IEEE Journal of Translational Engineering in Health and Medicine

Received 21 December 2016; revised 30 June 2017; accepted 9 July 2017. . Date of publication 14 November 2017; date of current version 14 November 2017.

Digital Object Identifier 10.1109/JTEHM.2017.2736559

Automated Assessment of Dynamic Knee Valgus and Risk of Knee Injury During the Single Leg Squat

REZVAN KIANIFAR[1], ALEXANDER LEE[2], SACHIN RAINA[2], AND DANA KULIĆ[1], (Member, IEEE)

[1]Electrical and Computer Engineering Department, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[2]MSK Metrics, Mississauga, ON L5J 1B1, Canada

CORRESPONDING AUTHOR: R. Kianifar (rkianifar@uwaterloo.ca)

This work was supported by the Natural Science and Engineering Research Council of Canada.

ABSTRACT Many clinical assessment protocols of the lower limb rely on the evaluation of functional movement tests such as the single leg squat (SLS), which are often assessed visually. Visual assessment is subjective and depends on the experience of the clinician. In this paper, an inertial measurement unit (IMU)-based method for automated assessment of squat quality is proposed to provide clinicians with a quantitative measure of SLS performance. A set of three IMUs was used to estimate the joint angles, velocities, and accelerations of the squatting leg. Statistical time domain features were generated from these measurements. The most informative features were used for classifier training. A data set of SLS performed by healthy participants was collected and labeled by three expert clinical raters using two different labeling criteria: ''observed amount of knee valgus'' and ''overall risk of injury''. The results showed that both flexion at the hip and knee, as well as hip and ankle internal rotation are discriminative features, and that participants with ''poor'' squats bend the hip and knee less than those with better squat performance. Furthermore, improved classification performance is achieved for females by training separate classifiers stratified by gender. Classification results showed excellent accuracy, 95.7 % for classifying squat quality as ''poor'' or ''good'' and 94.6% for differentiating between high and no risk of injury.

INDEX TERMS Human motion analysis, motion assessment protocols, single leg squat, classification, inertial measurement unit, feature selection.

I. INTRODUCTION

Many clinical assessment protocols rely on functional movement tests, where the patient is asked to perform a target movement while the clinician observes and assesses the movement. The single leg squat (SLS) is an example of a functional movement test, and is commonly used in rehabilitation, sports medicine and orthopedic settings [1]. Correct performance of the SLS can provide an indication of knee function and assessment of recovery. An important component of the rating of the quality of a performed SLS is the degree of inward movement of the knee, known as medial knee displacement or dynamic knee valgus (DKV), as shown in Fig. 1.

DKV correlates with non-contact Anterior Cruciate Ligament (ACL) injury and patellofemoral pain [2]. The SLS test helps with early screening of those at higher risk of ACL rupture, which happens frequently among athletes involved in high risk sports such as soccer, football, basketball, and lacrosse [3].

More than 120,000 ACL injuries occur annually, most during the high school years [3]. Treatment in 90% of patients includes reconstruction surgery, followed by a rehabilitation period [4]. The estimated average annual treatment cost of ACL rupture in the US is more than 2 billion dollars [5]. Return to play for professional athletes following ACL surgery can take almost one year [6]. More than 50% will not return to their pre injury level of performance [4] and between 50% to 100% develop osteoarthritis within 5 to 10 years after surgery. Moreover, the risk of re-injury increases up to 5 times in those who have undergone initial surgery [4]. All these statistics highlight the importance of early screening of individuals at

FIGURE 1. Left: "good" SLS performance. Right: inward movement of the knee during "poor" SLS called Dynamic Knee Valgus (DKV).

higher risk, through functional movement tests such as the SLS.

Current SLS assessment is based on visual observation, which is subjective and depends on the experience of the clinician. In addition, since clinicians see a large number of patients each day, it may be difficult for them to remember the previous condition of each patient without a quantitative history for each person.

The purpose of this study is to develop an automated assessment system for the SLS test. An IMU-based method is used for joint angle, velocity and acceleration estimation of the squatting leg. Statistical and time domain features are generated from these measurements. The most informative features are selected using a combination of different feature selection techniques and used as input for supervised classifier training. A dataset of SLS performed by healthy participants was collected and labeled by three expert clinical raters. The raters applied two different labeling criteria: 1) the observed amount of DKV that occurred during the SLS motion and 2) the clinician's judgement of the participant's risk of knee injury based on their SLS performance. The expert raters rated the DKV item as ''poor'', ''moderate'' or ''good''; and they rated the knee injury risk item as ''moderate to high'', ''mild'' and ''none''. The labeled data was used to train classifiers for each assessment criterion. The results showed excellent discrimination between ''good'' and ''poor'' SLS, and also between high risk and low risk participants.

The rest of the paper is organized as follows: Section II reviews related studies. Section III presents the proposed methodology including pose estimation, segmentation, feature extraction, feature selection, and classification. Section IV explains the data collection and labeling procedures. Section V provides the results and finally, section VI discusses the results and concludes the paper.

II. RELATED WORK

SLS and other functional movement tests such as the double leg squat and double leg jump have been widely investigated in clinical and sport medicine studies. The main purpose of the majority of these studies is to find relationships between the occurrence of knee valgus during the mobility test and factors such as age, gender, body mass index, history of injury, and kinematic or neuromuscular characteristics of the subjects (usually athletes) [2], [7], [9]–[11]. Studying these predictors aids in the development of appropriate injury prevention strategies. For example, if it is found that hip abductor weakness correlates with poor performance (DKV occurrence) in SLS, then specific exercises can be prescribed to improve the strength and function of this muscle group.

Zeller *et al.* [8] investigated the kinematics and muscular activity of nine men and nine women athletes during the SLS. According to their results, women exhibited more knee valgus, which was associated with greater ankle dorsiflexion and pronation, less trunk lateral flexion, and greater hip adduction (Add.), flexion (Flex.), and rotation. Rectus femoris muscle activation was also greater in women.

Hip and foot contributions to high DKV were investigated by Bittencourt et al. [9]. They examined 173 athletes during the SLS and at the landing moment of a double leg jump. Data was collected in a motion capture studio and the frontal plane knee projection angle was measured at 60 degrees of knee flexion and during a static single-leg stance. Four other measures, including the passive range of motion (ROM) of the hip internal rotation (IR), the isometric strength of the dominant-limb hip abductors, the shank-forefoot alignment and participants' gender were defined as features to be input into a classification and regression tree. Their results suggest that high DKV can be predicted by decreased hip abductor torque and increased passive ROM of the hip IR for both the SLS and double leg jump landing.

Padua *et al.* [2] compared the neuromuscular characteristics of a group of 18 individuals with excessive knee valgus with a control group of 19 healthy individuals during double leg squat performance. Electromyography (EMG) was used for muscle activation measurement. Individuals were assigned to either the control or DKV group based on an evaluation by an expert rater. A correlation between DKV and increased hip-adductor activation as well as increased coactivation of the gastrocnemius and tibialis anterior muscles was reported.

In a similar study, Stiffler *et al.* [10] compared kinematic characteristics including ROM and postural alignment of 97 healthy individuals during the double leg jump, in order to find differences between those with and without excessive DKV. Motion labeling was based on the total Landing Error Scoring System (LESS) [12]. Their results showed associations between DKV and less ankle dorsiflexion, as well as higher quadriceps angle (Q-angle).

The relationship between the occurrence of DKV with age, gender, and body mass index was studied by Ugalde *et al.* [11]. They investigated 142 middle and high school athletes while performing the SLS and drop jump tests. They defined knee-hip ratio as the distance between the knees at maximum flexion divided by the distance between hips at a quiet stance during the drop-jump test. Their results showed significantly lower knee-hip ratio for individuals with DKV during SLS. However, they found no relationship between DKV and age, gender, or body mass index during the SLS test.

The investigations described above have focused on identifying correlates of DKV. Generally, these studies detect positive DKV occurrence based on expert clinician observations or manual measurements extracted from video frames. Very few studies have tried to develop an algorithm for automated DKV detection. In one such study, Whelan *et al.* [13] classified SLS repetitions of 19 healthy participants into correct and incorrect categories using a single lumbar-mounted IMU. They extracted time domain features from accelerometer and gyroscope measurements, the IMU orientation (represented as roll, pitch, yaw), and accelerometer magnitude. Using the generated feature vector and labels provided by an expert, they trained a Random Forest classifier, which achieved 92.1 % accuracy with repeated random-sample validation. Despite these promising results, they may be difficult to interpret clinically, as features were defined based on direct acceleration and gyroscope output signals, whereas clinical assessment of the SLS includes the visual estimation and interpretation of kinematic joint parameters, especially the joint angles. Developing a classifier which works based on these parameters, therefore, has the advantage of interpretability. Furthermore, Whelan *et al.*, did not perform a Leave One Subject Out Cross Validation (LOSO-CV); therefore, it is not clear how well the classifier would generalize to subjects unseen during training, which is critical for clinical applications.

In our pilot study [14], we classified SLS performance as "poor," "moderate" or "good" based on clinically understandable features. In the study, 3 IMUs were attached to the shank, thigh and low back of 7 healthy volunteers who performed 5 consecutive repetitions of SLS. Using an Extended Kalman Filter (EKF) based method [15], joint angles, velocities and accelerations of the ankle, knee, and hip joints were estimated from the IMU data. Statistical features were then computed from estimated joint kinematics and feature selection was applied to find the best predictors of DKV during SLS. Three classifiers were applied to full dimensional features, the subset of selected features and extracted features based on Supervised Principal Component Analysis (SPCA). Classification results for both 10 fold cross validation (10F-CV) and LOSO-CV were reported. The results showed that the ankle internal rotation angle was the best predictor of DKV, with classification accuracies of 98% for 2-class ("good" versus "poor" squat) classification using LOSO-CV and 73% for 3- class ("good" versus "moderate" versus "poor") classification.

In the current study, we extend the proposed approach to a larger dataset including a similar number of male and female participants, whose performances are labeled by clinical experts using two different criteria: amount of knee valgus and risk of knee injury. Additional data analysis is also performed based on gender specific datasets and ankle only features.

III. PROPOSED METHODS

Since raw IMU outputs may not be intuitively interpretable to clinicians, joint angles, velocities and accelerations were extracted and used for classification, using an EKF-based pose estimation method [15].

The time series data of the estimated joint angles are then segmented into single squats, from which statistical time domain features are extracted and used for feature selection and classification.

A. POSE ESTIMATION

A set of three IMUs were employed to track lower body motion during the squats. An IMU is a compact package composed of an accelerometer measuring linear acceleration, a gyroscope measuring angular velocity and a magnetometer measuring the earth's magnetic field. The magnetometer is not usually used in pose estimation, as it is subject to interference by ferromagnetic objects [16].

Since the IMU data is noisy and can suffer from drift, similar to [15], a kinematic model of the lower leg was applied to calculate angular velocity and linear acceleration at each time step to be used for correction of sensor estimates of these values. The kinematic model was composed of a 3 Degree of Freedom (DOF) ankle joint, 1 DOF knee joint, and 3 DOF hip joint, depicted in Fig. 2.

The kinematic model predictions of the angular velocity and linear acceleration and sensor measurements of these parameters were then fused by the EKF [15]. The position, velocity, and acceleration of each DOF are defined as the states to be estimated by the EKF. A constant acceleration model was used for the state propagation. For more details see [15].

B. SEGMENTATION

To extract a single SLS repetition from continuous time series data, the joint angle trajectories needed to be segmented before feature extraction. For segmentation, a peak detection method developed by [17] was applied to the knee flexion angle. Knee flexion was chosen for segmentation because the knee has a large flexion ROM, and its peaks are easily detectable. A first order Butterworth filter with cutoff frequency of 0.3 Hz was applied to the knee joint trajectory prior to segmentation. Note that this filter is applied only for segmentation and not for the subsequent feature extraction. The midpoints between peaks were then calculated and used as segmenting points as depicted in Fig. 3. Fig. 4 shows an example of segmented joint angles used for feature extraction (without low pass filtering).

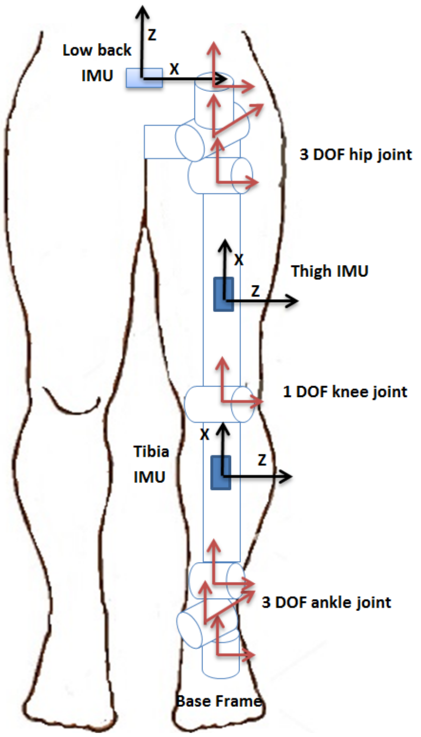

FIGURE 2. 7 DOF kinematic model of the left leg including the 3 DOF ankle joint, 1 DOF knee joint, and 3 DOF hip joint.

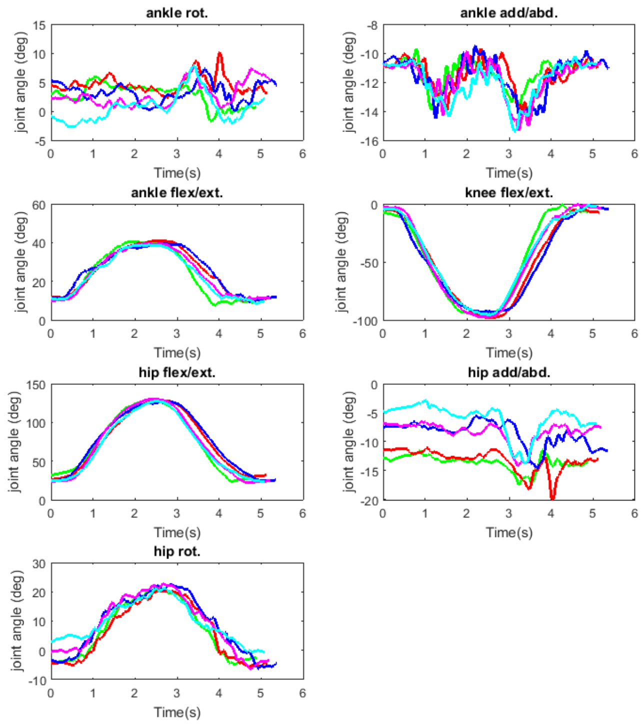

FIGURE 4. An example of segmented joint angles without low pass filtering used for feature extraction.

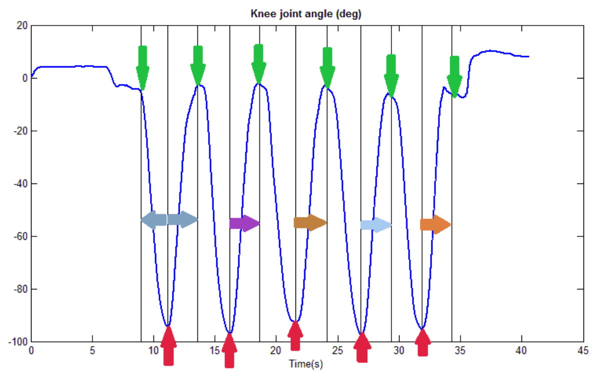

FIGURE 3. Segment points (top arrows) were found by detecting peaks (bottom arrows) of low pass filtered knee joint angle and computed the midpoint of the peak to peak distances (horizontal arrows).

C. FEATURE EXTRACTION

Feature extraction is used to transform raw time series data into a lower dimensional representation of the motion relevant for prediction of DKV. Various statistical feature extraction methods have been applied for human activity recognition [18]. These methods are categorized into time domain or frequency domain methods. The most common time domain methods are standard deviation (STD), mean, variance (VAR), mean absolute deviation (MAD), interquartile range (IQR), entropy, correlation between axes, and kurtosis. Common frequency domain methods include Fourier transform (FT) and discrete cosine transform (DCT).

Due to easier clinical interpretability and better temporal localization, we applied only time domain feature extraction methods including the root mean square (RMS), STD, VAR, mean, MAD, skewness, kurtosis, range, minimum, and maximum of the joint angle, velocity and acceleration of each DOF for each segment of the data. Therefore, for each repetition of the squat, a feature vector of 210 different features was extracted.

D. FEATURE SELECTION

We do not know which of the defined features better predicts DKV. Moreover, some features might be redundant or irrelevant, which may degrade the classification results. Selecting the most appropriate features not only helps with dimensionality reduction but also suggests the best predictors of DKV to clinicians.

A large number of feature selection techniques are available in the literature, usually categorized as filter, wrapper or embedded techniques [19]. Filter techniques select relevant features based on statistical tests. Wrapper techniques use the performance of a predefined learning algorithm as the selection criterion. In embedded techniques, feature selection occurs in parallel to model learning, so that feature selection is embedded within a classification model [19].

For this study, we applied 18 different feature selection techniques from all three categories. Matlab packages available from the Arizona State University [19] repository and from Pohjalainen et al. [20] were used for implementation. Wrapper methods included Random Subset Feature Selection, Sequential Forward Selection, and Sequential Floating Forward Selection.

Filter methods were Mutual Information, Statistical Dependency, Correlation based Feature Selection, Chi-square, Fast Correlation-Based Filter, Fisher Score, Gini Index, Information Gain, Kruskal- Wallis, Minimum-Redundancy-Maximum-Relevance selection, Relief-Feature selection strategy, and T-test.

From embedded techniques, Sparse Multinomial Logistic Regression via Bayesian L1 Regularization, Bayesian Logistic Regression, and Least Absolute Shrinkage and Selection Operator (LASSO) were utilized.

Features ranked among top ten by majority of the methods are reported as selected features. In addition to subset feature selection, feature extraction using SPCA was also applied. Matlab code developed by Barshan *et al.* [21] was used for SPCA implementation.

E. CLASSIFICATION

For classification purposes, six different methods were applied: Support Vector Machine (SVM), Linear Multinomial Logistic Regression (LMLR), Decision Tree (DT), Naïve Bayes (NB), K Nearest Neighborhood (KNN), and Random Forests.

All classification techniques were implemented using Matlab 2016a. The results showed that SVM, KNN, and NB always outperformed other classifiers for this dataset. Therefore, classification results are reported for these three classifiers only.

Support Vector Machine (SVM) is a statistical classification method which finds the optimal separating hyperplane by maximizing the margin between data points of the two classes [22]. SVM is suitable for small datasets and high dimensional data [23]. Combined with kernels, it can handle nonlinearly separable problems. However, the result may be difficult to interpret and there is no standard way for dealing with multi class problems.

Naïve Bayes is a probabilistic classifier based on Bayes rule which assumes conditional independence of the features [24]. It returns the probability that the input vector belongs to the target class. Naïve Bayes is a fast and simple algorithm that is not sensitive to irrelevant features. It can handle discrete and real data; however, the feature independence assumption may not be valid [25].

KNN is an instance-based learning algorithm which assigns a label to a new data point based on the labels of its k most similar neighbors in the training set (given a similarity metric) [26].

IV. EXPERIMENTS

14 participants including 7 males and 7 females with mean age of 30.8 ± 5.5, mean height of 173.8 ± 12 cm, and mean weight of 70.4 ± 10.4 kg participated in the study. For two participants, the dominant leg (the leg they would kick a soccer ball with) was the left; the other participants were right legged. To be included in the study, subjects had to be between the ages of 18-65 years, and must not have had a medical history that restricted participation in a standard musculoskeletal clinical examination of the lower extremity. This would include the clinical suspicion of an emergent health issue, severe neurological compromise, or the presence of an acute fracture, dislocation or severe knee ligament instability. Subjects were excluded from the study if they did not meet any of the above inclusion criteria.

Ethics approval from Institutional Review Board Services was obtained prior to the start of the study. All participants signed a consent form prior to the start of data collection.

Given our inclusion criteria, it is possible that some individuals with active knee pain or discomfort could have been recruited. Considering we were primarily interested in analyzing data from individuals without active knee injury, we identified subjects with active knee pain or discomfort by having participants fill out the International Knee Documentation Committee (IKDC) subjective knee evaluation form [27] for each knee. Data from all 14 subjects were used in the Inter and Intra-rater Reliability (IRR) analyses, while only data from knees that scored over 95% on the IKDC were included in training and cross validation.

FIGURE 5. Sensor placement during SLS data collection.

A. DATA COLLECTION

Three Yost [28] IMUs were attached to the participants' low back at the level of the first sacral vertebra, the anterior thigh 10 cm above the patella aligned with the sagittal plane, and the flat surface of the shank at the level of the tibial tuberosity using hypoallergenic tape. Sensor placement locations are depicted in Fig. 5. Data was communicated to a nearby computer via Bluetooth communication with an average sampling rate of 90 ± 10 Hz. Data were interpolated and resampled to the same rate of 200 Hz before subsequent analysis.

Participants were instructed to take off their shoes and perform five continuous cycles of SLS with their toes pointing forward and arms crossed in front of the body. They were asked to perform SLS with both the right and left legs without moving the foot or lifting the heel. In instances where subjects lost their balance, their legs contacted each other, or the non-weight bearing leg touched the ground, the trial was deemed unsuccessful and all cycles were repeated.

B. DATA LABELLING

The participants' performance was videotaped during the tests. Videos were then reviewed by three expert clinicians with advanced training in a sports sciences fellowship, with an average of 9 years clinical experience. Raters were asked to label each squat repetition. The clinical rating criteria were adapted and modified from [29] and included 2 items: "Knee Valgus" and "Rater's Subjective Overall Knee Injury Risk". We aimed to discriminate between "good" and "poor" squats or screen "high risk" subjects from "no risk" ones (2-class classification), and to assess if a finer grained assessment is possible by adding a "moderate" grading level (3-class classification). For this purpose, each criterion was comprised of a three-level rating scale of "0," "1" or "2." For the knee valgus criterion [29], a score of "0" was defined as no valgus, "1" as moderate knee valgus, and "2" as severe knee valgus. For the overall knee risk of injury criterion, a score of "0" was defined when the individual was at no risk and no intervention was required, a score of "1" was defined when there was mild/low risk and moderate intervention was required, and a score of "2" was defined when the individual was at moderate to high risk and significant intervention was required. To determine the rating for the "overall knee injury risk assessment" item, the clinical raters were asked to base their assessment of the overall whole body motion that occurred when the subject was performing the SLS, not just the subject's knee position.

Participants were asked to perform SLS using their natural movement. Many individuals in the population who do not have a knee injury exhibit DKV movement strategies of varying severity, DKV is correlated with potential risk of future injury.

The 14 participants performed 5 SLS repetitions with both left and right legs resulting in 140 squat repetitions to be labeled. Three categories were created from the labeled samples: samples which were unanimous (U) among raters, samples with a split (S) decision among raters, where two raters gave the same score and one gave a different score, and samples for which there was no consensus among raters, where each rater gave a different score. Labeled data statistics for each of the two criteria are summarized in Table I. Samples that came from participants who scored less than 95% on the IKDC are referred to as unhealthy in the table and are excluded during training; these data are used for final validation only.

For split decision ratings, a final label based on majority vote amongst the clinician raters was given to the samples. For feature selection and classification, 4 different datasets were generated: two with combinations of both healthy unanimous and healthy split decision samples (for the two different criteria) to be used for 3-class classification; 2-class classification datasets were generated by removing "moderate" exemplars from the previous sets. Unanimous only datasets were also generated, however, when split decision samples were removed, both the dataset size and the number of different participants contributing to each label were significantly reduced. For this reason, only the general datasets including both unanimous and split decision examples are used for analysis. Details of the training datasets are summarized in Table 2.

TABLE 1. **Labeled data information.**

U:unanimous S: split decision H: healthy	Labeled with knee valgus criterion		Labeled with overall risk of knee injury criterion	
	Male #	Female #	Male #	Female #
Good (U,H)	7	5	1	5
Good (S,H)	11	16	7	8
Moderate (U,H)	10	5	9	1
Moderate (S,H)	18	16	18	15
Poor (U,H)	6	4	5	14
Poor (S,H)	11	10	22	12
No-consensus (H)	2	4	3	5
Unhealthy	5	10	5	10
Total	70	70	70	70

Unhealthy samples came from participants who scored less than 95% on the IKDC.

TABLE 2. **Training dataset details.**

		Labeled with Knee Valgus criterion	Labeled with overall Risk of Injury criterion
Training and validation sets	Healthy – Unanimous or Split	119 exemplars (39 "good", 49 "moderate", 31 "poor")	117 exemplars (21 "good", 43 "moderate", 53 "poor")
Removed samples	Unhealthy and no-consensus	21 exemplars	23 exemplars

C. INTER AND INTRA-RATER RELIABILITY (IRR)

Since we have three raters in this study, the degree of agreement (inter-rater reliability), as well as consistency of the ratings by each of the raters (intra-rater reliability) have to be assessed.

IRR assessment was performed using the two-way mixed, consistency, average-measures ICC test [30]. Calculations were made using the *irr* package in R. The resulting ICC value was 0.80 for the knee valgus criterion and 0.84 for the risk of injury criterion. This indicates excellent agreement between raters according to CiCChetti guidelines [31]. To assess intra-rater reliability, 15 of the 140 squat samples were randomly selected and duplicated in the dataset provided to the raters for labeling. The two-way mixed, consistency, average-measures ICC test was applied to two ratings provided for the original and duplicated samples by each rater. Intra-rater reliability results for the three raters were 1, 0.96, and 0.88 suggesting excellent reliability for all raters. The IRR assessment results suggest that the measurement

TABLE 3. Feature selection results for 2-class problem and knee valgus criterion.

Knee Valgus criterion- 2class ("good" vs "poor") Healthy – Unanimous and Split decision data			
10Fold CV	Nr	LOSO CV	Nr
Max of **hip Flex**. angle RMS of **hip Flex**. angle	10/10	Max of **hip Flex**. angle RMS of **hip Flex**. angle	13/13
Mean of **hip Flex**. angle RMS of **hip IR** angle	9/10	Mean of **hip Flex**. angle Range of **hip Flex.** angle	11/13
Range of **hip Flex**. angle	8/10	RMS of **hip IR** angle	10/13
Mean of **knee Flex** angle	7/10		

Nr: Indicates for how many of the validation subsets the feature was selected

TABLE 4. Feature selection results for 2-class problem and injury risk criterion.

Injury Risk criterion- 2-class ("good" vs "poor") Healthy – Unanimous and Split decision data			
10Fold CV	Nr	LOSO CV	Nr
Mean of **hip Flex**. angle RMS of **hip Flex** angle	10/10	Mean of **knee Flex**. angle	12/13
Mean of **knee Flex**. angle	8/10	Mean of **hip Flex**. angle	10/13
MAX of **hip Flex** angle	7/10	RMS of **hip Flex**. angle	9/13

Nr: Indicates for how many of the validation subsets the feature was selected

error introduced by individual raters is minimal and that SLS ratings are suitable for the purpose of classification.

V. RESULTS

A. FEATURE SELECTION

Feature selection was performed together with classification in leave one out fashion. In each fold, feature selection was applied only to the training set and the majority of selected features by 18 methods for the training set was used for dimensionality reduction of the validation set. The same method was adopted for SPCA feature generation; the projection matrix resulting from the training set was utilized for dimensionality reduction of the validation set. Tables 3 to 6 show the feature selection results for the two criteria and for the two different classification problems (2-class versus 3-class). One of the subjects was excluded from analysis due to low IKDC score; therefore, the number of validation subsets are 13 in LOSO cross validation. As can be seen from Tables 3 – 6, the same key features tend to be selected for most of the validation subsets. For classification purposes, subset-specific selected features were used.

The feature selection results for both 2-class and 3-class problems reveal that hip Flex/IR and knee Flex angles are the best predictors of the knee valgus or risk of injury, across folds.

TABLE 5. Feature selection results for 3-class problem and knee valgus criterion.

Knee Valgus criterion- 3-class ("good" vs "moderate" vs "poor") Healthy – Unanimous and Split decision data			
10Fold CV	Nr	LOSO CV	Nr
Mean of **hip Flex**. angle RMS of **hip Flex**. angle MAX of **hip Flex**. angle Range of **hip Flex.** angle	10/10	Mean of **hip Flex**. angle MAX of **hip Flex**. angle	13/13
Min of **knee Flex**. angle	8/10	Range of **hip Flex**. angle	12/13
STD of **hip Flex**. angle	7/10	RMS of **hip Flex**. angle	11/13

Nr: Indicates for how many of the validation subsets the feature was selected

TABLE 6. Feature selection results for 3-class problem and injury risk criterion

Injury Risk criterion- 3-class ("good" vs "moderate" vs "poor") Healthy – Unanimous and Split decision data			
10Fold CV	Nr	LOSO CV	Nr
Mean of **hip Flex**. angle	10/10	Mean of **hip Flex**. angle	13/13
Mean of **knee Flex**. angle Range of **hip Flex**. angle Max of **hip Flex**. angle	8/10	Max of **hip Flex**. angle	12/13
		Range of **hip Flex**. angle	10/13
		Mean of **knee Flex**. angle	9/13

Nr: Indicates for how many of the validation subsets the feature was selected

TABLE 7. Classification results for 2-class problem and knee valgus criterion.

2-class problem accuracy (%) **Knee Valgus Criterion**- Unanimous+split						
	10F- CV			LOSO-CV		
Dim. Red. Method	All	Subset	SPCA	All	Subset	SPCA
SVM	**95.71**	92.86	85.71	71.42	88.57	67.14
KNN	94.28	94.28	74.28	68.57	81.42	70
NB	90	92.85	85.71	77.14	**90**	68.57

B. CLASSIFICATION

Classification results for both 10 fold and LOSO cross-validations are reported in Tables 7 to 10. Classifiers were applied to the full dimensional feature set (same features across all folds), the subset selected feature set and the SPCA extracted feature set (based on features selected/extracted in each fold). Classification results for 10F-CV showed that distinguishing between "good" and "poor" squats is achievable with a promising accuracy of 96%. For the 3- class problem; however, the best achieved accuracy was 66%. LOSO-CV

TABLE 8. Classification results for 2-class problem and risk of injury criterion.

2-class problem accuracy (%) **Risk of Injury Criterion** - Unanimous+split						
	10F- CV			LOSO-CV		
Dim. Red. Method	All	Subset	SPCA	All	Subset	SPCA
SVM	**94.59**	86.48	85.13	67.56	58.11	72.97
KNN	94.59	87.83	77.02	70.27	72.97	64.86
NB	86.48	85.13	85.13	**75.67**	70.27	72.97

TABLE 9. Classification results for 3-class problem and knee valgus criterion.

3-class problem accuracy (%) **Knee Valgus Criterion**- Unanimous+split						
	10F- CV			LOSO-CV		
Dim. Red. Method	All	Subset	SPCA	All	Subset	SPCA
SVM	47.90	**66.38**	47.05	41.17	**59.66**	29.41
KNN	64.70	55.46	42.01	57.98	47.89	43.69
NB	63.02	63.02	52.94	48.74	51.26	36.13

TABLE 10. Classification results for 3-class problem and risk of injury criterion.

3-class problem accuracy (%) **Risk of Injury Criterion** - Unanimous+split						
	10F- CV			LOSO-CV		
Dim. Red. Method	All	Subset	SPCA	All	Subset	SPCA
SVM	58.11	63.24	59.82	45.3	**58.97**	54.70
KNN	**71.79**	61.53	53.84	55.56	46.15	48.71
NB	69.23	59.82	58.11	58.12	50.42	47.86

results were lower, with the best accuracy of 90% for 2-class and 60% for the 3-class problems. With respect to predicting the risk of injury, the best achieved accuracy using 10F-CV was 95% for the 2-class and 72% for the 3-class problem. Using the LOSO-CV, the best accuracy for 2-class was 76% and for the 3-class problem was 59%.

C. UNHEALTHY TEST SET

The developed assessment method was trained using healthy participant data. However, the final system should work for both healthy subjects as well as subjects with active injury, whose movement characteristics may be different from healthy participants. To verify if the developed classifier for healthy subjects is also suitable for patients and also to test the performance of the assessment method on an unseen dataset; the best performing classifiers in Tables 9 and 10 were applied to the 15 squats that were set aside due to low IKDC score. 10 of these squats came from a female subject (including 1 poor and 9 moderate, same labels for both criteria) and 5 from a male subject (all poor, same labels for both criteria). The commonly selected features reported in Tables 5 and 6 were used for dimensionality reduction of the test set and the performance of the best performing classifiers on this test set is reported in Table 11.

TABLE 11. Prediction results for unhealthy test set.

	Accuracy %	
Best developed classifier	**10F-CV**	**LOSO-CV**
Knee Valgus Subject1: 5 poor Subject2: 1 good, 9 moderate	SVM-subset features Subject1:**100** Subject2:**80**	SVM -subset features Subject1:100 Subject2:50
Risk of Injury Subject1: 5 poor Subject2: 1 good, 8 moderate, 1 no-consensus (removed)	KNN-all features Subject1:100 Subject2:22.22	SVM -subset features Subject1:**100** Subject2:**88.88**

The results suggest that poor squats are detectable with very high accuracy but identifying moderate squats is more challenging. The SVM classifier using common subset selected features through LOSO including: Mean, Max, and Range of hip Flex angle (according to Table 6) can predict the Risk of Injury better while the SVM classifier using the subset of features selected through 10F-CV including: Mean, RMS, Max, and Range of hip Flex angle, and Min of knee Flex angle can predict the DKV better.

D. ANKLE ONLY FEATURES

In the pilot data analysis [14], we found the ankle IR features to be the best predictors of the DKV, which led us to suggest that it is possible to use only one sensor on the tibia (saving time and simplifying the test protocol) and still have good classification accuracy. To confirm this hypothesis with the larger datasets, we used feature selection on only ankle extracted features (90 out of 210 features) and found that ankle IR velocity, angle and acceleration, as well as ankle Add velocity features are the best predictors in the absence of hip or knee information. We also repeated the feature selection and classification using ankle only features using the leave one out method. The best achieved results using ankle only features and the percentage of change in accuracy in comparison to the best reported results using all joints' features are shown in Tables 12 and 13.

The results from Tables 12 and 13 indicate that there is less than 5.4% drop in accuracy for risk of injury detection using only ankle information (one tibia sensor), suggesting that one sensor can be used to simplify the data collection procedure, particularly if overall risk of injury is of interest.

IEEE Journal of Translational Engineering in Health and Medicine

TABLE 12. Best achieved classification results for 10F-CV using ankle features.

	10F- CV accuracy (%)			
	Knee Valgus Criterion		**Risk of Injury Criterion**	
Best results	ankle only features	change in accuracy	ankle only features	change in accuracy
2-class	90	-5.74	89.19	-5.4
3-class	63.86	-2.52	73.5	+1.71

TABLE 13. Best achieved classification results for loso-CV using ankle features.

	LOSO- CV accuracy (%)			
	Knee Valgus Criterion		**Risk of Injury Criterion**	
Best results	ankle only features	change in accuracy	ankle only features	change in accuracy
2-class	68.57	-21.43	83.78	+8.11
3-class	49.57	-10.09	57.26	-1.71

E. GENDER SPECIFIC ANALYSIS

We hypothesized that men and women might have different biomechanical characteristics and movement strategies which result in different predictors. To test this hypothesis, we separated the healthy subject data based on gender, resulting in female only data (60 samples) and male only data (60 samples). Unhealthy male and female subjects were set aside. Leave one out feature selection methods were applied to combined unanimous and split decision samples of both datasets separately. The results reported in Tables 14, 15, 16, 17 showed that different features are selected when the data is segregated by gender.

TABLE 14. Gender specific feature selection results for 2-class problem and knee valgus criterion.

Knee Valgus – 2-class (“good” vs “poor”)			
10 F CV			
Males	Nr	**Females**	Nr
Max of **hip Flex**. velocity	9/10	STD of **ankle IR**. velocity	10/10
Mean of **knee Flex**. angle	8/10	RMS of **ankle IR** velocity	9/10
		MAD of **ankle IR** velocity	8/10
LOSO CV			
Mean of **knee Flex.** angle	4/6	RMS of **ankle IR** velocity	5/6
Max of **hip Flex.** angle	3/6	MAD of **ankle IR** velocity	3/6

Nr: Indicates for how many of the validation subsets the feature was selected

For the male dataset, the features selected were the hip and knee flexion features. For females, hip Add/IR and ankle IR features were selected. Based on this finding, we also tested whether male-specific and female-specific classifiers might work better than a general classifier for both genders.

TABLE 15. Gender specific feature selection results for 2-class problem and injury risk criterion.

Injury Risk – 2-class (“good” vs “poor”)			
10F CV			
Males	Nr	**Females**	Nr
MAD of **hip Flex.** velocity	9/10	STD of **hip IR** velocity VAR of **hip IR** velocity	10/10
Kurtosis of **hip Flex.** velocity	8/10		
LOSO CV			
MAD of **hip Flex.** velocity	3/6	STD of **hip IR** velocity	6/6
		VAR of **hip IR** velocity	4/6

Nr: Indicates for how many of the validation subsets the feature was selected

TABLE 16. Gender specific feature selection results for 3-class problem and knee valgus criterion.

Knee Valgus – 3-class (“good” vs “poor” vs “moderate”)			
10F CV			
Males	Nr	**Females**	Nr
RMS of **hip Flex.** angle MAX of **hip Flex.** angle	10/10	RMS of **hip Add.** velocity	9/10
MAD of **hip Flex.** angle Range of **hip Flex.** angle	8/10		
LOSO			
Max of **hip Flex.** angle	5/6	RMS of **hip Add.** velocity	3/6
MAD of **hip Flex.** angle	3/6		

Nr: Indicates for how many of the validation subsets the feature was selected

TABLE 17. Gender specific feature selection results for 3-class problem and injury risk criterion.

Injury Risk – 3-class (“good” vs “poor” vs “moderate”)			
10F CV			
Males	Nr	**Females**	Nr
MAX of **hip Flex**. angle Kurtosis of **knee Flex**. acceleration MAX of **knee Flex**. acceleration	5/10	STD of **hip IR** velocity	10/10
		MAD of **ankle IR** acceleration	8/10
LOSO CV			
RMS of **hip Add**. velocity	3/6	STD of **hip IR** velocity	4/6
		MAD of **ankle IR** velocity	4/6

Nr: Indicates for how many of the validation subsets the feature was selected

The gender-specific classifiers were used for the two data sets and the best results are compared to the general classifier (developed in the previous section) in Tables 18 and 19. With respect to LOSO-CV, due to the small size of the gender-specific datasets, we performed feature selection using LOSO (all 10 squats of the test subject were left out during feature selection) but for classification, we used Leave One Leg Out (LOLO) (12 validation sets instead of 6). In this way, the testing subject remained blind to feature selection while providing sufficient data for the training sets. The best results of the LOLO gender-specific classifier are compared to the best accuracy achieved using the same method for all data in Table 19.

TABLE 18. Gender specific classification results for 10F-CV.

	10F-CV accuracy %			
Classifier type	2-Class		3-Class	
	Valgus	Risk	Valgus	Risk
Male only	94.28	97.14	71.42	67.74
Female only	94.28	100	69.64	78.18
General-best results	95.74	94.59	66.38	71.79

TABLE 19. Gender specific Classification Results for LOLO-CV.

	LOLO-CV accuracy %			
Classifier type	2-Class		3-Class	
	Valgus	Risk	Valgus	Risk
Male only	70	83.33	48.27	54.38
Female only	91.42	97.43	71.42	78.18
General-best results	91.42	81.08	63.86	62.39

Classification results show that for women, in all cases, the female-specific classifier works as well or better than the general classifier. For men, however, the male-specific classifier works no better than the general one using Leave One Out validation, and is comparable to the general one using 10F validation. This may be due to the fact that similar features are selected for the male only dataset (mostly flexion angle features), while a different set of features (IR angle features) are selected for the female only dataset. Further, in our dataset, we observed a larger variability within the squats performed by male participants compared to the female participants, which also may contribute to this finding.

VI. DISCUSSION AND CONCLUSIONS

In this study, we developed an automated assessment method to evaluate SLS quality. Two criteria were used for labeling by expert clinician raters: amount of inward knee movement that occurred during the task (knee valgus) and perceived overall knee injury risk. SLS data from 14 volunteers were collected and two data sets were generated for each index: one included the data with combination of full and partial agreement of all labeled data and the other by removing the moderate samples from the later. 18 feature selection methods were applied to the datasets to find the best predictors of knee valgus and risk of knee injury. The feature selection results suggested hip/knee flexion angle features as the main predictors of both DKV and risk of injury when the samples were not analyzed separately by gender. Hip and ankle IR features were the main predictors for DKV and risk of injury for females with gender-specific analysis.

The unanimous cases represent instances where 100% agreement between all three clinician raters occurred, and likely represent cases where the motion characteristics can be clearly identified. By combining both the unanimous and split decision cases, it is possible that some inaccuracies of labelling the cases may have been introduced. However, split decision cases may also represent borderline cases where a labelling judgment may be difficult. Considering borderline cases are likely to occur in the population, it is important to understand the impact these cases may have on feature selection and classification results.

The identification of knee Flex features for the determination of DKV and risk of injury is consistent with previous investigations that have identified a shallow knee flexion angle during single leg loading to be correlated with dynamic knee valgus loading, ACL injury, and patellofemoral pain syndrome [32]. Knee flexion angles less than 30 degrees have been shown to cause a large strain force on the ACL caused by quadriceps contraction, and shallow knee flexion angles coupled with hip internal rotation may also increase patellofemoral contact forces [33]–[36].

Analyzing the flexion joint angles of the SLS repetitions revealed that those labeled as "good" tended to have increased knee and torso flexion during the motion. Previous research investigating the effect of forward trunk lean on predicted anterior cruciate ligament (ACL) strains that occur during the SLS movement indicate that a more moderate forward trunk lean of approximately 40 degrees can lower ACL strains and increase muscle activation of the hamstring muscles that assist in preventing anterior tibial translation and lower activation of the quadriceps muscles that can increase anterior tibial translation [37]. It is possible that the torso flexion observed in the "good" subjects in the present study may be a result of a movement strategy utilized to minimize internal knee loads and optimize the co-contraction of the quadriceps and hamstring musculature. In the present study, the participants were not instructed to keep their torso upright during the data collection. The fact that hip flexion angle features appeared as best predictors of DKV and risk of injury in the full dataset indicates that other motion behaviors are also associated with knee valgus and knee injury risk, and that different test protocols and instructions can lead to different results. SLS test protocols do vary between studies, with some authors constraining the squat depth, time duration of the SLS, and upper and lower extremity position [8], [38], [39]. Since the purpose of the SLS test is to assess how an individual functions during single leg loading, which is a foundational movement that is encountered in everyday life and athletic instances, we chose not to constrain the rate and depth of the SLS, in attempts to study the subjects' own inherent movement preferences while performing the SLS. Our decision to not implement some of these constraints may have affected feature selection results, which has to be considered in the clinical application of the developed tool.

Three common classification techniques were applied to the datasets. The LOSO-CV results suggest that discriminating of ‘‘poor’’ squats from ‘‘good’’ ones is achievable with promising accuracy of 90%. Changing the problem to multiclass (adding ‘‘moderate’’ squats) drops the accuracy by 30%. Screening participants at high risk of injury from those at no risk can be fulfilled by 75% accuracy and adding mild risk subjects drops accuracy by 16%.

The achieved performance in the 2-class problem is comparable to Whelan *et al.* [13]. We further showed that the classification generalizes to unseen participants and investigate 3-class classification. Unlike Whelan *et al.*, joint angle, velocity, and acceleration features are used, which are clinically interpretable parameters.

To our knowledge, the present study is the first to investigate an automated SLS 3-class classification, which would be beneficial for clinicians, as this would allow a determination of not just the presence or absence of DKV and overall knee injury risk, but it would also provide an assessment of the severity of these parameters. This stratification could assist clinicians in developing interventions that could be tailored to an individual's severity level of DKV and knee injury risk. Our study is also unique in that we developed a classifier for the identification of overall knee injury risk based on the SLS movement test. The successful performance of a SLS requires the precise coordination and control of movement about multiple joints (the trunk, hip, knees and ankles) while simultaneously maintaining balance over a small base of support. Considering variables other than DKV, such as lateral trunk position [32] and control of the non-squatting leg [39], has been shown to influence knee loading during the SLS, we had expert clinical raters judge the entire composite SLS movement to rate individuals on knee injury risk.

The classification results for the unhealthy test set showed promising accuracy of 100% and 88.88% for the unseen subjects in the 3-class setting. However, this test sample is very small, confirmation of these results with a larger population of injured participants is needed.

One trade off to performing 3-class classification was a decrease in classification accuracy compared to the 2-class classifier. A possible contribution to this decrease may be due to potential labelling errors in clinician ratings of borderline cases. While the inter-rater reliability for the raters was excellent (ICC = 0.80 for DKV and 0.84 for knee injury risk), this process was not perfect. Improvements in labelling accuracy could be achieved by involving more clinician raters in the process of determining consensus ratings. In addition, future work could investigate the possible use of a clinical endpoint, such as sustaining a knee injury, as a label. This could be accomplished by conducting a prospective injury surveillance study where researchers can perform SLS assessments at an initial time point (e.g. pre-season) and track subjects over a time period (e.g. over the course of a season) to determine who sustains a knee injury.

Our exploration into a 3-class classification for SLS performance yielded accuracy rates of 66.4% for DKV, and 71.8% for knee injury risk, respectively. Our work provides initial evidence that merits future investigations that refine methods for 3-class SLS classification. Furthermore, we intend that this IMU assessment tool will be used as a quantitative aid to assist health care practitioners, and not to completely replace a clinician. Therefore, while the accuracy for 3-class classification was not as robust as 2-class, the classification will be used in conjunction with the joint kinematics provided by the IMU system, and other clinical information, such as patient history and physical examination findings.

A limitation of the present study includes the small sample size comprised of healthy individuals with no current knee injury. Previous work has suggested that the severity of DKV is more prevalent in samples with knee pathology compared to healthy control subjects [7], [32], [40], [41]. Since the prevalence of pathology in a study sample can influence the accuracy of an assessment tool [42], it is possible that larger clinical samples that have a greater prevalence of moderate to severe DKV and a smaller prevalence of borderline mild DKV cases, may improve classification compared to studying healthy individuals with no history of knee pain. Further clinical evaluation is warranted to determine the classifier's performance in symptomatic samples.

The results of gender specific classifiers suggest that developing separate classifiers improves classification results for females and strengthens our hypothesis about different biomechanical characteristics or movement strategies in males and females. Previous literature investigating gender differences in the SLS movement test has identified that females perform the SLS with more ‘‘valgus collapse’’ which involves more pelvic rotation, hip internal rotation, femoral adduction, knee external rotation and abduction, and ankle pronation compared to males [8], [32], [38], [43]. Females also perform the SLS with less trunk, hip and knee flexion compared to their gender counterparts [8], [32], [38]. As a result of these gender differences, it is not surprising that our gender-specific classifiers for females primarily involved hip and ankle IR/add. features, and the male classifier primarily involved hip and knee flexion features.

The development of an automated SLS knee assessment system has many clinical applications. Current clinical assessment of the SLS involves clinicians visually observing patients conducting the movement and qualitatively rating performance utilizing clinical rating tools [43]–[47]. While qualitative rating is common place, subjective evaluations are not always accurate, and assessor error can impact test validity and reliability [48], [49]. Moreover, qualitative assessment tends to rely primarily on the visual estimation of joint range of motion and limb position, and often neglects the assessment of higher order kinematics (such as velocity and acceleration), since it is difficult for clinicians to assess these parameters without instrumentation. An automated instrumented SLS assessment system can improve the accuracy of SLS clinical assessments, and can provide objective results that can be tracked and monitored over time to guide rehabilitation and determine an individual's response to an intervention. Such a system can also be used to perform large population screenings to identify individuals with DKV and those at risk of knee injury.

In conclusion, a method for the automated assessment of DKV and overall knee injury risk during SLS performance is reported. Classification performance for 2-class and 3-class classifiers are reported with 2-class performing better than 3-class classification. When gender-specific classifiers were created, overall performance of female subjects was improved. An automated SLS assessment system could be used to aid clinicians in screening individuals for DKV, knee injury risk and track recovery during the rehabilitation process.

ACKNOWLEDGMENT

This work was a collaboration with MSK Metrics.

REFERENCES

[1] P. Hattam and A. Smeatham, *Special Tests in Musculoskeletal Examination: An Evidence-Based Guide for Clinicians*. Amsterdam, The Netherlands: Elsevier, 2010.

[2] D. A. Padua, D. R. Bell, and M. A. Clark, "Neuromuscular characteristics of individuals displaying excessive medial knee displacement," *J. Athletic Training*, vol. 47, no. 5, pp. 525–536, 2012.

[3] A. L. Gornitzky, A. Lott, J. L. Yellin, P. D. Fabricant, J. T. Lawrence, and T. J. Ganley, "Sport-specific yearly risk and incidence of anterior cruciate ligament tears in high school athletes: A systematic review and meta-analysis," *Amer. J. Sports Med.*, vol. 44, no. 10, pp. 2716–2723, 2016.

[4] M. V. Paterno, "Incidence and predictors of second anterior cruciate ligament injury after primary reconstruction and return to sport," *J. Athletic Training*, vol. 50, no. 10, pp. 1097–1099, 2015.

[5] H. J. Silvers and B. R. Mandelbaum, "Prevention of anterior cruciate ligament injury in the female athlete," *Brit. J. Sports Med.*, vol. 41, pp. i52–i59, Aug. 2007.

[6] M. Waldén, M. Hägglund, H. Magnusson, and J. Ekstrand, "ACL injuries in men's professional football: A 15-year prospective study on time trends and return-to-play rates reveals only 65% of players still play at the top level 3 years after ACL rupture," *Brit. J. Sports Med.*, vol. 50, no. 12, pp. 744–750, 2016.

[7] L. Herrington, "Knee valgus angle during single leg squat and landing in patellofemoral pain patients and controls," *Knee*, vol. 21, no. 2, pp. 514–517, 2014.

[8] B. L. Zeller, J. L. McCrory, W. Ben Kibler, and T. L. Uhl, "Differences in kinematics and electromyographic activity between men and women during the single-legged squat," *Amer. J. Sports Med.*, vol. 31, no. 3, pp. 449–456, 2003.

[9] N. F. N. Bittencourt, J. M. Ocarino, L. D. M. Mendonça, T. E. Hewett, and S. T. Fonseca, "Foot and hip contributions to high frontal plane knee projection angle in athletes: a classification and regression tree approach," *J. Orthopaedic Sports Phys. Therapy*, vol. 42, no. 12, pp. 996–1004, 2012.

[10] M. R. Stiffler, A. P. Pennuto, M. D. Smith, M. E. Olson, and D. R. Bell, "Range of motion, postural alignment, and LESS score differences of those with and without excessive medial knee displacement," *Clin. J. Sport Med.*, vol. 25, no. 1, pp. 61–66, 2014.

[11] V. Ugalde, C. Brockman, Z. Bailowitz, and C. D. Pollard, "Single leg squat test and its relationship to dynamic knee valgus and injury risk screening," *Phys. Med. Rehabil.*, vol. 7, no. 3, pp. 229–235, 2015.

[12] D. A. Padua, S. W. Marshall, M. C. Boling, C. A. Thigpen, W. E. Garrett, Jr., and A. I. Beutler, "The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study," *Amer. J. Sports Med.*, vol. 37, no. 10, pp. 1996–2002, 2009.

[13] D. Whelan, M. O'Reilly, T. Ward, E. Delahunt, and B. Caulfield, "Evaluating performance of the single leg squat exercise with a single inertial measurement unit," in *Proc. 3rd Workshop ICTs Improving Patients Rehabil. Res. Tech.*, 2015, pp. 144–147.

[14] R. Kianifar, A. Lee, S. Raina, and D. Kulić, "Classification of squat quality with inertial measurement units in the single leg squat mobility test," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2016, pp. 6273–6276.

[15] J. F. S. Lin and D. Kulić, "Human pose recovery using wireless inertial measurement units," *Physiol. Meas.*, vol. 33, no. 12, pp. 2099–2115, 2012.

[16] T. Seel, J. Raisch, and T. Schauer, "IMU-based joint angle measurement for gait analysis," *Sensors*, vol. 14, no. 4, pp. 6891–6909, Jan. 2014.

[17] T. O. Haver. *A Pragmatic Introduction to Signal Processing*. Accessed: Nov. 5, 2017. [Online]. Available: https://terpconnect.umd.edu/ toh/spectrum/IntroToSignalProcessing.pdf

[18] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 3rd Quart., 2013.

[19] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, "Advancing feature selection research," ASU Featur. Sel. Repos., Arizona State Univ., Tempe, AZ, USA, Tech. Rep., 2010, pp. 1–28.

[20] J. Pohjalainen, O. Räsänen, and S. Kadioglu, "Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits," *Comput. Speech Lang.*, vol. 29, no. 1, pp. 145–171, 2015.

[21] E. Barshan, A. Ghodsi, Z. Azimifar, and M. Z. Jahromi, "Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds," *Pattern Recognit.*, vol. 44, no. 7, pp. 1357–1371, 2011.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[23] L. V. Ganyun, C. Haozhong, Z. Haibao, and D. Lixin, "Fault diagnosis of power transformer based on multi-layer SVM classifier," *Electr. Power Syst. Res.*, vol. 74, no. 1, pp. 1–7, 2005.

[24] T. M. Mitchell, *Machine Learning*. New York City, NY, USA: McGraw-Hill, 1997.

[25] E. Keogh. (2006). *Naïve Bayes Classifier*. Accessed: Nov. 5, 2017. [Online]. Available: http://www.cs.ucr.edu/~eamonn/CE/Bayesian%20Classification%20withInsect_examples.pdf

[26] C. Rudin. (2012). K-Nearest Neighbors. MIT OpenCourseWare. Accessed: Nov. 5, 2017. [Online]. Available: https://ocw.mit.edu/courses/sloan-school-of-management/15-097-prediction-machine-learning-and-statistics-spring-2012/lecture-notes/MIT15_097S12_lec06.pdf

[27] A. F. Anderson, J. J. Irrgang, M. Kocher, B. J. Mann, and J. J. Harrast, "The international knee documentation committee subjective knee evaluation form: Normative data," *Amer. J. Sports Med.*, vol. 34, no. 1, pp. 128–135, 2006.

[28] YOST. *Yostlabs Motion Sensors' Website*. Accessed: Nov. 5, 2017. [Online]. Available: https://yostlabs.com/

[29] L. Herrington and A. Munro, "A preliminary investigation to establish the criterion validity of a qualitative scoring system of limb alignment during single leg squat and landing," *J. Exercise, Sports Orthopedics*, vol. 1, no. 2, pp. 1–6, 2014.

[30] K. A. Hallgren, "Computing inter-rater reliability for observational data: An overview and tutorial," *Tutor Quant Methods Psychol.*, vol. 8, no. 1, pp. 23–34, 2012.

[31] D. V. Cicchetti, "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology," *Psychol. Assess.*, vol. 6, no. 4, pp. 284–290, 1994.

[32] K. Weiss and C. Whatman, "Biomechanics associated with patellofemoral pain and ACL injuries in sports," *Sports Med.*, vol. 45, no. 9, pp. 1325–1337, 2015.

[33] Y. Nagano, H. Ida, M. Akai, and T. Fukubayashi, "Gender differences in knee kinematics and muscle activity during single limb drop landing," *Knee*, vol. 14, no. 3, pp. 218–223, 2007.

[34] G. D. Myer *et al.*, "The incidence and potential pathomechanics of patellofemoral pain in female athletes," *Clin. Biomech.*, vol. 25, no. 7, pp. 700–707, 2010.

[35] R. B. Souza and C. M. Powers, "Differences in hip kinematics, muscle strength, and muscle activation between subjects with and without patellofemoral pain," *J. Orthopaedic Sports Phys. Therapy*, vol. 39, no. 1, pp. 12–19, 2009.

[36] G. Hughes, "A review of recent perspectives on biomechanical risk factors associated with anterior cruciate ligament injury," *Res. Sport. Med.*, vol. 22, no. 2, pp. 193–212, 2014.

[37] A. S. Kulas, T. Hortobágyi, and P. Devita, "Trunk position modulates anterior cruciate ligament forces and strains during a single-leg squat," *Clin. Biomech.*, vol. 27, no. 1, pp. 16–21, 2012.

[38] G. Graci, L. R. Van Dillen, and G. B. Salsich, "Gender differences in trunk, pelvis and lower limb kinematics during a single leg squat," *Gait Posture*, vol. 36, no. 3, pp. 461–466, 2013.

[39] A. Khuu, E. Foch, and C. L. Lewis, "Not all single leg squats are equal: A biomechanical comparison of three variations," *Int. J. Sports Phys. Therapy*, vol. 11, no. 2, pp. 201–211, 2016.

[40] P. Levinger, W. Gilleard, and C. Coleman, "Femoral medial deviation angle during a one-leg squat test in individuals with patellofemoral pain syndrome," *Phys. Therapy Sport*, vol. 8, no. 4, pp. 163–168, 2007.

[41] T. H. Nakagawa, T. U. Moriya, C. D. Maciel, and F. V. Serrão, "Trunk, pelvis, hip, and knee kinematics, hip strength, and gluteal muscle activation during a single-leg squat in males and females with and without patellofemoral pain syndrome," *J. Orthopaedic Sports Phys. Therapy*, vol. 42, no. 6, pp. 491–501, 2012.

[42] J. Boissoneault, L. Sevel, J. Letzen, M. Robinson, and R. Staud, "Biomarkers for musculoskeletal pain conditions: Use of brain imaging and machine learning," *Curr. Rheumatol. Rep.*, vol. 19, no. 1, p. 5, 2017.

[43] B. K. Weeks, C. P. Carty, and S. A. Horan, "Effect of sex and fatigue on single leg squat kinematics in healthy young adults," *BMC Musculoskeletal Disorders*, vol. 16, no. 1, p. 271, 2015.

[44] A. Almangoush, L. Herrington, and R. Jones, "A preliminary reliability study of a qualitative scoring system of limb alignment during single leg squat," *Phys. Therapy Rehabil.*, vol. 1, no. 1, pp. 1–7, 2014.

[45] D. R. Poulsen and C. R. James, "Concurrent validity and reliability of clinical evaluation of the single leg squat," *Physiother. Theory Pract.*, vol. 27, no. 8, pp. 586–594, 2011.

[46] E. Ageberg, K. L. Bennell, M. A. Hunt, M. Simic, E. M. Roos, and M. W. Creaby, "Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat," *BMC Musculoskeletal Disorders*, vol. 11, no. 1, p. 265, 2010.

[47] K. M. Crossley, W.-J. Zhang, A. G. Schache, A. Bryant, and S. M. Cowan, "Performance on the single-leg squat task indicates hip abductor muscle function," *Amer. J. Sports Med.*, vol. 39, no. 4, pp. 866–873, 2011.

[48] B. K. Weeks, C. P. Carty, and S. A. Horan, "Kinematic predictors of single-leg squat performance: A comparison of experienced physiotherapists and student physiotherapists," *BMC Musculoskeletal Disorders*, vol. 13, no. 1, p. 207, 2012.

[49] M. A. DiMattia, A. L. Livengood, T. L. Uhl, C. G. Mattacola, and T. R. Malone, "What are the validity of the single-leg-squat test and its relationship to hip-abduction strength?" *J. Sport Rehabil.*, vol. 14, no. 2, pp. 108–123, 2005.

REZVAN KIANIFAR received the B.S. degree in electrical engineering from the K. N. Toosi University of Technology, Tehran, Iran, in 2006, and the M.S. degree in biomedical engineering from the Amirkabir University of Technology (AUT), Tehran, in 2009. She is currently pursuing the M.A.Sc. degree with the Center for Pattern Analysis and Machine Intelligence, University of Waterloo (UW). She was a Researcher with the Biological Systems Modeling Laboratory, AUT, where she developed models for neural mechanisms including neural-glial interactions and human brain prediction mechanism in decision making. She is collaborating with MSK Metrics on human motion analysis as her M.S. research project. She received the NSERC-IPS Award from 2015 to 2017 and the UW Faculty of Engineering Award in 2016.

ALEXANDER LEE received the B.Sc. degree (Hons.) in kinesiology from the University of Waterloo, ON, Canada, and the Doctor of Chiropractic degree from the Canadian Memorial Chiropractic College (CMCC) in 2007. He further specialized in sports sciences at CMCC from 2007 to 2009, and received the Sports Sciences Fellowship from the Royal College of Chiropractic Sports Sciences, Canada, in 2010. He is currently a Faculty Member with the Clinical Education Department, CMCC, also in private practice at Velocity Sports Medicine and Rehabilitation and also a Director of MSK Metrics Inc.

SACHIN RAINA received the B.Kin. from McMaster University and the M.Sc. degree in biomechanics from the University of Waterloo. He was a Researcher with Harvard University and VU Amsterdam, where he studied the effects psychosocial factors on injury risk in the workplace. More recently, he has undertaken teaching appointments in biomechanics and biophysics through the Department of Kinesiology, University of Guelph-Humber. He is currently the Research Coordinator with MSK Metrics Inc.

DANA KULIĆ (M'98) received the combined B.A.Sc. and M.Eng. degree in electromechanical engineering and the Ph.D. degree in mechanical engineering from The University of British Columbia, Vancouver, BC, Canada, in 1998 and 2005, respectively. From 2006 to 2009, she was a Post-Doctoral Fellow of the Japan Society for the Promotion of Science and a Project Assistant Professor with the Nakamura Laboratory, The University of Tokyo, Japan. She is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. Her research interests include human motion analysis, rehabilitation, and human–machine interaction.