Quantifying Y Balance Test performance with multiple and single inertial sensors

William Johnston, James Davenport, Rachelle Connelly & Brian Caulfield *Member, IEEE*

***Abstract*— A growing body of evidence has highlighted that inertial sensor data can increase the sensitivity and clinical utility of the Y Balance Test, a commonly used clinical dynamic balance assessment. While early work has demonstrated the value of a single lumbar worn inertial sensor in quantifying dynamic balance control, no research has investigated if alternative (shank) or combined (lumbar and shank) sensor mounting locations may improve the assessments discriminant capabilities. Determining the optimal sensor set-up is crucial to ensuring minimal cost and maximal utility for clinical users The aim of this cross-sectional study was to investigate if single or multiple inertial sensors, mounted on the lumbar spine and/or shank could differentiate young (18-40 years [n = 41]) and middle-aged (40-65 years [n = 42]) adults, based on dynamic balance performance. Random-forest classification highlighted that a single lumbar sensor could classify age-related differences in performance with an accuracy of 79% (sensitivity = 81%; specificity = 78%). The amalgamation of shank and lumbar data did not significantly improve the classification performance (accuracy = 73-77%; sensitivity = 71-76%; specificity = 73-78%). Jerk magnitude root-mean-square consistently demonstrated predictor importance across the three reach directions: posteromedial (rank 1), anterior (rank 3) and posterolateral (rank 6).**

***Clinical Relevance*— This study showed the capacity of the Y Balance Test, instrumented with a single inertial-sensor, to discriminate healthy adults based on age-related differences in neuromotor control. This research particularly highlighting the potential clinical value of the jerk magnitude root-mean square feature, derived from the lumbar spine mounted sensor.**

I. Introduction

Advances in mobile sensing technology has led to the development of digital health technologies capable of quantifying balance and mobility within the clinical setting [1]. One such tool which has recently been developed is the inertial-sensor instrumented Y Balance Test (YBT). Recent research has highlighted that this digital health technology can provide a valid [2, 3] and reliable [4] measure of dynamic movement control. Furthermore, initial clinical validation within athletic populations has demonstrated that this approach is more sensitive to change than the traditional method of YBT scoring [2], can identify individuals at an increased risk of concussion [4] and may have value in injury recovery tracking [5, 6]. However, in order to ensure accessibility and clinical utility, it is important to consider the optimal sensor set-up to limit cost and maximize value for clinical users.

The sensor instrumented YBT research to date has focused on the use of a single lumbar mounted inertial sensor. While the lumbar sensor has demonstrated clinical value, it is possible that data derived from alternative or additional body locations (such as the shank of the stance leg) may prove more valuable when quantifying neuromotor control. For example, while the lumbar mounted sensor can provide a measure of performance based on the movement of the approximation of the bodies center of mass [7, 8], a shank mounted sensor may quantify performance by measuring movement occurring locally at the ankle [9]. Furthermore, combining data from multiple sensors mounted on different body parts may provide a more comprehensive measure of movement control during the YBT. Such an approach has previously been demonstrated by O'Reilly et al [10] who highlighted that lunge exercise technique classification could be performed with 90% accuracy (sensitivity = 80%; specificity = 92%) using five inertial sensors, while a single thigh-worn sensor achieved an accuracy of 82% (sensitivity = 78%; specificity = 83%). Furthermore, a recent systematic review recommended that researchers determine the optimal inertial sensor set-up for human movement analysis by systematically determining the cost-benefit of each sensor included in a system [11].

Therefore, the focus of this study is to investigate the role single or multiple inertial-sensor combinations may play in the quantification of dynamic balance performance, during the YBT. This will play a role in determining the most cost-effective and appropriate sensor set-up for clinical use-cases. As a large body of evidence has shown that the natural aging process results in the degradation of neuromotor function, leading to impairments in balance and mobility [12-14], we will investigate the capacity of the assessment to quantify age-related differences in performance. Establishing the optimal sensor set-up within this context will lay the groundwork for the development of future digital health technologies for use within older-adults and neurological populations.

Specifically, we aim to investigate which sensor mounting location (lumbar or shank) or combination of mounting locations (lumbar and shank) provides the greatest accuracy in discriminating young (18-40 years [n = 41]) and middle-aged (40-65 years [n = 42]) adults, based on dynamic balance performance. A secondary aim is to investigate the importance of the inertial sensor-based features in quantifying age-related differences in neuromotor control.

II. Methods

A. Participants

Two cohorts, healthy young (18-30 years) and middle-aged (40-65 years) adults, were recruited from the wider

*Research supported by the Science Foundation of Ireland (12/RC/2289_P2). W Johnston, J Davenport, R Connelly and B Caulfield are with the Insight Centre for Data Analytics, University College Dublin, Dublin 4, Ireland; phone +353 1 7166515; e-mail William.Johnston@ucdconnect.ie.

university community. Individuals were eligible to participate if they provided informed consent, met the age requirements and did not meet any exclusion criteria. Participants were excluded if they were in current receival of medical care for an illness, had a history of a neurological disorder, any major orthopedic surgery or fracture in the past twelve months, currently used any drugs that interfere with motor function or answered yes to any questions on the Physical Activity Readiness Questionnaire (PAR-Q). Ethical approval was sought and obtained from the University College Dublin Research Ethics Committee.

B. Experimental Procedure

Participants completed the PAR-Q questionnaire and provided their age. Following this, a measurement of height, weight and leg length were obtained. Three inertial sensors (Shimmer3, Dublin, Ireland) were secured to each participant; one at the level of the 4th lumbar vertebrae (in line with the superior aspect of the iliac crest) and one on each lateral shank, 10cm superior to the lateral malleolus. Each sensor was attached using a neoprene strap and secured using zinc oxide tape. The sensors were calibrated and configured to stream tri-axial accelerometer (±2g), gyroscope (±500°/s) and magnetometer (1.9 Gauss) data at a frequency of 51.2 Hz to an android tablet (Galaxy Note, Samsung, South Korea), operating a custom developed mobile application.

The YBT requires an individual to stand on one leg, place their hands on their hips, and slide a block as far as possible using the contralateral limb in three defined directions (anterior [ANT], posteromedial [PM] and posterolateral [PL]), prior to returning to bilateral stance (Fig. 1). Prior to completing the recorded trials, all individuals completed four practice trials in each reach direction to eliminate the training effect [15]. Following the practice trials, participants completed three recorded reach excursions (randomized order) in each direction, bilaterally. During each reach excursion, inertial sensor data was recorded for the duration the individual was in unilateral stance. The maximal reach distance for each excursion was recorded. A fail was recorded if the participant used the block for support, raised the stance heel from the platform, made ground contact, kicked the block forward for extra distance or removed one or both hands from their hips during the task. If a participant met one of the fail criteria, the individual repetition was repeated.

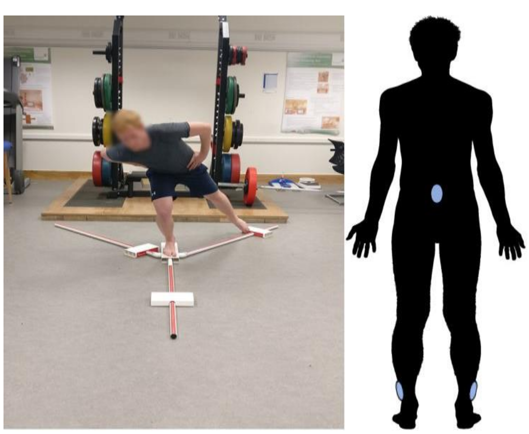

Fig. 1 : A participant completing the posteromedial reach direction of the Y Balance Test (left) and the sensor mounting location (right)

Data Processing

Only data obtained during the reach excursions performed on the right stance leg were used in this analysis. This was completed to reduce the number of random-forest models, as previous research has shown there is no significant difference in YBT performance between dominant and non-dominant sides in healthy adults [16]. Initially, tri-axial jerk signals were computed by obtaining the time-derivative of the tri-axial accelerometer (Accel) signals. The magnitudes (Mag) of the tri-axial jerk, Accel and gyroscope (Gyro) signals were then obtained by computing the vector Mag of the x, y and z signals. This resulted in the availability of three signals from each inertial sensor for feature extraction: Jerk Mag, Accel Mag and Gyro Mag. To ensure the data analyzed applied to the individual's movement, the signals were filtered using a 1st order high-pass filter with a frequency cut-off of 1Hz and a 1st order low-pass filter with a cut-off frequency of 20Hz. A series of descriptive features previously used to quantify YBT performance were obtained [17]. Accel and Gyro Mag features consisted of 'root-mean-square (RMS)', 'sample entropy (SEn)', 'variance' and the 'area under the fast-Fourier transform curve (AUC FFT)'. Additionally, the RMS of the jerk magnitude (Jerk Mag) signal was computed. Detailed descriptions of the methods used to describe these features have been reported elsewhere [17]. The mean of the three repeated YBT direction repetitions were obtained to ensure measurement reliability, in line with previously published reports [2, 4, 5]. This resulted in a total of nine features obtained from each sensor (lumbar and shank) describing the performance in each reach direction.

D. Statistical Methods

The random-forest machine learning method was used to perform the classification task [18]. Each observation was the mean of the three repetitions completed in each YBT direction, to ensure measurement reliability. Five-hundred decision trees were used in each random-forest classifier. Models were developed and evaluated for each of the three YBT reach directions (ANT, PM and PL) from the lumbar, shank and a combination of the lumbar and shank sensor features. Three additional models were developed and evaluated, amalgamating the features obtained from the three YBT reach directions. This resulted in the development and evaluation of a total of 12 random-forest classification models. To reduce model bias due to class imbalance, one participant was held out from the middle-aged cohort.

The quality of the classification model were evaluated using leave-one-subject-out-cross-validation (LOSOCV) [19]. Each trial corresponds to one-fold of the cross validation. At each fold, one participants data is held out as test data while the model is trained with the remaining data. The held-out data is then used to evaluate the classifiers capacity to correctly categorize the unseen data. LOSOCV ensures that there is no biasing of the classifier, meaning the test subjects data is completely unseen by the classifier prior to testing. Accuracy, sensitivity and specificity were used to evaluate the prediction capacity of the classification models. Accuracy was computed using the following formula:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Where TP is the number of true positives, TN is the true negatives, FP is the false positives and FN is the false negatives. Sensitivity and specificity were computed using formulas 2 and 3 below:

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

In this study, the classifiers were developed treating the middle-aged group as the 'positive class' and the young-cohort as the 'negative class'. Feature importance was evaluated for the model with the greatest classification accuracy by permutation of the out-of-bag predictor observations, using the Matlab '*oobPermutedPredictorImportance*' function. This method measures how important each predictor variable in the model is at predicting the response. The influence of this predictor increases with the value of this measure All data analysis was performed using Matlab (2018b, Matworks, Natwick, USA).

III. RESULTS

Eighty-nine participants were recruited as part of this study: 41 young adults and 48 middle-aged adults. Of the 89 adults, 83 had full datasets consisting of both lumbar and shank inertial-sensor data and were included in the analysis. The 83 adults consisted of 41 young adults (age: 21.5 years (1.3); height 176.1cm (8.8); weight: 73.5kg (10.8)) and 42 middle-aged adults (age: 51.5 years (5.8); height 170.4cm (19.7); weight: 78.4kg (19.9)).

Table 1 presents the accuracy, sensitivity and specificity for the various classification models. Across the three individual reach directions, the accuracy of the single sensor classification ranged from 71% to 78% for both the lumbar and shank sensors. Combining the lumbar and shank sensor data did not improve the classification, with an accuracy ranging from 73% to 77%, across the three reach directions. When merging the data obtained from the three YBT reach directions, the lumbar sensor demonstrated the greatest accuracy (accuracy = 79%; sensitivity = 81%; specificity = 78%), followed by the shank (accuracy = 74%; sensitivity = 74%; specificity = 75%) and combination of shank and lumbar (accuracy = 74%; sensitivity = 71%; specificity = 78%). Fig.2 presents the confusion matrices of the 12 random-forest classification models. Fig.3 illustrates the out-of-bag error predictor importance analysis for the model with the greatest accuracy (Lumbar sensor - all reach directions). The variables which were most influential at predicting age-group were PM Jerk Mag RMS (rank 1), PM Accel Mag SEn (rank 2), Ant Accel Mag SEn (rank 3), PL Gyro Mag RMS (rank 4), PL Gyro Mag variance (rank 5) and PL Jerk Mag RMS (rank 6).

Table 1: Accuracy, sensitivity and specificity from the LOSOCV random-forest models.

Reach Direction	Sensor Location	Accuracy (%)	Sensitivity (%)	Specificity (%)
Anterior	Lumbar	71	71	70
	Shank	71	71	70
	Lumbar & Shank	73	74	73
Posteromedial	Lumbar	78	81	75
	Shank	70	67	73
	Lumbar & Shank	76	74	78
Posterolateral	Lumbar	71	69	73
	Shank	78	76	80
	Lumbar & Shank	77	76	78
All Directions	Lumbar	79	81	78
	Shank	74	74	75
	Lumbar & Shank	74	71	78

IV. DISCUSSION

The results presented in this paper indicate that data obtained from a single inertial sensor, either mounted on the lumbar spine or lateral shank of the stance leg are capable of classifying young (18-30 years) and middle-aged (40-65 years) adults with an accuracy of 70 to 79%. Importantly, when combining the features obtained from the lumbar and

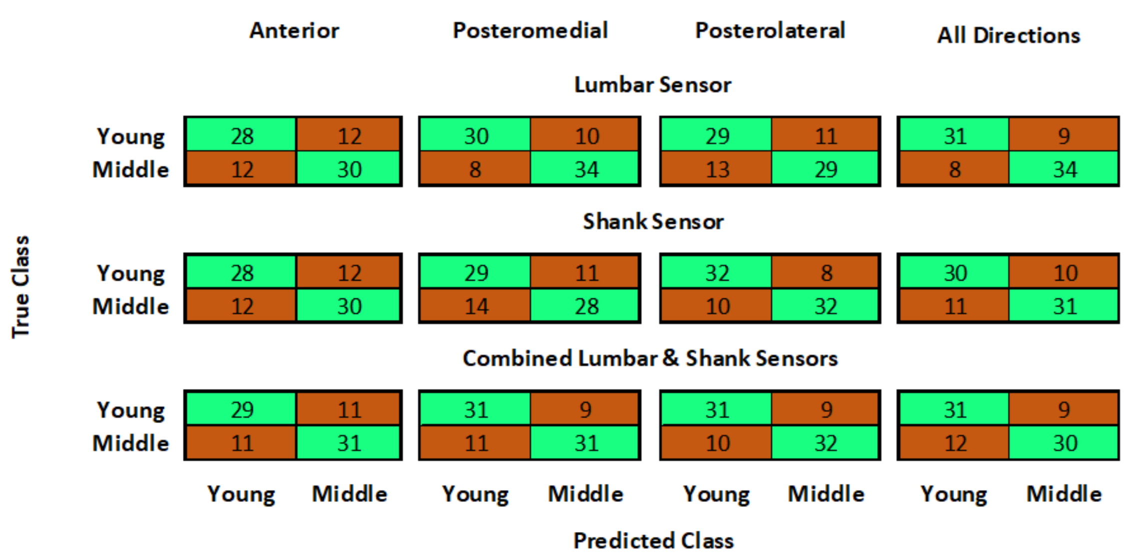

Fig. 2: Confusion matrices presenting the number of individuals correctly (green) and incorrectly (red) classified by each random-forest model. Rows represent the true class and columns represent the predicted class. N = 82 participants were used in the classification models to ensure balance classes.

shank sensors, the classification accuracy ranged from 73% to 77%. The similar classification accuracy observed for the single vs multi-sensor models arise because the features derived from both the lumbar and stance shank are highly likely to correlate [10]. For example, an individual who completes the reach excursion with a 'jerky' and unconstrained reaching strategy would likely possess a high Jerk Mag RMS of both the lumbar and shank sensors. This means that combining the sensor features provides minimal 'new' information to aid in the age-group classification. This is of importance as a minimal sensor set-up is advantageous for use within clinical practice, as it is easier for users to set-up, reducing the risk of sensor placement errors and lowers the associated costs [11]. This study is not the first to demonstrate that the addition of multiple sensors may not significantly improve the accuracy of exercise technique classification models [10]. O'Reilly et al [10] reported that a five sensor set-up resulted in a lunge exercise technique classification accuracy of 90% (sensitivity = 80%; specificity = 92%), while reduction to a three sensor set-up did not significantly reduce classification performance (accuracy 87%; sensitivity = 93%; specificity = 91%).

When merging the feature sets from the three individual YBT reach directions, the lumbar sensor model outperformed all other models, achieving the greatest accuracy of 79% (sensitivity = 81%; specificity = 78%). The improved classification accuracy is likely due to the additional information obtained when combining predictors from the three distinctly different reaching tasks (ANT, PM and PL reach directions). For example, the ANT direction requires an individual to perform a predominantly sagittal plane movement, while the PM and PL require a more complex multiplanar movement [20]. Thus, these three reach direction tasks challenge the sensorimotor subsystem to different extents, serving to sufficiently uncover neuromotor control differences between the two cohorts [21].

The out-of-bag predictor importance analysis helps demonstrate which particular features were most influential in predicting the participant age-group (Fig.3). This analysis highlighted that two to three features from each reach direction were consistently most influential in the classification model. Specifically, Jerk mag RMS ANT (rank 4), PM (rank 1) and PL (rank 7) were consistently important predictors. Interestingly, the PL Gyro Mag RMS (rank 5) and variance (rank 6) features were of greater importance to the classification model than the PL Accel Mag SEn and Jerk Mag RMS feature. The superior importance of the Gyro features in the PL reach direction is likely due to the requirements for the PL reach excursion. The PL reach requires the individual to reach posteriorly behind the stance leg, requiring rotation of the pelvis about the stance leg [20]. As such, the PL lumbar Gyro features, which measures rotational velocity about the lumbar spine, likely provide more valuable information than the linear acceleration information. This analysis contributes to the clinical understanding of what descriptive features may be of value when quantifying age-related differences in YBT performance. The middle-aged adults demonstrated a lower lumbar Jerk Mag RMS during the YBT reach excursions, signifying a more constrained movement strategy. These findings align with recent research which has indicated that individuals who have sustained a recent concussive injury will present with a reduction in lumbar Jerk Mag RMS during the YBT reach excursions [22].

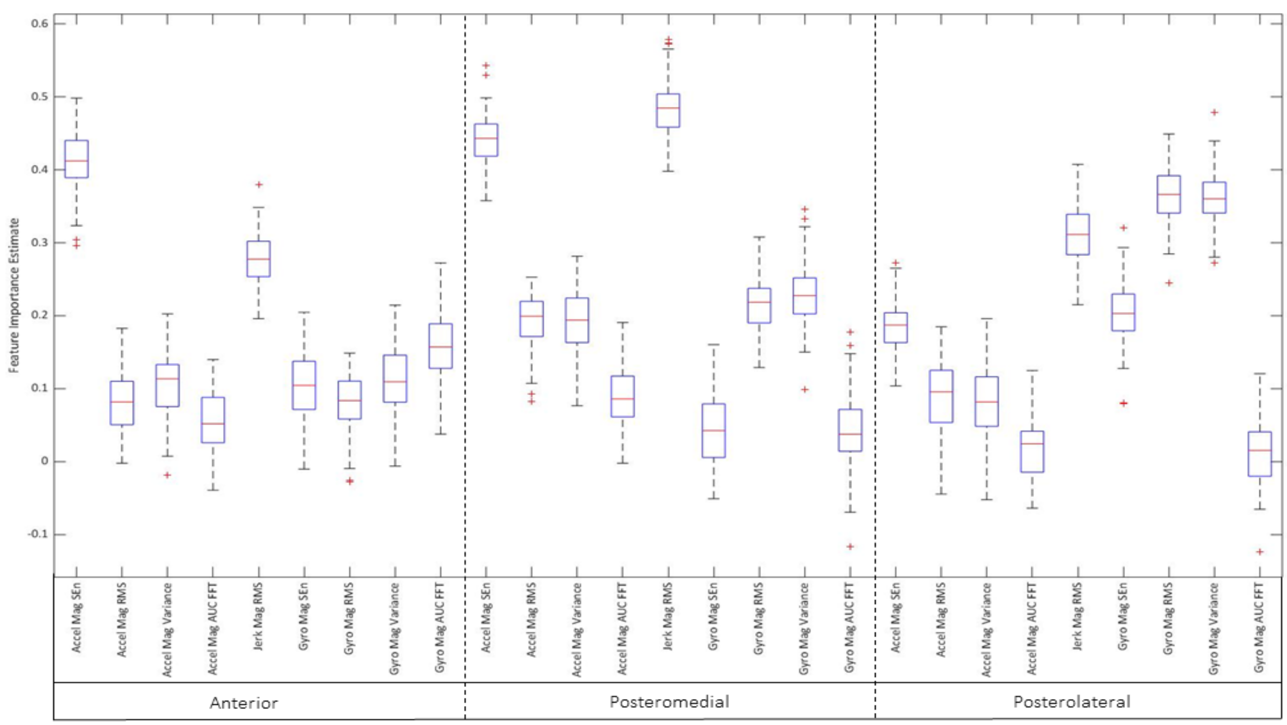

Fig. 3: The out-of-bag prediction error feature importance across the predictor variables for the random-forest model (classifying age-group) consisting of lumbar sensor data from the three YBT reach directions (accuracy = 79%; sensitivity 81%; specificity = 78%).

The results of this study are of importance for clinical practice as they indicate that features such as Jerk Mag RMS, derived from a lumbar inertial sensor, may provide valuable clinical information pertaining to age-related changes in neuromotor function. This information may be of value as a marker of frailty, falls risk and response to interventions in older-adult and neurological populations. While the shank or a combination of the lumbar and shank sensor data did not improve the classification accuracy when differentiating age-groups, it is important to consider that this may be context specific. For example, age-related changes in dynamic balance likely result in a global reduction in neuromotor control which is reflected by changes in movement control throughout the kinetic chain. As such, the movement quantified by the shank and lumbar sensor may be highly correlated, resulting in a negligible difference in classification performance. Conversely, it is possible that conditions such as lateral ankle sprains may result in local changes in neuromotor control at the level of the ankle. Thus, within this context, a shank mounted sensor may provide more valuable information than a lumbar mounted sensor. However, further research is required to further investigate this hypothesis.

There are several contextual factors which should be considered related to this study. Firstly, this study involved a relatively small sample of healthy young and middle-aged adults, not recruiting older or infirm individuals. As such, the results should not be generalized beyond this population. Further research involving a large cohort of adults aged >65 years should be conducted. Secondly, this study was cross sectional in nature and did not longitudinally track changes in balance over time. Thirdly, only the random-forest algorithm was leveraged in this study. It is possible that other classification methods may have outperformed the random-forest models. However, this technique was chosen as it has previously been shown to outperform support vector machine, k-nearest neighbor and Naïve Bayes methods in classifying YBT performance [23].

V. Conclusion

The study highlights that a single inertial sensor, mounted at the lumbar or shank of the stance leg is capable of correctly classifying age-related differences in balance performance with an accuracy of 79% and 78%, respectively. The amalgamation of the lumbar and shank sensor data did not improve classification performance, highlighting the capacity of a single-sensor system for dynamic balance evaluation. Furthermore, this study showed the potential clinical value of the lumbar sensor Jerk Mag RMS feature, highlighting its importance in the quantification of age-related balance performance.

References

[1] B. R. Greene, K. McManus, S. J. Redmond, B. Caulfield, and C. C. Quinn, "Digital assessment of falls risk, frailty, and mobility impairment using wearable sensors," *npj Digital Medicine,* vol. 2, no. 1, p. 125, 2019.

[2] W. Johnston, M. O'Reilly, G. F. Coughlan, and B. Caulfield, "Inertial Sensor Technology Can Capture Changes in Dynamic Balance Control during the Y Balance Test," *Digital Biomarkers,* vol. 1, no. 2, pp. 106-117, 2017.

[3] W. Johnston, M. O'Reilly, K. Dolan, N. Reid, G. Coughlan, and B. Caulfield, "Objective classification of dynamic balance using a single wearable sensor," in *4th International Congress on Sport Sciences Research and Technology Support 2016, Porto, Portugal, 7-9 November 2016*, 2016, pp. 15-24: SCITEPRESS–Science and Technology Publications.

[4] W. Johnston, M. O'Reilly, G. F. Coughlan, and B. Caulfield, "Inter-session test-retest reliability of the quantified Y balance test," presented at the 6th International Congress on Sports Sciences Research and Technology Support, Seville, Spain, 2018.

[5] W. Johnston, M. O'Reilly, M. Liston, R. McLoughlin, G. F. Coughlan, and B. Caulfield, "Capturing concussion related changes in dynamic balance using the Quantified Y Balance Test – a case series of six elite rugby union players," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 2063-2067.

[6] W. Johnston, B. Heiderscheit, J. Sanfilippo, A. Brooks, and B. Caulfield, "Athletes with a concussion history in the last two years have impairments in dynamic balance performance " *Scand J Med Sci Sport,* vol. in Press, 2020.

[7] W. Zijlstra and A. L. Hof, "Assessment of spatio-temporal gait parameters from trunk accelerations during human walking," (in eng), *Gait Posture,* vol. 18, no. 2, pp. 1-10, 2003.

[8] R. Moe-Nilssen, "A new method for evaluating motor control in gait under real-life environmental conditions. Part 1: The instrument," (in eng), *Clin Biomech (Bristol, Avon),* vol. 13, no. 4-5, pp. 320-327, 1998.

[9] D. C. Mackey and S. N. Robinovitch, "Mechanisms underlying age-related differences in ability to recover balance with the ankle strategy," *Gait & Posture,* vol. 23, no. 1, pp. 59-68, 2006.

[10] M. A. O'Reilly, D. F. Whelan, T. E. Ward, E. Delahunt, and B. Caulfield, "Classification of lunge biomechanics with multiple and individual inertial measurement units," (in eng), *Sports Biomech,* vol. 16, no. 3, pp. 342-360, 2017.

[11] M. O'Reilly, B. Caulfield, T. Ward, W. Johnston, and C. Doherty, "Wearable Inertial Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review," (in eng), *Sports Med,* vol. 48, no. 5, pp. 1221-1246, 2018.

[12] A. J. Strang, A. DiDomenico, W. P. Berg, and R. W. McGorry, "Assessment of differenced center of pressure time series improves detection of age-related changes in postural coordination," *Gait & Posture,* vol. 38, no. 2, pp. 345-348, 2013.

[13] S. Thurner, C. Mittermaier, and K. Ehrenberger, "Change of Complexity Patterns in Human Posture during Aging," *Audiology and Neurotology,* vol. 7, no. 4, pp. 240-248, 2002.

[14] R. P. Hubble, G. A. Naughton, P. A. Silburn, and M. H. Cole, "Wearable sensor use for assessing standing balance and walking stability in people with Parkinson's disease: a systematic review," *PloS one,* vol. 10, no. 4, p. e0123705, 2015.

[15] P. A. Gribble, J. Hertel, and P. Plisky, "Using the Star Excursion Balance Test to assess dynamic postural-control deficits and outcomes in lower extremity injury: a literature and systematic review," (in eng), *J Athl Train,* vol. 47, no. 3, pp. 339-57, 2012.

[16] W. Johnston, C. Duignan, G. F. Coughlan, and B. Caulfield, "Dynamic balance performance varies by position but not by age group in elite Rugby Union players – a normative study," *Journal of Sports Sciences,* vol. 37, no. 11, pp. 1308-1313, 2019.

[17] W. Johnston, "Validity, Reliability & Clinical Utility of the Quantified Y Balance Test [dissertation]," PhD, School of Public Health, Physiotherapy & Sport Science, University College Dublin, Dublin (Ireland), 2019.

[18] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, no. 1, pp. 5-32, 2001.

[19] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statistics and Computing,* vol. 21, no. 2, pp. 137-146, 2011.

[20] M.-H. Kang, G.-M. Kim, O.-Y. Kwon, J.-H. Weon, J.-S. Oh, and D.-H. An, "Relationship Between the Kinematics of the Trunk and Lower Extremity and Performance on the Y-Balance Test," *PM&R,* vol. 7, no. 11, pp. 1152-1158, 2015.

[21] W. Johnston, G. F. Coughlan, and B. Caulfield, "Challenging concussed athletes: the future of balance assessment in concussion," *QJM: An International Journal of Medicine,* vol. 110, no. 12, pp. 779-783, 2016.

[22] W. Johnston *et al.*, "Concussion recovery evaluation using the inertial sensor instrumented Y Balance Test," *Journal of Neurotruama,* vol. In Press, 2020.

[23] M. A. O'Reilly, W. Johnston, C. Buckley, D. Whelan, and B. Caulfield, "The influence of feature selection methods on exercise classification with inertial measurement units," in *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2017, pp. 193-196.