Gait & Posture 70 (2019) 211–217

Contents lists available at ScienceDirect

Gait & Posture

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

Adherence monitoring of rehabilitation exercise with inertial sensors: A clinical validation study

Luckshman Bavan[a,*], Karl Surmacz[b], David Beard[a], Stephen Mellon[a], Jonathan Rees[a]

[a] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Old Road, Oxford, OX3 7LD, United Kingdom

[b] McLaren Applied Technologies, McLaren Technology Centre, Chertsey Road, Woking, GU21 4YH, United Kingdom

A R T I C L E I N F O

Keywords:
Inertial sensor
Shoulder
Rehabilitation
Activity classification

A B S T R A C T

Background: Rehabilitation has an established role in the management of a wide range of musculoskeletal conditions. Much of this treatment relies on self-directed exercises at home, where adherence of execution is unknown. Demonstrating treatment fidelity is necessary to draw conclusions about the efficacy of rehabilitation interventions in both clinical and research settings. There is a lack of tools and methods to achieve this.

Research question: This study aims to evaluate the feasibility of using a single inertial sensor to recognise and classify shoulder rehabilitation activity using supervised machine learning techniques.

Methods: Twenty patients with shoulder pain were monitored performing five rehabilitation exercises routinely prescribed for their condition. Accelerometer, gyroscope and magnetometer data were collected via a device mounted onto an arm sleeve. Non-specific motion data was included in the analysis. Time and frequency domain features were calculated from labelled data segments and ranked in terms of their predictive importance using the ReliefF algorithm. Selected features were used to train four supervised learning algorithms: decision tree, k-nearest neighbour, support vector machine and random forests. Performance of algorithms in accurately classifying exercise activity was evaluated with ten-fold cross-validation and leave-one-subject-out-validation methods.

Results: Optimal predictive accuracies for ten-fold cross-validation (97.2%) and leave-one-subject-out-validation (80.5%) were achieved by support vector machine and random forests algorithms, respectively. Time domain features derived from accelerometer, magnetometer and orientation data streams were shown to have the highest predictive value for classifying rehabilitation activity.

Significance: Classification models performed well in differentiating patient exercise activity from non-specific movement and identifying specific exercise type using inertial sensor data. A clinically useful account of home rehabilitation activity will help guide treatment strategies and facilitate methods to improve patient engagement. Future work should focus on evaluating the performance of such systems in natural and unsupervised settings.

1. Introduction

Musculoskeletal disorders are prevalent amongst all age groups and are a major contributor to global disability [1]. Exercise-based rehabilitation has an established role in the management of a wide range of musculoskeletal conditions and is commonly used as first line treatment [2–4]. Although most rehabilitation programs are guided by a specialist, patients are typically expected to perform the majority of their prescribed treatment independently in their home environment.

It is well established that adherence to home musculoskeletal rehabilitation is poor, with studies reporting non-compliance to be in the region of 50–65% [5]. With such variability and uncertainty in treatment implementation, it is almost impossible to draw meaningful conclusions about the efficacy of rehabilitation in both clinical and research settings. Existing self-reported measures of adherence perform poorly under psychometric testing and do not provide a detailed account of exercise activity [6,7]. There is therefore a clear need for robust methods to measure implementation of exercise-based treatments [8].

Recent developments in inertial sensor technology have created new

* Corresponding author.
E-mail addresses: Luckshman.Bavan@ndorms.ox.ac.uk (L. Bavan), Karl.Surmacz@mclaren.com (K. Surmacz), David.Beard@ndorms.ox.ac.uk (D. Beard), Stephen.Mellon@ndorms.ox.ac.uk (S. Mellon), Jonathan.Rees@ndorms.ox.ac.uk (J. Rees).

https://doi.org/10.1016/j.gaitpost.2019.03.008
Received 10 December 2018; Received in revised form 21 February 2019; Accepted 11 March 2019
0966-6362/

Fig. 1. Wireless mIMU mounted to arm sleeve positioned on the upper arm facing posteriorly.

opportunities for the objective monitoring and analysis of human motion [9]. Supervised machine learning algorithms can be used to classify patient activity based on inertial sensor data [10], and it is felt that these methods could provide a practical and reliable approach to monitoring rehabilitation activity [11–14]. These techniques involve the use of labelled training data, containing samples of inertial data sets and their corresponding activity labels, to build predictive classification models capable of determining what exercises are being performed by the patient. These tools could facilitate the evaluation of treatment fidelity within clinical and research settings, and have the potential to provide patients with meaningful feedback [15,16]. Qualitative studies also indicate that this form of feedback could drive patient motivation and help maintain or improve adherence to home exercise programs [17].

Results from preliminary investigations appear promising, however most classification systems have been validated with inertial data acquired from healthy study participants [15]. Patients with pathology will naturally display greater variation in movement pace and trajectory, and trained classifiers will need to accommodate for this variation. Additionally, existing studies do not evaluate classifiers' ability to differentiate target activity from non-specific or irrelevant movements typically present within inertial recordings. This process is usually achieved by manual or semi-automated extraction of target activity data, which increases user burden and would not be compatible with real-time monitoring applications.

This study aims to evaluate the performance of a fully automated classification pathway for recognition of analytical shoulder rehabilitation activity carried out by patients with subacromial shoulder pain. Supervised machine learning models will be used to classify shoulder activity from inertial data acquired through a single wireless magneto-inertial measuring unit (mIMU).

2. Methods

2.1. Study participants and data collection

Ethical approval for the study was granted by a UK NHS Research Ethics Committee (17/SW/0217). Patients with a diagnosis of subacromial shoulder pain and who were deemed suitable for outpatient physiotherapy were recruited from NHS orthopaedic clinics. Criteria for exclusion included adhesive capsulitis, inflammatory arthropathy, any neuromuscular condition, cervical spondylosis or a history of recent trauma. Twenty consecutive patients (eight males, twelve females) meeting the eligibility criteria provided informed consent for their participation in this study. The recruitment target was chosen to ensure collection of sufficient movement data to represent that of the target population, without over-recruiting. Five patients had partial or full degenerative rotator cuff tears and 18 patients had previously received some form of treatment in the form of physiotherapy or steroid injections. Mean participant age was 58.7 years (range 48–80 years) and 18 participants were right arm dominant. Assessments were performed on the affected arm only (10 left 10 right) and the mean Oxford Shoulder Score was 30.9 (range 11–42).

Participants were asked to review a patient information booklet detailing five rehabilitation exercises routinely prescribed for the treatment of subacromial shoulder pain: shoulder abduction, shoulder flexion, wall slide, wall press and shoulder rotation. Under supervision, participants were then instructed to carry out 10 repetitions of each exercise at a self-selected pace and advised to stop if they experienced significant discomfort. All exercises were carried out with the patient standing, except the shoulder rotations, which were carried out from a seated position with the elbow supported to keep the upper arm abducted close to 90°.

Raw nine-axis (three-axis accelerometer, gyroscope and magnetometer) motion data was acquired wirelessly via a single mIMU (MetaMotion R, MbientLab, San Francisco, California, USA) mounted to an arm sleeve worn above the level of the elbow (Fig. 1). Accelerometer, gyroscope and magnetometer sampling rates were set to 100 Hz, 100 Hz and 25 Hz, respectfully. Data was collected wirelessly via Bluetooth Low Energy to a smart device and processing was performed offline in MATLAB (R2018a from MathWorks).

2.2. Pre-processing

Pre-processing was performed to create a consistently sampled data set suitable for feature extraction. Magnetometer data was up-sampled to 100 Hz and a low pass 6th order Butterworth filter (10 Hz cut-off frequency) was applied to the accelerometer data to remove high frequency noise. In addition to the individual pre-processed nine-axis data streams, individual sensor vector magnitude values and three-dimensional estimates of sensor orientation were also calculated. The Madgwick orientation filter was used to fuse nine-axis sensor data to calculate sensor orientation [18], and features were calculated from both quaternion and Euler angle (*ZYX* sequence) representations of orientation.

2.3. Segmenting and labelling

Segmentation of pre-processed data streams was performed to create unique data segments representing each exercise sequence. The employed method involved identifying the dominant gyroscope axis zero-velocity time points to define the boundaries of each movement

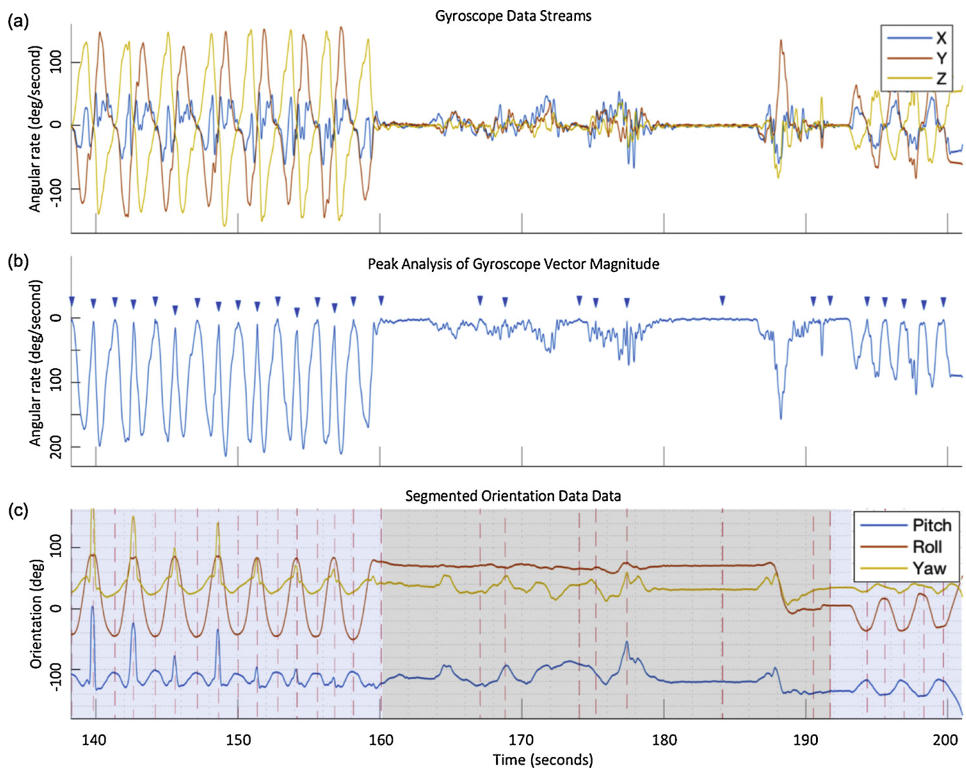

Fig. 2. (a) 3-axis gyroscope plot; (b) Peak analysis of (inverted) gyroscope vector magnitude identifies locations close to 0; (c) Locations of identified peaks superimposed on orientation data indicates segmentation points.

cycle. The dominant axis naturally varies depending on movement trajectory, therefore a peak analysis function was applied to the gyroscope vector magnitude data stream to identify time points when velocity was closest to zero (Fig. 2). Unlike conventional fixed-size sliding windowing techniques, this method potentially ensures each segment represents a defined portion of an activity, irrespective of the time taken to carry out the movement. This stage was fully automated and all recorded data segments (exercise and non-exercise activity) were included.

A unique exercise label was attributed to each activity segment from ground truth data collected during performance of the exercises. Two adjacent segments defined a single exercise due to the zero-velocity state at maximum shoulder excursion (e.g. peak of abduction). Both segmented components of a single exercise (e.g. abduction and adduction) were assigned the same label (*Abduction*) providing five activity classes. A 6th class (*Junk*) represented non-specific activity recorded during the session, which was unrelated to the execution of the prescribed exercises. This included inertial data recorded while patients adjusted body or arm position in between exercises, adjusted clothing or performed upper limb gestures while communicating. Accuracy of ground truth data sets were verified with offline visualisation of orientation data streams, which provide an interpretable representation of arm position and movement, allowing the investigator to infer exercise type and confirm label.

2.4. Feature extraction

The feature extraction process calculates unique and distinguishing characteristics of each activity segment, which are used as input for the classification algorithms. This study evaluates a wide range of time and frequency domain features derived from accelerometer, gyroscope, magnetometer and orientation data [19]. Time domain features included mean, root mean square, standard deviation, variance, range, inter-quartile range, percentiles (5th, 25th, 50th, 75th, 95th) and vector pair (*xy*/*xz*/*yz*) Pearson correlation coefficients (bivariate features). Frequency domain features included maximum frequency component, mean frequency component, energy spectral density, entropy and kurtosis. A total of 237 features were calculated for each activity segment.

2.5. Feature selection

Feature selection aims to reduce the number of features used by identifying those that are of most independent predictive importance [20]. This step optimises the learning process and improves classification accuracy by reducing the risk of model overfitting [10]. The ReliefF algorithm was applied [21] to assign weight values to features, penalizing those that give similar values to segments of a different class and rewarding those that give different values to segments of different classes. A subset of features was selected after setting an empirical threshold based on the calculated average of attributed weights, all features below the threshold were excluded. All features included for classifier inputs were scaled to zero-mean and norm-standard deviation.

2.6. Classification and validation

Four supervised machine learning classification models were chosen for evaluation: decision tree (DT), support vector machine (SVM), k-nearest neighbour (k-NN) and random forests (RF). Hyperparameter settings were determined following preliminary investigations. All models were trained using the same selected subset of high-ranking features and predictive performance was evaluated with two different validation methods. *Ten-fold cross-validation* (10FCV) randomly distributes all labelled data segments evenly amongst ten folds. Training is performed on data contained within nine folds and testing is performed on the omitted fold to calculate the predictive accuracy. Ten iterations of this process are performed with different partitions of the data set, to ensure all data is used for training and testing. *Leave-one-subject-out-validation* (LOSOV) partitions the data set into subject specific folds with each fold containing only the data from a single participant. Similar to the previous method, a single fold is retained each round for testing and a single performance estimate is reported as the average of all testing results.

Accuracy is used to provide an overall indicator of model performance and reflects the proportion of correctly classified segments within testing. The accuracy metric does not take unbalanced datasets into account and can be biased to favour the majority classes. Therefore, individual label *sensitivity* (true positive rate or recall), *positive predictive value* (PPV; precision) and *specificity* (1-false positive rate) are also calculated.

3. Results

The segmentation protocol yielded 2568 samples. Segments representing exercise activity had a mean duration of 1.78 s. Not all patients were able to perform 10 repetitions of all five exercises due to movement restrictions related to their shoulder pain. Any additional repetitions and limited attempts at exercise performance recorded within the session were labelled and included in the analysis.

Two sources of forced segmenting errors were identified during labelling: *oversegmenting* resulted in single activity cycles being divided into two or more segments and occurred in 2.7% of activity segments; *undersegmenting* resulted in a registered segment incorporating more than one half of an exercise cycle and occurred in 1.8% of activity segments. In both cases segments were assigned labels according to the activity being performed. Exercise activity labels were fairly evenly distributed; 344, 355, 348, 314 and 418 segments represented abduction, flexion, wall press, wall slide and rotation activity, respectively. Junk activity formed the largest class with 789 segments, representing 31% of all data segments.

The feature selection strategy resulted in the top 120 features being included; these were predominantly time domain features derived from accelerometer, magnetometer and processed orientation data. Feature ranking identified bivariate features of all data streams as offering the highest discriminative value for exercise classification. Only a small proportion of frequency domain and gyroscope derived features were weighted sufficiently for inclusion (Fig. 3).

Performance metrics for each model with both validation methods are detailed in Table 1. The 10FCV results indicate how the system would perform if a subject-specific training set was created for each new patient. For this validation method models achieved accuracies between 90.9% (DT) and 97.2% (SVM). SVM sensitivity scores, measuring the effectiveness of the classifier to identify a specific activity, exceeded 96% for all classes. PPVs were equally high and all specificity scores exceeded 99%.

The LOSOV evaluates the system's ability to generalise across to entirely new patient activity. For all classifiers, the LOSOV method yielded lower overall accuracy scores and higher variation than 10FCV. DT classifiers were particularly poor at generalising to new subject data, with specific class sensitivity scores as low as 50.9% (abduction). The RF classifier performed significantly better than all other models with 80.5% overall accuracy. There was large variation in model performance depending on the activity being classified; sensitivity scores for wall slides and abductions were 93.4% and 62% respectively. LOSOV results also showed significant variation in individual subject fold accuracies.

The proportions of correctly and incorrectly predicted activity labels can be reviewed within a confusion matrix (Fig. 4). This demonstrates the tendency for abduction, flexion and wall press activity to be misclassified as junk activity, and for abduction and flexion activity to be confused for one another.

4. Discussion

This study reports on the performance of supervised machine learning techniques in classifying shoulder rehabilitation exercises from inertial data acquired from a single nine-axis mIMU device worn on the upper arm. Classifier performance was evaluated with the use of real patient data and the inclusion of non-specific activity, thus providing representative estimates of their capabilities in real world scenarios. The predictive performance demonstrated suggests these systems can provide a clinically useful account of exercise adherence.

The classifiers chosen for this study represent a range of supervised machine learning models successfully implemented in previous human activity classification (HAC) studies. The choice of classifier for future clinical platforms will rely not only on predictive accuracy but also on factors related to training time, prediction speed and memory usage. The SVM performed well for both validation methods and are highly regarded for their ability to handle high-dimensional data. Choosing appropriate model parameters for SVM classifiers can be difficult however, and like k-NN classifiers, they can be memory intensive. Simple DTs are quick to train and easy to interpret, but are typically less accurate as this study has demonstrated. RF classifiers build multiple DTs using data bootstrapped from the original data set (*bagging*) and randomly selects predictors at each decision split. This classifier performed well for both validation methods and has shown excellent predictive accuracy in previous studies, outperforming SVM and k-NN algorithms [11,22]. The improved predictive capabilities and relatively fast prediction speeds make this algorithm a promising candidate for sensor-based rehabilitation monitoring [23]. Deep learning techniques, utilising neural network based algorithms, provide an alternative method for activity classification [11,12]. A key advantage of these methods is their ability to learn directly from sensor output data, removing the need for manual feature engineering. The models do however require substantially longer training times compared to traditional supervised classification methods and typically demand larger volumes of training data.

Accurate classification of human activity from sensor data is reliant on appropriate segmentation methods. The employed technique identified gyroscope zero-velocity crossing locations to define activity boundaries. This alternative method to conventional fixed-size overlapping windows was chosen for this study as establishing an optimal fixed window length can be challenging, and presents a trade-off between algorithm responsiveness and model performance [24]. Furthermore, short and long-term factors such as disease progression, recovery, pain, fatigue and stress can cause significant variation in the time taken to carry out certain exercises. Segmenting based on gyroscope zero-velocity crossings ensures the windows capture a defined activity cycle and system accuracy will not be directly related to inter and intra-subject variability in exercise performance time. Additionally, this method provides a direct measurement of activity duration and number of repetitions performed; both valuable metrics in the assessment of exercise adherence and performance [25].

Feature sets explored in this study were selected based on previous investigations evaluating HAC with inertial sensor data. The feature selection strategy employed an established *filter* method to evaluate the intrinsic properties of the calculated features [26]. Our investigation supports previous findings on the relatively higher value of accelerometer over gyroscope derived features, and time domain over frequency domain features for these applications [27]. Features derived from magnetometer and processed orientation data (9-axis sensor fusion) were also shown to have high predictive value. Although, it must be noted that magnetometer readings are susceptible to distortions caused by ferromagnetic materials and the testing environment could influence the reliability of these features [28]. Bivariate features of all data streams were shown to have the highest predictive importance, however a recent study evaluating similar activities in healthy subjects found correlation coefficients of vector pair combinations to have poor predictive value [11]. This discrepancy demonstrates the influence of data collection methods and windowing techniques on the differentiating qualities of calculated features [29].

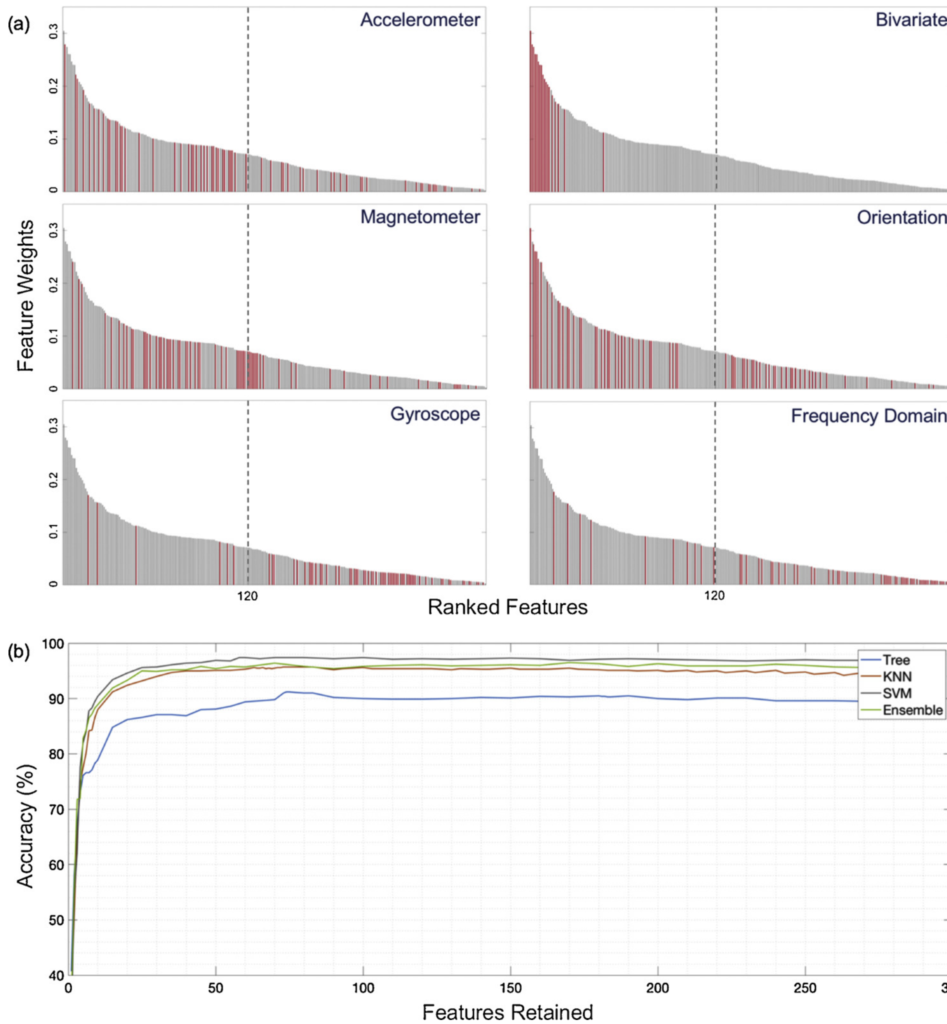

Fig. 3. (a) All features ranked according to ReliefF algorithm weights (high to low) with individual plots highlighting specific feature subsets. (b) Relationship between feature inclusion (in order of ranking) and classifier performance (10FCV).

4.1. Limitations

The exercises chosen for analysis only incorporate single plane movements. While these exercises are derived from recognised strategies for the treatment of subacromial pain, some multi-planar movement exercises can also be prescribed which have not been evaluated in this study. Data was collected from a relatively small group of patients and improved performance is likely to be achieved with larger amounts of training data. [13]. The relatively poor differentiation between abduction and flexion movements within this study could also potentially be improved with an additional body sensor or a functional calibration procedure. Additional sensors could enable measurement of posture, scapulohumeral rhythm and movement trajectory, allowing an assessment of exercise performance [30]. This would however increase patient burden and future work should be aimed at understanding the trade-off between system performance and usability [15].

Activity was recorded during a single observed session for each subject, and it is uncertain how well the trained algorithms would generalise to subsequent independently performed exercise sessions where recorded movement patterns may change. A disadvantage to supervised learning methods is the necessity for labelled training data. Labelling is an expensive process, and there are significant challenges around collecting labelled data from home environments. Unsupervised learning systems have been implemented for HAC studies, however predictions are less accurate and interpreting the output can be challenging [10]. Semi-supervised learning approaches, which attempt to utilise large volumes of data with reduced labelling requirements may offer the best balance between process efficiency and system accuracy [22]. These methods will be explored in further work investigating home rehabilitation monitoring.

5. Conclusion

Demonstrating adherence and program differentiation is critical in the development and evaluation of evidence-based rehabilitation interventions. This study demonstrates the feasibility of patient monitoring and automated classification of single plane exercises with wireless sensor technology. Future work should focus on optimising technical performance of these systems in natural settings and understanding factors around usability.

Funding

This work was supported by the NIHR Oxford Biomedical Research Centre and the Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford.

Conflict of interest statement

Authors have no conflicts of interest to declare.

Table 1
Performance of all four classification models using 10FCV and LOSOV.

	10-Fold Cross-Validation (%)				Leave-One-Subject-Out-Validation (%)			
	Accuracy	Sensitivity	PPV	Specificity	Accuracy	Sensitivity	PPV	Specificity
DT	**90.9** (± 2.0)				**71.0** (± 16.9)			
Abduction		86.9	86.4	97.9		50.9	62.9	95.4
Flexion		89.9	87.2	97.9		51.8	63.2	95.2
Wall Slide		96.3	95.4	99.3		82.5	83.2	97.4
Wall Press		93.6	92.2	98.9		67.2	75.6	97.0
Rotation		93.8	94.5	98.9		74.6	87.6	97.5
Junk		88.2	90.3	95.8		87.7	67.9	81.6
K-NN	**95.7** (± 1.2)				**76.1** (± 13.5)			
Abduction		97.4	94.4	99.1		63.7	67.4	95.2
Flexion		96.9	94.2	99.1		61.4	71.0	96.0
Wall Slide		97.1	97.7	99.6		85.3	81.1	96.9
Wall Press		98.1	93.9	99.1		74.5	82.7	97.8
Rotation		97.4	95.1	99.0		74.2	92.3	98.8
Junk		91.9	97.2	98.8		88.7	73.6	85.9
SVM	**97.2** (± 1.2)				**74.9** (± 17.0)			
Abduction		96.8	96.0	99.4		56.4	68.8	96.0
Flexion		96.6	96.9	99.5		55.5	75.2	97.1
Wall Slide		97.1	98.0	99.7		89.1	83.8	97.3
Wall Press		98.4	98.4	99.8		58.0	96.8	99.7
Rotation		97.8	97.4	99.5		74.9	97.5	99.6
Junk		96.8	97.0	98.7		95.7	65.9	78.1
RF	**96.4** (± 1.7)				**80.5** (± 15.6)			
Abduction		95.9	94.6	99.1		62.2	76.4	95.2
Flexion		94.6	94.4	99.1		67.6	75.9	96.0
Wall Slide		96.6	98.0	99.7		93.4	99.1	96.9
Wall Press		97.8	97.5	99.6		65.3	94.0	97.8
Rotation		98.6	97.2	99.4		90.0	95.4	98.8
Junk		95.7	96.7	98.5		94.6	72.3	85.9

PPV: positive predictive value; DT: decision tree; K-NN: k-nearest neighbour; SVM: support vector machine, RF: Random Forests.

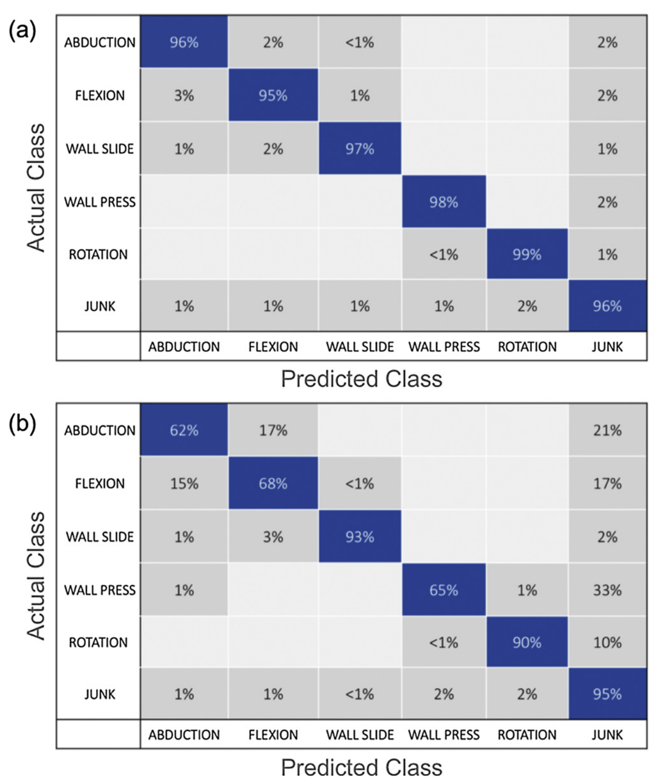

Fig. 4. Confusion matrix of (a) 10FCV and (b) LOSOV validation results of single model (*Random Forests*).

Declarations of interest

None.

References

[1] K. Storheim, J.A. Zwart, Musculoskeletal disorders and the global burden of disease study, Ann. Rheum. Dis. 73 (6) (2014) 949–950 http://www.ncbi.nlm.nih.gov/pubmed/24790065.

[2] P.H. England, Return on Investment of Interventions for the Prevention and Treatment of Musculoskeletal Conditions, (2017).

[3] N.N.I.f.H. Research, Moving Forward - Physiotherapy for Musculoskeletal Health and Wellbeing, (2018) https://http://www.dc.nihr.ac.uk/themed-reviews/Moving-Forward-Final.pdf.

[4] N. Smidt, H.C. de Vet, L.M. Bouter, J. Dekker, J.H. Arendzen, R.A. de Bie, et al., Effectiveness of exercise therapy: a best-evidence summary of systematic reviews, Aust. J. Physiother. 51 (2) (2005) 71–85 http://www.ncbi.nlm.nih.gov/pubmed/15924510.

[5] S.F. Bassett, The assessment of patient adherence to physiotherapy rehabilitation, NZ J. Physiother. 31 (2003) 60–62.

[6] J.C. Bollen, S.G. Dean, R.J. Siegert, T.E. Howe, V.A. Goodwin, A systematic review of measures of self-reported adherence to unsupervised home-based rehabilitation exercise programmes, and their psychometric properties, BMJ Open 4 (6) (2014) e005044http://www.ncbi.nlm.nih.gov/pubmed/24972606.

[7] K. Jack, S.M. McLean, J.K. Moffett, E. Gardiner, Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review, Man. Ther. 15 (3) (2010) 220–228 http://www.ncbi.nlm.nih.gov/pubmed/20163979.

[8] J. Room, E. Hannink, H. Dawes, K. Barker, What interventions are used to improve exercise adherence in older people and what behavioural techniques are they based on? A systematic review, BMJ Open 7 (12) (2017) e019221http://www.ncbi.nlm.nih.gov/pubmed/29247111.

[9] R. Khusainov, D. Azzi, I.E. Achumba, S.D. Bersch, Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications, challenges and limitations, Sensors 13 (10) (2013) 12852–12902 http://www.ncbi.nlm.nih.gov/pubmed/24072027.

[10] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, Y. Amirat, Physical human activity recognition using wearable sensors, Sensors 15 (12) (2015) 31314–31338 http://www.ncbi.nlm.nih.gov/pubmed/26690450.

[11] D.M. Burns, N. Leung, M. Hardisty, C.M. Whyne, P. Henry, S. McLachlin, Shoulder physiotherapy exercise recognition: machine learning the inertial signals from a smartwatch, Physiol. Meas. 39 (7) (2018) 075007http://www.ncbi.nlm.nih.gov/pubmed/29952759.

[12] H.C. Lin, S.Y. Chiang, K. Lee, Y.C. Kan, An activity recognition model using inertial

sensor nodes in a wireless sensor network for frozen shoulder rehabilitation exercises, Sensors 15 (1) (2015) 2181–2204 http://www.ncbi.nlm.nih.gov/pubmed/25608218.

[13] M. O'Reilly, B. Caulfield, T. Ward, W. Johnston, C. Doherty, Wearable inertial sensor systems for lower limb exercise detection and evaluation: a systematic review, Sport. Med. 48 (5) (2018) 1221–1246 http://www.ncbi.nlm.nih.gov/pubmed/29476427.

[14] O.M. Giggins, K.T. Sweeney, B. Caulfield, Rehabilitation exercise assessment using inertial sensors: a cross-sectional analytical study, J. Neuroeng. Rehabil. 11 (2014) 158 http://www.ncbi.nlm.nih.gov/pubmed/25431092.

[15] Q. Wang, P. Markopoulos, B. Yu, W. Chen, A. Timmermans, Interactive wearable systems for upper body rehabilitation: a systematic review, J. Neuroeng. Rehabil. 14 (1) (2017) 20 http://www.ncbi.nlm.nih.gov/pubmed/28284228.

[16] J. Verbrugghe, E. Knippenberg, S. Palmaers, T. Matheve, W. Smeets, P. Feys, et al., Motion detection supported exercise therapy in musculoskeletal disorders: a systematic review, Eur. J. Phys. Rehabil. Med. 54 (4) (2018) 591–604 http://www.ncbi.nlm.nih.gov/pubmed/29517185.

[17] C. Palazzo, E. Klinger, V. Dorner, A. Kadri, O. Thierry, Y. Boumenir, et al., Barriers to home-based exercise program adherence with chronic low back pain: patient expectations regarding new technologies, Ann. Phys. Rehabil. Med. 59 (2) (2016) 107–113 http://www.ncbi.nlm.nih.gov/pubmed/27050664.

[18] S.O. Madgwick, A.J. Harrison, A. Vaidyanathan, Estimation of IMU and MARG Orientation Using a Gradient Descent Algorithm, IEEE … International Conference on Rehabilitation Robotics : [Proceedings] 2011, (2011) 5975346http://www.ncbi.nlm.nih.gov/pubmed/22275550.

[19] S. Gonzalez, J. Sedano, J.R. Villar, E. Corchado, B. Herrero, Baruque, Features and models for human activity recognition, Neurocomputing 167 (2015) 52–60.

[20] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.

[21] I. Kononenko, Estimating attributes: analysis ad extensions of RELIEF, Lect. Notes Comput. Sci. (784) (1994) 171–182.

[22] B. Schiele, M. Stikic, K. Van Laerhoven, Exploring semi-supervised and active learning for activity recognition, 2008 12th IEEE International Symposium on Wearable Computers(ISWC) (2008) 81–88.

[23] Z. Feng, L. Mo, M. Li, A random forest-based ensemble method for activity recognition, Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference 2015, (2015), pp. 5074–5077 http://www.ncbi.nlm.nih.gov/pubmed/26737432.

[24] O. Banos, J.M. Galvez, M. Damas, H. Pomares, I. Rojas, Window size impact in human activity recognition, Sensors 14 (4) (2014) 6474–6499 http://www.ncbi.nlm.nih.gov/pubmed/24721766.

[25] A. de los Reyes-Guzman, I. Dimbwadyo-Terrer, F. Trincado-Alonso, F. Monasterio-Huelin, D. Torricelli, A. Gil-Agudo, Quantitative assessment based on kinematic measures of functional impairments during upper extremity movements: a review, Clin. Biomech. 29 (7) (2014) 719–727 http://www.ncbi.nlm.nih.gov/pubmed/25017296.

[26] E.R. Dougherty, J. Hua, C. Sima, Performance of feature selection methods, Curr. Genomics 10 (6) (2009) 365–374 http://www.ncbi.nlm.nih.gov/pubmed/20190952.

[27] F.R. Allen, E. Ambikairajah, N.H. Lovell, B.G. Celler, Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models, Physiol. Meas. 27 (10) (2006) 935–951 http://www.ncbi.nlm.nih.gov/pubmed/16951454.

[28] C. Schiefer, R.P. Ellegast, I. Hermanns, T. Kraus, E. Ochsmann, C. Larue, et al., Optimization of inertial sensor-based motion capturing for magnetically distorted field applications, J. Biomech. Eng. 136 (12) (2014) 121008http://www.ncbi.nlm.nih.gov/pubmed/25321344.

[29] O. Laguna, A. Olaya, D. Borrajo, A dynamic sliding window approach for activity recognition, Lect. Notes Comput. Sci. 6787 (2011).

[30] L. De Baets, R. van der Straaten, T. Matheve, A. Timmermans, Shoulder assessment according to the international classification of functioning by means of inertial sensor technologies: a systematic review, Gait Posture 57 (2017) 278–294 http://www.ncbi.nlm.nih.gov/pubmed/28683420.