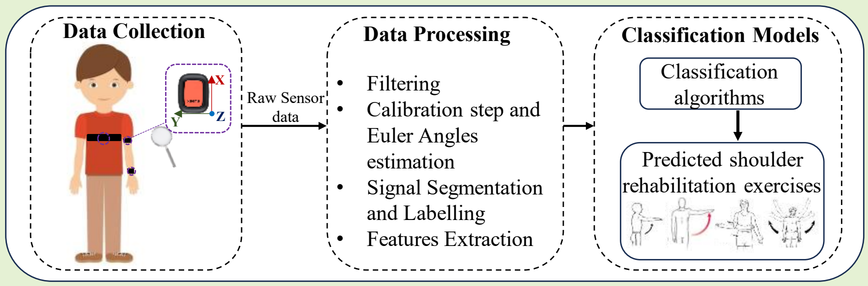
This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3406138

Classification of shoulder rehabilitation exercises by using wearable systems and machine learning algorithms

Martina Sassi, Arianna Carnevale (Member, IEEE), Matilde Mancuso, Emiliano Schena (Senior Member, IEEE), Leandro Pecchia (Member, IEEE) and Umile Giuseppe Longo

***Abstract*— Shoulder rehabilitation is considered one of the most effective treatments for restoring functional abilities, reducing shoulder pain, and enabling the leading of an active life, improving mobility, strength, and endurance. However, the burdens of travel and time may prevent patients from taking part in such rehabilitation programs. The increased availability of wearable sensors and the development of machine learning (ML) algorithm has shown the feasibility of remote home-based rehabilitation therapy. In this study, we proposed a wearable system based on 3 magneto-inertial sensors to classify shoulder rehabilitation exercises. The classification has been performed by 5 different supervised ML algorithms (i.e., k-Nearest Neighbours, Support Vector Machine, Naïve Bayes, Decision Tree, and Random Forest) to find out the most performant one.**

The feasibility of the wearable system was assessed on nineteen healthy subjects during six rehabilitation exercises. Each exercise was performed six times, for a total of 684 samples. The data were analysed and classified using the five mentioned classification models. Performances of the algorithms in accurately classifying exercise activity were evaluated with the k-fold cross-validation method and the nested validation method. The results demonstrated the effectiveness of the proposed algorithms in recognizing all the exercises. Features derived from acceleration, angular velocity, and orientation data were shown to reach the optimal predictive accuracies. Future work should focus on evaluating the performance of such systems on data acquired on patients with musculoskeletal disorders and on the inclusion of more shoulder rehabilitation exercises in the protocol.



***Index Terms*— Activity recognition, classification, inertial measurement unit, machine learning, rehabilitation exercises, shoulder, wearable sensors.**

I. Introduction

SHOULDER disorders (SD) represents the most frequently reported musculoskeletal disorders, entailing pain, reduced functionalities, and a decreased quality of life [1]. An adequate rehabilitation protocol represents the primary therapeutic protocol to guarantee the return of complete shoulder function [2], [3], [4]. Different treatments methods exist to execute medical rehabilitation. Among these, physical therapy, also known as physiotherapy, aims to restore functional abilities, enable the leading of an active life, improving mobility, strength and endurance [5], [6]. Traditional shoulder rehabilitation methods consist of a therapist-patient one-to-one activity, and on the execution of physical exercises [7]. Physical therapists actively monitor and direct patients through their rehabilitation process while they are in a hospital or clinical setting [6]. The traditional rehabilitation process is time-consuming, requires going directly to the physiotherapy centre for each session, is restricted by the availability of trained clinicians and places a

MS and LP were supported in part by the ODIN project, as part of the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 101017331. AC, MS and GUL have been funded by the Italian Ministry of Health in the framework of RICERCA FINALIZZATA 2021, grant agreement Nº RF-2021-12372810. (*Corresponding author: E. Schena).*

Martina Sassi, and Leandro Pecchia, Unit of Intelligent Health Technologies, Sustainable Design Management and Assessment, Department of Engineering, Università Campus Bio-Medico di Roma, Via Álvaro del Portillo, 21, 00128 Rome, Italy (e-mail: martina.sassi@unicampus.it, leandro.pecchia@unicampus.it).

Arianna Carnevale, and Matilde Mancuso, Fondazione Policlinico Universitario Campus Bio-Medico di Roma, Via Álvaro del Portillo, 200, 00128 Rome, Italy (e-mail: arianna.carnevale@policlinicocampus.it, m.mancuso@policlinicocampus.com).

Emiliano Schena, Laboratory of Measurement and Biomedical Instrumentation, Department of Engineering, Università Campus Bio-Medico di Roma, Via Álvaro del Portillo, 21, 00128 Rome, Italy (e-mail: e.schena@unicampus).

Umile Giuseppe Longo, Fondazione Policlinico Universitario Campus Bio-Medico di Roma, Via Álvaro del Portillo, 200, 00128 Rome, Italy; Research Unit of Orthopaedic and Trauma Surgery, Department of Medicine and Surgery, Università Campus Bio-Medico di Roma, Via Álvaro del Portillo, 21, 00128 Roma, Italy (e-mail: g.longo@policlinicocampus.it).

significant economic burden on patients [8]. Therefore, the effectiveness of rehabilitation is primarily dependent on the patient's engagement, which can be affected by a variety of factors [8]. Considering the increasing incidence of SD and progressive population ageing, during the last decades there is a demand for an efficient home-based rehabilitation therapy [9]. Patients perform the prescribed treatment independently in their home environment. However, unlike sessions conducted under the supervision of a therapist, successful self-home therapy demands a significant increase in commitment from the patients [6]. According to the European Musculoskeletal Conditions Surveillance and Information Network, the breakthrough for an effective treatment of muscular skeletal disorders is the proactive participation of the patient. Challenges that impact the effectiveness of home-based programs include adherence to the prescribed rehabilitation program, and exercise correctness. Evidence suggests that patients often do not fully comply to the prescribed program of exercise [6], [10]. Consequences of a non-adherence to the prescribed rehabilitation program are the prolongation of the duration of treatment and the risk of relapse. Additionally, without the supervision of their therapist, many patients perform their exercises incorrectly [6], [11]. Therefore, objective and quantitative assessment of adherence to exercise programs and of exercise performance are necessary to improve rehabilitation outcomes [12], [13]. Quantitatively assessing adherence and the execution of exercises offers several advantages in monitoring and improving the overall efficacy of rehabilitative treatments. Such measures enable clinicians to evaluate the extent to which patients are adhering to therapeutic prescriptions, allowing timely interventions to provide feedback, patient engagement, and adjustments to the ongoing rehabilitation program based on individual needs. A variety of sensors have been introduced to address the demand for gathering objective data of movement quality in the home settings (see Fig. 1) [14], [15]. However, most of them are often not suitable for home-based rehabilitation [16], [17]. Optical sensors are widely used to monitor human activities, but the effective use of these systems is not practical in many indoor environments since they suffer from lighting variations, environmental occlusion and space constraints [18], [19]. Nowadays, wearable systems can be directly attached to the user ensuring all-time data collection [13], [20], [21], [22], [23], [24]. These solutions may allow the tracking of patient functioning and recovery during rehabilitation protocol [25]. Among several sensors, magneto-inertial measurement units (M-IMUs) are spreading to develop wearable systems since they are portable, inexpensive, and unobtrusive [1], [21], [26], [27]. Data recorded by IMUs components (e.g., accelerometers [28], [29], [30], [31], a combination of accelerometers and gyroscopes [17], [24], [32], [33] , orientation data [32], [34], [35], [36]) are used to the automatic detection of physical activities with different algorithms . Regarding applications to shoulder motion, several studies had used Machine Learning (ML) algorithms based on M-IMUs' data [11], [16], [17]. Heterogeneity among studies is relative to the type, the number and the placement location of the sensors on the human body, as well as the executed shoulder exercises and the implemented ML algorithms [37]. Regarding the set of exercises executed, only a few movements have been investigated. Some studies limit their analysis only to planar motion movements, such as flexion/extension [29], [33], [38], [39], [40], abduction/adduction [29], [33], [38], [40], internal/external rotations [29], [36], [38], [40],; instead, other studies include also more complex functional tasks, such as touch ear, use fingers to climb wall, pendulum, hand-to-back [28], [33], [36], [38], [39], [40]. However, it is still challenging to recognize the exercise performed by the subjects in unstructured environment.

The objective of this study is to combine a custom wearable system based on 3 M-IMUs with supervised ML algorithms to classify six of the most relevant exercises in shoulder rehabilitation [41]. The innovative configuration of the proposed wearable system allowed for a comfortable solution with an easy and fast setup, offering a practical solution for monitoring shoulder rehabilitation sessions. To understand how the selected algorithm influences the performance of the system in terms of exercise classification, we analised the experimental data recorded on nineteen subjects performing six exercises of shoulder rehabilitation with 5 ML algorithms (i.e., k-Nearest Neighbours, Support Vector Machine, Naïve Bayes, Decision Tree, and Random Forest). This study poses the bases for the possible application of the proposed system for monitoring home-based rehabilitation sessions. The ease of setup and modularity of the proposed system enhance the patient's ability to self-position the sensing units without requiring operator support. In the future, this solution may provide complete and useful data to the clinicians to monitor patient progress remotely, correct the ongoing rehabilitation process if needed. This enables to customize rehabilitation programs based on individual patient needs, improving the patients' outcomes.

The paper is structured as follows: Section 2 describes the experimental setup used, the dataset, and the human activity recognition workflow; Section 3 present the results; and Section 4 discuss the results and concludes the paper.

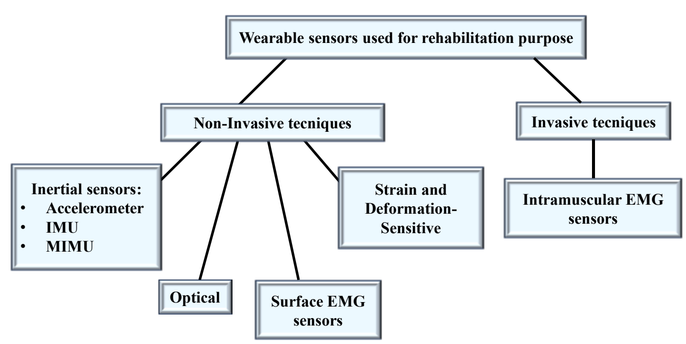


Fig. 1. Wearable sensors for rehabilitation purpose.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3406138

II. Methods And Materials

A. M-IMU Based Wearable System

A wearable system equipped with three M-IMUs (Xsens DOT, Xsens Technologies, Enschede, the Netherlands) has been used in this study [42]. Each Xsens DOT incorporates 3D-gyroscopes, accelerometers, and magnetometers. Xsens DOTs are small (36.3 x 30.4 x 10.8 mm - length x width x height), lightweight (10.8 g), and wireless sensors. The embedded processor in the sensors handles sampling, calibration, and Strap-Down Integration (SDI) of inertial data. Raw data are initially collected at high frequency, and then down sampled to 60 Hz for transmission.

The Xsens DOTs communicated wirelessly via Bluetooth 5.0 with a smartphone (OnePlus 8T - 8GB RAM + 128GB ROM, processor Snapdragon™865) running the Xsens DOT App for Android. Synchronization of the sensors is initiated through the application, requiring approximately 14 seconds. This process ensures that all sensor data are accurately time-synchronized to a common sensor time base.

The wearable system is characterized by an easy and fast setup. The three M-IMUs were fastened to body districts using elastic straps provided by Xsens to ensure reliable positioning by preventing slippage with the underlying skin. Each sensor was firstly placed horizontally inside the pocket of the corresponding strap, with the Y-axis pointing upwards. Then the three straps were wrapped around the segments of interest. Fig. 2 shows the final positions of the sensors in the wearable system. One sensor was positioned on the thorax over the flat portion of sternum, with the Y-axis pointing upward cranially, the Z-axis pointing away from the body, and the X-axis pointing laterally to the left. Another sensor was placed slightly posterior on the upper arm near the elbow, with the Y-axis pointing upward, the X-axis pointing laterally to the right, and the Z-axis to complete the right-handed coordinate system. The remaining sensor was placed on the forearm's dorsal side near the wrist, with the Y-axis pointing upward, the X-axis pointing away from the body, and the Z-axis pointing laterally to complete the right-handed coordinate system.

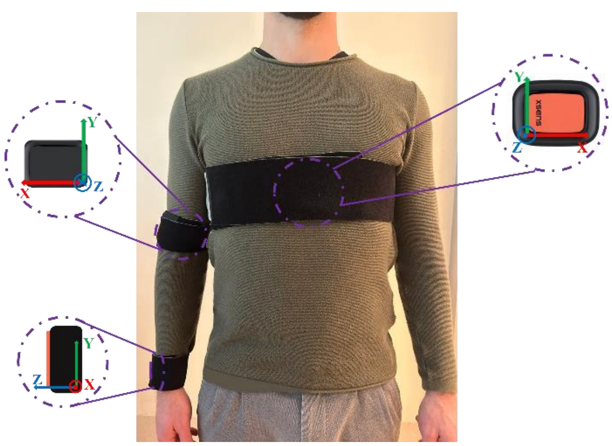


Fig. 2. Xsens DOT placement. The three straps were wrapped around the segments of interest. The circles show the coordinate systems of the three sensors: red, green, and blue arrows represent X-axis, Y-axis, and Z-axis, respectively. The dot indicates an outgoing arrow, while the cross indicates an incoming arrow.

B. Experimental Protocol

Nineteen healthy volunteers (5 male and 14 female) with no shoulder musculoskeletal disorders were enrolled in this study All participants were right-handed. The characteristics of the younger cohort are (mean ± standard deviation): age, 25.2 ±1.7 years; height, 167. 9 ± 8.5 cm; weight, 61.6 ± 11.9 kg. Specifically, for female volunteers, the age ranged from 23 to 28 years, the height from 156 to 170 cm, and the weight from 46 to 80 kg. Instead, for male volunteers, the age ranged from 24 to 26 years, the height from 170 to 187 cm, and the weight from 60 to 88 kg.

The experiments have been carried out at the biomechanical laboratory of the Fondazione Policlinico Universitario Campus Bio-Medico of Rome. Before experimental sessions, all volunteers read and signed an informed consent approved by the Ethical Committee of University Campus Bio-Medico of Rome (protocol code: 09/19 OSS ComEt UCBM). Then, the volunteers were instructed on the protocol consisting of a static trial and six dynamic tasks. The static recording, known as N-pose, corresponds to an anatomic stance with the arms at the sides and the palms of the hands facing internally (Fig. 3). Six shoulder rehabilitation exercises were selected from the guidelines developed by the American Society of Shoulder and Elbow Therapists [41]: Task 1) upright active flexion/extension; Task 2) upright active flexion/extension with a weight (2 kg); Task 3) external rotation with the shoulder at 90° of adduction, holding a weight (2 kg); Task 4) towel slide; Task 5) external/internal rotation self-assisted with a stick; Task 6) abduction/adduction (see Table I). Under supervision, each subject was required to complete six consecutive repetitions of each task at a comfortable and self-selected speed. Therefore, a total of 684 shoulder movements (19 subjects x 6 tasks x 6 repetitions) were analysed.

C. Data Analysis

The data analysis was performed offline in MATLAB environment (version R2022b, TheMathWorks® Inc., Natick, MA, USA). The ML approach is composed of the following steps: 1) signal pre-processing, 2) signal segmentation and labelling, 3) features extraction, 4) features standardization and selection, and 5) classification and validation.

Fig. 3. N-pose: anatomic stance with arms along the sides, and the palms of the hands facing internally. (a) Frontal view; (b) Right side view.

TABLE I
SHOULDER PHYSIOTHERAPY EXERCISES FOR DATA COLLECTION

Task ID	Shoulder Task	Abbreviation
1	Flexion/extension without a weight	FE
2	Flexion/extension with a weight (2 kg)	FEd
3	External rotation with the shoulder at 90° of adduction, holding a weight (2 kg)	ERs
4	Towel slide	SL
5	External/internal rotation self-assisted with a stick	EIR
6	Abduction/adduction	AA

1) *Pre-processing*

The data of delta angle, delta velocity, and 3D orientation (expressed by quaternions) were collected with a sampling frequency of 60 Hz. A low pass fifth order Butterworth filter with a cut off frequency of 2 Hz was applied to delta angle and delta velocity data to remove high-frequency noise. Angular velocity and acceleration were obtained from the filtered data of delta angle and delta velocity, respectively.

2) *Calibration and Euler angle estimation*

To estimate joint angles, it is necessary to measure the relative orientation of two adjacent body segments forming the joint [43]. The output quaternion from each M-IMU represents the orientation of the sensor coordinate system with respect to the Local Earth-fixed reference coordinate system. These outputs cannot be converted into clinically interpretable data because the coordinate frames of the sensors are not aligned with the anatomical coordinate frame of the respective body segment to which they are attached [26], [27]. The aim of the sensor-to-segment calibration is to express the relative orientation of each sensor to the segment to which it is attached [44]. In this study, the static sensor-to-segment calibration algorithm was performed by preprocessing data acquired from Xsens DOTs over the static N-pose acquisition [43], [45]. After the calibration quaternions have been calculated, joint rotations were estimated as the relative orientation of two adjacent body segments [23]. Specifically, humerothoracic (HT) joint angles were defined as the orientation of the humerus body segment relative to the thorax body segment, whereas elbow joint angles were defined as the orientation of the forearm body segment relative to humerus body segment [46]. Subsequently, a conversion from quaternion to rotation angles was performed using different Euler rotation sequences. The HT joint angles were evaluated using the Cardan sequence XZY for Task 6, and the Cardan sequence ZXY for all the other tasks [47]. Whereas the elbow joint angles were assessed using the Cardan sequence ZXY during all the exercises [27].

3) *Signal Segmentation and Labelling*

A manual segmentation was firstly performed to isolate every single repetition performed by each subject. The signal considered in subsequent analysis was the one between the beginning of each repetition of the movement and the end of the same repetition (Fig. 4). Each isolated repetition of each task was considered as a sample, resulting 684 overall (114 samples for each task). Since a supervised learning was implemented, a unique label was attributed to each sample (see Table I): FE for Task 1, FEd for Task 2, ERs for Task 3, SL for Task 4, EIR for Task 5, and AA for Task 6, providing six activity classes in total.

4) *Features Extraction*

Afterward, a feature extraction process was performed, which consist of providing the most relevant information that will have a crucial role in the classification process [48]. Specifically, the following features were extracted: variance, mean, standard deviation, median, maximum value, minimum value, range, root mean square, interquartile range (between 25th and 75th percentiles), correlation coefficient, kurtosis, and skewness. These calculations were automated and carried out for each sample. Since one of the goals is to show and compare the effectiveness of different data in human activity recognition when they are used separately, these features were extracted from the triaxial accelerometer data, the triaxial gyroscope data, the quaternion data, and from the Euler Angles of humerothoracic and elbow joints.

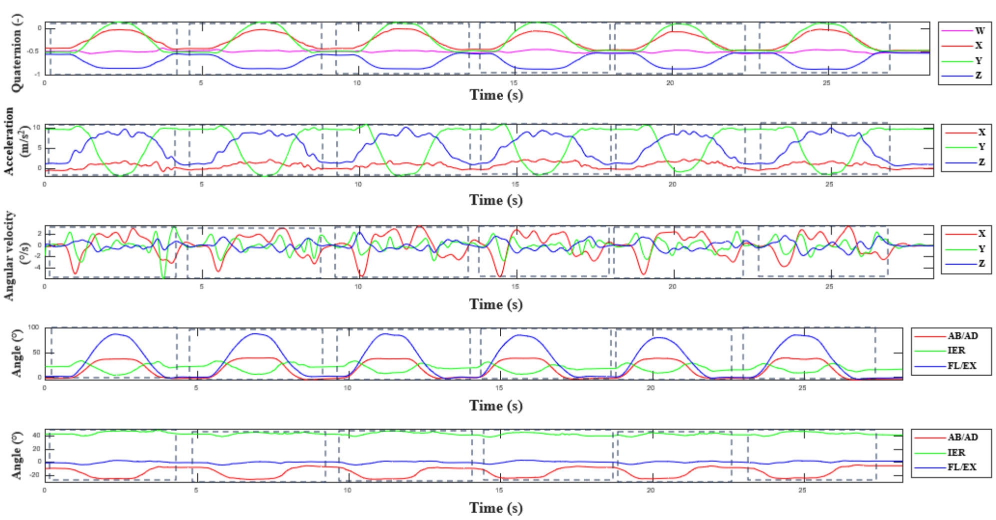


Fig. 4. Signal segmentation of: (a) quaternion data, (b) accelerometer data, (c) angular velocity data, (d) Euler angles of humerothoracic joint, (e) Euler angles of elbow joint, acquired by forearm M-IMU during the six repetitions of Task 1.

5) *Features Standardization and Selection*

Then, features standardization was conducted according to (1), so they were scaled to zero mean and unit variance. Standardization allows all features contribute equally to the classification process.

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

where x and x' are the original and normalized features, respectively, and μ and σ are the mean and standard deviation of the x signals, respectively. Feature selection constitutes an essential phase for improving classification accuracy [33], [48]. Although all the features could be useful to represent the data, it is not a good procedure to employ a large number of features. The objective of features selection is to identify a subset of relevant features which are highly informative regarding classification process, and eliminate irrelevant and redundant attributes. This involves reducing the complexity of the model, obtaining good generalization to avoid overfitting and avoid the curse of dimensionality [6], [49]. The Relief-F feature filtering method was implemented in this study to determine the most appropriate feature sets. This algorithm assigns a weight value W to each feature depending on how well its value distinguish between instances, and it ranks them according to feature relevance scores [38], [49]. After the setting of an empirical threshold, only features that have a weight greater than it are selected, whereas those below the threshold are excluded.

6) *Classification and Validation Method*

The classifiers chosen for this study represent a range of supervised machine learning models successfully implemented in previous shoulder motion classification studies. The five supervised machine learning classification models were: k-Nearest Neighbours (k-NN), Support Vector Machine (SVM) extended for multiple class classification scenario (using the one-versus-one method), Naïve Bayes (NB), Decision Tree (DT), and Random Forest (RF) with an ensemble of 180 trees [50], [51].

The entire dataset was divided into two portions, a training part (90% of the dataset) and a test part (10% of the dataset). The data of the remaining 10% of the dataset (2 subjects) were extracted to further validate effectiveness of the classification models. The training of the classifiers on the other 17 subjects was performed using two different validation methods. *K-folds cross validation* (CV) randomly distributes all labelled samples into K folds of equal size [52]. Stratification was employed to assure each fold was representative of the cohort. Training is performed on data contained within K-1 folds, and testing is performed on the remaining fold. This process is repeated K times, to ensure all data is used for training and testing once. At the end, the K results obtained for all the experiments were averaged to provide a single estimation of training performances (Fig. 5). *Nested cross validation* (NCV), also known as *double cross-validation*, consists of splitting data into K outer folds: each fold is held out for the test, while the remaining K-1 folds are merged and further split obtaining a sub-training and validation datasets (Fig. 6). Within each of these sub-folds, the classification models are trained on the sub-training dataset and tested on the validation dataset. Then, the best subset of features with the best performance across the validation datasets is selected and used to train the classification model on the entire set of the outer training dataset. The model is then tested on the outer testing dataset [49], [52].

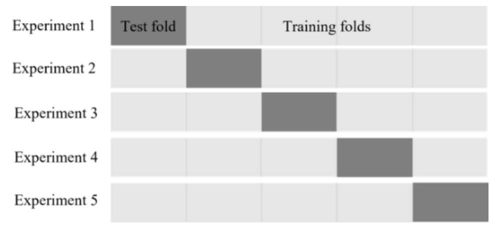


Fig. 5. K-folds cross validation. Split the data (90% of the entire dataset) into K folds (K = 5 as example). Training is performed on data of K-1 folds, and testing is performed on the remaining fold.

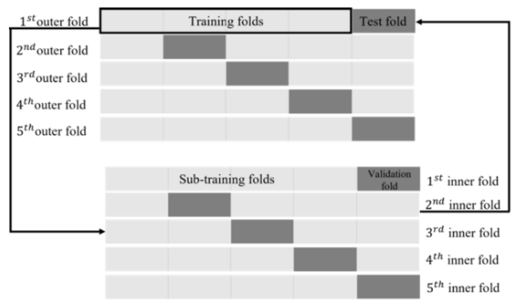


Fig. 6. Nested cross validation. Split the data (90% of the entire dataset) into K outer folds (K = 5 as example). Then, all the K-1 folds are merged and split into inner folds (5 inner folds as example). Feature selection and training are performed using the inner sub-training folds, and testing is performed on the remaining inner fold. Use the best inner training model including features extracted for train the classification model on the entire outer training dataset and test on the outer testing fold.

7) *Performance Metrics*

A confusion matrix (CM) is a table that enables the visualization of the classifier's performance in classifying the label of a test set for which the true labels are knows. The rows of the CM correspond to the true classes, whereas the columns correspond to the predicted classes. The diagonal cells represent those samples that are correctly classified, while the off-diagonal values are the incorrectly classified samples. In particular: TP is true positive, which represents the number of positive observations that were predicted as positive by the model; FP is false positive, which represents the number of negative observations that have been predicted as positive by the model; FN is false negative, which represents the number of positive observations that were predicted as negative by the model; TN is true negative, which represents the number of negative observations that were predicted as negative by the model [53].

The classification performances of the models were assessed in terms of different metrics, which are based on the CM. The quality measures evaluated were accuracy (Acc), both overall and balanced, specificity (Sp), sensitivity (Se) or recall, precision (Pr). Accuracy measures the overall effectiveness of a classifier, and is computed as the ratio of correctly classified samples and the total numbers of samples. Specificity measures the ability of the classifier to detect negative labels, whereas the sensitivity measures the ability of the classifier to detect a desired label [6].

In addition, the F_β-score was also calculated and defined as follows:

$$F_\beta score = \frac{(1+\beta^2)\cdot(\mathrm{Pr}\cdot\mathrm{Se})}{(Pr+Se)} \quad (2)$$

In (2), β is a weighting factor that controls the degree of importance of sensitivity and precision. This parameter is a positive real number. In this paper β was set equal to 1, to give the same importance to both sensitivity and precision.

Other metrics, including the Matthews correlation coefficient (MCC), the Fowlkes–Mallows index, the Youden index (or informedness), the Prevalence Threshold (PT), were also computed to provide a more comprehensive assessment of the model's performance [54], [55].

Furthermore, the Receiver Operating Characteristic (ROC) curve provides a graphical representation of the classification performances [56], [57]. It represents the relation between the false positive rate (FPR) and the true positive rate (TPR), that can be calculated from the sensitivity and the specificity:

$$FPR = 1 - Sp \quad (3)$$

$$TPR = Se \quad (4)$$

It has been demonstrated that the area under the ROC curve (AUC) is an excellent indicator of the classification performance, because it visualizes classifier performance as a curve rather than a single scalar number, which conveys more information than many scoring measures.

III. Results

The implementation of the Relief-F algorithm as feature selection method involves the setting of a threshold. The features with a weight superior to this threshold are selected, whereas all features lower the threshold are excluded.

In general, the higher the threshold, the lower the number of selected features. To examine the impact of the number of the features on the performance of motion recognition of the shoulder exercises, different experiments were executed including a different number of features. Starting from a low threshold value, this was incremented by 0.010. The performance of all the classifiers were evaluated for each i-th iteration. Fig. 7 shows the relationship between the number of retained features at each iteration and the accuracy values of the five classification models. The trends for all the classifiers are generally similar. An increase in number of features implicates an increase on the classification accuracy of all the classifiers. At the end, the threshold was set at 0.09 because the further addition of features did not provide great improvement to the classifiers' performance.

Accurate classification of shoulder exercises is reliant on suitable type of sensor data. To figure out the most accurate sensor data for activities' recognition, a comparison was made between results obtained with different set of features as input. Table II and Table III summarize respectively the performances employing features extracted from acceleration, angular velocity, and quaternion data and employing features extracted from Euler angles of the humerothoracic and elbow joints. Then, the averages of these metrics were calculated considering all the classifiers (Table IV).

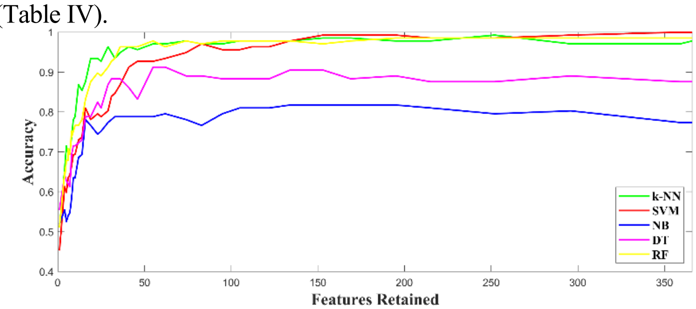


Fig. 7. Relationship between the features selected by Relief-F algorithm at different threshold's values and classifier accuracies. SVM = Support Vector Machine, k-NN = k-Nearest Neighbors, NB = Naïve Bayes, DT = Decision Tree, RF = Random Forest.

TABLE II
METRICS OF PERFORMANCE USING FEATURES EXTRACTED FROM ACCELERATION, ANGULAR VELOCITY, AND QUATERNION DATA, IMPLEMENTING 5 FOLDS CROSS VALIDATION

Classifier	Overall Accuracy	Balanced Accuracy	F1 score	Sensitivity	Specificity	Precision	MCC	FM	PT	Youden
k-NN	0.8472	0.9083	0.8355	0.8472	0.9694	0.8420	0.8214	0.8400	0.1157	0.8167
SVM	0.9306	0.9583	0.9304	0.9306	0.9861	0.9312	0.9169	0.9307	0.0618	0.9167
NB	0.9444	0.9667	0.9429	0.9444	0.9889	0.9583	0.9368	0.9471	0.0342	0.9333
DT	0.9444	0.9667	0.9459	0.9444	0.9889	0.9312	0.9836	0.9864	0.0342	0.9333
RF	0.9861	0.9917	0.9861	0.9861	0.9972	0.9872	0.8681	0.8881	0.0191	0.9833

K-NN = K-NEAREST NEIGHBOURS, SVM = SUPPORT VECTOR MACHINE, NB = NAÏVE BAYES, DT = DECISION TREE, RF = RANDOM FOREST.

TABLE III
METRICS OF PERFORMANCE USING FEATURES EXTRACTED FROM EULER ANGLES DATA OF THE HUMEROTHORACIC AND ELBOW JOINTS, IMPLEMENTING 5 FOLDS CROSS VALIDATION

Classifier	Overall Accuracy	Balanced Accuracy	F1 score	Sensitivity	Specificity	Precision	MCC	FM	PT	Youden
k-NN	0.7222	0.8333	0.6732	0.7222	0.9444	0.6830	0.6882	0.6875	0.2628	0.6667
SVM	0.7500	0.85	0.7444	0.7500	0.9500	0.7778	0.7059	0.7539	0.1431	0.7
NB	0.6389	0.7833	0.5820	0.6389	0.9278	0.5548	0.5871	0.5892	0.1672	0.5667
DT	0.6250	0.775	0.6134	0.6250	0.9250	0.7778	0.5567	0.6316	0.1125	0.55
RF	0.7778	0.8667	0.7690	0.7778	0.9556	0.8125	0.7466	0.7817	0.1620	0.7333

K-NN = K-NEAREST NEIGHBOURS, SVM = SUPPORT VECTOR MACHINE, NB = NAÏVE BAYES, DT = DECISION TREE, RF = RANDOM FOREST.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3406138

The use of inertial data obtained 93.05% of overall accuracy, 92.82% of F1 score, 93.05% sensitivity, 98.61% specificity, and 93% precision. Instead, the use of Euler angles obtained lower values, i.e., 70.28% of overall accuracy, 67.64% of F1 score, 70.28% sensitivity, 94.06% specificity, and 72.12% precision.

To better investigate the approach of using the features extracted from acceleration, angular velocity, and quaternion data as input to the classifiers, a comparison of the performances was carried out implementing different validation methods: 5-folds cross validation, 10-folds cross validation, and the nested cross validation (with 10 inner and outer folds). Table II, Table V, and Table VI detailed the performance metrics for each classifier with all these validation methods. High values of average performance metrics were obtained in every cases. Results point out that the employed classification protocol is efficient at recognizing the six shoulder exercises with overall accuracies values ranging between 84.72% and 98.61% implementing the 5 folds CV, between 83.33% and 100% implementing the 10 folds CV, and between 87.50% and 100% implementing the nested cross validation.

In addition, Table VII, Table VIII, and Table IX compares the accuracies in classifying each class separately, and then the averages of those values. Fig. 8, Fig. 9 and Fig. 10 show the related confusion matrices, whereas Fig. 11, Fig. 12 and Fig. 13 show the ROC curves graphs and the related values of the areas under the curve of all the classifiers.

Fig. 14 shows the features extracted with the proposed feature selection method. The three different columns indicate how many features are related to which data. In terms of percentage, by averaging the values obtained from the three validation methods, the Relief-F algorithm selected the 53.67% of features from the quaternion data, followed by the 42.33% of features extracted from acceleration data. Only the 4% of the features were the ones related to angular velocity data.

TABLE IV
OVERALL METRICS OF PERFORMANCE USING DIFFERENT FEATURES SETS

Type Data	Overall Accuracy	Balanced Accuracy	F1 score	Sensitivity	Specificity	Precision	MCC	FM	PT	Youden
Acceleration + angular velocity + quaternion	0.9305	0.9583	0.9282	0.9305	0.9861	0.9300	0.9054	0.9185	0.053	0.9167
Euler angles	0.7028	0.8217	0.6764	0.7028	0.9406	0.7212	0.6569	0.6888	0.1695	0.6433

TABLE V
PERFORMANCE METRICS FOR EACH CLASSIFIER AND FOR ALL THE TASKS PERFORMED USING 10-FOLD CROSS VALIDATION METHOD.

Classifier	Overall Accuracy	Balanced Accuracy	F1 score	Sensitivity	Specificity	Precision	MCC	FM	PT	Youden
k-NN	0.8750	0.925	0.8669	0.8750	0.9750	0.8797	0.8550	0.8721	0.0979	0.85
SVM	0.8889	0.9333	0.8881	0.8889	0.9778	0.8905	0.8675	0.8889	0.0794	0.8667
NB	0.9444	0.9667	0.9429	0.9444	0.9889	0.9583	0.9368	0.9471	0.0342	0.9333
DT	0.8333	0.9	0.8205	0.8333	0.9667	0.8905	0.8139	0.8383	0.0824	0.8
RF	1	1	1	1	1	1	1	1	0	1

K-NN = K-NEAREST NEIGHBOURS, SVM = SUPPORT VECTOR MACHINE, NB = NAÏVE BAYES, DT = DECISION TREE, RF = RANDOM FOREST.

TABLE VI
PERFORMANCE METRICS FOR EACH CLASSIFIER AND FOR ALL THE TASKS PERFORMED USING THE NESTED CROSS VALIDATION METHOD (WITH 10 INNER AND OUTER FOLDS).

Classifier	Overall Accuracy	Balanced Accuracy	F1 score	Sensitivity	Specificity	Precision	MCC	FM	PT	Youden
k-NN	0.8750	0.925	0.8669	0.8750	0.9750	0.8797	0.8550	0.8721	0.0979	0.85
SVM	0.8750	0.925	0.8730	0.8750	0.9750	0.8778	0.8518	0.8747	0.0846	0.85
NB	0.9444	0.9667	0.9429	0.9444	0.9889	0.9583	0.9368	0.9471	0.0342	0.9333
DT	0.8750	0.925	0.8655	0.8750	0.9750	0.8778	0.8556	0.8728	0.0964	0.85
RF	1	1	1	1	1	1	1	1	0	1

K-NN = K-NEAREST NEIGHBOURS, SVM = SUPPORT VECTOR MACHINE, NB = NAÏVE BAYES, DT = DECISION TREE, RF = RANDOM FOREST.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3406138

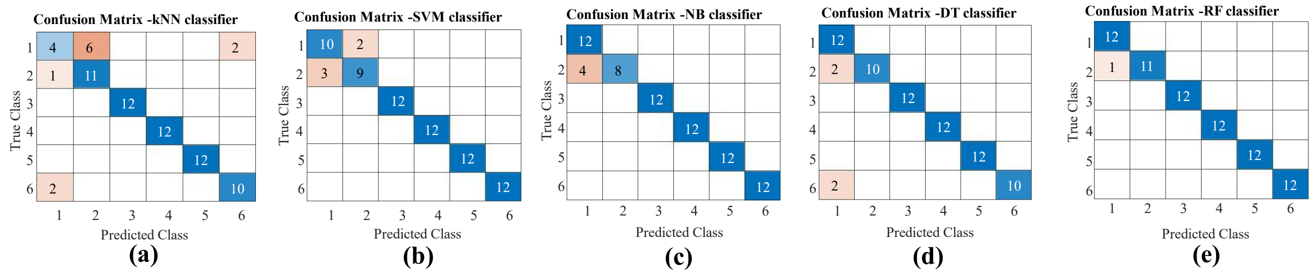


Fig. 8. Confusion matrices for activity recognition using the 5 folds cross validation. Shoulder exercises are as follow, 1: flexion/extension without a weight; 2: flexion/extension with a weight; 3: External rotation with the shoulder at 90° of adduction, holding a weight (2 kg); 4: Towel slide; 5: External/internal rotation self-assisted with a stick; 6: Abduction/adduction. (a), (b), (c), (d), and (e) represent confusion matrices for the kNN classifier, SVM classifier, NB classifier, DT classifier, and RF classifier, respectively.

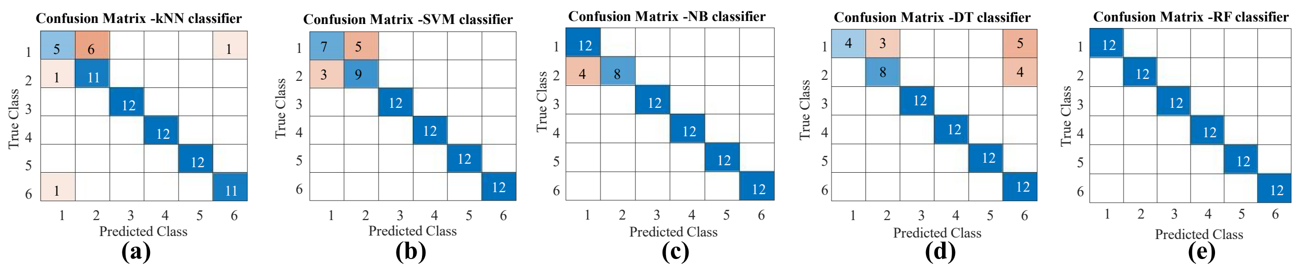


Fig. 9. Confusion matrices for activity recognition using the 10 folds cross validation. Shoulder exercises are as follow, 1: flexion/extension without a weight; 2: flexion/extension with a weight; 3: External rotation with the shoulder at 90° of adduction, holding a weight (2 kg); 4: Towel slide; 5: External/internal rotation self-assisted with a stick; 6: Abduction/adduction. (a), (b), (c), (d), and (e) represent confusion matrices for the kNN classifier, SVM classifier, NB classifier, DT classifier, and RF classifier, respectively.

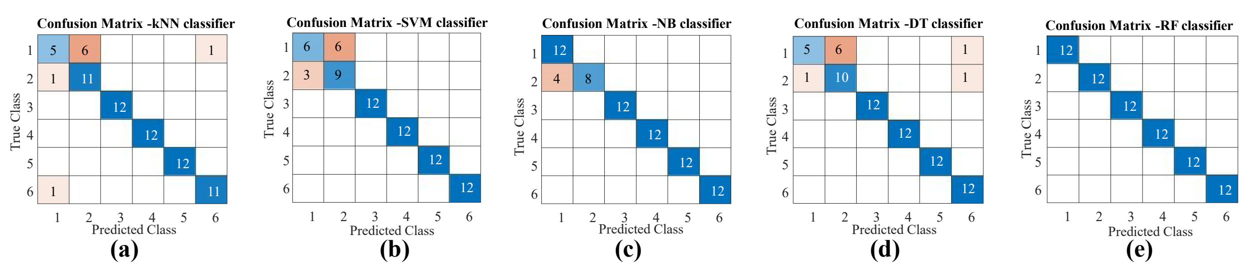


Fig. 10. Confusion matrices for activity recognition using the nested cross validation (with 10 inner and outer folds). Shoulder exercises are as follow, 1: flexion/extension without a weight; 2: flexion/extension with a weight; 3: External rotation with the shoulder at 90° of adduction, holding a weight (2 kg); 4: Towel slide; 5: External/internal rotation self-assisted with a stick; 6: Abduction/adduction. (a), (b), (c), (d), and (e) represent confusion matrices for the kNN classifier, SVM classifier, NB classifier, DT classifier, and RF classifier, respectively.

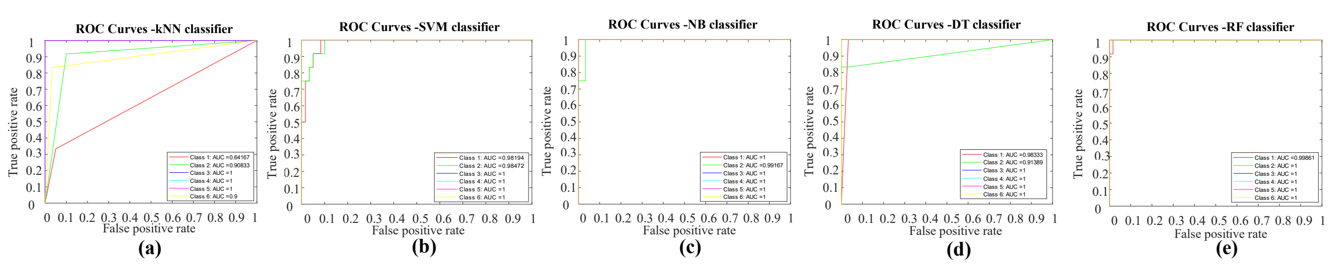


Fig. 11. ROC curves and areas under the curves (AUC) values using the 5 folds cross validation. (a), (b), (c), (d), and (e) represent ROC curves for the kNN classifier, SVM classifier, NB classifier, DT classifier, and RF classifier, respectively.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3406138

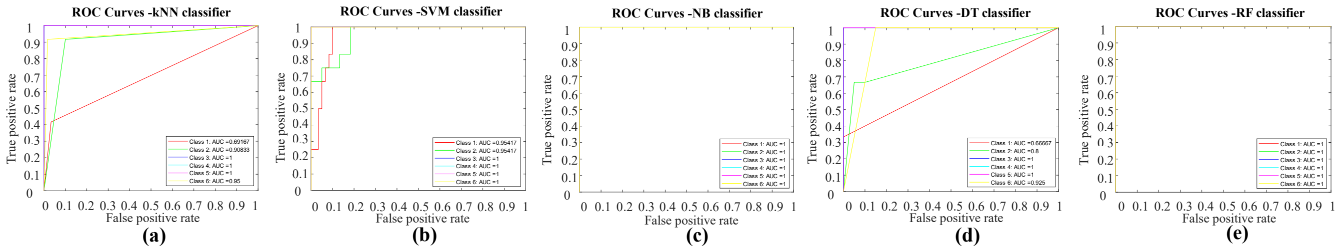


Fig. 12. ROC curves and areas under the curves (AUC) values using the 10 folds cross validation. (a), (b), (c), (d), and (e) represent ROC curves for the kNN classifier, SVM classifier, NB classifier, DT classifier, and RF classifier, respectively.

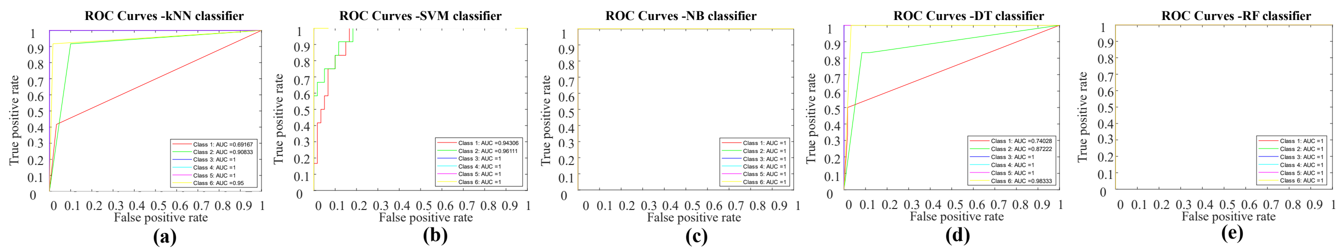


Fig. 13. ROC curves and areas under the curves (AUC) values using the nested cross validation (with 10 inner and outer folds). (a), (b), (c), (d), and (e) represent ROC curves for the kNN classifier, SVM classifier, NB classifier, DT classifier, and RF classifier, respectively.

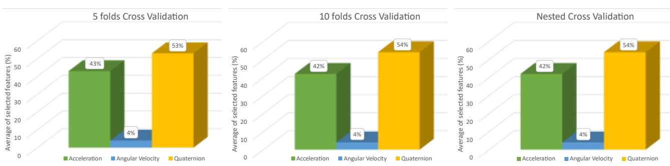


Fig. 14. Percentage of selected features with the Relief-F algorithm. In sequence, from left to right: implementing 5 folds cross validation, implementing 10 folds cross validation, and implementing nested cross validation (10 inner and outer folds).

TABLE VII
RECOGNITION ACCURACY FOR ALL 6 SHOULDER EXERCISES IMPLEMENTING 5-FOLD CROSS VALIDATION METHOD.

Classifier	FE	FEd	ERs	SL	EIR	AA	Averaged accuracy
k-NN	0.8472	0.9028	1	1	1	0.9444	0.9491
SVM	0.9306	0.9306	1	1	1	1	0.9769
NB	0.9444	0.9444	1	1	1	1	0.9815
DT	0.9444	0.9722	1	1	1	0.9722	0.9815
RF	0.9861	0.9861	1	1	1	1	0.9954

K-NN = K-NEAREST NEIGHBOURS, SVM = SUPPORT VECTOR MACHINE, NB = NAÏVE BAYES, DT = DECISION TREE, RF = RANDOM FOREST.

TABLE VIII
RECOGNITION ACCURACY FOR ALL 6 SHOULDER EXERCISES IMPLEMENTING 10-FOLD CROSS VALIDATION METHOD.

Classifier	FE	FEd	ERs	SL	EIR	AA	Averaged accuracy
k-NN	0.8750	0.9028	1	1	1	0.9722	0.9583
SVM	0.8889	0.8889	1	1	1	1	0.9630
NB	0.9444	0.9444	1	1	1	1	0.9815
DT	0.8889	0.9028	1	1	1	0.8750	0.9444
RF	1	1	1	1	1	1	1

K-NN = K-NEAREST NEIGHBOURS, SVM = SUPPORT VECTOR MACHINE, NB = NAÏVE BAYES, DT = DECISION TREE, RF = RANDOM FOREST.

TABLE IX
RECOGNITION ACCURACY FOR ALL 6 SHOULDER EXERCISES IMPLEMENTING THE NESTED CROSS VALIDATION METHOD (WITH 10 INNER AND OUTER FOLDS.

Classifier	FE	FEd	ERs	SL	EIR	AA	Averaged accuracy
k-NN	0.8750	0.9028	1	1	1	0.9722	0.9583
SVM	0.8750	0.8750	1	1	1	1	0.9583
NB	0.9444	0.9444	1	1	1	1	0.9815
DT	0.8889	0.9028	1	1	1	0.9722	0.9583
RF	1	1	1	1	1	1	1

K-NN = K-NEAREST NEIGHBOURS, SVM = SUPPORT VECTOR MACHINE, NB = NAÏVE BAYES, DT = DECISION TREE, RF = RANDOM FOREST.

IV. DISCUSSION AND CONCLUSION

This study investigated the potential application of a wearable system based on 3 M-IMUs in classifying six shoulder rehabilitation exercises. The use of Euler angles of humerothoracic and elbow joints can be interesting as they can better discriminate the assessed exercises. However, this attitude representation suffers from gimbal lock problem: orientation singularities can make Euler angles unsuited to correctly represent the different activities in some cases and consequently they will produce less accurate results. Indeed, Table III and Table IV shows that the use of the features set related to Euler angles decreases the overall recognition accuracy (70.28%). This indicates that the inertial data contains more discriminant information than the Euler angles in human activity recognition. Zmitri et al. performed the same analysis implementing the leave-one out cross validation technique, obtaining lower accuracy value when using Euler angles data (80.3%) than when using quaternion data (87.9%) [32].

Results shared above demonstrate the effectiveness of the proposed ML algorithms in classifying shoulder rehabilitation exercises. High recognition performances were obtained with all the implemented validation methods. Experimental results indicate an excellent recognition rate and a high level of agreement between the classification results and the true labels. Table II shows the performances of all the classifiers implementing the 5 folds CV. In this validation approach, the k-NN classifier achieved an overall accuracy of 84.72%, the SVM classifier attained 93.06%, both the NB and DT classifiers achieved an accuracy of 94.44%, and the RF achieved an overall accuracy of 98.61%. Averaged accuracies were notably higher for most classifiers, with the k-NN reaching 94.91%, and both the NB and DT achieving 98.15%, while the RF reached an outstanding 99.54%. The SVM classifier demonstrated an averaged accuracy of 97.69%, exceeding the reported accuracy in a comparable study (96.85%) where 5 exercises were classified [28]. The RF classifier performed significantly better than all the other ones with an overall and an averaged accuracy equal to 98.61% and 99.54% respectively. It achieves high values also for the other metrics, such as 98.61%, 98.61%, 99.72%, and 98.72% for F1 score, sensitivity score, specificity score, and precision score respectively. Additionally, the other metrics, including 86.81% for MCC, 88.81% for FM, 0.0191 for PT, and 0.9833 for Youden, indicated perfect classification results.

The implementation of the 10 folds cross validation method improves classification performances of almost all the classifiers. Table V shows that an overall accuracy equal to 87.50%, 88.90%, 94.44%, and 100% was achieved by the k-NN, the SVM, the NB, and the RF respectively. The other metrics were also high for these four classifiers, as all specificity scores exceeded 97.50%, all precision scores exceeded 87.97%, all F1 scores exceeded 86.69%, and all sensitivity scores exceeded 87.50%. The RF classifiers demonstrated exceptional performance, accurately classifying all the labels of the test dataset and achieving 100% for all the metrics, with the PT metric of 0, indicating perfect classification.

The results obtained by the RF in this study surpassed those reported in other studies [29], [36], [38]. The highest performances achieved by Bavan et al. were 97.2% of accuracy, reporting more challenges in classifying flexion and abduction tasks [38]. Specifically, the RF model yielded metrics of 98.40% of accuracy and precision, 96.5% of sensitivity, and 99.23% of specificity. Alhammad et al. reported lower values achieved by the RF compared to this study: 96.86% of accuracy and sensitivity, 97.2% of precision, and 97.02% of F1score [29]. Lastly, Hua et al. achieved 97.4% of accuracy using the kNN classifier and 98.6% accuracy with the RF classifier [36].

The nested cross validation method yielded to similar results obtained with the 10 folds cross validation. The DT classifier improves its performances, from 83.33% to 87.50% of overall accuracy. Table VI shows overall accuracies equal to 87.50%, 87.50%, 94.44%, and 100% achieved by the k-NN, the SVM, the NB, and the RF respectively.

Table VII, Table VIII, Table IX, and Fig. 8-13 highlight the ability of all the classifiers in recognize each single shoulder exercise. Most of the six rehabilitation exercises considered in this study were classified correctly. In particular, 100% of prediction accuracy was always obtained for Task 3, Task 4 and Task 5, and also for Task 6 with the SVM, NB, and RF classifiers. Since these exercises were extremely different from each other, each one presented easily recognizable and classifiable features. All the classifiers encountered the greatest difficulties for the classification of two exercise, i.e., Task 1 (flexion/extension without a weight) and Task 2 (flexion/extension with a weight). These two movements are the same, differing only in the use of a dumbbell (2 kg). This misclassification could be related to the involvement of healthy participants with no shoulder musculoskeletal diseases that executed the two movements in the same way. For this reason, there were no significant differences between the sensor data acquired while performing these two tasks. However, the averaged accuracies ranged between 94.91% and 99.54% using the 5 folds CV, between 94.44% and 100% using the 5 folds CV, and between 95.83% and 100% using the nested cross validation. These results are coherent with the AUC ones. The AUC values for Task 3, Task 4, and Task 5 were always equal to 1, meaning that the TPR was equal to 1 and the FPR was equal to 0. In most cases also Task 6 was classified correctly by all classifiers, with AUC ranging between 0.9 and 1. Lower values were obtained for Task 1 (AUC ranging between 0.6417 and 1) and Task 2 (AUC ranging between 0.8 and 1).

Results obtained in this work are promising for the application of the proposed wearable system for shoulder home-based remote monitoring. The ease of setup and modularity of the proposed wearable system increase the ability for the patient to self-position the sensing units without operator support, increasing the variety of contexts in which it can be used. The relatively poor distinction between flexion/extension movements without and with a weight could potentially be improved integrating more sensors. For example, electromyography (EMG) sensors can determine which muscles are being activated to enhance classifiers' models including additional features (such as muscles' fatigue), and to more accurately evaluate the efficiency of rehabilitation exercises.

Some limitations are evident in this study. Firstly, only six rehabilitation exercises were selected from the guidelines provided by the American Society of Shoulder and Elbow Therapists. While these exercises are representative of commonly practiced ones, there exist additional movement exercises that were not examined in this study. Secondly, the study sample exclusively consisted of younger and healthy subjects, potentially compromising the representativeness of observed characteristics for older age groups. Patients with shoulder musculoskeletal disorders (such as rotator cuff tears) are expected to exhibit greater variability in the pace and trajectory of movements compared to healthy subjects, presenting heightened challenges for classifiers in accurately categorizing the performed exercises. Thirdly, the experimental data were recorded during supervised sessions. The extraction of features from continuous exercise sessions conducted in uncontrolled environments poses greater challenges. Future endeavors will explore unsupervised or semi-supervised learning approaches, along with the inclusion of a larger sample size, the assessment of proposed algorithms on data acquired from patients with musculoskeletal disorders, and the incorporation of additional shoulder rehabilitation exercises in the protocol.

REFERENCES

[1] Y. C. Lin, Y. J. Tsai, Y. L. Hsu *et al.*, "Assessment of Shoulder Range of Motion Using a Wearable Inertial Sensor Network," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 15330-15341, 2021.

[2] U. G. Longo, S. De Salvatore, A. Carnevale *et al.*, "Optical Motion Capture Systems for 3D Kinematic Analysis in Patients with Shoulder Disorders," *International Journal of Environmental Research and Public Health*, 19, 2022].

[3] G. R. Regterschot, G. M. Ribbers, and J. B. J. Bussmann, "Wearable Movement Sensors for Rehabilitation: From Technology to Clinical Practice," *Sensors*, 21, 2021].

[4] U. A.-O. Longo, A. Berton, L. Risi Ambrogioni *et al.*, "Cost-Effectiveness of Supervised versus Unsupervised Rehabilitation for Rotator-Cuff Repair: Systematic Review and Meta-Analysis. LID - 10.3390/ijerph17082852 [doi] LID - 2852," no. 1660-4601 (Electronic).

[5] Y. Ke, "Location and Tracking Mode of Sports Rehabilitation Training With Self-Powered Sensors Based on Particle Swarm Optimization Algorithm," *IEEE Sensors Journal*, vol. 23, no. 18, pp. 20894-20903, 2023.

[6] K. T. Giggins O Fau - Sweeney, B. Sweeney Kt Fau - Caulfield, and B. Caulfield, "The use of inertial sensors for the classification of rehabilitation exercises," no. 2694-0604 (Electronic).

[7] R. M. Viglialoro, S. Condino, G. Turini *et al.*, "Review of the Augmented Reality Systems for Shoulder Rehabilitation," *Information*, 10, 2019].

[8] S.-C. Yeh, S. Lee, Y. Fank *et al.*, "A Cloud-Based Tele-Rehabilitation System for Frozen Shoulder," *Advanced Materials Research*, vol. 717, pp. 766-771, 07/01, 2013.

[9] A. González, P. Fraisse, and M. Hayashibe, "Adaptive Interface for Personalized Center of Mass Self-Identification in Home Rehabilitation," *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2814-2823, 2015.

[10] R. A.-O. Argent, A. A.-O. X. Daly, and B. A.-O. Caulfield, "Patient Involvement With Home-Based Exercise Programs: Can Connected Health Interventions Influence Adherence?," no. 2291-5222 (Print).

[11] Y.-P. Chen, C.-Y. Lin, M. Tsai, Jr. *et al.*, "Wearable Motion Sensor Device to Facilitate Rehabilitation in Patients With Shoulder Adhesive Capsulitis: Pilot Study to Assess Feasibility," *J Med Internet Res*, vol. 22, no. 7, pp. e17032, 2020/7/23, 2020.

[12] A. Carnevale, U. G. Longo, E. Schena *et al.*, "Wearable systems for shoulder kinematics assessment: a systematic review," *BMC Musculoskelet Disord*, vol. 20, no. 1, pp. 546, Nov, 2019.

[13] U. G. Longo, S. De Salvatore, M. Sassi *et al.*, "Motion Tracking Algorithms Based on Wearable Inertial Sensor: A Focus on Shoulder," *Electronics*, vol. 11, no. 11, 2022.

[14] S. Patel, R. Hughes, T. Hester *et al.*, "A Novel Approach to Monitor Rehabilitation Outcomes in Stroke Survivors Using Wearable Technology," *Proceedings of the IEEE*, vol. 98, no. 3, pp. 450-461, 2010.

[15] P. Bonato, "Advances in wearable technology and applications in physical medicine and rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 2, no. 1, pp. 2, 2005/02/25, 2005.

[16] R. Sharma, A. Dasgupta, R. Cheng *et al.*, "Machine Learning for Musculoskeletal Modeling of Upper Extremity," *IEEE Sensors Journal*, vol. 22, no. 19, pp. 18684-18697, 2022.

[17] L. A.-O. Brennan, A. Bevilacqua, T. Kechadi *et al.*, "Segmentation of shoulder rehabilitation exercises for single and multiple inertial sensor systems," no. 2055-6683 (Electronic).

[18] M. Rana, and V. Mittal, "Wearable Sensors for Real-Time Kinematics Analysis in Sports: A Review," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1187-1207, 2021.

[19] A. Wang, G. Chen, J. Yang *et al.*, "A Comparative Study on Human Activity Recognition Using Inertial Sensors in a Smartphone," *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4566-4578, 2016.

[20] S. C. Mukhopadhyay, "Wearable Sensors for Human Activity Monitoring: A Review," *IEEE Sensors Journal*, vol. 15, no. 3, pp. 1321-1330, 2015.

[21] I. H. López-Nava, and A. Muñoz-Meléndez, "Wearable Inertial Sensors for Human Motion Analysis: A Review," *IEEE Sensors Journal*, vol. 16, no. 22, pp. 7821-7834, 2016.

[22] T. L. Baldi, F. Farina, A. Garulli *et al.*, "Upper Body Pose Estimation Using Wearable Inertial Sensors and Multiplicative Kalman Filter," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 492-500, 2020.

[23] A. RajKumar, F. Vulpi, S. R. Bethi *et al.*, "Wearable Inertial Sensors for Range of Motion Assessment," *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3777-3787, 2020.

[24] R. Jain, V. B. Semwal, and P. Kaushik, "Deep ensemble learning approach for lower extremity activities recognition using wearable sensors," *Expert Systems*, vol. 39, no. 6, pp. e12743, 2022/07/01, 2022.

[25] C. N. Teague, J. A. Heller, B. N. Nevius *et al.*, "A Wearable, Multimodal Sensing System to Monitor Knee Joint Health," *IEEE Sensors Journal*, vol. 20, no. 18, pp. 10323-10334, 2020.

[26] T. Li, and H. Yu, "Upper Body Pose Estimation Using a Visual–Inertial Sensor System With Automatic Sensor-to-Segment Calibration," *IEEE Sensors Journal*, vol. 23, no. 6, pp. 6292-6302, 2023.

[27] L. Meng, M. Chen, B. Li *et al.*, "An Inertial-Based Upper-Limb Motion Assessment Model: Performance Validation Across Various Motion Tasks," *IEEE Sensors Journal*, vol. 23, no. 7, pp. 7168-7177, 2023.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3406138

[28] J.-I. Pan, H.-W. Chung, and J.-J. Huang, "Intelligent Shoulder Joint Home-Based Self-Rehabilitation Monitoring System," *International Journal of Smart Home*, vol. 7, pp. 395-404, 09/30, 2013.

[29] N. A.-O. Alhammad, and H. Al-Dossari, "Recognizing Physical Activities for Spinal Cord Injury Rehabilitation Using Wearable Sensors. LID - 10.3390/s21165479 [doi] LID - 5479," no. 1424-8220 (Electronic).

[30] V. B. Semwal, N. Gaud, P. Lalwani *et al.*, "Pattern identification of different human joints for different human walking styles using inertial measurement unit (IMU) sensor," *Artificial Intelligence Review*, vol. 55, no. 2, pp. 1149-1169, 2022/02/01, 2022.

[31] P. Patil, K. S. Kumar, N. Gaud *et al.*, *Clinical Human Gait Classification: Extreme Learning Machine Approach*, 2019.

[32] M. Zmitri, H. Fourati, and N. Vuillerme, "Human Activities and Postures Recognition: From Inertial Measurements to Quaternion-Based Approaches," *Sensors*, 19, 2019].

[33] Q. A.-O. X. Li, Y. A.-O. Liu, J. A.-O. Zhu *et al.*, "Upper-Limb Motion Recognition Based on Hybrid Feature Selection: Algorithm Development and Validation," no. 2291-5222 (Electronic).

[34] L. Truppa, E. Bergamini, P. Garofalo *et al.*, "Magnetic-Free Quaternion-Based Robust Unscented Kalman Filter for Upper Limb Kinematic Analysis," *IEEE Sensors Journal*, vol. 23, no. 3, pp. 3212-3219, 2023.

[35] S. Alavi, D. Arsenault, and A. Whitehead, "Quaternion-Based Gesture Recognition Using Wireless Wearable Motion Capture Sensors," *Sensors*, vol. 16, no. 5, 2016.

[36] A. Hua, P. Chaudhari, N. Johnson *et al.*, "Evaluation of Machine Learning Models for Classifying Upper Extremity Exercises Using Inertial Measurement Unit-Based Kinematic Data," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 9, pp. 2452-2460, 2020.

[37] M. Cornacchia, K. Ozcan, Y. Zheng *et al.*, "A Survey on Activity Detection and Classification Using Wearable Sensors," *IEEE Sensors Journal*, vol. 17, no. 2, pp. 386-403, 2017.

[38] L. Bavan, K. Surmacz, D. Beard *et al.*, "Adherence monitoring of rehabilitation exercise with inertial sensors: A clinical validation study," *Gait Posture*, vol. 70, pp. 211-217, May, 2019.

[39] S. H. Chae, Y. Kim, K. S. Lee *et al.*, "Development and Clinical Evaluation of a Web-Based Upper Limb Home Rehabilitation System Using a Smartwatch and Machine Learning Model for Chronic Stroke Survivors: Prospective Comparative Study," *JMIR Mhealth Uhealth*, vol. 8, no. 7, pp. e17216, Jul 9, 2020.

[40] D. M. Burns, N. Leung, M. Hardisty *et al.*, "Shoulder physiotherapy exercise recognition: machine learning the inertial signals from a smartwatch," *Physiological Measurement*, vol. 39, no. 7, pp. 075007, 2018/07/23, 2018.

[41] C. A. Thigpen, M. A. Shaffer, B. W. Gaunt *et al.*, "The American Society of Shoulder and Elbow Therapists' consensus statement on rehabilitation following arthroscopic rotator cuff repair," no. 1532-6500 (Electronic).

[42] Xsens Technologies, Xsens DOT User Manual, Movella, Pantheon, 7521 Enschede, The Netherlands, 2021. [Online]. Available https://www.xsens.com/hubfs/Downloads/Manuals/Xsens%20DOT%20User%20Manual.pdf.

[43] L. S. Vargas-Valencia, A. Elias, E. Rocon *et al.*, "An IMU-to-Body Alignment Method Applied to Human Gait Analysis," *Sensors (Basel)*, vol. 16, no. 12, Dec 10, 2016.

[44] L. Pacher, C. Chatellier, R. Vauzelle *et al.*, "Sensor-to-Segment Calibration Methodologies for Lower-Body Kinematic Analysis with Inertial Sensors: A Systematic Review," *Sensors*, vol. 20, no. 11, 2020.

[45] R. Laura Susana Vargas-Valencia and Arlindo Elías and Anselmo Frizera Neto and Eduardo, "Body to sensor calibration procedure for lower limb joint angle estimation applied to imu-based gait analysis," 2014.

[46] G. Wu, F. C. van der Helm, H. E. Veeger *et al.*, "ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion--Part II: shoulder, elbow, wrist and hand," *J Biomech*, vol. 38, no. 5, pp. 981-992, May, 2005.

[47] A. G. Cutti, A. Giovanardi, L. Rocchi *et al.*, "Ambulatory measurement of shoulder and elbow kinematics through inertial and magnetic sensors," *Med Biol Eng Comput*, vol. 46, no. 2, pp. 169-78, Feb, 2008.

[48] Aboudi NE, Benhlima L: Review on wrapper feature selection approaches. In: 2016 International Conference on Engineering & MIS (ICEMIS): 22-24 Sept. 2016 2016; 2016: 1-5.

[49] S. Parvandeh, H. W. Yeh, M. P. Paulus *et al.*, "Consensus features nested cross-validation," no. 1367-4811 (Electronic).

[50] A. Shdefat, A. Halimeh, and H. Kim, "Human Activities Recognition Via Smartphones Using Supervised Machine Learning Classifiers," *Primary Health Care Open Access*, vol. 08, 01/01, 2018.

[51] B. Mahesh, *Machine Learning Algorithms -A Review*, 2019.

[52] J. Wainer, and G. Cawley, "Nested cross-validation when selecting classifiers is overzealous for most practical applications," *Expert Systems with Applications*, vol. 182, pp. 115222, 2021/11/15/, 2021.

[53] M. Grandini, E. Bagli, and G. Visani, *Metrics for Multi-Class Classification: an Overview*, 2020.

[54] D. Chicco, and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Mining*, vol. 16, no. 1, pp. 4, 2023/02/17, 2023.

[55] D. Chicco, and G. Jurman, "A statistical comparison between Matthews correlation coefficient (MCC), prevalence threshold, and Fowlkes–Mallows index,"

Journal of Biomedical Informatics, vol. 144, pp. 104426, 2023/08/01/, 2023.

[56] M. de Figueiredo, C. B. Y. Cordella, D. Jouan-Rimbaud Bouveresse *et al.*, "A variable selection method for multiclass classification problems using two-class ROC analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 177, pp. 35-46, 2018/06/15/, 2018.

[57] J. S. Aguilar-Ruiz, and M. Michalak, "Multiclass Classification Performance Curve," *IEEE Access*, vol. 10, pp. 68915-68921, 2022.