Technology in Rehabilitation: Comparing Personalised and Global Classification Methodologies in Evaluating the Squat Exercise with Wearable IMUs

D.F. Whelan*[┼╫], M.A. O'Reilly*[┼╫], T.E. Ward[╪¥], E. Delahunt[╫], B. Caulfield[┼╫]
[┼] Insight Centre for Data Analytics, University College Dublin
[╫] School of Public Health, Physiotherapy and Sports Science, University College Dublin
[╪] Insight Centre for Data Analytics, Maynooth University
[¥] Biomedical Engineering Research Group, Department of Electronic Engineering, Maynooth University

Joint lead authors

<u>*Contact details:*</u>
Name: Darragh Whelan
Address: Insight UCD,
Science Centre EAST
Belfield
Dublin 4
EIRE
Email: darragh.whelan@insight-centre.org
Tel: +353 85 7566606

1 Summary

Background: The barbell squat is a popularly used lower limb rehabilitation exercise. It is also an integral exercise in injury risk screening protocols. To date athlete/patient technique has been assessed using expensive laboratory equipment or subjective clinical judgement; both of which are not without shortcomings. Inertial measurement units (IMUs) may offer a low cost solution for the objective evaluation of athlete/patient technique. However, it is not yet known if global classification techniques are effective in identifying naturally occurring, minor deviations in barbell squat technique.

Objectives: The aims of this study were to: (a) determine if in combination or in isolation, IMUs positioned on the lumbar spine, thigh and shank are capable of distinguishing between acceptable and aberrant barbell squat technique; (b) determine the capabilities of an IMU system at identifying specific natural deviations from acceptable barbell squat technique; and (c) compare a personalised (N=1) classifier to a global classifier in identifying the above.

Methods Fifty-five healthy volunteers (37 males, 18 females, age = 24.21 +/- 5.25 years, height = 1.75 +/- 0.1 m, body mass = 75.09 +/- 13.56 kg) participated in the study. All participants performed a barbell squat 3-repetition maximum max strength test. IMUs were positioned on participants' lumbar spine, both shanks and both thighs; these were utilized to record tri-axial accelerometer, gyroscope and magnetometer data during all repetitions of the barbell squat exercise. Technique was assessed and labelled by a Chartered Physiotherapist using an evaluation framework. Features were extracted from the labelled IMU data. These features were used to train and evaluate both global and personalised random forests classifiers.

Results: Global classification techniques produced poor accuracy (AC), sensitivity (SE) and specificity (SP) scores in binary classification even with a 5 IMU set-up in both binary (AC: 64%, SE: 70%, SP: 28%) and multi-class classification (AC: 59%, SE: 24%, SP: 84%). However, utilising personalised classification techniques even with a single IMU positioned on the left thigh produced good binary classification scores (AC: 81%, SE: 81%, SP: 84%) and moderate-to-good multi-class scores (AC: 69%, SE: 70%, SP: 89%).

Conclusions: There are a number of challenges in developing global classification exercise technique evaluation systems for rehabilitation exercises such as the barbell squat. Building large, balanced data sets to train such systems is difficult and time intensive. Minor, naturally occurring deviations may not be detected utilising global classification approaches. Personalised classification approaches allow for higher accuracy and greater system efficiency for end-users in detecting naturally occurring barbell squat technique deviations. Applying this approach also allows for a single-IMU set up to achieve similar accuracy to a multi-IMU setup, which reduces total system cost and maximises system usability.

Keywords: Exercise Therapy; Biomedical Technology; Lower Extremity; Physical Therapy Speciality; Inertial Measurement Units

2 Introduction

The squat is a compound full-body exercise, whose constituent movements are integral to activities of daily living. The barbell squat (squat with a weighted barbell placed across the upper shoulders) often features as a fundamental exercise in resistance training and rehabilitation programs. Furthermore, it is incorporated into musculoskeletal injury risk screening/identification protocols (1). Aberrant squat technique has been shown to increase stress on the joints of the lower extremity (2), potentiating the risk of injury. Thus, the reliable assessment of technique is necessary to mitigate injury risk.

The assessment of squat technique is typically undertaken using one of two distinct methods: (a) 3-D motion capture; (b) subjective visual analysis. Both of these have a number of limitations. 3-D motion capture systems are expensive and the application of skin-mounted markers may hinder normal movement (3, 4). Furthermore, data processing can be time intensive and specific expertise is often required to interpret the processed data and make recommendations on the observed results. Therefore, these systems are not frequently used to assess squat technique beyond the research laboratory (5). In clinical and gym-based settings, subjective visual assessment is typically used to assess technique. This subjective visual assessment of human movement is not always reliable even amongst experts, as the need to visually assess the numerous constituent components of the movement simultaneously is challenging (6).

Wearable inertial measurement units (IMUs) may offer the potential to bridge the gap between laboratory and day-to-day “real-world” acquisition and assessment of human movement. These IMUs are small, inexpensive sensors that consist of accelerometers, gyroscopes and magnetometers. They are able to acquire data pertaining to the linear and angular motion of individual limb segments and the centre of mass of the body. Self-contained, wireless IMU devices are easy to set-up and allow for the acquisition of human movement data in unconstrained environments (7). In this paper the term IMU system will be used to describe the IMU sensors, the sensor signals, the associated signal processing applied to them and the output of the exercise classification algorithms.

IMU systems can robustly track the variety of postures and environmental complexities associated with training, unlike camera-based systems, which are hampered by location, occlusion and lighting issues in such settings (8). IMUs have also been shown to be as effective as marker-based systems at measuring joint angles (5, 9, 10). There are many commercially available examples of IMU systems that monitor physical activity (e.g. Jawbone™ and Fitbit™). However, using IMU systems to assess gym-based exercises such as the barbell squat is less common. Researchers have demonstrated the ability of IMU-based systems to distinguish different exercises and count exercise repetitions with moderate to good levels of accuracy (11-15). Whilst these systems are capable of counting exercise repetitions, they do not provide instruction on technique and performance quality. A holistic exercise tracking system should not only recognise the exercise completed and count repetitions, but should also provide technique feedback. Furthermore, in order for IMU systems to assess human movement data as part of a musculoskeletal injury risk screening protocol, they need to be able to identify aberrant movement patterns and provide easily interpretable data to clinicians/coaches who use them.

A growing body of scientific literature has investigated the ability of IMU systems to assess technique in order to provide this holistic exercise analysis (15-22). The majority of authors have developed these IMU systems by employing the following steps: (a) collection of labelled dataset; (b) pre-processing of data; (c) data segmentation; (d) feature extraction; (e) classification development and evaluation (23). These studies have demonstrated the ability of IMU systems in identifying deviations with moderate to excellent levels of accuracy in exercises such as the biceps curl, military press, squat and lunge. However the majority of these IMU systems were developed using a dataset consisting of induced deviations (i.e. deviations that were intentionally performed by participants). When deviations occur naturally, the exact way in which they present may be more nuanced and subsequently more difficult to classify. This means that these systems may not be suitable for a real world environment where deviations present in a natural manner. When collecting data in a natural environment a variety of deviations may present in different quantities with some deviations occurring less frequently than others. This means collecting a large and balanced data set of natural deviations is challenging. This is necessary to allow for the development of a robust global classification system (24-26). In these situations a personalised classifier may be more appropriate.

A personalised classifier is a classifier developed on data provided by a single person (N of 1). The data used to develop this classifier is collected from participants as they complete exercises wearing IMUs. Each individual exercise repetition is assessed and labelled by a movement expert through live or post-hoc video analysis. This means that IMU signals for each exercise repetition can be associated with this repetition's movement pattern. When the data set used for training the IMU system is collected in this way it means the system can be individualised. While this may prove more labour intensive than using an IMU system based on a global classifier it may be more appropriate in some situations.

3 Objectives

The barbell squat is a compound full-body exercise that is typically a constituent component of resistance training, rehabilitation programs and musculoskeletal injury risk screening protocols. Incorrect technique can increase the risk of sustaining a musculoskeletal injury. Traditionally, exercise technique has been evaluated using expensive motion capture systems or via subjective visual inspection from trained professionals. IMU systems offer an opportunity to provide low-cost exercise technique assessment. However to date, no research has evaluated the capability of IMU systems to identify natural deviations in barbell squat technique. In this setting the use of an individualised classifier based on an N of 1 data set may prove more appropriate than global classifiers.

Therefore, the research question that this study seeks to address is: "how well can an IMU-based system quantify barbell squat technique?" The aims of this study were to: (a) determine if in combination or in isolation, IMUs positioned on the lumbar spine, thigh and shank are capable of distinguishing between acceptable and aberrant barbell squat technique; (b) determine the capabilities of an IMU system at identifying specific natural deviations from acceptable barbell squat technique; (c) compare a personalised to a global classifier in identifying the above.

4 Methods

4.1 Experimental Approach to Problem

This study employed an opportunistic approach to the development of a wearable sensor system for automatically assessing barbell squat technique. Participants were required to perform a 3-repetition maximum barbell squat test. This test was recorded in HD video. A Chartered Physiotherapist then assessed each repetition video and labelled the labelled it appropriately (i.e. acceptable or containing one of the deviations identified in Table 1). In order to ensure standardisation, form was considered acceptable if it was completed as defined by the National Strength and Conditioning Association (NSCA) (27). The deviations from this acceptable form are detailed in Table 1. During performance of the barbell squats, data was acquired from 5 IMUs (SHIMMER, Shimmer Research, Dublin, Ireland) placed on the lumbar spine, right and left thigh and right and left shank. The IMUs were positioned on each participant by the same researcher using a standardised and repeatable protocol. Participants were allowed a rest interval between performances of each set of repetitions. Following data collection, a total of 306 variables were extracted from the sensor signals for every repetition from each IMU. These variables were used to develop and evaluate the quality of an automated classification system for the analysis of barbell squat technique. This was undertaken using data derived from each individual IMU and combinations of multiple IMUs. A global classification system was evaluated as well as separate (N of 1) personalised classifier for each participant.

Table 1. List and description of barbell squat exercise deviations used in this study

Label	Description
Acceptable	Acceptable technique
Knee Valgus	Knees coming together during downward phase
Knee Varus	Knees coming apart during downward phase
Knees Too Forward	Knees ahead of toes during downward phase
Heels Elevated	Heels raising off the ground during exercise
Bent Over	Excessive flexion of hip and torso during exercise
Other	Other deviation, not highlighted in NSCA guidelines

NSCA = National Strength and Conditioning Association

4.2 Participants

Fifty-five healthy volunteers (37 males, 18 females, age = 24.21 +/- 5.25 years, height = 1.75 +/- 0.1 m, body mass = 75.09 +/- 13.56 kg) participated in the study. No participant reported having a current or recent musculoskeletal injury that would impair his or her performance of the exercise. All participants reported a level of familiarity with the barbell squat exercise. The University College Dublin Human Research Ethics Committee approved the study protocol and written informed consent was obtained from all participants before testing. In cases where participants were under the age of 18, written informed consent was also obtained from a parent or guardian.

4.3 Procedures

The testing protocol was explained to participants upon their arrival at the laboratory. Prior to formal testing all participants performed a ten-minute warm-up on an exercise bike (Lode B.V., Groningen, The Netherlands) maintaining a power output of 100W and constant cadence of 75-85 revolutions per minute. Following completion of the warm-up, a Chartered Physiotherapist secured the IMUs to pre-determined specific anatomic locations on the participant as follows: the spinous process of the 5th lumbar vertebra, the mid-point of both the right and left femurs (determined as half way between the greater trochanter and lateral femoral condyle), and on both shanks 2cms above the lateral malleolus (Figure 1). The orientation and location of the IMUs was consistent across participants.

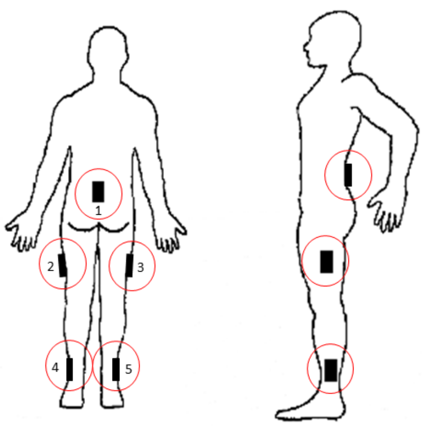

Figure 1: Image showing the five IMU positions: (1) the spinous process of the 5th lumbar vertebra, (2&3) the mid-point of both femurs on the lateral surface (determined as half way between the greater trochanter and lateral femoral condyle), (4&5) both shanks 2cm above the lateral malleolus

A pilot study was undertaken to determine the most appropriate sampling rate and the ranges for the accelerometer and gyroscope on board the IMUs. For the pilot study, data were acquired (512 samples/s) during performance of the squat, lunge, deadlift, single-leg squat and tuck jump exercises. A Fourier transform was then used to estimate the spectral extent of the signals which was found to be less than 20 Hz. Therefore, a sampling rate of 51.2 samples/s was chosen based upon the Shannon sampling theorem and the Nyquist criterion (28). Each IMU was configured to stream triaxial accelerometer (± 2 g), gyroscope (± 500 °/s) and magnetometer (± 1.9 Ga) data with the sensor ranges chosen based upon data from the pilot study. Each IMU was calibrated for these specific sensor ranges using the Shimmer 9DoF Calibration application (29).

Participants were required to complete a full 3-repetition maximum (3RM) strength test for the barbell squat (29). Following a warm-up on an exercise bike, participants completed a set of barbell squat exercises with a resistance that allowed for 8-12 repetitions comfortably. After resting for 1-minute, the load was increased by 10-20% and they performed a further 4-6 repetitions. This was followed by a 2-minute rest period. Following this they performed 3 repetitions with near maximum load. They then rested for 2-4 minutes. If they passed the previous set, the weight was incremented by 5-10% and another 3-repetition set was completed. This load increment was repeated until the participant could no longer lift the weight in a safe manner for three repetitions.

4.4 Data Labelling

All repetitions were recorded using a HD video camera placed in front of the participants. The video recordings of each exercise repetition were reviewed by a Chartered Physiotherapist with over seven years experience in musculoskeletal and sports physiotherapy. Each exercise repetition was separated and reviewed on multiple occasions systematically. For each repetition, the Chartered Physiotherapist first deemed if exercise technique was "acceptable". The criteria for acceptable technique were based upon the recommendations detailed in National Strength and Conditioning Association guidelines (27). For safety reasons participants completed the exercise in a squat rack. The barbell was placed on the rack just above shoulder level and loaded appropriately. The participant then stepped under the bar and placed it on the back of their shoulders, slightly below their neck. The bar was held with both arms and lifted off the rack by pushing with the legs and straightening the torso. The participant then stepped away from the rack and completed the squatting movement. Their chest was held up and out with their head tilted slightly up. As participants moved into the squat position they were instructed to allow hips and knees to flex while keeping their torso to floor angle constant. They were required to keep their heels on the floor and knees aligned over their feet. Participants continued flexing at the hips and knees until their thighs were parallel to the floor. As they moved upward a flat back was maintained and their chest was held up and out. Hips and knees were to be extended at the same rate with heels on floor and knees aligned over feet until the starting position was reached. The bar was then placed back on the rack. If a repetition was not completed as above, then the Chartered Physiotherapist selected the most dominant deviation from a pre-defined list (Table 1). This method of data labelling replicates methods from recently published work in the field of IMU based exercise technique classification systems (21).

4.5 Signal Processing and Statistical Analysis

Nine signals were collected from each IMU; accelerometer *x, y, z,* gyroscope *x, y, z* and magnetometer *x, y, z*. Data were analysed using MATLAB (2012, The MathWorks, Natwick, USA). To eliminate unwanted high-frequency noise during each repetition, the nine signals were low pass filtered at f_c = 20 Hz using a Butterworth filter of order $n = 8$. Whilst classification is solely possible using features derived from the accelerometer, gyroscope and magnetometer signals, the use of additionally derived signals improves system accuracy, sensitivity and specificity. As such, nine additional signals were then calculated as follows: The 3-D orientation of the IMU was computed using the gradient descent algorithm developed by Madgwick *et al.* (30). The resulting W, X, Y and Z quaternion values were also converted to pitch, roll and yaw signals. The pitch, roll and yaw signals describe the inclination, measured in radians, of each IMU in the sagittal, frontal and transverse plane respectively. The magnitude of acceleration was also computed using the vector magnitude of accelerometer *x, y, z*. The magnitude of acceleration describes the total acceleration of the IMU in any direction. This is the sum of the magnitude of inertial acceleration of the lumbar spine and acceleration due to gravity. Additionally, the magnitude of rotational velocity was computed using the vector magnitude of gyroscope *x, y, z*.

Each exercise repetition was extracted from the IMU data and resampled to a length of 250 samples. This time-normalisation was undertaken in an attempt to minimise the influence a participant's repetition tempo had on signal feature calculations. It also ensured consistent computational efficiency in applications for end users and has been used in recently published, similar work (19, 21, 22). Repetitions completed by the participant where the IMU's Bluetooth signal dropped were excluded from analysis. The total number of repetitions belonging to each class are shown in Table 2. Time-domain and frequency-domain descriptive features were computed in order to describe the pattern of each of the eighteen signals when the barbell squats were completed. These features were namely 'Mean', 'RMS', 'Standard Deviation', 'Kurtosis', 'Median', 'Skewness', ' Range', 'Variance', 'Max', 'Min', 'Energy', '25th Percentile', '75th Percentile', 'Level Crossing Rate', 'Fractal Dimension' (31) and the 'variance of both the approximate and detailed wavelet coefficients using the Daubechies 5 mother wavelet to level 7' (32). This resulted in 17 features for each of the 18 available signals producing a total of 306 features per IMU.

Figure 2 summarises the above whereby, 5 IMUs recorded 9 signals each, 9 more signals were derived from these resulting in a total of 18 signals per IMU. 17 features were computed per repetition for each signal from each IMU resulting in a total of 1530 features (306 per IMU, 17 per signal). These features were then used to develop and evaluate a variety of classifiers as described below.

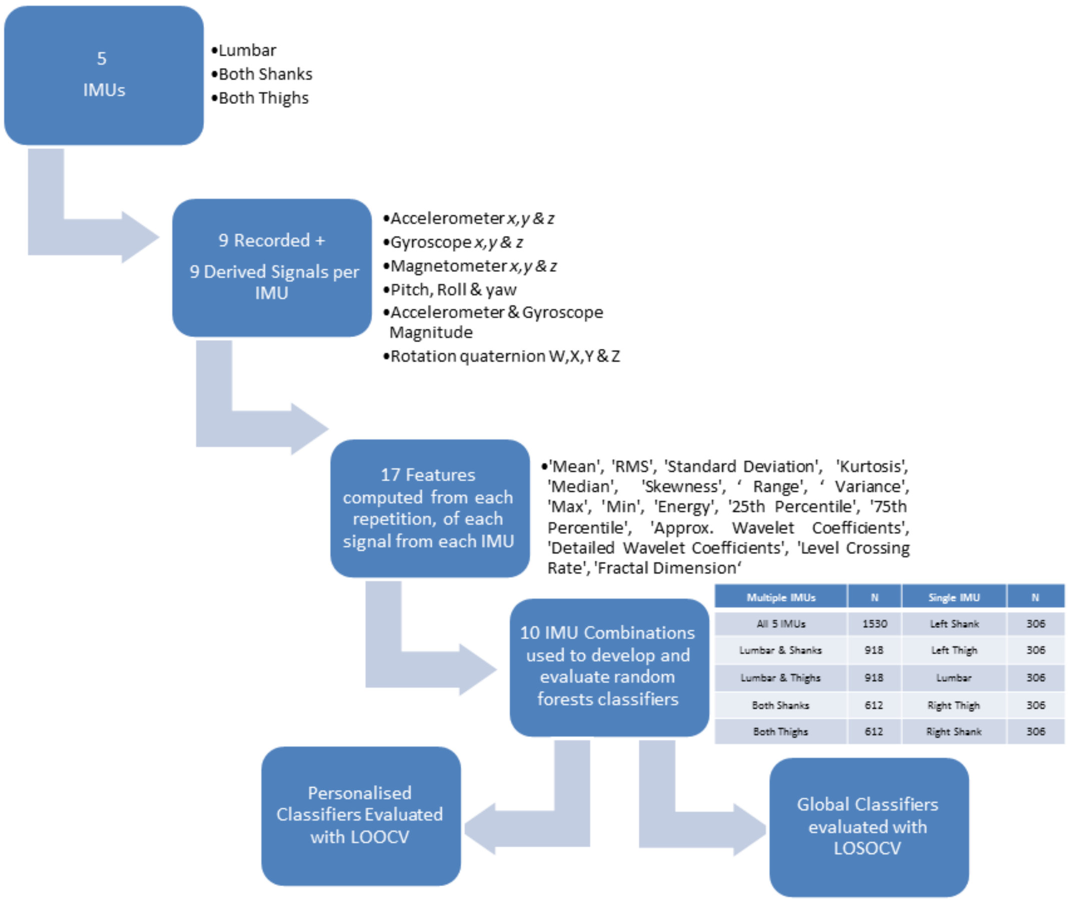

Figure 2: Diagram linking number of IMUs, number of recorded and derived signals, number of features extracted and the variety of feature combinations used to test classifiers

The random-forests method was employed to perform classification (33). This technique was chosen as it has been shown to be effective in analysing exercise technique with IMUs when compared to the Naïve-Bayes and Radial-basis function network techniques (34). 128 decision trees were used in each random-forest classifier. Classifiers were developed and evaluated for the ten combinations of IMUs as shown in Figure 2.

Initially, binary classification was evaluated to establish how effectively each individual IMU and each combination of IMUs could distinguish between acceptable and aberrant barbell squat technique. All repetitions of acceptable technique were labelled '0' and all repetitions performed with one of the pre-defined deviations as outlined in Table 1 were labelled '1'. Multi-label classification was then evaluated on the IMU data to investigate how effectively each individual IMU and each IMU combination could be used to discriminate between acceptable barbell squat technique and each of the six pre-defined deviations from acceptable technique as described in Table 1. All repetitions of acceptable performance remained labelled as '0' and each of the different deviations were labelled '1-6'.

The quality of the global exercise classification system was established using leave-one-subject-out-cross-validation (LOSOCV) and the random-forests classifier with 128 trees (35). Each participant's data corresponds to one fold of the cross validation. At each fold, one participant's data is held out as test data while the random forests classifier is trained with all other participants' data. Where each class in the training data did not have an equal number of instances (i.e. equal number of acceptable and aberrant repetitions in binary classification), random instances of the over-represented class(es) were removed in order to balance the training data. The held out data is used to assess the classifier's ability to correctly categorise new data it is presented with. The use of LOSOCV ensures that there is no biasing of the classifiers, because the test subjects data is completely unseen by the classifier prior to testing.

The quality of the personalised exercise classification systems was established using leave-one-out-cross-validation and a random forests classifier with 128 trees. Each repetition corresponds to one fold of the cross validation. At each fold, one repetition is held out as test data while the random forests classifier is trained with the same participant's other completed repetitions. Where each class in the training data did not have an equal number of instances (i.e. equal number of acceptable and aberrant repetitions in binary classification), random instances of the over-represented class(es) were removed in order to balance the training data. The held out data is used to assess the classifier's ability to correctly categorise new data it is presented with. Participants were not included for this analysis if they did not have at least 2 repetitions belonging to each class being classified as this would not allow for training and test data for that class.

The scores used to measure the quality of classification were total accuracy, average sensitivity and average specificity. Accuracy is the number of correctly classified repetitions of all the exercises divided by the total number of repetitions completed; this is calculated as the sum of the true positives (TP) and true negatives (TN) divided by the sum of the true positives, false positives (FP), true negatives and false negatives (FN):

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

In binary classification acceptable technique was considered the 'positive' class and aberrant technique was considered the 'negative' class. As such, single sensitivity and specificity values were computed to establish binary classification quality for each IMU combination. In multi-label classification, the sensitivity and specificity were calculated for each of the six class labels as outlined in Table 1. Each label was sequentially treated as the 'positive' class, and then the mean and standard deviation across the six values was taken. Sensitivity and specificity were computed using the formulas below. Sensitivity measures the effectiveness of a classifier at identifying a desired label, while specificity measures the classifier's ability to detect other labels.

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

In addition to these measures, receiver operating characteristic (ROC) curves were plotted to compare the quality of global and individualised binary classifiers. A single ROC curve was created for individualised classifiers and global classifiers by pooling the true label score and predicted labels together for all participants. The MATLAB 'perfcurve' function was used to generate the X and Y points for both ROC curves [https://uk.mathworks.com/help/stats/perfcurve.html].

In reviewing the accuracy, sensitivity and specificity scores produced by each classifier, 90% or higher was considered an 'excellent' quality result, 80%-89% was considered a 'good' quality result, 60-79% was considered a 'moderate' result and anything less than 59% was deemed a poor result. The authors chose these values after reviewing the aforementioned literature on identifying deviations from acceptable exercise performance using data derived from IMUs. In reviewing such literature, an existing accepted standard for an excellent, good, moderate or poor classifier could not be found.

Therefore, the above system was agreed on by the authors to facilitate interpretation of results.

5 Results

Table 2 shows the total number of repetitions collected for each class, as labelled by the Chartered Physiotherapist. For binary classification, there were 884 acceptable repetitions and 606 aberrant repetitions recorded.

Table 2. List and description of barbell squat exercise labels used in this study and the number of repetitions extracted of each class as labelled by the Chartered Physiotherapist

Label	Description	Total reps
Acceptable	Acceptable technique	884
Knee Valgus	Knees coming together during downward phase	22
Knee Varus	Knees coming apart during downward phase	183
Knees Too Forward	Knees ahead of toes during downward phase	50
Heels Elevated	Heels raising off the ground during exercise	7
Bent Over	Excessive flexion of hip and torso during exercise	96
Other	Other deviation, not highlighted in NSCA guidelines	250

NSCA = National Strength and Conditioning Association

Table 3 demonstrates the accuracy, sensitivity and specificity of the global classification methods in binary classification.

Table 3. Overall accuracy, sensitivity and specificity in binary classification (acceptable or aberrant technique) for each combination of IMUs following LOSOCV using global classifiers

Sensor(s)	Accuracy (%)	Sensitivity (%)	Specificity (%)
All 5 Sensors	64	70	28
Lumbar & Shanks	65	69	34
Lumbar & Thighs	62	68	21
Both Shanks	66	70	38
Both Thighs	63	75	26
Left Shank	62	70	31
Left Thigh	63	69	24
Lumbar	61	68	21

Right Thigh	63	70	27
Right Shank	62	69	45

Table 4 shows the total accuracy, mean sensitivity and mean specificity of the global classification methods in multi-class classification (detection of exact deviation).

Table 4. Overall accuracy, average sensitivity and average specificity in multi-label classification (exact deviation) for each combination of IMUs following LOSOCV using global classifiers

Sensor(s)	Accuracy (%)	Sensitivity (%)	Specificity (%)
All 5 Sensors	59	24	84
Lumbar & Shanks	57	25	85
Lumbar & Thighs	57	22	84
Both Shanks	53	20	85
Both Thighs	52	15	82
Left Shank	48	19	85
Left Thigh	48	15	82
Lumbar	52	19	83
Right Thigh	51	14	82
Right Shank	55	21	86

Table 5 demonstrates the mean accuracy, sensitivity and specificity scores for each individual participant's personalised barbell squat technique binary classifier that was evaluated with LOOCV.

Table 5. Average accuracy, sensitivity and specificity in binary classification (acceptable or aberrant technique) for each combination of IMUs following LOOCV using personalised, N of 1 classifiers

Sensor(s)	Accuracy (%) ± SD	Sensitivity (%) ± SD	Specificity (%) ± SD
All 5 Sensors	82 ± 13	83 ± 14	84 ± 14
Lumbar & Shanks	80 ± 14	81 ± 16	82 ± 14
Lumbar & Thighs	82 ± 12	82 ± 13	87 ± 11
Both Shanks	79 ± 16	80 ± 19	81 ± 15
Both Thighs	83 ± 11	84 ± 12	88 ± 12
Left Shank	79 ± 6	81 ± 17	80 ± 20
Left Thigh	81 ± 13	81 ± 13	84 ± 16
Lumbar	80 ± 14	81 ± 15	83 ± 16
Right Thigh	80 ± 16	84 ± 12	82 ± 17

Right Shank	80 ± 15	78 ± 17	82 ± 15

Figure 3 shows an ROC curve for all participants when both global and individualised classification methodologies were used for a binary classification system based on data from the left thigh IMU. The area under the curve (AUC) for the global method was 0.52 and the AUC for the personalised method was 0.98.

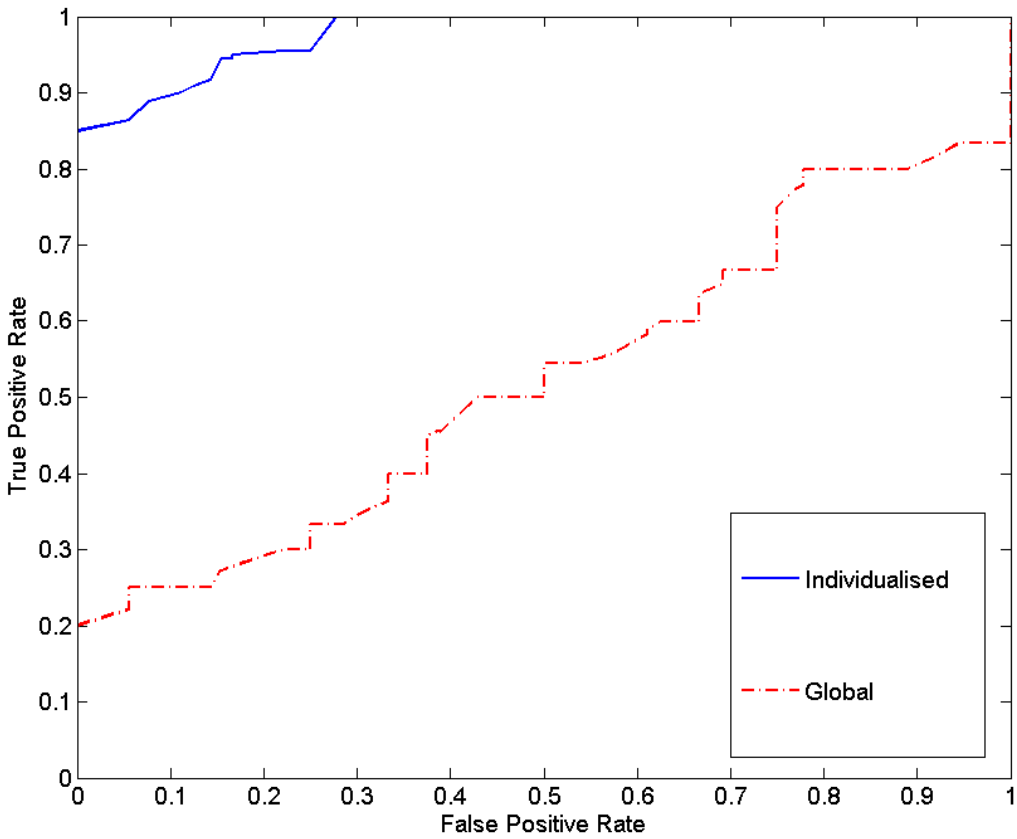

Figure 3: ROC curves comparing binary classification systems when using global and personalised classification methodologies using data from the left thigh IMU. 'Acceptable' technique was considered the 'true' class.

Table 6 demonstrates the mean accuracy, sensitivity and specificity scores for each individual participant's personalised barbell squat technique multi-class classifier that was evaluated with LOOCV.

Table 6. Overall accuracy, average sensitivity and average specificity in multi-label classification (exact deviation) for each combination of IMUs following LOOCV using personalised, N of 1 classifiers

Sensor(s)	Accuracy (%) ± SD	Sensitivity (%) ± SD	Specificity (%) ± SD

All 5 Sensors	70 ± 20	73 ± 17	88 ± 12
Lumbar & Shanks	69 ± 20	71 ± 18	90± 8
Lumbar & Thighs	70 ± 17	70 ± 15	87 ± 9
Both Shanks	70 ± 18	71 ± 17	89 ± 7
Both Thighs	70 ± 16	72 ± 13	88 ± 11
Left Shank	67 ± 20	71 ± 17	86 ± 12
Left Thigh	69 ± 18	70 ± 18	89 ± 9
Lumbar	67 ± 20	70 ± 19	89 ± 10
Right Thigh	70 ± 16	72 ± 13	86 ± 12
Right Shank	67 ± 20	71 ± 15	88 ± 8

6 Discussion

The aims of this study were to: (a) determine if an IMU system is capable of distinguishing between acceptable and aberrant barbell squat technique; (b) determine the capabilities of an IMU system at identifying specific natural deviations from acceptable barbell squat technique; and (c) compare a personalised (N of 1) classifier to a global classifier in identifying the above. The results of this paper indicate that an IMU system is not capable of detecting aberrant barbell squat technique using global classifiers as demonstrated by the low specificity scores (Table 3). However, good levels of accuracy, sensitivity and specificity are achieved using a personalised classifier (Table 5). Similarly, the ability of an IMU system to identify specific deviations in technique is poor using a global classifier (Table 4) however these results are improved to moderate levels using a personalised classifier (Table 6).

To the best of the authors' knowledge this is first paper to demonstrate the ability of an IMU system to identify natural deviations during performance of the barbell squat exercise. To date there has been a lack of research investigating the ability of IMU systems to classify technique in lower limb compound exercises. Whilst global classification techniques replicating those used in this paper have been shown to successfully classify naturally occurring deviations in the single leg squat (21, 36), they were shown to be ineffective in classifying barbell squat technique. Additionally, we have demonstrated that a personalised classifier out performs a global classifier in assessing barbell squat technique (Figure 3, Tables 3-6). This is likely due to a number of factors. As outlined in Table 2 the number of acceptable repetitions far outnumbers any other label. This unbalanced data set makes it difficult to create global classifiers that can be used for all individuals (24, 25). As many deviations were seen sporadically, the use of a global classifier to identify specific deviations in the barbell squat may require the collection of a data set consisting of larger amounts of each deviation. The inter-subject variability in movement patterns that are considered acceptable in barbell squat technique may also exceed the intra-subject variability between acceptable technique and aberrant technique. This would make the creation of global classifiers exceptionally difficult. It is likely that this is not the case for the single leg squat and hence global classification methodologies worked better for classifying deviations in this exercise.

It is difficult to directly compare results with previous work in the area due to differences in exercises investigated, sensor positions and classifier techniques employed. However, the results presented in this paper using a personalised classifier compare favourably to other research in the area (16-19). The majority of research to date has investigated the ability of IMU systems to monitor technique in simple exercises such as straight leg raises (16), dumbbell curls (18), or heel slides (19). This paper describes an evaluation of an IMU system's ability to quantify barbell squat technique, a more complex exercise that involves multiple joints. This system has also demonstrated the ability to identify a total of seven different classes (Table 2). The lower number of classes in some of the studies (16, 18, 19) may make it easier for classifiers to identify specific deviations and subsequently produce higher accuracy, sensitivity and specificity scores. However, it must be noted that all of these systems used a global classifier in distinguishing between exercise technique and many of the studies classified deviations that were deliberately induced. As shown in Table 4 the ability of a global classifier to identify specific deviations in barbell squat technique is poor. Therefore, a personalised classifier may be more suitable when assessing this exercise in a clinical setting where technique deviations are natural.

The results presented in Table 5 and Table 6 show that a single IMU system is comparable to a multiple IMU system in determining barbell squat technique using a personalised classifier. Multiple IMU systems are more expensive than a single IMU system due to the need to purchase additional sensors. Furthermore, they are less practical for end users as there is an increased risk of placement error in addition to power usage and

Bluetooth™ connectivity issues. For these reasons a reduced IMU set-up is more desirable for daily environment applications (37). Therefore, the single IMU system results presented in this paper increase the likelihood of clinical adoption.

A personalised classifier offers a number of benefits compared to a global classifier when assessing barbell squat technique. Most obviously, the higher levels of accuracy would mean an improved user experience in a clinical setting. A personalised classifier also allows for analysis to be performed on data sets that are unbalanced, like the one shown in Table 2. Furthermore, personalised classifiers are also more computationally efficient than global classifiers as they are developed using less training data and therefore require less memory. This would improve processing time and increase battery life.

The main disadvantage associated with a personalised classifier is that the user must collect and label data sets from individual patients. This means clinicians must monitor exercise technique in real time or use post-hoc video analysis and label this appropriately. This may prove time consuming. Furthermore, this does not lend itself to a "set-up and go" approach that involves minimal interaction with the user interface, which is more preferable for end-users (8). However, as clinicians often monitor exercise technique prior to allowing patients complete their exercises it may fit into clinical practice without issue, with clinicians labelling repetitions as they analyse exercise completion. Furthermore, the labelled data set developed using this method could be used to build global classifiers better equipped at identifying natural deviations in the future. This is because all labelled data that is collected by practitioners could be stored and used to build the large data set necessary to improve global classifier scores.

A challenging aspect of this work is to ascertain whether the results presented in this paper are sufficient for real-life applications. It is likely that the classification accuracy achieved using a global classifier is too low for use in healthcare environments, while those produced by a personalised classifier may be acceptable. However, it is important to note that what is considered an acceptable level of classification accuracy is likely to be influenced by application domain (injury rehabilitation, strength and conditioning, musculoskeletal injury risk screening, etc.) and end user profile (rehabilitation professionals, sports coaches, strength and conditioning staff, recreational gym users). Our research team is undertaking further projects to determine usability, functionality and user perceptions of wearable technology to assess exercise biomechanics. This information is being gathered from a range of professionals and patients, who incorporate exercises such as the barbell squat in their rehabilitation programme, exercise routine and injury risk screening protocols. It is envisaged that this will provide greater indication as to the levels of accuracy end users would define as acceptable. Furthermore, this work will contribute new information regarding how best to provide actionable feedback to these users that allows for safe and effective exercise completion.

7 Conclusion

Our results show that a system based on data derived from body worn IMUs can classify acceptable and aberrant barbell squat biomechanics with good overall accuracy, sensitivity and specificity using a personalised classifier. These classification scores are maintained even with a single IMU. The ability to identify specific deviations is more difficult but can be achieved with a moderate level of overall accuracy using a personalised classifier. Our results are comparable with other research in the area, despite the barbell squat being a more complex exercise then many of those previously investigated. However, most of this research has been carried out using global classifiers. While this may allow for less user interaction, it produces poor levels of accuracy when attempting to identify specific natural deviations during performance of the exercise. As a result, the use of a personalised classifier may be more appropriate for identifying natural deviations in barbell squat technique.

8 References

1. Cook G, Burton L, Hoogenboom B. Pre-participation screening: the use of fundamental movements as an assessment of function - part 1. North American journal of sports physical therapy. 2006;1(2):62-72.
2. Hall M, Nielsen JH, Holsgaard-Larsen A, Nielsen DB, Creaby MW, Thorlund JB. Forward lunge knee biomechanics before and after partial meniscectomy. The Knee. 2015;22(6):506-9.
3. Ahmadi A, Mitchell E, Destelle F, Gowing M, O'Connor NE, Richter C, et al. Automatic Activity Classification and Movement Assessment During a Sports Training

Session Using Wearable Inertial Sensors. Proceedings of the 11th International Conference on Wearable and Implantable Body Sensor Networks (BSN): IEEE; 2014. p. 98-103.
4. Bonnechere B, Jansen B, Salvia P, Bouzahouene H, Omelina L, Moiseev F, et al. Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry. Gait & posture. 2014;39(1):593-8.
5. Bonnet V, Mazza C, Fraisse P, Cappozzo A. Real-time estimate of body kinematics during a planar squat task using a single inertial measurement unit. IEEE Transactions on Biomedical Engineering. 2013;60(7):1920-6.
6. Whiteside D, Deneweth JM, Pohorence MA, Sandoval B, Russell JR, McLean SG, et al. Grading the functional movement screen: A comparison of manual (real-time) and objective methods. The Journal of Strength & Conditioning Research. 2016;30(4):924-33.
7. McGrath D, Greene BR, O'Donovan KJ, Caulfield B. Gyroscope-based assessment of temporal gait parameters during treadmill walking and running. Sports Engineering. 2012;15(4):207-13.
8. Morris D, Saponas TS, Guillory A, Kelner I. RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: ACM; 2014. p. 3225-34.
9. Leardini A, Lullini G, Giannini S, Berti L, Ortolani M, Caravaggi P. Validation of the angular measurements of a new inertial-measurement-unit based rehabilitation system: comparison with state-of-the-art gait analysis. Journal of neuroengineering and rehabilitation. 2014;11(1):1-7.
10. Tang Z, Sekine M, Tamura T, Tanaka N, Yoshida M, Chen W. Measurement and Estimation of 3D Orientation using Magnetic and Inertial Sensors. Advanced Biomedical Engineering. 2015;4:135-43.
11. Muehlbauer M, Bahle G, Lukowicz P. What can an arm holster worn smart phone do for activity recognition? Proceedings of the International Symposium on Wearable Computers (ISWC): IEEE; 2011. p. 79-82.
12. Chang K-H, Chen MY, Canny J. Tracking free-weight exercises. Ubiquitous Computing. 4717: Springer; 2007. p. 19-37.
13. Seeger C, Buchmann A, Van Laerhoven K. myHealthAssistant: a phone-based body sensor network that captures the wearer's exercises throughout the day. Proceedings of the 6th International Conference on Body Area Networks: ICST; 2011. p. 1-7.
14. Pernek I, Hummel KA, Kokol P. Exercise repetition detection for resistance training based on smartphones. Personal and ubiquitous computing. 2013;17(4):771-82.
15. Pernek I, Kurillo G, Stiglic G, Bajcsy R. Recognizing the intensity of strength training exercises with wearable sensors. Journal of Biomedical Informatics. 2015;58:145-55.
16. Taylor PE, Almeida GJ, Hodgins JK, Kanade T. Multi-label classification for the analysis of human motion quality. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC): IEEE; 2012. p. 2214-8.
17. Melzi S, Borsani L, Cesana M. The virtual trainer: supervising movements through a wearable wireless sensor network. 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops, 2009: IEEE; 2009. p. 1-3.
18. Velloso E, Bulling A, Gellersen H, Ugulino W, Fuks H. Qualitative activity recognition of weight lifting exercises. Proceedings of the 4th Augmented Human International Conference: ACM; 2013. p. 116-23.

19. Giggins OM, Sweeney KT, Caulfield B. Rehabilitation exercise assessment using inertial sensors: a cross-sectional analytical study. Journal of Neuroengineering and Rehabilitation. 2014;11(1):158-68.
20. O'Reilly M, Whelan D, Chanialidis C, Friel N, Delahunt E, Ward T, et al. Evaluating squat performance with a single inertial measurement unit. IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN): IEEE; 2015. p. 1-6.
21. Whelan D, O'Reilly M, Ward T, Delahunt E, Caulfield B. Evaluating Performance of the Single Leg Squat Exercise with a Single Inertial Measurement Unit. Proceedings of the 3rd Workshop on ICTs for Improving Patients Rehabilitation Research Techniques: ACM; 2015. p. 144-7.
22. Whelan D, O'Reilly M, Ward T, Delahunt E, Caulfield B. Evaluating Performance of the Lunge Exercise with Multiple and Individual Inertial Measurement Units. Pervasive Health 10th EAI International Conference on Pervasive Computing Technologies for Healthcare 2016. p. 101-8.
23. Whelan D, O'Reilly M, Huang B, Giggins O, Kechadi T, Caulfield B. Leveraging IMU data for accurate exercise performance classification and musculoskeletal injury risk screening. IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)2016. p. 659-62.
24. Chawla NV. Data mining for imbalanced datasets: An overview. Data mining and knowledge discovery handbook: Springer; 2005. p. 853-67.
25. He H, Garcia EA. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering. 2009;21(9):1263-84.
26. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques. Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies: IOS Press; 2007. p. 3-25.
27. Baechle TR, Earle RW. Resistance Training Exercise Techniques. NSCA's Essentials of Personal Training: Champaign, IL; 2004.
28. Jerri AJ. The Shannon sampling theorem—Its various extensions and applications: A tutorial review. Proceedings of the IEEE. 1977;65(11):1565-96.
29. Shimmer 9DOF calibration [Available from: http://www.shimmersensing.com/shop/shimmer-9dof-calibration (last accessed 13 March 2017).
30. Madgwick SOH, Harrison AJL, Vaidyanathan R. Estimation of IMU and MARG orientation using a gradient descent algorithm. Proceedings of the IEEE International Conference on Rehabilitation Robotics (ICORR): IEEE; 2011. p. 1-7.
31. Katz MJ, George EB. Fractals and the analysis of growth paths. Bulletin of Mathematical Biology. 1985;47(2):273-86.
32. Single-level discrete 1-D wavelet transform [Available from: http://uk.mathworks.com/help/wavelet/ref/dwt.html (last accessed 13 March 2017).
33. Breiman L. Random forests. Machine Learning. 2001;45(1):5-32.
34. Mitchell E, Ahmadi A, O'Connor NE, Richter C, Farrell E, Kavanagh J, et al. Automatically detecting asymmetric running using time and frequency domain features. Proceedings of the 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN): IEEE; 2015. p. 1-6.
35. Fushiki T. Estimation of prediction error by using K-fold cross-validation. Statistics and Computing. 2011;21(2):137-46.

36. Whelan D, O'Reilly M, Ward T, Delahunt E, Caulfield B. Technology in Rehabilitation: Evaluating the Single Leg Squat Exercise with Wearable Inertial Measurement Units. Methods of Information in Medicine. 2016;55(6):[Epub ahead of print].
37. Bonnet V, Mazza C, Fraisse P, Cappozzo A. A least-squares identification algorithm for estimating squat exercise mechanics using a single inertial measurement unit. Journal of biomechanics. 2012;45(8):1472-7.