instructions, as shown in Figure 6. Figure 7 shows that the assessed finger movement encompassing MF, ME, HG, TA, PG, CG, and SG. The participant was instructed to perform an FMA-based finger movement for 5 s with five repetitions. This repetition approach was intended to observe the stability of the patient's finger movement to achieve a reasonable assessment. Simultaneously, the doctor observed the exhibited finger movement and performed the assessment accordingly.

Additionally, the patient was instructed to move only the finger part while the doctor performed a tug of the experimental object. This measure was intended to produce a proper assessment by the doctor. This study utilized the doctor's assessment as the ground truth or true label for the employed machine learning models. The experimental environment is shown in Figure 6.

One session of finger movement assessment encompassed five repetitions of one finger movement. A 12 s relaxation or resting period was allowed between movement repetitions. However, a preferable resting period between repetitions may be set arbitrarily based on the doctor's instruction. Subsequently, the participant was allowed a preferable resting period before moving to another finger movement task. This setting was implemented to avoid muscle fatigue, which can influence the stability of the measured EMG [32]. Additionally, an experimenter held a timer to guide the movement repetitions to avoid any mistakes. The participant's EMG signal and an experimental video were also recorded throughout the data collection process. Illustrations of movement repetitions with the corresponding resting periods are shown in Figure 8.

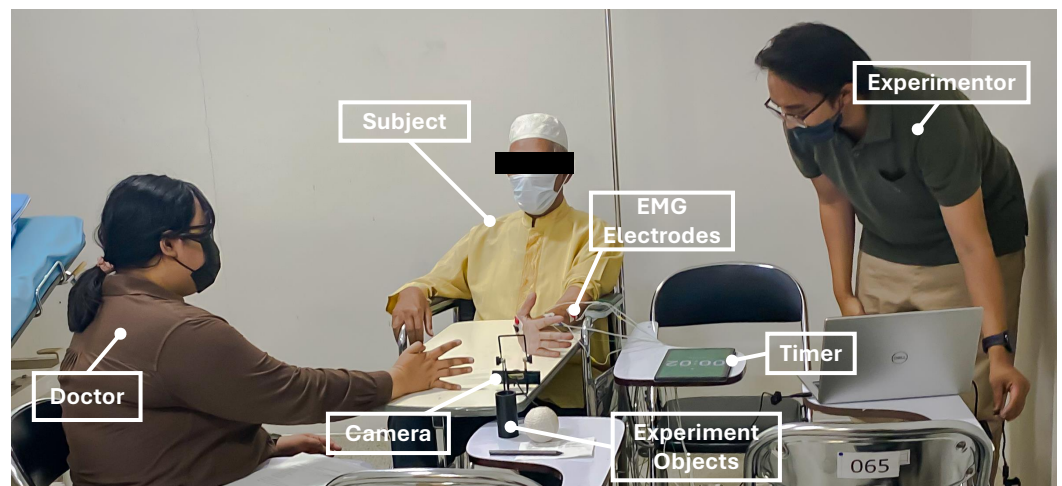


Figure 6. Experimental environment.

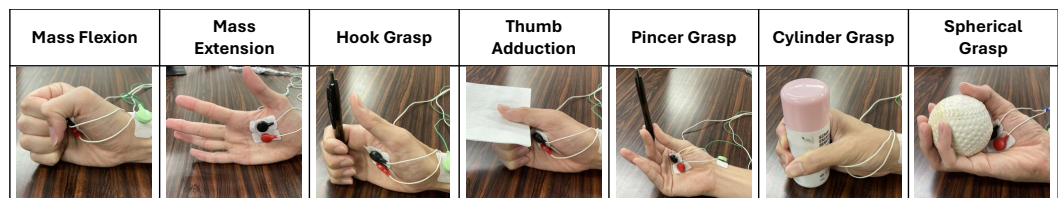


Figure 7. Finger movement tasks of the FMA.

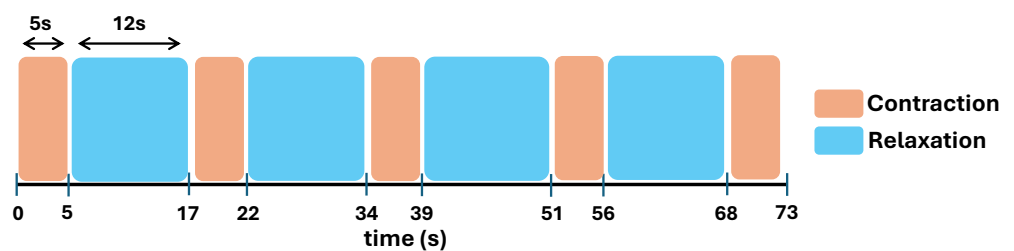


Figure 8. One session of finger movement assessment.

4. Results

This section demonstrates the presence of an inherently imbalanced dataset in the actual patient experiment. Consequently, recall scores between ISCV and DS-ISCV are compared. Furthermore, the recall score of non-majority classes is evaluated. Finally, detailed cross-validation results of the best machine learning model for each finger movement task are shown. Each result includes the recall score of the actual target class and the misclassification rate of the other target classes.

4.1. Movement Event Data

As mentioned in Section 3.7, there were 28 participants, and each participant was instructed to perform each finger movement task with five repetitions. Ultimately, the total number of movement events for one task with five repetitions across 28 participants was 140, and the total number repetitions of the 7 FMA finger movement tasks across the 28 participants was 980. However, some movement events were removed due to the participant's undesirable hand movement, with unstable conditions occurring in the EMG signal, resulting in a poor EMG waveform shape. The final movement events after data removal are shown in Table 2.

Table 2. Summary of Movement Events (F: Full, P: Partial, N: None; ME: Mass Extension, MF: Mass Flexion, HG: Hook Grasp, TA: Thumb Adduction, PG: Pincer Grasp, CG: Cylinder Grasp, SG: Spherical Grasp; Black Square: Removed Movement Event.)

Subject	ME			MF			HG			TA			PG			CG			SG			
	F	P	N	F	P	N	F	P	N	F	P	N	F	P	N	F	P	N	F	P	N	
P1	5			5			5				5			5			5			5		
P2	5			3			4				3				■		5			4		
P3	5			5			5				5			5			5			5		
P4	5			5			3				5			5			5			5		
P5				5			5	2	2		1	3		2			2					■
P6		5		4	1		5				5			5			5			5		
P7				5			5				5			5			5			5		
P8	■						3	2			5			5			5			5		
P9		5		5			5				5			5			5			5		
P10	5			5			5				5			5			5			5		
P11	5			5			5				5			5			5			5		
P12	5			5			5			4	1		2	3		5			5			
P13				4			5			5			5			2	3		5			5
P14	5			5			5				5			5			5			5		
P15	5			5			5				5			5			5			5		
P16	5			5			5				5			5			5			5		
P17	4	1		5			5				5			2	3		5			5		
P18				5			5				5			5			5			5		
P19	5			5			5				5			5			5			5		
P20	5			5			5				5			5			5			5		
P21	5			5			5				5			5			5			5		
P22	5			5			5				5			5			5			5		
P23				5			5			2			5			5			5			5
P24	5			5			5				5			5			5			5		
P25	5			5			5				5			5			5			5		
P26				5			5	2			3			5			2					5
P27	5			5			5				5			5			5			5		
P28	5			5			5				5			5			5			5		
Total	94	11	29	102	9	25	104	9	14	97	25	13	94	25	13	117	7	10	109	10	15	

As shown in Table 2, the movement events of participant P8 in ME, P2 in PG, and P5 in SG were removed entirely. Consequently, the classification of these movements comprised only 27-fold cross-validation. On the other hand, several participants had fewer than five events in some movement tasks after data removal. The total number of movement events in Table 2 shows an imbalanced dataset, where the majority class is the full level in all movement tasks. The partial level was the minority class in ME, MF, HG, CG, and SG, whereas the none level was the minority in TA and PG. Furthermore, there is a significant difference in the number of majority and minority classes, with an average of 92 ± 10 events.

Following the data collection experiment, the EMG data underwent a data processing step, as shown in Figure 2. Following the feature extraction step with a 500 ms window size and a 100 ms window step, the processed EMG data were extracted to be prepared as machine learning input. The numbers of data points for each movement task and impairment level are shown in Table 3.

Table 3. Total data points after the feature extraction process across subjects.

Movement	Impairment Level		
	Full	Partial	None
ME	4246	491	1288
MF	4601	431	1145
HG	4522	382	640
TA	4083	1001	614
PG	4013	1156	604
CG	5264	320	476
SG	4825	438	689

4.2. Recognition Performance

A well-known metric such as accuracy may provide misleading information in an imbalanced dataset. The bias of the score toward the majority class accuracy causes this issue. Thus, this study evaluated recognition performance using an average of recall metrics to minimize bias due to an imbalanced dataset. The average of recall encompassed calculating the recall for each impairment level after concatenating the actual and predicted levels in inter-subject cross-validation, followed by calculating the overall average across impairment levels. Additionally, the average recall score of minority classes is presented.

In Table 4, gray cells represent the highest recall score for each movement. Four out of seven movements surpassed a recall of 0.50, namely ME, HG, PG, and SG. Additionally, the ISCV and DS-ISCV machine learning models performed well depending on the movement, with the highest recall achieved by SVM for the second mechanism (SVM_DS-ISCV) in MF, the first mechanism of MLP (MLP_ISCV) in HG, SVM_DS-ISCV in TA, SVM_ISCV in PG, MLP_ISCV in CG, and MLP_ISCV in SG. However, in ME, SVM_DS-ISCV and MLP_ISCV shared the same recall score of 0.70. In this case, the average recall score of the minority classes of the partial and none levels must be observed.

Table 4. Average recall score of inter-subject cross-validation (gray: highest recall score of each movement).

Movement	ISCV			DS-ISCV		
	SVM	RF	MLP	SVM	RF	MLP
ME	0.61	0.61	0.70	0.70	0.62	0.64
MF	0.40	0.44	0.40	0.49	0.46	0.44
HG	0.46	0.46	0.50	0.46	0.43	0.49
TA	0.27	0.33	0.25	0.35	0.31	0.29
PG	0.60	0.50	0.45	0.48	0.50	0.50
CG	0.32	0.33	0.40	0.34	0.33	0.33
SG	0.49	0.42	0.73	0.46	0.42	0.56

The average recall scores of non-minority classes presented in Table 5 were calculated from the partial and none levels. In the ME movement, SVM_DS-ISCV achieved the highest score, making it the best model for recognizing non-majority classes. However, different outcomes occurred in HG and TA, with MLP_DS-ISCV achieving the highest recall score for non-majority classes in both movements. There was a marginal difference in the recall scores for the overall (Table 4) and non-majority (Table 5) classes in the HG movement due to the trade-off between the recognition of majority and non-majority classes. In the TA movement, despite the highest score being achieved by SVM_DS-ISCV (Table 4), for the non-majority class, MLP_DS-ISCV achieved the highest score. This was due to the high

recall score for the majority class in TA. Considering the importance of the non-majority class, this study indicates the best model for each movement task with a dagger (†) symbol in Table 5.

Table 5. Average recall score of inter-subject cross-validation in non-majority classes (gray: highest recall score for each movement; †: selected machine learning model).

Movement	ISCV			DS-ISCV		
	SVM	RF	MLP	SVM	RF	MLP
ME	0.5	0.49	0.64	0.72 [†]	0.49	0.57
MF	0.15	0.21	0.18	0.37 [†]	0.24	0.27
HG	0.25	0.24	0.35 [†]	0.25	0.19	0.36
TA	0.07	0.12	0.08	0.10	0.10	0.14 [†]
PG	0.49 [†]	0.29	0.28	0.26	0.29	0.37
CG	0.03	0.00	0.16 [†]	0.05	0.00	0.09
SG	0.29	0.18	0.65 [†]	0.26	0.18	0.42

Figure 9 shows detailed information in concatenated confusion matrices for each movement after inter-subject cross-validation. In general, SVM_DS-ISCV of ME and MLP_ISCV of SG achieved good recognition performance, with the predicted labels achieving more than 50% correctness compared with all true labels, as shown in Figure 9a,g respectively. Subsequently, SVM_DS-ISCV of MF, MLP_ISCV of HG, and SVM_ISCV of PG achieved more than 50% for two labels, as shown in Figure 9b,c,e respectively. In contrast, the remainder achieved correct classification of the majority class only. However, the true labels were the doctor’s manual inspection assessment results. Therefore, the aforementioned good recognition performance represented the ability of the machine learning model to produce the same assessment result as the doctor.

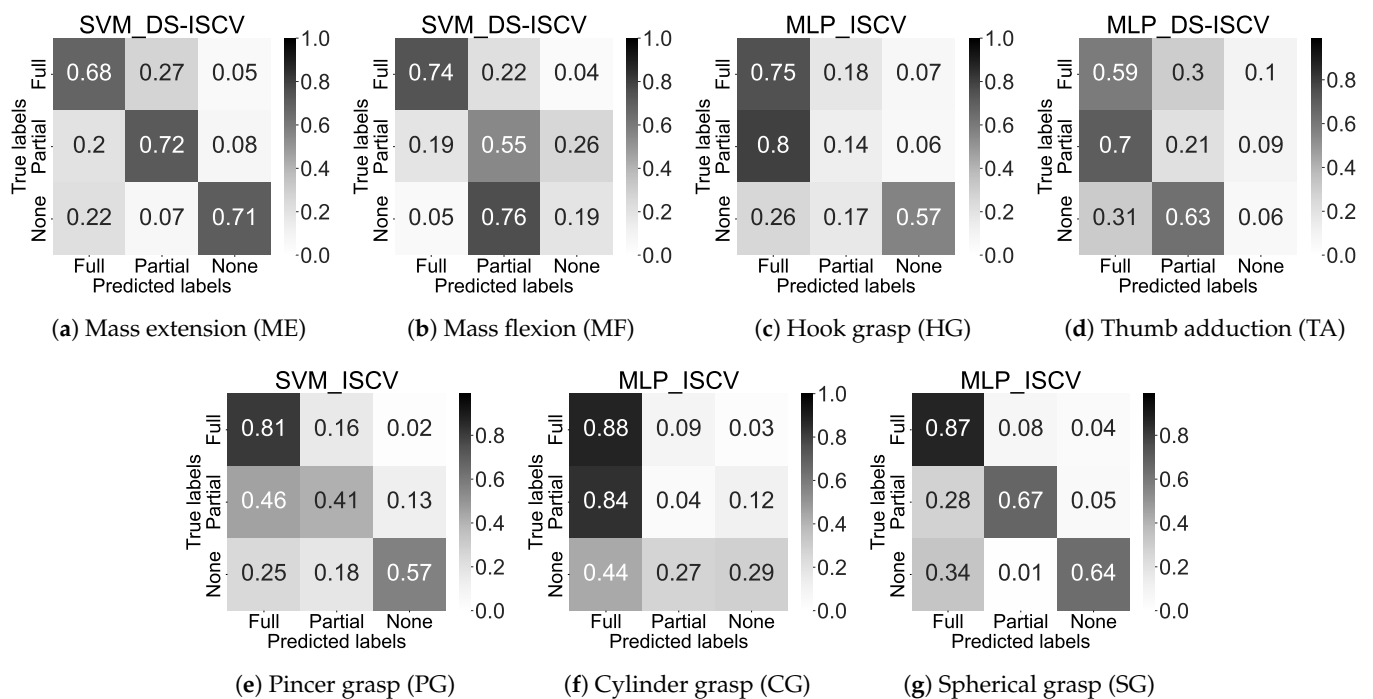


Figure 9. Concatenated Normalized confusion matrices of the best machine learning models for each movement after inter-subject cross-validation.

4.3. Portion of Classification Outcome Across Impairment Levels in Inter-Subject Cross-Validation

Figure 9 shows the general information of correct-incorrect classification after concatenating the classification outcome across the participants. This subsection presents the detailed classification outcome of each cross-validation fold, including the misclassification rate across impairment levels. In this study, the misclassification rate of each cross-validation fold was calculated from the false positive rate, and the classification outcome of the corresponding impairment level was the recall score. Therefore, the classification portion can be observed using these metrics as shown in Table 6. The recall score for the actual level is shown in black, while the false positive rate for other levels is shown in red for each fold. Each row corresponds to a specific fold in the inter-subject cross-validation, as illustrated in Figure 5.

Table 6. Portion of Classification Across Impairment Levels in Inter-Subject Cross-Validation (Red: False Positive Rate Score; Black Square: Removed Movement Event).

Test Data (Fold)	Actual Level	ME			Actual Level	MF			Actual Level	HG			Actual Level	TA			Actual Level	PG			Actual Level	CG			Actual Level	SG		
		Predicted Level				Predicted Level				Predicted Level				Predicted Level				Predicted Level				Predicted Level				Predicted Level		
		F	P	N		F	P	N		F	P	N		F	P	N		F	P	N		F	P	N		F	P	N
P1	F	0.78	0.19	0.03	F	1.00	0.01	0.00	F	0.92	0.04	0.04	P	0.65	0.28	0.06	F	0.97	0.02	0.02	F	0.92	0.08	0.00	F	0.79	0.12	0.09
P2	F	0.05	0.00	0.95	F	0.13	0.83	0.04	F	0.53	0.12	0.35	F	0.10	0.28	0.61					F	0.82	0.02	0.17	F	0.93	0.07	0.00
P3	F	0.99	0.01	0.00	F	0.70	0.25	0.05	F	0.32	0.68	0.00	F	0.04	0.13	0.83	F	0.17	0.64	0.19	F	0.23	0.77	0.00	F	0.98	0.02	0.00
P4	F	0.71	0.20	0.09	F	0.14	0.84	0.02	F	1.00	0.00	0.00	F	0.00	0.99	0.00	F	0.68	0.32	0.00	F	0.99	0.00	0.01	F	0.69	0.08	0.23
P5	N	0.10	0.04	0.86	N	0.03	0.88	0.09	P	0.00	0.00	1.00	P	0.00	1.00	0.00	P	0.00	1.00	0.00	P	0.66	0.00	0.34				
					F	0.19	0.42	0.34	N	0.00	0.01	0.99	N	0.00	0.99	0.01												
P6	P	0.23	0.77	0.00	P	0.10	0.77	0.34	F	0.13	0.25	0.61	F	0.75	0.19	0.05	P	0.91	0.09	0.00	F	0.69	0.22	0.10	F	0.67	0.33	0.00
P7	N	0.26	0.00	0.74	N	0.02	0.86	0.12	N	0.00	0.17	0.83	N	0.00	1.00	0.00	N	0.00	0.38	0.62	N	0.47	0.52	0.01	N	0.05	0.01	0.94
P8					P	0.03	0.89	0.08	F	0.09	0.00	0.91	F	0.00	0.89	0.11	F	0.97	0.02	0.00	F	0.94	0.00	0.06	F	0.76	0.00	0.24
P9	P	0.22	0.75	0.03	F	1.00	0.00	0.00	F	0.91	0.07	0.02	P	0.66	0.30	0.03	F	0.77	0.21	0.03	F	1.00	0.00	0.00	F	0.74	0.09	0.16
P10	F	0.11	0.89	0.00	F	0.77	0.23	0.00	F	0.67	0.33	0.00	F	0.03	0.97	0.00	F	0.99	0.01	0.00	F	1.00	0.00	0.00	F	0.98	0.03	0.00
P11	F	1.00	0.00	0.00	F	0.68	0.32	0.00	F	0.98	0.01	0.01	F	0.02	0.98	0.00	F	1.00	0.00	0.00	F	1.00	0.00	0.00	F	1.00	0.00	0.00
													F	0.92	0.06	0.02	F	1.00	0.00	0.00					F	0.84	0.05	0.12
P12	F	0.81	0.19	0.00	F	0.98	0.02	0.00	F	0.98	0.00	0.02	P	1.00	0.00	0.02	P	1.00	0.00	0.00					F	1.00	0.00	0.00
													P	0.69	0.00	0.26												
P13	N	0.70	0.04	0.26	N	0.03	0.97	0.00	N	0.88	0.00	0.12	P	0.83	0.07	0.10	N	0.69	0.05	0.31	F	0.91	0.00	0.09	N	1.00	0.00	0.00
P14	F	0.95	0.05	0.00	F	0.97	0.02	0.00	F	1.00	0.00	0.00	F	0.99	0.01	0.00	F	1.00	0.00	0.00	F	0.98	0.02	0.00	F	1.00	0.00	0.00
P15	F	1.00	0.00	0.00	F	1.00	0.00	0.00	F	1.00	0.00	0.00	F	1.00	0.00	0.00	F	0.67	0.14	0.19	F	1.00	0.00	0.00	F	1.00	0.00	0.00
P16	F	0.42	0.58	0.00	F	0.36	0.61	0.03	F	0.98	0.01	0.01	F	0.25	0.75	0.00	F	0.78	0.19	0.03	F	0.19	0.81	0.00	F	0.61	0.39	0.00
	F	0.08	0.81	0.23													F	1.00	0.00	0.00								
P17	P	0.00	0.34	0.23	F	0.05	0.70	0.25	F	0.46	0.54	0.00	F	0.62	0.38	0.00	P	0.87	0.13	0.00	F	0.76	0.06	0.18	F	0.81	0.02	0.16
P18	N	0.00	0.04	0.96	P	0.32	0.27	0.41	P	0.88	0.11	0.01	P	0.65	0.12	0.23	P	0.08	0.37	0.55	P	0.92	0.06	0.02	P	0.23	0.76	0.01
P19	F	0.19	0.74	0.07	F	1.00	0.00	0.00	F	0.78	0.00	0.22	F	0.99	0.01	0.00	F	1.00	0.00	0.00	F	1.00	0.00	0.00	F	0.95	0.02	0.04
P20	F	0.68	0.32	0.00	F	0.93	0.07	0.00	F	0.97	0.03	0.00	F	0.32	0.61	0.07	F	0.86	0.14	0.00	F	0.96	0.00	0.04	F	0.84	0.16	0.00
P21	F	1.00	0.00	0.00	F	1.00	0.00	0.00	F	1.00	0.00	0.00	F	0.94	0.06	0.00	F	0.84	0.16	0.00	F	1.00	0.00	0.00	F	1.00	0.00	0.00
P22	F	0.97	0.00	0.04	F	0.97	0.03	0.00	F	0.48	0.00	0.52	F	0.94	0.00	0.06	F	0.99	0.01	0.00	F	1.00	0.00	0.00	F	0.60	0.40	0.00
P23	N	0.17	0.00	0.83	N	0.06	0.51	0.44	N	0.50	0.47	0.03	N	0.81	0.03	0.15	N	0.25	0.06	0.69	N	0.42	0.01	0.57	N	0.02	0.03	0.95
P24	F	0.99	0.00	0.00	F	0.97	0.03	0.00	F	0.86	0.12	0.01	F	0.60	0.40	0.00	F	1.00	0.00	0.00	F	1.00	0.00	0.00	F	0.97	0.02	0.01
P25	F	0.93	0.07	0.00	F	1.00	0.00	0.00	F	0.97	0.03	0.00	F	0.99	0.01	0.00	F	0.07	0.93	0.00	F	1.00	0.00	0.00	F	1.00	0.00	0.00
P26	N	0.21	0.27	0.51	N	0.11	0.59	0.30	P	0.15	0.85	0.00	P	0.79	0.21	0.00	P	0.00	1.00	0.00	F	0.93	0.07	0.00	P	0.33	0.58	0.09
P27	F	0.86	0.12	0.02	F	0.94	0.06	0.00	F	0.96	0.00	0.04	F	0.91	0.09	0.00	F	0.76	0.24	0.00	F	0.96	0.03	0.01	F	0.94	0.06	0.00
P28	F	0.00	1.00	0.00	F	0.89	0.09	0.02	F	1.00	0.00	0.00	F	0.96	0.04	0.00	F	1.00	0.00	0.00	F	1.00	0.00	0.00	F	0.98	0.02	0.00

In contrast to Figure 9, which showcased the concatenated confusion matrices after cross-validation, the classification portion offers detailed insight into discerning the misclassification trends of each movement. Consequently, this allowed for a comprehensive evaluation of the performance of the best machine-learning models on each movement. Most movements showed the misclassification that resulted in the recall score below 0.5 occurring between Full and Partial levels in ME, HG, TA, PG, and CG. On the other hand,

misclassification between Partial and None levels occurred in MF, whereas misclassification between None and Full levels in SG.

4.4. Comparison of Recognition Performance with Previous Experiment

This study presents an improved recognition method for finger movement impairment levels based on the FMA. In the previous experiment, the F1 score of each impairment level was utilized as an evaluation metric. The previous experiment only involved the application a holdout method in the data-splitting process, and the dataset was split into 50% of training and testing datasets. However, this study included leave-one-out cross-validation, with the participant data representing the number of observations for data splitting. Accordingly, despite making this experiment slightly less comparable to the previous experiment, we employed the F1 score of the best fold of inter-subject cross-validation for comparison.

In the previous study, the impairment level of the none level was not acquired, as shown in Table 7 [8]. Therefore, only full and partial levels were utilized in this study. MF achieved the lowest F1 score, whereas TA achieved the highest. In the current study, the lowest F1 score was that of CG, whereas PG achieved the highest, as shown in Figure 8. The F1 scores presented in Table 8 have higher values than the recall score in Table 6. Figure 10 shows that the F1 scores achieved in this study are superior to those achieved in the previous experiment.

Table 7. F1 scores achieved in the previous experiment (Black Square: Unobtained Level) [8].

Movement	Best Machine Learning Model	Impairment Level		
		Full	Partial	None
ME	SVM	0.60	0.58	
MF	SVM	0.32	0.54	
HG	SVM	0.54	0.52	
TA	SVM	0.98	0.83	
PG	RF	0.90	0.40	
CG	RF	0.89	0.60	
SG	SVM	0.76	0.74	
Average		0.71	0.60	

Table 8. Highest F1 scores of a selected inter-subject cross validation fold in the present study.

Movement	Best Machine Learning Model	Impairment Level		
		Full	Partial	None
ME	SVM_DS-ISCV	1.00	0.87	0.98
MF	SVM_DS-ISCV	1.00	0.94	0.99
HG	MLP_ISCV	1.00	0.92	0.91
TA	MLP_DS-ISCV	1.00	0.47	0.27
PG	SVM_ISCV	1.00	1.00	0.82
CG	MLP_ISCV	1.00	0.11	0.72
SG	MLP_ISCV	1.00	0.86	0.97
Average		1.00	0.74	0.81

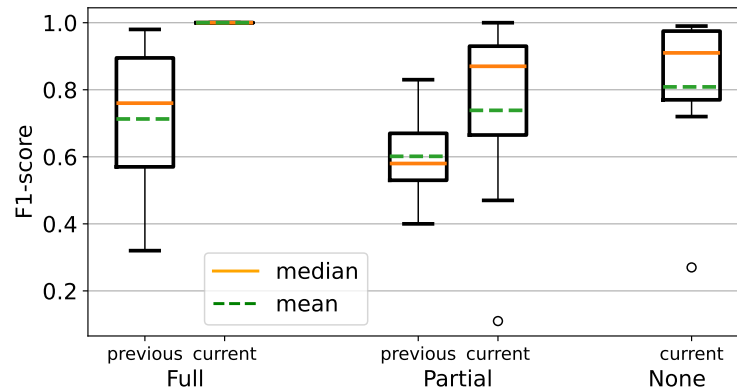


Figure 10. Box plot of F1 scores achieved in the current and previous study.

5. Discussion

The subjective nature of assessment presents a significant challenge in the context of the Fugl-Meyer Assessment (FMA). This challenge arises from the reliance on manual evaluation conducted by a doctor through visual inspection and manual tugging to determine the level of impairment of the patient. Thus, the current study presents a method for recognizing finger movement impairment levels in the FMA using EMG. The discussion delves into the recognition performance of FMA impairment levels within an imbalanced dataset and an inherently subjective environment. Subsequently, this section discusses how the inherent subjectivity of conventional assessment leads to misclassification by the machine learning model.

5.1. Recognition Performance

According to Table 3, the minority class of the partial level appeared in ME, MF, HG, CG, and SG, while TA and PG accounted for the minority class at the none level. Theoretically, an imbalanced dataset drives the recognition of the minority class toward the majority class [50]. Based on the results shown in Figure 9 and Table 6, there was no sign of majority-class bias in terms of misclassification of the minority class. As shown in Figure 6, several full levels in ME were misclassified as the minority class of the partial level. Subsequently, the partial level of P17 in ME was misclassified as the none level instead of the full level. This finding infers that the proposed method performed well in addressing the imbalanced condition of the dataset. In contrast, no related research has specifically evaluated performance in the recognition of finger movement impairment level under the condition of an imbalanced dataset [8,10–13]. Therefore, the resulting recognition performance represents the actual condition of the FMA.

In the previous experiment, the data processing method comprised data scaling only [8]. Since the current study employed an improved recognition method, the results are not comparable with those achieved in the previous study. Consequently, the best F1 score of each impairment level from the cross-validation fold was selected to match with the holdout method of the previous experiment. In addition, this study evaluated recognition performance in inter-subject cross-validation settings and addressed the imbalanced dataset issue, which was not explicitly addressed in the aforementioned related research [10–13]. Two related studies that investigated finger movement recognition only employed accuracy as an evaluation metric [12,13]. Considering the shortcomings of the accuracy metric under the condition of an imbalanced dataset, this study only compared the F1 scores with the results of the previous experiment [8].

Compared to the recall metric, which is suitable for imbalanced datasets, several F1 scores in this study showed higher values due to the contrary nature of precision relative to the F1 score. Consequently, it might not represent the actual performance of the employed machine learning model. According to this metric, the experimental results achieved in this study are superior to those of the previous experiment, as shown in Figure 10. In the previous experiment, the employed features focused on the statistical

and amplitude characteristics of EMG signals[8]. In contrast, this study emphasized the complex characteristics of EMG signals that improve recognition.

5.2. Subjectivity Issue of Fugl-Meyer Assessment

In this study, the doctor's manual assessment result served as the ground truth when training the employed machine learning models. Therefore, a high recall score corresponds to good machine learning model performance in terms of matching the doctor's assessment result. However, as mentioned in related research on FMA impairment level recognition, the conventional FMA is prone to subjectivity issues [8,10–13]. The interaction between the full and none levels exhibited distinct boundaries, whereas interactions involving the partial level tended to yield uncertain and subjective assessment results. This ambiguity arises from the indistinct boundary at the partial level, which behaves like a transitional phase between the none and full levels, as shown in Table 1.

Interesting findings arose with respect to the portion of classification across impairment levels. According to Table 6, the proposed method mostly showed an interaction of the misclassification rate between full and partial levels in ME, HG, TA, PG, and CG. Meanwhile, misclassification between the partial and none levels mostly appeared in MF. Misclassification cases were rare between the full and none levels, indicating that the system could easily recognize the clear boundaries between these levels. This finding supports the rule of thumb according to which uncertainty is rare between the full and none levels.

Furthermore, the result also indicates that the proposed method differs from the doctor's decisions in areas with a high risk of subjectivity. This difference mainly appeared because the machine learning model used muscle information to decide upon the impairment level. In addition, differences appeared in the results involving the partial impairment level, which is associated with a high subjectivity risk. Therefore, the proposed method may assist doctors or physiotherapists in clarifying the patient's impairment level under uncertain assessment conditions, objectifying the FMA assessment.

5.3. Limitations and Future Work

This study addressed the inherent subjectivity of the FMA using the EMG impairment level recognition method. Various uncontrollable anatomical and physiological factors, such as age and gender variations, anatomical characteristics influencing electrode attachment, and the potential presence of atrophy or fibrosis in the muscle unit, pose challenges in EMG settings. However, aside from muscle information, the FMA requires the dexterity of the fingers in executing tasks, as shown in Table 1. Consequently, several participants with different finger dexterity profiles may be recognized as identical due to marginal differences in EMG signals. Therefore, to comprehensively recognize the impairment level in finger movement, an additional modality should be considered to capture the finger movement profile of each participant.

6. Desktop Application

This study also presents a use case of the proposed method in actual assessment using a desktop application. This study aimed to assist doctors in making more accurate judgments by employing an EMG biosignal modality to automatically recognize the finger movements of the FMA. The implementation of the method proposed in this study requires the doctor to perform a manual assessment, especially in the object-grasping task. Simultaneously, the system automatically assesses the impairment level based on EMG muscle information. Finally, the doctor can double-check the system's output before deciding on the impairment level of the patient.

Figure 11 shows the desktop application's assessment window with several features. The movement list feature consists of a list of FMA finger movements, where each item corresponds to the selected machine learning model. The connect and disconnect buttons control the connections between sensors and the application. The start and stop record buttons control impairment level recognition and camera recording processes. The im-

impairment level is displayed in the impairment level display. A camera display shows and records finger movements. Lastly, the signal obtained from the sensors is displayed in the EMG signal display. The assessment window mainly implements the proposed method to output an impairment level for the FMA finger movements. As shown in Figure 11, a display shows impairment at the none level, with a red background. This impairment level was produced through several processes in the proposed method that were adjusted for real-time assessment inside the application.

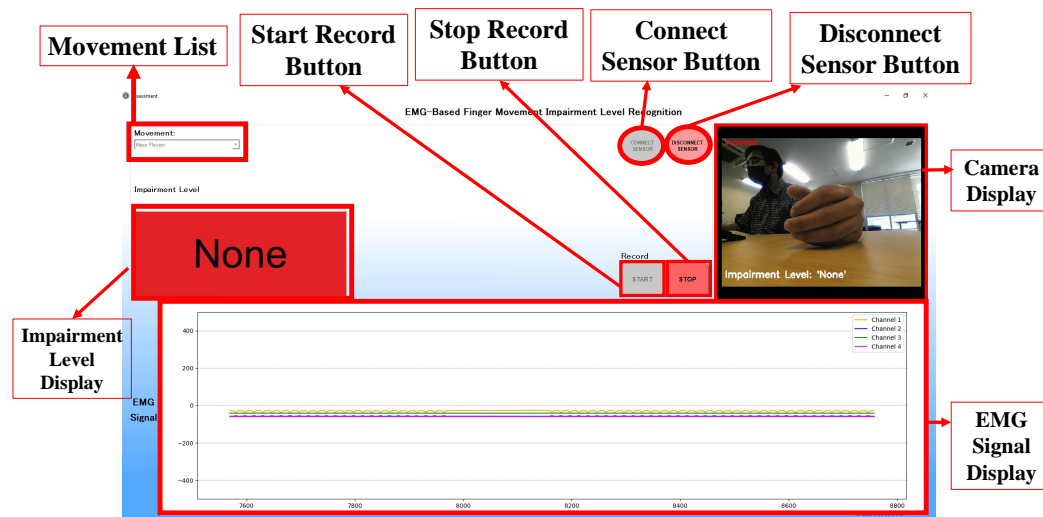


Figure 11. Assessment window of the desktop application.

The first process of the application workflow was to store the EMG signal from four channels individually in an object with a size of 1000 instances. This process was intended to fix the number of EMG data to 1000 instances. The stored EMG data were then filtered using both Butterworth and wavelet filters. Subsequently, the data were extracted with five time-domain features and one frequency-domain feature, as explained in Section 3.4.3. The last process was to input the data into the established machine learning model to output the impairment level. The established machine learning models were connected with the finger movement item inside the movement list features. Therefore, the model can automatically select when the user chooses a finger movement item inside the feature.

The assessment window consists of several stages—the early stage, the pre-recognition stage, and the recognition stage—intended to help the user run the assessment comfortably, as shown in Figure 12. At this stage, the camera display, the green push button for connecting to sensors, and the combo box for finger movement selection are enabled. Meanwhile, the other features are automatically disabled. This stage has a function to facilitate the user adjusting the camera position for the assessment, connecting the application to the EMG sensors and selecting a finger movement to be assessed. In addition, each item in the list of finger movements is connected with the selected machine learning model to be utilized in the recognition stage.

The next stage is the pre-recognition stage. This stage starts after the user clicks the connect-sensor push button and the connection of the sensor with the application is successfully established. At this stage, several features that correspond to several functions are enabled. The first enabled feature is the plot widget, which is used to display the EMG signal. In this feature, the user may check the displayed EMG signal for each channel. With this feature, the doctor may also check for an error in the resulting EMG signal when the patient performs any movement. Therefore, an adjustment in the position of the electrodes or other necessary action to obtain a proper EMG signal is feasible. In addition, the selection of the finger movement is still feasible before the recognition stage. Another enabled feature is the green start-record push button, which is used to start the

recognition and video recording process. If the user wishes to finish the assessment, the red disconnect-sensor push button is enabled at this stage.

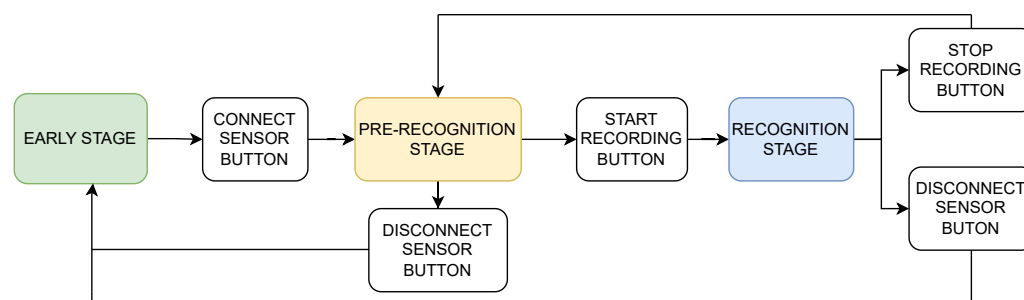


Figure 12. User flow of the assessment window with the corresponding trigger button.

Following the command from the start-record push button, the recognition stage is started. At this stage, the corresponding impairment-level information is also transferred to the displayed video. Therefore, the video displayed in the recognition stage simultaneously shows the assessment date and the impairment level. Furthermore, the video recorded on the PC allows the doctor to review the impairment level of the corresponding finger movement outside the desktop application.

The user has two options to proceed with the assessment window at this stage. The first option is to stop the assessment for the current movement and start another movement. In this option, the user can press the stop-record button and enter the pre-recognition stage. Subsequently, the user can select different movements from the movement list and press the green start-record button to start the assessment. The other option is to finish all assessments. The user can press the red stop-record button and disconnect-sensor buttons.

7. Conclusions

In this study, we proposed a recognition method for finger movement impairment levels in the FMA, using EMG to address the subjectivity issue under the condition of an imbalanced dataset. This study also introduces a desktop application that outputs an assessment video embedded with the system's impairment-level result. The evaluation result of inter-subject cross-validation showed that both data processing mechanisms addressed the imbalanced dataset condition and the subjectivity issue of the FMA. The recognition results show that the SVM and MLP of both mechanisms outperformed the RF model in all FMA tasks. Subsequently, the misclassification rate mainly occurred between the full and partial levels. Despite moderate recall scores in some movement tasks, this finding highlights the subjectivity issue in conventional assessments. Therefore, the output of the proposed method shows significant potential to assist doctors in making proper judgments through the used of EMG-based recognition, especially under uncertain assessment conditions.

Author Contributions: Conceptualization, A.R.S.S.P., A.O., Y.D.P., R.R., T.T. and M.T.; methodology, A.R.S.S.P., A.O., and T.T.; software, A.R.S.S.P.; validation, A.R.S.S.P.; formal analysis, A.R.S.S.P. and T.T.; investigation, A.R.S.S.P.; resources, A.R.S.S.P.; data curation, A.R.S.S.P. and Y.D.P.; writing—original draft preparation, A.R.S.S.P.; writing—review and editing, A.R.S.S.P., A.O., R.R., and T.T.; visualization, A.R.S.S.P.; supervision, A.O., T.T., and M.T.; project administration, A.R.S.S.P., A.O., R.R., T.T., and M.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted following the guidelines of the Declaration of Helsinki, ICH-GCP standards, CIOMS-WHO guidelines 2016 (Protocol No. UA-02-23144), and was approved by the Ethical Committee of Airlangga University Hospital, Surabaya, Indonesia (No. 125/KEP/2023, 29 August 2023).

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: The data are not publicly published due to ethical restrictions.

Acknowledgments: The authors appreciate the cooperation of all patients, clinicians, and all parties supporting the experiment conducted at Airlangga University Hospital.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ISCV	Inter-subject Cross Validation
DS-ISCV	Data-Scaled Inter-Subject Cross Validation
SVM	Support Vector Machine
RF	Random Forest
MLP	Multi-layer Perceptron
FMA	Fugl-Meyer Assessment
ME	Mass Extension
MF	Mass Flexion
HG	Hook Grasp
TA	Thumb Adduction
PG	Pincer Grasp
CG	Cylinder Grasp
SG	Spherical Grasp

References

- GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **2019**, *18*, 459–480.
- Marciniak, C. Poststroke Hypertonicity: Upper Limb Assessment and Treatment. *Top. Stroke Rehabil.* **2011**, *18*, 179–194.
- Gladstone, D.J.; Danells, C.J.; Black, S.E. Poststroke Hypertonicity: The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties. *Neurorehabil. Neural Repair* **2002**, *16*, 232–240.
- Nam, H.U.; Huh, J.S.; Yoo, J.N.; Hwang, J.M.; Lee, B.J.; Min, Y.S.; Kim, C.H.; Jung, T.D. Effect of Dominant Hand Paralysis on Quality of Life in Patients With Subacute Stroke. *Ann. Rehabil. Med.* **2014**, *38*, 450–457.
- Dombovy, M.L.; Sandok, B.A.; Basford, J.R. Rehabilitation for Stroke: A Review. *Stroke* **1986**, *17*, 363–369.
- Kotov-Smolenskiy, A.M.; Khizhnikova, A.E.; Klochkov, A.S.; Suponeva, N.A.; Piradov, M.A. Surface EMG: Applicability in the Motion Analysis and Opportunities for Practical Rehabilitation. *Hum. Physiol.* **2021**, *47*, 237–247.
- Jaramillo-Yáñez, A.; Benalcázar, M.E.; Mena-Maldonado, E. Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review. *Sensors* **2020**, *20*, 2467.
- Sugiharto, A.R.; Tsuchida, S.; Pawana, I.P.A.; Terada, T.; Tsukamoto, M. EMG-Based Recognition Method of Finger Movement Impairment Level in Post-Stroke Patients Based on Fugl-Meyer Assessment. *JBINS* **2022**, *8*, 425–433.
- Pradipta, G.A.; Wardoyo, R.; Musdholifah, A.; Sanjaya, I.N.H.; Ismail, M. SMOTE for Handling Imbalanced Data Problem: A Review. In Proceedings of the 2021 Sixth International Conference of Informatics and Computing (ICIC), Jakarta, Indonesia, 3–4 November 2021.
- Wang, J.; Yu, L.; Wang, J.; Guo, L.; Gu, X.; Fang, Q. Automated Fugl-Meyer Assessment Using SVR Model. In Proceedings of the 2014 IEEE International Symposium on Bioelectronics and Bioinformatics, Chung Li, Taiwan, 11–14 April 2014.
- Alsayed, A.; Kamil, R.; Ramli, H.; As'array, A. An Automated Data Acquisition System for Pinch Grip Assessment Based on Fugl Meyer Protocol: A Feasibility Study. *J. Appl. Sci.* **2020**, *10*, 3436.
- Formstone, L.; Pucek, M.; Wilson, S.; Bentley, P.; McGregor, A.; Vaidyanathan, R. Myographic Information Enables Hand Function Classification in Automated Fugl-Meyer Assessment. In Proceedings of the 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER), San Fransisco, CA, USA, 20–23 March 2019.
- Lee, S.; Lee, Y.S.; Kim, J. Automated Evaluation of Upper-Limb Motor Function Impairment Using Fugl-Meyer Assessment. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 125–134.
- Vijayvargiya, A.; Prakash, C.; Kumar, R.; Bansal, S.; Tavares, J.M.R. Human knee abnormality detection from imbalanced sEMG data. *Biomed. Signal Process. Control.* **2021**, *66*, 102406.
- Coi, S.; Seo, H.C.; Cho, M.S.; Joo, S.; Nam, G.B. Performance Improvement of Deep Learning Based Multi-Class ECG Classification Model Using Limited Medical Dataset. *IEEE Eng. Med. Biol. Soc. Sect.* **2023**, *11*, 53185–53194.

16. Hasni, H.; Yahya, N.; Asirvadam, V.S.; Jatoi, M.A. Analysis of Electromyogram (EMG) for Detection of Neuromuscular Disorders. In Proceedings of the 2018 International Conference on Intelligent and Advanced System (ICIAS), Kuala Lumpur, Malaysia, 13–14 August 2018.
17. Anatomy, Shoulder and Upper Limb, Hand Muscles. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK537229/> (accessed on 30 May 2024).
18. Adewuyi, A.; Hargrove, L.; Kuiken, T. An Analysis of Intrinsic and Extrinsic Hand Muscle EMG for Improved Pattern Recognition Control. *J. IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 485–494.
19. Hermens, H.J.; Freriks, B.; Disselhorst-Klug, C.; Rau, G. Development of recommendations for SEMG sensors and sensor placement procedures. *J. Electromyogr. Kinesiol.* **2000**, *10*, 361–374.
20. Luca, C.J.D.; Gilmore, L.D.; Kuznetsov, M.; Roy, S.H. Filtering the Surface EMG Signal: Movement Artifact and Baseline Noise Contamination. *J. Biomech.* **2010**, *43*, 1573–1579.
21. Wang, J.; Tang, L.; Bronlund, J.E. Surface EMG Signal Amplification and Filtering. *Int. J. Comput. Appl.* **2013**, *82*, 15–22.
22. Bansal, M.; Sharma, R.; Grover, P. Performance evaluation of Butterworth Filter for Signal Denoising. *Int. J. Electron. Commun. Technol. (IJECT)* **2010**, *1*, 59–62.
23. Ahsan, M.R.; Ibrahimy, M.I.; Khalifa, O.O. VHDL Modelling of Fixed-point DWT for the Purpose of EMG Signal Denoising. In Proceedings of the 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks, Bali, Indonesia, 26–28 July 2011.
24. Phinyomark, A.; Limsakul, C.; Phukpattaranont, P. Optimal Wavelet Functions in Wavelet Denoising for Multifunction Myoelectric Control. *Ecti Trans. Electr. Eng. Electron. Commun.* **2009**, *8*, 43–52.
25. Sousa, A.S.; Tavares, J.M.R. Surface electromyographic amplitude normalization methods: A review. *Electromyogr. New Dev. Proced. Appl.* **2012**, *20*, 85–102.
26. Sabri, M.; Miskon, M. F.; Yaacob, M. R.; Basri, A.; Soo, Y.; Bukhari, W. MVC based normalization to improve the consistency of emg signal. *J. Theor. Appl. Inf. Technol.* **2014**, *65*, 336–343.
27. Botelho, A.L.; Gentil, F.H.U.; Sforza, C.; Da Silva M.A.M.R. Standardization of the Electromyographic Signal Through the Maximum Isometric Voluntary Contraction. *J. Craniomandib. Sleep Pract.* **2011**, *29*, 23–31.
28. Tanaka, T.; Nambu, I.; Maruyama, Y.; Wada, Y. Sliding-Window Normalization to Improve the Performance of Machine-Learning Models for Real-Time Motion Prediction Using Electromyography. *Sensors* **2022**, *22*, 5005.
29. Shuman, B.R.; Schwartz, M.H.; Steele, K.M. Electromyography Data Processing Impacts Muscle Synergies during Gait for Unimpaired Children and Children with Cerebral Palsy. *Front. Comput. Neurosci.* **2017**, *11*, 1–9.
30. Beck, T.W.; Housh, T.J.; Cramer, J.T.; Mielke, M.; Hendrix, Russell. The influence of electrode shift over the innervation zone and normalization on the electromyographic amplitude and mean power frequency versus isometric torque relationships for the vastus medialis muscle. *J. Neurosci. Methods* **2008**, *169*, 100–108.
31. AlQudah, A.; Bariouli, R.; Lweesy, K.; Elkhalil, H.; Ibbini, M.; Kanoun, O. Electrical Impedance Myography Measurements for Gesture Recognition Data Normalization. In Proceedings of the 2022 International Workshop on Impedance Spectroscopy (IWIS), Chemnitz, Germany, 27–30 September 2022.
32. Qing, Z.; Lu, Z.; Cai, Y.; Wang, J. Elements Influencing sEMG-Based Gesture Decoding: Muscle Fatigue, Forearm Angle and Acquisition Time. *Sensors* **2021**, *21*, 7713.
33. Lee, K.H.; Min, J.Y.; Byun, S. Electromyogram-Based Classification of Hand and Finger Gestures Using Artificial Neural Networks. *Sensors* **2021**, *22*, 225.
34. Phinyomark, A.; Limsakul, C.; Phukpattaranont, P. EMG Feature Extraction for Tolerance of White Gaussian Noise. In Proceedings of the International Workshop and Symposium Science Technology (I-SEEC), Nong Khai, Thailand, 15–16 December 2008.
35. Wang, J.; Cao, D.; Wang, J.; Liu, C. Action Recognition of Lower Limbs Based on Surface Electromyography Weighted Feature Method. *Sensors* **2021**, *21*, 6147.
36. Thongpanja, S.; Phinyomark, A.; Quaine, F.; Laurillau, Y.; Limsakul, C.; Phukpattaranont, P. Probability Density Functions of Stationary Surface EMG Signals in Noisy Environments. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 2016.
37. Oo, T.; Phukpattaranont, P. Signal-to-Noise Ratio Estimation in Electromyography Signals Contaminated with Electrocardiography Signals. *Fluctuations Noise Lett.* **2020**, *19*, 2050027.
38. Bein, B. Entropy. *Best Pract. Res. Clin. Anaesthesiol.* **2006**, *20*, 101–109.
39. Bendat, J.; Piersol, A. *Random Data: Analysis and Measurement Procedures*, 3rd ed.; Wiley: New York, NY, USA, 2000.
40. Dendamrongvit, S.; Kubat, M. Undersampling Approach for Imbalanced Training Sets and Induction from Multi-label Text-Categorization Domains. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Bangkok, Thailand, 27–30 April 2009.
41. Tusell-Rey, C.C.; Camacho-Nieto, O.; Yáñez-Márquez, C.; Villuendas-Rey, Y. Customized Instance Random Undersampling to Increase Knowledge Management for Multiclass Imbalanced Data Classification. *Sustainability* **2022**, *14*, 14398.
42. Ratnasari, A.P. Performance of Random Oversampling, Random Undersampling, and SMOTE-NC Methods in Handling Imbalanced Class in Classification Models. *Int. J. Sci. Res. Manag. (IJSRM)* **2024**, *12*, 494–501.
43. Al-Faiz, M.Z.; Ibrahim, A.A.; Hadi, S.M. The effect of z-score standardization on binary input due the speed of learning in back-propagation neural network. *Iraqi J. Inf. Commun. Technol. (IJICT)* **2018**, *1*, 42–48.

44. Suma, V.R.; Renjith, S.; Ashok, S.I. Judy, M.V. Analytical Study of Selected Classification Algorithms for Clinical Dataset. *Indian J. Sci. Technol.* **2016**, *9*, 1–9.
45. Long, X.; Fonseca, P.; Haakma, R.; Foussier, J.; Aarts, R.M. Automatic detection of overnight deep sleep based on heart rate variability: A preliminary study. In the Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014.
46. Toledo-Pérez, D.C.; Rodríguez-Reséndiz, J.; Gómez-Loenzo, R.A.; Jauregui-Correa, J.C. Support Vector Machine-Based EMG Signal Classification Techniques: A Review. *J. Appl. Sci.* **2019**, *9*, 4402.
47. Oskoei, M.A.; Hu, H. Support Vector Machine-Based Classification Scheme for Myoelectric Control Applied to Upper Limb. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 1956–1965.
48. Al-Sharu, W.N.; Alqudah, A.M. Enhancing Prediction of Prosthetic Fingers Movement Based on sEMG using Mixtures of Features and Random Forest. *Int. J. Recent Technol. Eng.* **2019**, *8*, 289–294.
49. Li, Z.; Guan, X.; Zou, K.; Xu, C. Estimation of Knee Movement from Surface EMG Using Random Forest with Principal Component Analysis. *Electronics* **2019**, *9*, 43.
50. KrishnaVeni, C.V.; Rani, T.S. On the Classification of Imbalanced Datasets. *Int. J. Comput. Sci. Technol.* **2011**, *2*, 145–148.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.