Classification of Squat Quality with Inertial Measurement Units in the Single Leg Squat Mobility Test

Rezvan Kianifar[1], Alex Lee[2], Sachin Raina[2] and Dana Kulić[1]

Abstract— **Many assessment and diagnosis protocols in rehabilitation, orthopedic surgery and sports medicine rely on mobility tests like the Single Leg Squat (SLS). In this study, a set of three Inertial Measurement Units (IMUs) were used to estimate the joint pose during SLS and to classify the SLS as poor, moderate or good. An Extended Kalman Filter pose estimation method was used to estimate kinematic joint variables, and time domain features were generated based on these variables. The most important features were then selected and used to train Support Vector Machine (SVM), Linear Multinomial Logistic Regression, and Decision Tree classifiers. The results of feature selection highlight the importance of the ankle internal rotation (IR) angle in classifying SLS. Classification results on a human motion dataset achieved an accuracy of 98% for the two-class problem using SVM, while for 3 class classification, the maximum accuracy was 73% using Decision Tree.**

I. INTRODUCTION

The Single Leg Squat is a mobility test usually used for orthopedic knee surgery assessment, sports medicine, and rehabilitation [1]. During the SLS test, a key indicator of performance is the degree of knee movement out of the sagittal plane. Inward movement of the knee is known as medial knee displacement or Dynamic Knee Valgus (DKV). DKV is a risk factor for non-contact Anterior Cruciate Ligament (ACL) injury and patellofemoral pain [2]. Almost 250,000 ACL injuries occur in the USA yearly with an average annual cost of more than 2 billion dollars [3]. An automated assessment method can help with early detection of DKV among young athletes and with identifying those at higher risk of injury, and assist orthopedic and rehabilitation professionals with patient assessment and provide a record of past performance, leading to better treatment protocols. The goal of this study is to develop an automated assessment system to distinguish between good, moderate, and poor squats.

SLS analysis has been the subject of several clinical and sport medicine studies. Bittencourt et al. [4] investigated the Frontal Plane Knee Projection Angle (FPKPA) during SLS and at the moment of double leg jump landing for 173 young athletes using a motion capture system. Other measurements included the isometric strength of the dominant-limb hip abductors, the passive range of motion (ROM) of the hip internal rotation (IR), and shank-forefoot alignment. These measurements together with participant sex were input into a Decision Tree classifier in order to determine which of these factors predict high FPKPA. High FPKPA was detected by decreased hip abductor torque and increased passive ROM of the hip IR.

Zeller et al. [5] analyzed 18 athletes (9 male, 9 female) performing 5 consecutive SLS while wearing both markers and surface electrodes to measure their hip muscle activity. The ROM was obtained from markers and was analyzed using one-way analysis of variance. Women showed more medial knee displacement in SLS than men, which was associated with more ankle dorsiflexion, ankle pronation, hip adduction, hip flexion, hip external rotation, and less trunk lateral flexion.

While the above studies focused on identifying the correlates of DKV, fewer studies have attempted to automatically classify SLS. Whelan et al. [6] applied a single IMU worn on the lumbar vertebra to classify SLS as correct or incorrect for a dataset of 19 healthy volunteers. Labels were provided by an expert physiotherapist for each repetition. Time domain characteristics of accelerometer and gyroscope outputs, accelerometer magnitude, and sensor orientation were used as descriptive features for the classifier, which resulted in 92.1% accuracy with repeated random-sample validation. Although they showed promising results for the two class problem, they do not report any results for Leave One Subject Out cross validation (LOSO-CV), which is a requirement for clinical applications where previously unseen participants are to be assessed. Moreover, defining features based on direct acceleration and gyroscope output signals makes clinical interpretation and analysis difficult.

II. APPROACH

In this study, an approach for automated SLS classification based on joint kinematics is proposed. First, an Extented Kalman Filter based method [7] is used to estimate ankle, knee, and hip kinematic parameters during SLS from IMU measurements. Time domain features are then extracted from these measurements; the most informative ones are selected via feature selection. Based on an expert labeled dataset, classifiers are then trained to distinguish between good, poor and moderate squats.

A. Pose Estimation

To develop an automated DKV assessment system suitable for clinical use, it is preferable to measure joint angles, as they best describe the occurrence of DKV in clinically interpretable terms. For this purpose, we adapted the pose estimation algorithm proposed in [7] to estimate the joint

[1]R.Kianifar and D. Kulić are with the University of Waterloo, Canada (email:rkianifar@uwaterloo.ca; dana.kulic@uwaterloo.ca)

[2]A. Lee and S. Raina are with MSK Metrics, Canada.

*This work was supported by Canada's Natural Sciences and Engineering Research Council and MSK Metrics

978-1-4577-0220-4/16/$31.00 ©2016 IEEE

angles, velocities and accelerations during SLS using IMUs. A kinematic model of the lower leg, consisting of a 3 Degree of Freedom (DOF) ankle joint, 1 DOF knee joint, and 3 DOF hip joint (as depicted in Fig. 1) (left) and the IMU measurements were fused via an Extended Kalman Filter to recover the joint angle, velocity, and acceleration of each DOF. See [7] for details.

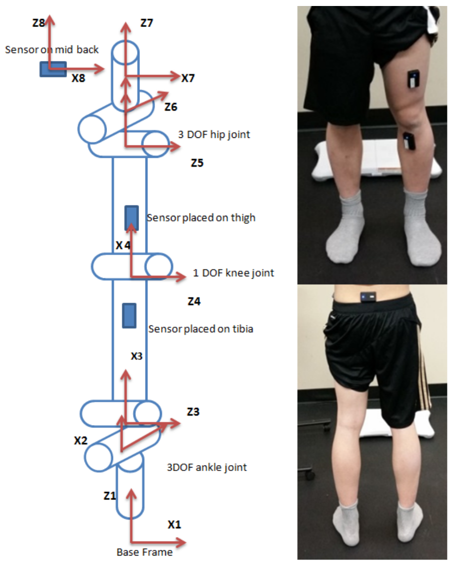

Fig. 1. 7 DOF kinematic model of the left leg (left) and sensor placement (right), The kinematic link lengths were either measured or obtained from the motion capture marker information.

B. Feature Generation

Various feature extraction methods have been used for human activity recognition [8]. The mean, standard deviation (STD), variance (VAR), interquartile range (IQR), mean absolute deviation (MAD), correlation between axes, entropy, and kurtosis are among the time domain features commonly used for activity recognition from the acceleration signal [8]. In a similar review, Preece et al. [9] have identified the mean, median, variance, skewness, kurtosis and interquartile range as commonly used time domain features.

In the present study, we generate all of the commonly used features and use feature selection techniques to identify the best features from the data. The features generated in this study include: the mean, root mean square (RMS), STD, VAR, MAD, skewness, kurtosis, range, minimum, and maximum of the joint angle, velocity and acceleration of each DOF during each repetition of a SLS.

C. Feature Selection

The purpose of feature selection in this study is not only to reduce the dimensionality, but also to identify which factors are best predictors of DKV. Due to the importance of feature selection, 18 different feature selection techniques were tried and those identified by the majority of the methods were chosen as the selected features. To identify the majority, features which were among the top 10 ranked by each algorithm and repeated more than 8 times (selected by at least half of the methods as the top ten) were reported as the best predictors of DKV. In addition to the feature selection methods, SPCA was also applied to the data for comparison.

For feature selection, available MATLAB packages from Arizona State University [10] and from Pohjalainen et al. [11] were used. Pohjalainen's package included five different techniques: Mutual Information, Statistical Dependency, Random Subset Feature Selection, Sequential Forward Selection, and Sequential Floating Forward Selection. The ASU package included 12 techniques: Correlation based Feature Selection, ChiSqaure, Fast Correlation-Based Filter, Fisher Score, Gini Index, Information Gain, Kruskal-Wallis, Minimum-Redundancy-Maximum-Relevance selection, Relief-Feature selection strategy, Sparse Multinomial Logistic Regression via Bayesian L1 Regularization, T-test, and the Bayesian logistic regression. Least Absolute Shrinkage and Selection Operator (LASSO) was also implemented using MATLAB's default function. For SPCA, the MATLAB code developed by Barshan et al. [12] was utilized.

D. Classification

Three different classifiers, the Support Vector Machine, Linear Multinomial Logistic Regression, and Decision Tree, were tried for both the 2 class and 3 class classification problems. All classifiers were implemented using MATLAB R2014b. SVM was selected as it is robust to small training data size. For 3 class classification, one-versus-all and one-versus-one SVM with linear kernel were implemented. SVM multi-label results were computed by majority vote between one-vs-one classification results. The Decision Tree is beneficial as it provides threshold values (cutoff points) in the selected features which can be informative for clinical interpretation.

III. Experiments

Seven participants (6 male, 1 female, mean age 32.3±11.6 years) took part in this study. Inclusion criteria were adults not having any lower back or leg injuries in the past six months. The experiment was approved by the University of Waterloo Research Ethics Board, and all participants signed a consent form prior to the start of data collection.

A. Data Collection

Three Yost [13] IMU sensors were affixed to the participant using hypoallergenic tape. Sensor placement sites included the 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 lower leg on the flat surface of the tibia at the level of the tibial tubercle, as illustrated in Fig.1(right). Due to wireless communication, sampling rates were not consistant or identical for all sensors. The average sampling rate was 90±10 Hz. All sensors were interpolated and resampled to the same rate (100 HZ). Participants were instructed to remove their shoes and socks, and stand on their dominant leg (the leg they would kick a ball with) with toes pointing straight ahead, while keeping their weight centered over the ball of the foot and their arms crossed in front of their body. In each trial, participants performed five consecutive cycles of the SLS movement. For the SLS collection to be deemed successful, the subject had to perform the squat without allowing the legs to contact each other, and without losing balance (ie. without having the non-weight bearing leg touch the ground).

B. Data Labeling

Three of the participants replicated good, poor, and moderate squats under the instruction and supervision of an expert clinician; the other participants performed the squats naturally. The naturally performed squats were labelled by an experienced movement scientist using a modified qualitative SLS clinical rating tool [14]. A SLS was rated ”good” if DKV did not occurr during the squat or if DKV occured, the patella did not have a trajectory that pointed towards the second toe; ”moderate” if the patella pointed toward or past the second toe, but did not point past the inside aspect of the foot; and ”poor” if the patella pointed past the inside aspect of the foot. To ensure a balanced dataset, we made use of all the natural squats (which were mostly bad or moderate) and supplemented with the replicated exemplars.

The number of trials was not the same for all participants. There were 7 labeled trials available from participant 2 (3 good, 1 moderate and 3 poor), 6 from participant 1 and participant 3 (1 good, 1 poor, and 1 moderate for each), 1 from participant 4 (poor), 2 from participant 7 (moderate), 3 from participant 5 (2 poor and 1 good) and 1 from participant 6 (moderate). Each trial consisted of 5 consecutive squats, which resulted in 100 examples of SLS including 30 examples of good, 30 examples of moderate, and 40 examples of poor squats.

Given the 7 DOF kinematic model, where each DOF includes an estimate of its position, velocity, and acceleration, the total number of features for each segment or observation was 210. Therefore, our final data set had 100×210 dimensions. Another dataset was also produced with the same features, but including only good and poor data (i.e., excluding the moderate SLS data) which had 70 observations. All data was normalized to bring values in [0 1] range. Zero velocity crossing criteria [15] were used to segment continuous time series data into five squats.

IV. Results and Discussion

The feature selection results are summarized in Table I. The feature selection results highlight the importance of the ankle IR angle features for differentiating good, moderate and poor squats. Although according to clinical studies [4], [5], the hip plays an important role in DKV, the feature selection results in this study suggest that good classification can be performed based on only the ankle kinematics. Possible explanations for the finding that the hip data is not as informative as the ankle for classification include the large variability in hip joint movements between different subjects (independent of squat quality) or a larger error in the pose estimation for the hip parameters. Further analysis with a larger dataset and a larger number of participants is needed to confirm this finding.

TABLE I

FEATURES RANKED AS TOP TEN BY MORE THAN 8 FEATURE SELECTION TECHNIQUES.

Selected features For 2 class problem	N_r	Selected features For 3 class problem	N_r
ROM of ankle IR	14	STD of ankle IR angle	13
STD of ankle IR angle	11	VAR of ankle IR angle	13
MAD of ankle IR angle	11	MAD of ankle IR angle	13
VAR of ankle IR angle	10	ROM of ankle IR	12
RMS of ankle IR vel.	9		
MAD of ankle IR vel.	9		
RMS of ankle adduc. acc.	9		

N_r: Number of times ranked as top ten features

Classification results for 2 class and 3 class problems are reported in Table II for both 10 fold CV and LOSO cross validations. For reporting the accuracy, the number of selected features or Principal Components (PCs) in SPCA was set to one first and accuracy was calculated. Then, the number of features or PCs was increased one by one up to the point that further increases did not improve performance. The reported accuracies are the best performance each classifier achieved. Matrix inversion with the full dimensional dataset was not possible with LMLR; therefore no results are reported for this condition.

Analysis of the decision tree results using majority selected features shows that for both LOSO and 10 fold CV, the best performance was achieved using only the ankle IR ROM feature for the 2 class problem, while for the three class problem, ROM and MAD of the ankle IR angle resulted in best accuracy for LOSO CV and STD and MAD of the ankle IR angle for 10 fold CV. The decision tree structure for the 2 and 3 class problems is shown in Fig. 2.

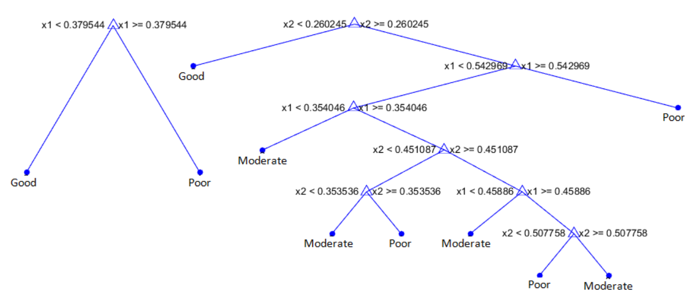

Fig. 2. Decision Tree structure for LOSO cross validation for the 3 class (right) and 2 class (left) problems, where x1 corresponds to ankle IR ROM and x2 corresponds to MAD of ankle IR angle. For the 2-class problem, poor squats are detected when ankle IR ROM (x1) is greater than 0.38 rad (20.63°). For the 3-class problem, MAD (x2) of ankle IR angle greater than 0.26 rad (14.9°) identifies good squats. MAD of ankle IR angle less than 14.9° indicates either moderate or poor squats, which are again differentiated based on ankle IR ROM (x1).

The classification results in Table II show very good accuracy for almost all of the classifiers in the 2 class classification problem, which suggests that differentiation

TABLE II

ACCURACIES (%) FOR THE 2-CLASS AND 3-CLASS CLASSIFICATION PROBLEMS USING THREE CLASSIFIERS AND TWO DIFFERENT CROSS-VALIDATION METHODS

# of Classes	Validation method	10 Fold CV			LOSO CV		
	Dimensionality reduction method	**Majority Selected features**	**SPCA**	**No reduction**	**Majority Selected features**	**SPCA**	**No reduction**
2 class	SVM	95.7143	98.5714	**99.7143**	88.5714	**98.5714**	75.7143
	Logistic Regression	93.5714	98.5714	—	91.4286	**98.5714**	—
	Decision Tree	92.4286	98.5714	95.8571	87.1429	**98.5714**	81.4286
3 class	SVM good vs all	91.8	84.7	98.1	91	83	79
	SVM poor vs all	77.8	87.4	96.9	75	82	72
	SVM moderate vs all	64.8	77.5	87.2	62	74	42
	SVM good vs moderate	91.6667	100	99	91.6667	73.3333	62.3333
	SVM poor vs moderate	70.1429	90	96.7143	67.1429	80	50
	SVM good vs poor	93.7143	97.5714	99.4286	91.4286	97.1429	75.7143
	SVM majority vote	74.2	93.2	**96.6**	72	70	46.4
	Logistic Regression	73.6	93.1	—	68	68	—
	Decision Tree	70.2	83.5	77.6	62	**73**	68

between good and poor squats is achievable. For 10 fold CV, the best performance was obtained with SVM using the full dimensional data. SPCA resulted in the best performance for all three classifiers in LOSO CV, indicating that the 10-fold CV results using all the features may be overfitted. With regard to dimensionality reduction, SPCA in combination with all three classifiers resulted in better accuracy than subset selection methods; however, features extracted by SPCA are difficult to interpret clinically. For the three class problem, again SVM using the full dimensional data outperformed other classifiers in 10 fold CV. However, for the LOSO cross validation, the combination of Decision Tree and SPCA (first four PCs) resulted in the best accuracy. As expected, classification of the moderate squat is most difficult, showing the lowest accuracy in the one-vs-all and moderate-vs-poor SVM results.

V. DISSCUSSION AND CONCLUSIONS

For the dataset in this study, good and poor squats of an unseen subject were classified with 98.6% accuracy using SVM and SPCA for dimensionality reduction. In the 3 class case, 73% accuracy was achieved with a decision tree and SVM. There was no significant difference in classification performance between subjects who performed natural squats versus those who replicated good, poor and moderate squats, suggesting that replicated movements were similar to natural movements. Feature selection results emphasized the ankle internal rotation joint angle features for determining squats quality, suggesting that it may be possible to achieve good classification of the SLS by using only a simple 3 DOF model to estimate ankle joint kinematics. This is advantageous, as it simplifies the pose estimation and reduces the number of sensors from 3 to 1, reducing the complexity of the measurement apparatus and the setup and computation procedure. Similar clinical studies [4] used time consuming manual measurements and focused on only the feature selection part, while the proposed method in this study is completely automated and simple to apply, and therefore more easy to apply in the clinical setting. For future work, the proposed approach will be implemented and tested with a larger dataset of natural squats labeled by an expert.

REFERENCES

[1] R. Bailey, J. Selfe, and J. Richards, "The single leg squat test in the assessment of musculoskeletal function: a review," *Physiother. Pract. Res.*, vol. 32, no. 2, pp. 18–23, 2011.

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

[3] H. J. Silvers and B. R. Mandelbaum, "Prevention of anterior cruciate ligament injury in the female athlete," *Br. J. Sports Med.*, vol. 41, no. suppl 1, pp. i52–i59, 2007.

[4] N. F. Bittencourt, J. M. Ocarino, L. D. Mendonça *et al.*, "Foot and hip contributions to high frontal plane knee projection angle in athletes: a classification and regression tree approach," *J. Orthop. Sport Phys.*, vol. 42, no. 12, pp. 996–1004, 2012.

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

[6] D. Whelan, M. O'Reilly, T. Ward *et al.*, "Evaluating performance of the single leg squat exercise with a single inertial measurement unit," in *Workshop on ICTs for improving Patients Rehabilitation Research Techniques*. ACM, 2015, pp. 144–147.

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

[8] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 3, pp. 1192–1209, 2013.

[9] S. J. Preece, J. Y. Goulermas, L. P. Kenney *et al.*, "Activity identification using body-mounted sensorsa review of classification techniques," *Physiol. Meas.*, vol. 30, no. 4, p. R1, 2009.

[10] Z. Zhao, F. Morstatter, S. Sharma *et al.*, "Advancing feature selection research," *ASU feature selection repository*, pp. 1–28, 2010.

[11] 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.

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

[13] YOST LAB. [Online]. Available: https://www.yostlabs.com/

[14] A. Almangoush, L. Herrington, and R. Jones, "A preliminary reliability study of a qualitative scoring system of limb alignment during single leg squat," *Physical Therapy and Rehabilitation*, vol. 1, no. 1, p. 2, 2014.

[15] J. F.-S. Lin, M. Karg, and D. Kulic, "Movement primitive segmentation for human motion modeling: A framework for analysis," *IEEE Trans. Hum.-Mach. Syst.*, 2015.