THE OBJECTIVE ASSESSMENT OF MOVEMENT QUALITY USING MOTION CAPTURE AND MACHINE LEARNING

by

Gwyneth Butler Ross

A thesis submitted to the School of Human Kinetics
in conformity with the requirements for
the degree of Doctor of Philosophy

School of Human Kinetics
Faculty of Health Sciences
University of Ottawa
Ottawa, Ontario, Canada
(September, 2021)

Supervisor:
Ryan B. Graham, PhD

Contents

Acknowledgements

To my thesis committee, Drs. T. Blaine Hoshizaki and Adrian Chan, thank you for your continual support, feedback and constructive criticism over the past 4 years. You have not only strengthened this thesis but bettered me as a researcher.

To the STOH community, thanks for ensuring that I got my weekly dose of exercise, chirps, laughter, beer, and humble pie. Maybe one day I'll score a goal or finally realize I can't hold my ground against men 60+ pounds heavier than me, but then again probably not.

To the Ultimate team, thanks for bringing me into the fold whole-heartedly, entertaining my 'oldies' music on long car rides, teaching me the way of the game (even though it involves a lot of running), and teaching me the true meaning of 'Never Enough'!

To my student colleagues at the University of Ottawa, thank you for your continuous help and support in and out of the lab. To RDT, from daily euchre and cribbage tournaments to brewing beer to ski trips to spontaneous sleepovers, I couldn't have asked for a better bunch to develop with professionally, emotionally, and especially, socially. To Kenneth/Brent, thanks for sharing my self-deprecating humor; being my personal book, TV, and movie recommender; making me look less selfish; bettering my oral hygiene; and, always taking my calls, years after you left, even though you knew it is just me asking yet another research-related question.

To my friends near and far, thank you for always being there with a phone call, meal, couch to sleep on, tissue, visit, zoom workout, python lesson and so much more. Between Lyme, COVID, and failed studies, these have not been the easiest five years in any sense of the word, so thank you for supporting me through them.

To my family, words will never properly convey the magnitude and breadth of all that each of you do for me or how fortunate I feel to have such a strong (and very large) army behind me. So, to keep it concise, I love you and thank you.

Lastly, to my supervisor, Ryan Graham, thank you for taking a risk on that struggling biomechanics student who needed weekly help all those years ago! I feel so lucky to not only have worked under you, but also beside you, these past 10 years. Your 'play-hard, work hard' mentality, knowledge of the field, drive for continual improvement, and care for each and every one of your students is truly admirable and something I strive to emulate every day. Most importantly, thank you for highlighting my strengths, acknowledging my weaknesses, and always believing in me, even when I didn't believe in myself. I can't wait for the next venture to begin, but hope it brings more physical activity, googling, name reminding, GIFs, bar hopping, OCD PowerPoint tendencies, and of course, ski chalets!

In the words of the great Amy Poehler, "Find a group of people who challenge and inspire you, spend a lot of time with them, and it will change your life." To all of you, thank you for being my people. It's been one hell of a life-changing ride!

General Abstract

Background: Movement screens are frequently used to identify abnormal movement patterns that may increase risk of injury and/or hinder performance. However, abnormal patterns are often detected visually based on the observations of a coach or clinician leading to poor inter- and intra-rater reliability. In addition, they have been criticized for having poor validity and sensitivity. Quantitative, or data-driven methods can increase objectivity, remove issues related to inter-rater reliability and offer the potential to detect new and important features that may not be observable by the human eye. The combination of motion capture data, pattern recognition and machine learning could provide a quantitative method to better assess movement competency.

Purpose: The purpose of this doctoral thesis was to create the foundation for the development of an objective movement screening tool that combines motion capture data, pattern recognition and machine learning. This doctoral thesis is part of a larger project to bring an objective movement screening tool for use in the field to market.

Methods: This thesis is comprised of four studies based on a single data collection and a common series of pre-processing steps. Data from 542 athletes were collected by Motus Global, a for-profit biomechanics company, with athletes ranging in competition level from youth to professional and competing in a wide-range of sports. For the first study of this thesis, an online software program was developed to examine the inter- and intra-reliability of a movement screen, with intra-reliability being further examined to compare reliability when body-shape was and was not modified. The second study developed the objective movement screen framework that utilized motion capture, pattern recognition and machine learning. Study 3 and 4 assessed different types of input data, classification goals (e.g., skill level and sport played), feature reduction and selection methods, and increasingly complex machine learning algorithms.

Results: For Study 1, when looking at inter- and intra-rater reliability of expert assessors during subjective scoring of movements, intra-rater reliability was better than inter-rater reliability. When assessing the effects of body-shape, on average, reliability worsened when body-shape was manipulated. Study 2 provided proof-of-principle that athletes were able to be classified based on skill level using marker-based optical motion capture data, principal component analysis (PCA)

and linear discriminant analysis. For Study 3, PCA in combination with linear classifiers outperformed non-linear classifiers when classifying athletes based on skill level; feature selection increased classification rates, and classification rates when using simulated inertial measurement unit data as the input data were on average better than when using marker-based optical motion capture data. In Study 4, athletes were able to be differentiated based on sport played and recurrent neural nets (RNNs) and PCA in combination with traditional linear classifiers were the optimal machine learning algorithms when classifying athletes based on skill level and sport played.

Conclusion: This thesis demonstrates that objective methods can differentiate athletes based on desired demographics using motion capture, pattern recognition and machine learning. This thesis is part of a larger project to bring an objective movement screening tool for field-use to market and provides a solid foundation to use in the continued development of an objective movement screening tool.

Contributions

Study 1: Assessing inter- and intra-rater reliability of movement scores and the effects of body-shape using a custom visualisation tool *(submitted to British Journal of Sports Medicine)*

The first study of this thesis assessed the inter- and intra-rater reliability of expert assessors subjectively scoring a movement screen. The theoretical conception of the study was a combined effort among all co-authors. The raw data for the animations were collected, labelled, and cleaned by Motus Global, preprocessed by GB Ross, and transformed using sparse motion and shape capture by Dr. Nikolaus Troje. The subjective scoring data collection with the expert assessors was conducted by GB Ross and all data resulting from the data collection were processed and analyzed by GB Ross. The manuscript was written by GB Ross and edited by all co-authors.

Study 2: Objectively differentiating movement patterns between elite and novice athletes *(published in Medicine and Science in Sport and Exercise)*

The second study of this thesis was focused on differentiating elite and novice athletes' movement patterns using PCA and linear discriminant analysis during a movement screen. The data were collected by Motus Global, Inc, a for-profit company specializing in quantitative biomechanical movement screens in sport, where Brittany Dowling was a full-time employee and GB Ross interned in 2013. The theoretical conception of the study was a concerted effort among all co-authors. Cortex labelling was completed by Motus Global and all data cleaning was performed by GB Ross. With respect to the data analysis, GB Ross was responsible for writing all processing scripts with collaboration with Dr. Nikolaus Troje. The manuscript was written by GB Ross and edited by all co-authors.

Study 3: Classifying elite from novice athletes using simulated wearable sensor data *(published in Frontiers in Bioengineering and Biotechnology)*

The third study of this thesis used the previous collected data to test a wider breadth of machine learning algorithms, feature selection methods, and to assess whether simulated inertial measurement unit data were able to better differentiate athlete movement patterns than optical motion capture data. The conception of the study was spearheaded by GB Ross, with input from all collaborators. GB Ross was responsible for all processing scripts and the writing of the manuscript. All co-authors edited the manuscript before submission.

Study 4: Comparison of machine learning classifiers for differentiating level and sport using movement data *(submitted to Journal of Sport Sciences)*

The fourth study of this thesis used the previously collected data to test increasingly complex machine learning algorithms (i.e., algorithms specific for time-series data) and to test whether athletes' movement patterns were able to be differentiated based on sport. Drs. Alistair Boyle and Ryan Graham were responsible for the initial conception of the study design, which was later modified by GB Ross. GB Ross was responsible for the writing of the processing scripts with collaboration from Drs. Alistair Boyle and Allison Clouthier. GB Ross wrote the manuscript with contribution of feedback from all co-authors.

List of Abbreviations

2D	2-dimensional
3D	3-dimensional
ACL	Anterior Cruciate Ligament
ANN	Artificial neural network
ASIS	Anterior superior iliac spine
BD	Bird-dog
BDL	Bird-dog left
BDR	Bird-dog right
BLR	Binary logistic regression
BMI	Body-mass index
cm	Centimetre
CoG	Centre of gravity
CR	Correct rejection
DJ	Drop-jump
DT	Decision tree
EMG	Electromyography
F	Main effect
FA	False alarm
FIFA	Federation Internationale de Football Association
FMS	Functional Movement Screen
GB	Gigabyte
HD	Hop-down
HDL	Hop-down left
HDR	Hop-down right
Hz	Hertz
ICC	Intraclass correlation coefficient
IMU	Inertial measurement unit
kg	kilogram
***k*NN**	*k* nearest neighbors
LDA	Linear discriminant analysis
LESS	The Landing Error Scoring System
LG	Lunge
LH	L-hop
LHL	L-hop left
LHR	L-hop right
LL	Lunge left
LR	Lunge right
LSTM	Long-short term memory
MLB	Major League Baseball
MLR	Multiple logistic regression
mm	Millimetre
MoSh	Motion and shape capture from sparse markers
n	number of
NB	Naïve Bayes
NBA	National Basketball Association
NFL	National Football League
OMAT	Objective movement assessment tool
OPT	Optical motion capture data
PC	Principal component
PCA	Principal component analysis

PCHIP	Piecewise Cubic Hermine Interpolating Polynomial
PSIS	Posterior superior iliac spine
RBF	Radial basis function
RC	Reservoir computing
RMSE	Root mean squared error
RNN	Recurrent neural network
SD	Step-down
SDL	Step-down left
SDR	Step-down right
SDT	Signal Detection Theory
sIMU	Simulated inertial measurement unit data
STD	Standard deviation
SVM	Support vector machine
T2	2nd thoracic vertebrae
T8	8th thoracic vertebrae
TB	T-balance
TBL	T-balance left
TBR	T-balance right

List of Tables

List of Figures

Chapter 1: General Introduction

Movement screens are widely used across many disciplines including in ergonomic, clinical, and athletic settings to identify aberrant movement patterns in hopes of decreasing risk of injury, identifying talent, and/or improving performance (Cook et al., 2014; Donà et al., 2009; Kritz et al., 2009; McCall et al., 2014; McCunn et al., 2016; Padua et al., 2009). Researchers have previously shown that young athletes who perform better on non-sport specific movement batteries are more likely in the future to compete at a more elite level compared to athletes that perform more poorly on the movement batteries (Faber et al., 2016; Mostaert et al., 2016; Pion et al., 2015; Vandorpe et al., 2012). Performance on a non-sport specific movement battery was a better indicator for future sport performance than subjective coach rankings (Vandorpe et al., 2012), anthropometrics (Pion et al., 2015; Vandorpe et al., 2012) or current competition rankings (Vandorpe et al., 2012). Most commonly, during a movement screen, an individual's movement is evaluated based on visual appraisal (McCunn et al., 2016); however, there is agreement within the literature that inter- and intra-rater reliability of these subjective movement screens is poor (Gulgin and Hoogenboom, 2014; Onate et al., 2012; Smith et al., 2013).

A commonly used quantitative movement screen amongst coaches and clinicians is the functional movement screen (FMS; Functional Movement Systems, Chatham, VA, USA). The FMS is comprised of mobility and stability tasks that are designed to highlight possible weaknesses and imbalances by putting the participant into extreme positions (Cook et al., 2014, 2006). The test is a subjective, quantitative movement screen, meaning that the participant is given a numerical score based on subjective observations (Cook et al., 2014, 2006). Limitations of these tests include: 1) the difference between scores needs to be large enough for the human eye to detect; 2) the scores are ordinal and discrete allowing for only a few scores to be available (e.g., FMS only has 4 options, 0-3), which may not be sensitive enough to capture variability between athletes nor dysfunction, 3) individuals are able to increase their scores when aware of the scoring criteria, and 4) task-specific criteria are not linked to any known risk factors (Frost et al., 2015a, 2015b). Motion capture systems are able to detect differences in movements much smaller than can be seen and processed by a human observer and give measures that can provide coaches, clinicians and athletes with quantitative, objective feedback.

In order to give quantitative, objective feedback regarding movement patterns, a motion capture system is needed. Traditionally, motion capture systems comprise of multiple high-speed infra-red cameras and specialized software that track 3D position of reflective markers to describe human motion. However, due to the high cost of both the equipment and software to use these systems in terms of both time and money, recent research has been investigating the use of alternative motion capture systems such as inertial measurement units (IMUs) and markerless motion capture to capture movement in the field (Johnston et al., 2019, 2016; Sgro et al., 2017). IMUs allow for motion capture to be more time- and cost-effective and transportable, making it easier to implement motion capture analyses in non-laboratory settings.

With access to motion capture systems that are more affordable and requiring less time to set-up and post-process data, more data are able to be collected resulting in larger data sets. With the ability to have larger datasets, more advanced analyses involving pattern recognition and machine learning to classify or predict movement are being applied. Various machine learning algorithms have been used in the biomechanics field to predict: the glenohumeral joint rotation centre based on scapular bony landmarks (Meskers et al., 1997); tibial stress fractures in female runners based on tibial shock (Milner et al., 2006); low back pain in patients based on three-dimensional (3D) trunk kinematics (Ashouri et al., 2017); high- and low- mileage runners based on lower limb joint angles (Clermont et al., 2019), different human activities using an IMU (Altun and Barshan, 2010), and lower extremity torques in the sagittal plane during a counter-movement-jump and squat-jump using ground reaction force (Liu et al., 2009), among many other uses. Traditionally, analyzing time-series data has been considered one of the most challenging problems for machine learning as time-series data tends to be complex, high-dimensional, highly-correlated, inherently noisy, and prone to overfitting (Esling and Agon, 2012; Yang and Wu, 2006). Previous research has focused on the development of techniques and classifiers to combat traditional issues with time-series data, for example: using data-reduction techniques such as principal components analysis (PCA) and the development of recurrent neural networks (RNNs).

PCA is a multivariate statistical technique, which reduces high-dimensional datasets by identifying redundancies in the data, where a subset of data will often explain the majority of variance (Troje, 2002a). PCA detects linear combinations of the original data called principal components (PCs); where the first principal component accounts for the largest variance within

the movement and each following component descends in variance under the constraints of the previous components (Abdi and Williams, 2010). Previous research has used PCA to examine differences in gait patterns based on a participant's age, sex, and feelings (happy/sad, nervous/relaxed; Troje, 2002a); to develop an objective judging tool for competitive diving (Young and Reinkensmeyer, 2014); and to develop an objective training tool for analyzing technique for downhill skiing (Federolf et al., 2014). RNNs are a type of neural network where the connections between nodes form a directed graph along a temporal sequence, which allows the network to learn dynamic, temporally-varying behaviour (Hochreiter and Schmidhuber, 1997; Lukoševicius and Jaeger, 2009; Troje, 2002a; Weng and Shen, 2008). RNNs use state variables that allow for past information to be stored, creating a "memory" (Hochreiter and Schmidhuber, 1997). The unit can then use past information along with current input to calculate an output. By reducing the data or through learning dynamic behaviour, PCA and RNNs allow for machine learning to be applied to time-series, kinematic biomechanical data that has not been traditionally possible.

To this end, this doctoral thesis investigated the development of an objective movement screening tool that combines whole-body kinematic biomechanical data with PCA and machine learning. This thesis is a portion of a larger project with the end goal of bringing an objective screening tool to market. The global objectives of this doctoral thesis were:

1. Determine the inter- and intra-rater reliability of a movement screen without task-specific scoring criteria
2. Develop an objective movement screening tool that is able to objectively score an athlete's movement competency using optical motion capture data
3. Refine the object movement screening tool to allow for multiple data types and classifiers to be used and to determine the optimal machine learning algorithm.

The first objective was addressed through Study 1, which looked at the inter- and intra-rater reliability of the movement screen using a custom designed, online visualization tool. The second objective was addressed in Study 2, which developed an objective movement screening framework using optical motion capture data that was able to classify elite and novice athletes as well as calculate a movement competency score to each athlete using PCA and linear discriminant analysis (LDA). To address the third objective, this framework was then refined

in Study 3 and Study 4 to assess the optimal combination of pattern recognition, feature selection, and machine learning algorithms (Study 3 & 4); the framework's ability to classify athletes based on simulated IMU data (Study 3); and its ability to classify on other classes such as sport (Study 4).

Chapter 2: Literature Review

2.1 Movement Screens

Fitness assessments are used to profile and categorize one's physical capabilities and movement competency and is popular amongst applied practitioners such as clinicians, physiotherapists, ergonomists, and strength and conditioning coaches. Fitness tests are typically quantitative in nature and designed to assess an athletes physical and aerobic capabilities, whereas movement screens are subjective in nature and designed to assess one's movement competency by identifying movement patterns that are associated with an increased risk of injury and/or decreased performance (Cook et al., 2014; Donà et al., 2009; Kritz et al., 2009; McCall et al., 2014; McCunn et al., 2016; Padua et al., 2009). There are many different types of movement screens with some being sport-specific, such as the Soccer Injury Movement Screen (SIMS; McCunn et al., 2017), and some being meant for more general use, such as the Functional Movement Screen (FMS; Cook et al., 2014, 2006) and the Movement Competency Screen (MCS; Kritz, 2012). While each assessment has their own unique battery of movements and scoring criteria, the one constant is they are all scored subjectively based on visual appraisal (McCunn et al., 2016).

Among the literature, there is no consensus on what defines movement quality, with each movement screen consisting of its own unique scoring criteria, however, movement quality generally encapsulates performing correct posture and joint alignment while maintaining balance through the movement battery. There are many movement screens available to use, each with their own definition of movement quality, including the: FMS, MCS, SIMS, physical performance measures screen (16-PPM), Foundation Matrix, Landing Error Scoring System (LESS), athletic ability assessment, netball movement screening tool, star excursion balance test movement quality screen and single-leg squat, drop vertical jump, and tuck jump screen variations, with the FMS being the most well-known (McCunn et al., 2016).

The FMS is designed to assess one's fundamental movement patterns in order to highlight movement deficiencies to decrease the risk of injury and provide individualized, specific, fundamental recommendations for physical protocols for athletic and active groups (Cook et al., 2014, 2006). The FMS consists of seven unique movements requiring one to balance mobility and stability and requiring appropriate function of the kinetic linking system (Cook et al., 2014, 2006).

The movements place one into extreme positions to highlight potential poor, compensatory, and/or inefficient movements. The seven movements are the deep squat, hurdle step, in-line lunge, shoulder mobility, active straight-leg raise, trunk stability push-up, and rotary stability with the majority of the movements being performed bilaterally (Cook et al., 2014, 2006). Each movement is scored between zero and three, with three being the highest score. A score of zero is given if there is pain felt in any part of the body during any part of the movement, a one is given if one is not able to complete the movement, a two is given if one is able to complete the movement but utilizes compensatory movements, and a three is given if the movement is performed properly without compensatory movements. Individual scores are then summed to create a composite score with a highest possible score of 21. If the movement is performed bilaterally, the lower of the two scores is used for the composite score (Cook et al., 2014, 2006). A composite score of $\leq$14 is commonly associated in the literature with an increased risk of injury (Chorba et al., 2010; Garrison et al., 2015; Kiesel et al., 2007, 2014; Knapik et al., 2015)

2.1.1 Reliability

Due to the subjective nature and the relative recent adoption of movement screens by practitioners, there is relatively limited scientific research on movement screens (McCunn et al., 2016). Within the small amount of available scientific research, there are differing results in regards to inter- and intra-rater reliability, even within the same movement screen and/or study. When looking between studies, this is most likely attributed to the small number of raters being compared (with the majority being two) within each study, rater experience, real-time scoring versus scoring from videos, and/or the qualitative interpretation of reliability measures (McCunn et al., 2016).

With respect to the FMS, there is agreement within the literature that there is strong inter-rater reliability when looking at composite scores (Gribble et al., 2013; Gulgin and Hoogenboom, 2014; Minick et al., 2010; Onate et al., 2012; Smith et al., 2013), however, when looking at individual movement scores, inter- and intra-rater reliability is poor (Gulgin and Hoogenboom, 2014; Onate et al., 2012; Smith et al., 2013). For the Foundation Matrix, a movement screen consisting of 9 motor control tests scored based on 3-5 yes/no scoring questions, the inter- and intra-rater reliability ranged from no agreement to perfect agreement depending on the task and raters being compared (Mischiati et al., 2015). For a 16-item physical performance measure

screening battery, a movement screen consisting of both quantitative and qualitative-scored movements, the inter-rater reliability ranged from poor to near perfect agreement for the qualitative-scored movements depending on the task and the experience of the rater (Tarara and Hegedus, 2014).

In terms of rater experience, the majority of research agrees that the influence of rater experience has a negligible effect on inter- and intra-rater reliability (Gulgin and Hoogenboom, 2014; Leeder et al., 2013; Smith et al., 2013); however, Gribble et al. (2013) found that rater experience did impact inter- and intra-rater reliability. Gribble et al. (2013) compared athletic trainer students, athletic trainers with no FMS experience, and athletic trainers with at least one year of experience with the FMS and found a clear trend highlighting the importance of rater experience with intra-class correlation coefficient (ICC) values of 0.37 (95% CI -0.79-0.78), 0.76 (95% CI 0.32-0.92), and 0.95 (95% CI 0.68-0.99), respectively. Gribble et al. (2013) suggests that their stronger experimental foundation, due to the greater number of raters included in the study, likely contributed to the differing results to previous studies. However, although there were an increased number of raters, the ICC values were calculated on only three raters (Gribble et al., 2013; McCunn et al., 2016). Therefore, McCunn et al (2016) concludes that rater experience does not play a significant factor in inter- and intra-rater reliability.

When comparing the effects of real-time scoring versus video-based scoring, it is suggested that video-based scoring could increase the inter- and intra-rater reliability; however, there is very limited research studying the effects of real-time scoring versus video-based scoring. The only study to date studying the effects of real-time versus video-based scoring found that video-based scoring had better reliability (ICC = 0.92, 95% CI 0.85-0.96) compared to live scoring (ICC = 0.60, 95% CI 0.35-0.77; Shultz et al., 2013). However, the results are based on a single participant and therefore it is hard to make firm conclusions on the effects of real-time versus video-based scoring on inter- and intra-rater reliability (McCunn et al., 2016; Shultz et al., 2013).

The variability in inter- and intra-rater reliability within the same study is thought to be due to the dynamic nature of the movements, the rater's perspective, and/or rater bias (Onate et al., 2012; Smith et al., 2013). The movements can be dynamic and fast-paced in nature involving multiple joints, making it difficult for the rater to evaluate all parts of the movement across the whole-body (Onate et al., 2012). In addition, the rater's perspective may have an influence on the score, as they may only see the performance from one vantage point, making it difficult for the

rater to see scoring criteria that are either out of view or occluded by the athlete's body (Onate et al., 2012; Smith et al., 2013). Lastly, the rater's own inherent bias could impact movement scores.

2.1.2 Predicting Injury Risk

Movement screens are marketed as being able to identify movement deficiencies and predict injury risk. However, similar to inter- and intra-rater reliability, there are conflicting results within the literature in terms of the ability of movements screens to predict injury risk (Bardenett et al., 2015; Butler et al., 2013; Chorba et al., 2010; Garrison et al., 2015; Kiesel et al., 2007; Mcgill et al., 2015; McGill et al., 2012; Warren et al., 2015). For the FMS, even within the studies that reported associations between composite scores of ≤14 and an increase in injury risk, the degree of the relationship varied across studies (Butler et al., 2013; Chorba et al., 2010; Garrison et al., 2015; Kiesel et al., 2007, 2014; Knapik et al., 2015), with many similar studies reporting no association (Bardenett et al., 2015; Mcgill et al., 2015; McGill et al., 2012). The wide range in results is hypothesized to be due the differences in the definition of injury, duration of injury tracking, and the number of participants followed (McCunn et al., 2016). In addition, major confounding variables such as exposure time and training load, the heterogeneity of populations studied both in skill/activity level and sport/occupation is largely ignored, making drawing conclusions difficult (McCunn et al., 2016).

The FMS has been criticized for not being sensitive enough to predict injury risk, due to vague scoring criteria (Frost et al., 2015b). The scores are graded on task-specific criteria on a scale of zero to three, however, many of the criteria have no connection (epidemiologically or biomechanically) to injury mechanisms or risk factors (Frost et al., 2015b). Furthermore, due to the large amount of movement variability between athletes, the FMS scoring criteria may be insensitive to potentially risky movement behaviour, with previous research recommending that whole-body segment and joint kinematics should be incorporated when administering movement screens (Frost et al., 2015b). In addition, due to the task-specific scoring criteria, Frost et al. (2015a) found that individuals are able to significantly increase their movement scores strictly by being aware of the scoring criteria. Therefore, psychological factors such as understanding the task, focus of attention, motivation, and awareness of grading criteria, could be playing just as big of a role on performance as deficits in joint mobility and stability (Frost et al., 2015a). Due to the lack of consistent research and the limitations with the scoring criteria, the FMS nor any other

movement screen can be heralded as an injury prevention/prediction tool (Frost et al., 2015b; McCunn et al., 2016).

2.2 Motion Capture Technology

Motion capture is the digital tracking and modelling of movement of a desired subject(s) or object(s) in space (Menolotto et al., 2020). The use of motion capture technology dates back to the mid-to-late 19th century (Marey, 1874; Muybridge, 1887) with the desire to better understand human movement driving the technological advances in motion capture that have since been seen. Some common motion capture technologies include 3D optical motion capture and inertial measurement units (IMUs).

2.2.1 Optical Motion Capture

Optical motion capture refers to motion capture technology that is camera-based, with marker-based systems being the current gold standard. For marker-based optical motion capture systems, markers, either passive-reflective or active, are affixed to the skin or to clusters on key anatomical landmarks. Using multiple infra-red, high-speed cameras and custom algorithms, the markers are able to be located within a 3D global coordinate system. Although marker-based systems are the current gold standard, the system has been criticized for: 1) being expensive, 2) requiring large amounts of time to use, 3) having a limited capture volume, and 4) being susceptible to marker occlusion, errors associated with marker placement, and skin motion artefact. The systems are generally expensive to purchase, creating a barrier for wide use in the field (Carse et al., 2013). Another barrier to use in the field, is that they are time costly both for set-up before data collection and to post-process the data, limiting the scalability to the general population. Due to the need for multiple cameras to see the markers (>3 on average), in order to be able to calculate the 3D position of the marker in space, the capture volume area is limited. In addition, since the markers need to be visible to the cameras, any clothing, accessories, equipment and/or movements may need to be modified so that markers are not being occluded (Mavor et al., 2020). For these three reasons, optical motion capture is typically confined to a laboratory setting. In terms of associated error, there is ranging reliability of kinematics due to variability in marker placement depending on the plane of interest with frontal and transverse planes having lower reliability compared to the sagittal plane (Noehren et al., 2010). In addition, even though markers are often placed on bony landmarks, there is still skin motion artefact that occurs due to the skin moving

atop the skeletal system. Stagni et al. (2005) found that ab/adduction and internal/external rotation angles were the most affected by skin motion artefact with root mean square errors up to 192% and 117% of the corresponding range, respectively. Due to the limitations of marker-based optical motion capture, other motion capture technologies have started to become popular within the biomechanics field.

2.2.2 Inertial Measurement Units

Wearable systems are gaining popularity in clinical, sport, and ergonomic settings offering an inexpensive alternative to optical motion capture systems (Hadjidj et al., 2013; Patel et al., 2012). The wearable systems are easily transportable, require minimal post-processing, are able to collect data in larger capture volumes compared to optical systems, and are immune to problems associated with optical systems such as occlusion and line-of-sight problems (Zhou and Hu, 2008). A common type of wearable sensor is the IMU. IMUs most often contain an accelerometer, gyroscope, and magnetometer, allowing measurement of linear accelerations and angular velocities in three axes and the triaxial magnetic fields of the earth. Although there are many benefits to using IMUs in the field, there are still limitations with the technology. Limitations of IMUs include: 1) susceptibility to drift, 2) lack of reference to a global position and to other IMUs, and 3) lack of environmental information. IMUs are susceptible to drift, especially when close to metal, although more robust algorithms are continuously being developed to mitigate these effects (Madgwick et al., 2011; Wittmann et al., 2019). Each sensor has its own local coordinate system, making it hard to integrate data from multiple IMUs into a global coordinate system, decreasing the interpretability of results consisting of data from multiple IMUs. In addition, IMUs are not able to incorporate environmental information, such as other objects or obstacles, that may be pertinent to the results of the data (Chen et al., 2014). In recent years, companies such as Xsens Motion Technologies (Enschede, Netherlands) have been developing technology to create detailed underlying biomechanical models based on IMU outputs to easily calculate joint kinematics (Roetenberg et al., 2013), however, these systems are roughly $20,000 USD, creating a barrier for wide-use within the field.

2.3 Principal Component Analysis (PCA)

Principal Component Analysis (PCA), a multivariate statistical technique that can be applied to reduce the dimensionality of high-dimensional data sets, is one of the most popular multivariate statistical techniques used across all scientific disciplines (Abdi and Williams, 2010). The main purpose of PCA is to extract the most pertinent information from a dataset, and therefore, decrease the size of the dataset by only retaining the pertinent information (Abdi and Williams, 2010). To achieve this, principal components are calculated as linear combinations of the original data using singular value decomposition, with the first principal component representing the largest possible variance within the data. The following components are constrained to being orthogonal to the previous principal component and having the next largest possible variance (Abdi and Williams, 2010). Scores for each sample are then able to be calculated for each principal component, which can be interpreted as the projections of the sample onto each principal component (Abdi and Williams, 2010). If there are redundancies in the data, a subset of the data will explain the majority of the variance (Troje, 2002a). The motion data can be modelled with fewer parameters based on the following equation:

$$Pj = pj,0 + pj,1cj,1 + pj,2cj,2 + \ldots pj,ncj,n \quad (1)$$

where *Pj(t)* is the modelled motion, *pj,0* is the time-series mean posture, *pj,n* is the n^{th} principal component, and *cj,n* is the score associated with the n^{th} principal component (Troje, 2002). The higher the sum of the explained variance, the more accurate the reconstructed data will be.

While PCA has been applied to joint waveforms (e.g., angles and moments; Deluzio and Astephen, 2007; Wrigley et al., 2006), it can also be applied directly to 3D motion capture marker data (Troje, 2002a). This allows for whole-body kinematics to be analyzed instead of individual joints, allowing for the data to be reconstructed in a visually meaningful way. PCA-based movement pattern recognition techniques have been used effectively to detect and explain differences in whole-body movements. For example differences in gait patterns were detected and explained based on factors including age, body mass index, biological sex, and feelings (happy/sad, nervous/relaxed; Baydal-Bertomeu et al., 2016; Troje, 2008, 2002a). When applied to ergonomics, egress/ingress motion patterns were detected and explained based on vehicle design (Masoud et al., 2017). In sport examples, PCA was applied objectively to analyze down-hill

(Federolf et al., 2014) and cross-country skiing (Gløersen et al., 2017) techniques, and to develop an objective judging tool for competitive diving (Young and Reinkensmeyer, 2014). PCA applied to whole-body 3D motion capture data offers researchers, coaches, and judges with a method to objectively assess and even score movement patterns of athletes as they perform their specific movement skills, removing subjectivity/bias from traditional observational-based assessment.

Although most biomechanics researchers who use PCA as a data reduction tool continue on using traditional statistical methods, biomechanists are starting to use PCA outputs as inputs for machine learning algorithms (Halilaj et al., 2018). Time-series data have been considered one of the most challenging problems for machine learning as they tend to be complex, high-dimensional, highly-correlated, inherently noisy, and prone to overfitting (Esling and Agon, 2012; Yang and Wu, 2006). By reducing dimensionality and extracting meaningful patterns in the dataset, PCA is an excellent tool to better prepare time-series data for machine learning algorithms.

2.4 Machine Learning

With access to larger data sets and computational resources, the use of machine learning is on the rise and being incorporated into research programs not only in biomechanics but across a wide spectrum of fields of investigation. Machine learning grew out of a sub-field of computer science known as artificial intelligence (AI) with the focus of creating computer systems and algorithms so that machines can learn from previous experiences (Izenman, 2013). Machine learning is defined as the ability to use previous experience to develop new knowledge so performance improves on a specific task over time (Izenman, 2013). The field is multidisciplinary with influences from statistics, pattern recognition, neural networks, symbolic machine learning, computational learning theory and AI (Izenman, 2013).

Machine learning can be divided into multiple categories with the two most recognized being supervised and unsupervised learning (Izenman, 2013). Supervised learning is defined as learning that occurs when the machine is able to train with a given set of continuous or discrete input variables and a known output variable (Hastie et al., 2008; Izenman, 2013). If the output variable is continuous, it is known as a supervised regression model, whereas if the output variable is discrete, it is known as a supervised classification model (Hastie et al., 2008; Izenman, 2013). For unsupervised learning, there are no known outputs given to the machine (Izenman, 2013). The

studies in this thesis only utilized supervised learning, therefore, only supervised learning will be discussed from here on out.

For supervised learning, there are three main components: 1) representation, 2) evaluation, and 3) optimization. Representation is the hypothesis space language used to express all possible classifiers (Domingos, 2012). The hypothesis space refers to the set of allowed classifiers (hypotheses) for all possible outputs that might be learned (Blockeel, 2011; Domingos, 2012). The number of hypotheses in the hypothesis space is determined by the number/types of input features and possible outputs. In a model where the features and outputs are binary, if the number of input features is n, there are 2^n possible unique input combinations. Since each input can have two possible outputs, the total number of unique hypotheses can be calculated as 2^{2^n} (Blockeel, 2011; Domingos, 2012). The hypothesis language (or machine learning model) is the language in which hypotheses are described and are dependent on the learning task (Blockeel, 2011). For supervised learning, the output is typically a function and therefore the hypothesis language must include functions, whereas for unsupervised learning, the language must include constructs for representing clusters (Blockeel, 2011). Different languages can include the same function(s) with the differences between the languages being syntactic and multiple languages can be used for the same problem (Blockeel, 2011). Lastly, multiple languages can be combined to create a stronger machine learner (i.e., model ensembles; Domingos, 2012). Common representations in biomechanics include: support vector machines (Altun and Barshan, 2010; Ashouri et al., 2017; Begg and Kamruzzaman, 2005; Clermont et al., 2019; Eskofier et al., 2013; Rekhi et al., 2009; Subasi, 2013; Wu and Wang, 2008), discriminant analysis (Fortier et al., 2005; Lee et al., 2009; Smith and Spinks, 1995), linear (De Wit et al., 2000; Guskiewicz et al., 2007; Lelas et al., 2003; Meskers et al., 1997; Owings and Grabiner, 2004) and logistic (Hewett et al., 2005; Kerr et al., 2001; Marras et al., 1995, 1993; Milner et al., 2006) regression, *k*-nearest neighbours (Altun and Barshan, 2010; Subasi, 2013), Naïve Bayes (Caggiari et al., 2020; Karuc et al., 2021), decision trees (Armand et al., 2007; Ermes et al., 2008; Girard et al., 2020; Hong et al., 2008), and neural networks (Bačić, 2016; Clouthier et al., 2020).

Evaluation refers to the objective function (also known as the scoring function) used to judge classifiers (Domingos, 2012). Common evaluation functions are: accuracy/error rate (percentage of correctly classified observations), margin (distance between separating hyperplane

and closest observations), squared-error (error between predicted output versus known output), and likelihood (the estimated probability of a model given the known observed data; Domingos, 2012). Results from the evaluation are then used to optimize the model.

Optimization is the method used to search the hypothesis space to identify the highest-scoring classifier based on the evaluation. Ideally, using the optimization method, the optimal classifier will be identified, however, this is often not the case with the optimization function identifying the local minimum rather than the global minimum (Hastie et al., 2008) or when there is more than one global minima (Domingos, 2012). Examples of optimization methods include: beam search, greedy search, and gradient descent. For beam search, the different hypotheses are organized into a hierarchical tree based on the objective function (Ow and Morton, 1988). At each level of the tree, only the n^{th} (determined by the beam width) most promising hypotheses are kept and the other hypotheses and their branches are pruned (Ow and Morton, 1988). For greedy optimization methods, like beam search, the hypotheses are arranged in a hierarchical tree based on the objective function. However, instead of the n^{th} number of hypotheses being retained, only one hypothesis from each level is retained (Emken et al., 2007; Norouzi et al., 2015). Due to retaining one hypothesis at each level, greedy optimization has low computational costs, however, a local minimum, not the global minimum, is frequently found. Gradient descent is often used for the optimization of neural networks. The gradient of the objective function gives the slope of the function along every dimension. For gradient descent, at each iteration the gradient of the function is determined, a step is taken along the steepest slope, and the gradient is reassessed until a local minimum is reached (Bishop, 2006). The step distance can be modified based on the problem.

Although a certain evaluation and optimization technique is often used in conjunction with a certain representation (e.g., support vector machines commonly utilize margin and quadratic programming for evaluation and optimization, respectively), a representation can use many evaluation and optimization methods (Domingos, 2012). Common evaluation and optimization techniques for representations will be discussed further as the different machine learning algorithms are introduced.

2.4.1 Regression

Regression as a scientific method was first applied in 1885 by Francis Galton to study heredity stature by exploring the relationship between the height of parents and children (Izenman,

2013). However, the application of least squares to regression was first shown by George Udny Yule in 1897 (Izenman, 2013). Since then, regression analyses have been applied widely and have grown into multiple different forms. Regression analyses predict or classify an output variable(s) from one or more predictor variables (Maroof, 2012). The output variable(s) can be either continuous (linear regression) or discrete (logistic regression; Maroof, 2012).

Linear regression attempts to fit a linear model with the equation y = mx + b, where x is the input variable(s), y is the predicted output(s), m is the slope of the line and b is the constant/bias, to data so that the squared distance between the known output values and the predicted output values are minimized (Maroof, 2012). Multiple iterations are performed until the linear line with least amount of error (shortest distance between observed and predicted values) is identified (Maroof, 2012). Linear regression can be split into simple, multiple, or multivariate models depending on the number of input and output variables in the model (Izenman, 2013). Simple linear regression requires only one input and output variable (Izenman, 2013). Multiple regression deals with more than one input variable and a single output variable (Izenman, 2013). Lastly, multivariate regression includes multiple input and output variables (Izenman, 2013). For all linear regressions, the regression model can be defined mathematically by:

$$Y = \beta_0 + \sum_{j=1}^{r} \beta_j X_j + e \tag{2}$$

where Y is the output variable or response, X is the input variable(s) or predictor(s), r is the number of input variables, e is an unknown error, and β_0, $\beta_1 \ldots \beta_r$ are unknown parameters (Izenman, 2013). Linear regression is easy to implement; however, linear regression does not perform well for data that are not linearly related.

Logistic regression is very similar to linear regression, however, unlike linear regression which predicts continuous outputs, logistic regression classifies samples into discrete categories and assesses the likelihood of falling into one of the output categories using logistic sigmoid functions (Bishop, 2006; Maroof, 2012). For example, researchers previously used lower limb neuromuscular moment and force variables as predictors in a logistic regression model to predict whether individuals would tear their anterior cruciate ligament (ACL; yes/no output; Hewett et al., 2005). The model not only classifies individuals into the tear/no tear category, but also predicts the percent likelihood that the individual would tear their ACL (Hewett et al., 2005). The probability equation is denoted as:

$$p = \frac{1}{1 + e^{-(b_0 + \Sigma b_j x_j)}} \quad (3)$$

where p is the probability that an individual belongs to a certain category, b_0 is the unknown bias, b_j are the unknown parameters being solved for and x_j are the input variables (Dodson et al., 2014). Like linear regression, logistic regression is simple to implement, however, does not perform well with complex, non-linearly separable data.

Regression has been used widely across the biomechanics field using both linear (De Wit et al., 2000; Guskiewicz et al., 2007; Lelas et al., 2003; Meskers et al., 1997; Owings and Grabiner, 2004) and logistic (Hewett et al., 2005; Kerr et al., 2001; Marras et al., 1995, 1993; Milner et al., 2006) regression. Linear regression has been used to try and predict the glenohumeral joint rotation centre based on scapular bony landmarks (Meskers et al., 1997). Researchers used a step-wise regression model based on minimization of the root mean squared error (RMSE; Meskers et al., 1997). The model was able to predict the joint rotation centre with a RMSE of 2.32 mm, 2.68 mm, and 3.04 mm in the x-,y-, and z-coordinate, respectively (Meskers et al., 1997). For logistic regression, the algorithm has been applied to determine if tibial shock (peak positive acceleration) can predict tibial stress fractures in female runners (Milner et al., 2006). The binary logistic regression model was able to accurately classify 70% of the female runners based on tibial shock and had an R-squared value of 0.169 suggesting that 17% of the variance between injured and non-injured runners is due to tibial shock (Milner et al., 2006).

2.4.2 Support Vector Machines

Support vector machines (SVMs) are one of the conceptually simple machine learning algorithms, yet also one of the best methods especially for binary classification (Xanthopoulos et al., 2012). Support vector machines were first introduced by Vladimir Vapnik in 1995 (Subasi, 2013; Vapnik, 1995; Xanthopoulos et al., 2012). The support vector machine aims to find the optimal hyperplane that separates the classes of data (Subasi, 2013; Vapnik, 1995; Xanthopoulos et al., 2012). The optimal hyperplane is defined as the plane that separates the data into classes with the maximal possible margin (largest distance) between the hyperplane and closest of the data points for each class (support vectors; Subasi, 2013; Vapnik, 1995; Xanthopoulos et al., 2012). The hyperplane can be mathematically defined as $w^T x + b = 0$, where w represents the parameters of the hyperplane and is perpendicular to the hyperplane, b represents the bias of the hyperplane and x represents the known input features (Mammone et al., 2009; Subasi, 2013). With

the output being represented as $y = \{1,-1\}$, the upper and lower margin boundaries can then be mathematically represented as $w^T x + b = 1$ and $w^T x + b = -1$, respectively (Mammone et al., 2009; Subasi, 2013). Therefore, the distance between the two boundaries (margin) can be defined as:

$$m = \frac{w^T}{\|w\|}(x^+ - x^-) = \frac{2}{\|w\|} \tag{4}$$

where m is the margin, $\|w\|$ is the length of w, x^+ is a point on the upper boundary and x^- is a point on the lower boundary (Mammone et al., 2009). SVMs try to maximize the margin ($\frac{2}{\|w\|}$), while still classifying samples correctly, which can be denoted as the following optimization problem:

$$\max \frac{2}{\|w\|} \qquad \text{s.t. } y_i(w^T x_i + b) \geq 1 \tag{5}$$

where y_i is the class label (Cristianini and Shawe-Taylor, 2000; Gunn, 1998; Mammone et al., 2009). The optimization problem can then be solved using the Lagrangian formulation so that:

$$\Phi(\alpha) = \sum_i \alpha_i - \frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j \qquad \text{s.t. } \alpha_i \geq 0 \qquad \sum_j \alpha_i y_i = 0 \tag{6}$$

where α are Lagrangian multipliers (Cristianini and Shawe-Taylor, 2000; Mammone et al., 2009). w can then be calculated by:

$$w = \sum_i \alpha_i y_i x_i \tag{7}$$

therefore, only samples with an $\alpha_i \neq 0$ contribute to the margin (Cristianini and Shawe-Taylor, 2000; Gunn, 1998; Mammone et al., 2009). These samples are known as the support vectors and are the critical elements of the training sample (Mammone et al., 2009). If all other training samples were removed and the model was retrained, the hyperplane would not change, therefore, only a few samples are needed to reconstruct the hyperplane, minimizing the computational resources needed to preserve the classification function (Mammone et al., 2009). Another benefit of SVMs is that an infinite number of input features can be used regardless of the number of samples without the risk of overfitting (Domingos, 2012). When a linear hyperplane is unable to separate the data, data can be mapped onto a higher dimensional space (feature space) by replacing

$x_i^T x_j$ in equation 6 with a kernel function ($K(w_i^T x_j)$), to get a potentially better representation of the data, which is then able to be linearly separated by a hyperplane (Mammone et al., 2009). The shape of the feature space is determined by the kernel function (Mammone et al., 2009). A benefit of multiple kernel options is the ability to classify data using non-linear decision boundaries, however, this increases the difficulty of tuning the model.

SVMs are commonly used in the biomechanics field to classify data (Altun and Barshan, 2010; Ashouri et al., 2017; Begg and Kamruzzaman, 2005; Clermont et al., 2019; Eskofier et al., 2013; Rekhi et al., 2009; Subasi, 2013; Wu and Wang, 2008). Two of these applications include using SVM to classify higher- and lower- mileage runners based on their running kinematics (Clermont et al., 2019) and to predict low-back pain in patients based on three-dimensional (3D) trunk kinematics (Ashouri et al., 2017). To classify high- and low- mileage runners, 57 (male and female), 47 (female only) and 12 (male only) principal component scores based on lower limb joint angles were retained as input features for the SVM (Clermont et al., 2019). The model was able to accurately classify 92.6%, 89.8% and 100% of the male and female, female only, and male only runners, respectively as high- or low- mileage with 10-fold cross-validation (Clermont et al., 2019). To predict low-back pain in patients, because the data were not linearly separable, a Gaussian radial basis function (RBF) was used as the kernel function, with percent accuracy, specificity, and sensitivity as the evaluation method. 96% of patients were able to be accurately classified as either low-back pain (LBP) or non-LBP with a specificity of 92% and a sensitivity of 100% based on inertial measurement unit (IMU) data from the thorax during a simple flexion/extension task in the sagittal plane (Ashouri et al., 2017).

2.4.3 Linear Discriminant Analysis

Linear discriminant analysis (LDA) was first developed by Fisher in 1936 to discriminate different types of flowers (Fisher, 1936). In LDA, a subspace of lower dimension than the original data sample is determined where the classes are maximally separable, where separability is defined by mean and variance (Xanthopoulos et al., 2013). For a two-class problem, this is done by finding a hyperplane that maximizes the distance between the two classes, while minimizing the scatter, also known as within-class variance, within in each class (Xanthopoulos et al., 2013). Mathematically, the distance between the two classes can be described as:

$$s_B = (\mu_2 - \mu_1)(\mu_2 - \mu_1)^T \tag{8}$$

where S_B is the distance between the two classes and μ_1 and μ_2 are the means of the two classes. The following equation is then used to compute the scatter:

$$S_w = \sum_{j=1}^{2} \sum_{x} (x - \mu_j)(x - \mu_j)^T \tag{9}$$

where, S_w is the within-class variance and μ is the class mean. The following optimization equation can then be used to find the optimal hyperplane:

$$w = \max_w \frac{w^T S_B w}{w^T S_W w} \tag{10}$$

where w is the hyperplane defined by a vector (Xanthopoulos et al., 2013). For multi-class problems, the equations are quite similar, however, instead of finding the difference in mean between two classes, the mean of each class is compared to the mean of all data. In addition, C-1 hyperplanes are calculated, where C is the number of classes (Xanthopoulos et al., 2013).

Linear discriminant analysis is popular among biomechanics research (Cavalheiro et al., 2009; Fortier et al., 2005; Lee et al., 2009; Schilaty et al., 2020; Smith and Spinks, 1995). Highlighting two studies, the first by Smith and Spinks (1995) used LDA to determine differences in biomechanical variables between novice, good, and elite rowers. LDA was able to predict skill level for novices, good, and elite athletes with 88.9%, 73.9%, and 100% accuracy, respectively. From the linear discriminant functions, they were able to determine that mean propulsive power per kilogram of body mass was the greatest predictor of skill level followed by stroke-to-stroke consistency and stroke smoothness (Smith and Spinks, 1995). The second study by Schilaty and colleagues (2020) used LDA to classify knee injury outcome based on demographic and biomechanical input variables. When using a 5 factor model, there was good prediction for 4 of the 5 classifications, whereas the 10 and 15 factor models had excellent classification for 4 and 5 of the 5 classifications, respectively (Schilaty et al., 2020). The greatest predictors were peak knee adduction moment and vertical ground-reaction force at initial contact (Schilaty et al., 2020).

2.4.4 Naïve Bayes

Naïve Bayes classifiers are based on the Bayes Theorem. For Naïve Bayes, it is assumed that each sample only belongs to one class and that each input feature is independent (Berrar,

2019). However, in most real-world applications, independence cannot be assumed, hence the name naïve. The classifier can be mathematically denoted as:

$$\hat{y} = arg\max_{y_j} \prod_{k=1}^{P} P(x_k|y_j)P(y_j) \tag{11}$$

where $\hat{y}$ is the winning class, also known as the maximum a posteriori, y is the class, and x are the set of input features (Berrar, 2019). For features that are continuous, the likelihoods can be calculated as:

$$p(x_k|y_j) = \frac{1}{\sqrt{2\pi\sigma_{y_j}^2}} \frac{-(x_i-\mu_{y_j})^2}{2\sigma_{y_j}^2} \tag{12}$$

where μ and σ are the mean and variance of the class, respectively (Caggiari et al., 2020).

Within biomechanics research, researchers looking to automate the detection of a range of static laying postures and transition postures from pressure map data to reduce the risk of pressure ulcers found that Naïve Bayes in conjunction with PCA performed the best compared to *k*-nearest neighbors and support vector machine classifiers with accuracies ranging from 82-100% (Caggiari et al., 2020). Another study looked at using FMS scores as input to predict injury in adolescents using machine learning. For the non-athletic group, Naïve Bayes had the highest area under the receiver operating characteristic curve at 0.58 (Karuc et al., 2021).

2.4.5 Decision Trees

First introduced in the 1960s, decision trees are able to discover and extract patterns from large datasets for classification or predictive modelling, while being one of the most intuitive machine learning models to interpret (Myles et al., 2004; Song and Lu, 2015). A decision tree is built by recursively partitioning the feature space of the training data, with models consisting of nodes and a hierarchy of branches (Myles et al., 2004). The goal is to find a set of decision rules that naturally partition the feature space to provide a robust hierarchical classification or regression model (Myles et al., 2004). There are three types of nodes: root node, internal nodes, and leaf nodes. The root node represents the initial binary choice, which all further subdivisions result from; internal nodes represent one of the possible choices within the tree, with each internal node being

connected to a parent node on the upper edge and a child node on the bottom edge; and, leaf nodes, also known as end nodes, are the final results based on the decisions made through the tree (Song and Lu, 2015). Branches are the chance outcomes or occurrences that stem from root nodes and internal nodes.

When building the model, splitting occurs to determine the different branches and nodes. During the splitting process, a set of possible partitions are identified and the partition that is pure or the least impure is selected (Myles et al., 2004). Partitions that differentiate the classes with 100% accuracy are known as pure and all other partitions are referred to as impure. To determine the impurity of a partition, different scoring techniques are available with popular options including: information gain, Gini index, gain ratio, and twoing criteria (Myles et al., 2004; Song and Lu, 2015). Splitting continues until the stopping criteria is met, such as the minimum number of records in a leaf, the minimum number of records in a node prior to splitting, the depth of any leaf from the root node, or further splitting does not increase the evaluation criteria (Myles et al., 2004; Song and Lu, 2015). In some situations, the stopping criteria is not adequate, which results in overfitting (Myles et al., 2004; Song and Lu, 2015). In that case, the model can be pruned by either pre-pruning or post-pruning techniques. Pre-pruning uses Chi-square tests or multiple-comparison adjustment methods to prevent the model from generating non-significant branches, whereas, post-pruning removes branches that improves evaluation criteria when applied to the validation dataset (Song and Lu, 2015).

Within biomechanics research, one study looked at classifying injured and non-injured ACLs in adolescent girls. A decision tree was able to identify two clinically relevant variables from a possible 30 that were able to classify healthy and injured participants with a sensitivity and specificity of 69% and 78%, respectively (Girard et al., 2020). A second study was able to identify 15 daily living activities with 88-100% accuracy depending on the task using a decision tree with data from a wearable sensor placed on either the thigh and waist or the wrist as the input data (Hong et al., 2008).

2.4.6 k-Nearest Neighbours (kNN)

K-nearest neighbours uses known information in the training set to predict future samples' outputs (Hastie et al., 2008). The model uses *k* number of samples (neighbours) around the new

sample to predict the output class (Hastie et al., 2008). The model can be mathematically represented by:

$$\hat{Y}(x) = \frac{1}{k} \sum_{x_i \epsilon N_k(x)} y_i \quad (13)$$

where $N_k(x)$ is the neighbourhood of x defined by the pre-selected k number of points closet to x_i in the training sample (Hastie et al., 2008). Commonly, the Euclidean distance is used to quantify the distance between x_i and the neighbors (Hastie et al., 2008). $\hat{Y}$ represents the proportion of one of the classes in the neighbourhood (Hastie et al., 2008). If $\hat{Y}$ is > 0.5, then x_i will be predicted as that class (Hastie et al., 2008). Although computationally simple, due to the necessity of retaining all samples, *k*NN is memory intensive.

Previous research has used *k*NN to recognize nineteen different human activities (e.g., standing, walking on a treadmill, rowing, playing basketball) from inertial measurement unit (IMU) data (Altun and Barshan, 2010). Depending on the validation technique used, the *k*NN model was able to correctly classify 86.9-98.7% of movements when using the seven closest neighbours (Altun and Barshan, 2010). In addition, *k*NN has been used to accurately classify 95.17% of patients as normal, myopathic or neurogenic based on bicep brachii electromyography (EMG) data (Subasi, 2013).

2.4.7 Neural Networks

The development of the computational model of a neuron, based on mathematics and algorithms developed my McCulloch and Pitts in 1943, effectuated the application of artificial neural networks (ANN) for machine learning and AI (McCulloch and Pitts, 1943). Originally, ANNs were developed to mimic brain activity, however, ANNs are now more abstract and have evolved to be networks of highly interconnected nonlinear computing elements (Izenman, 2013). ANNs are nonlinear, adaptive models that perform well with high-dimensional data and large sample sizes (Hastie et al., 2008; Izenman, 2013; Schöllhorn et al., 2008). ANNs learn with the addition of new data and are able to generalize learnings to similar populations (Agatonovic-Kustrin and Beresford, 2000; Schöllhorn et al., 2008). Because ANNs retain only the essential characteristics of the data, the model inherently reduces the amount of stored data (Schöllhorn et al., 2008).

ANNs are comprised of three types of layers: the input layer, the hidden layer(s) and the output layer. There is always one input and output layer per ANN, however, there can be multiple hidden layers. Each layer is made-up of simple processing units (neurons) that are interconnected between and within layers by different weighted connections (Agatonovic-Kustrin and Beresford, 2000; Schöllhorn et al., 2008). Each neuron contains an activation and transformation function (Figure 2.4.1; Agatonovic-Kustrin and Beresford, 2000; Schöllhorn et al., 2008). The activation function is the weighted sum of activation (*ws)* calculated by summing the inputs (x_i) times the weight of the connection (w_i; Figure 2.4.1; Agatonovic-Kustrin and Beresford, 2000; Schöllhorn et al., 2008). The output (y) for each neuron is calculated using the transformation function (e.g., threshold, piecewise linear, sigmoid, Gaussian) and the weighted sum of activation (Figure 2.4.1; Agatonovic-Kustrin and Beresford, 2000; Jain et al., 1996; Schöllhorn et al., 2008). A neuron can either inhibit or excite another neuron. If the weight is negative, the neuron can be said to inhibit the neuron, whereas if the weight is positive, it activates the neuron (Agatonovic-Kustrin and Beresford, 2000). Commonly, the weights are initialized randomly and the network then learns by manipulating the weights of the network until the predicted outputs match the known outputs (Agatonovic-Kustrin and Beresford, 2000).

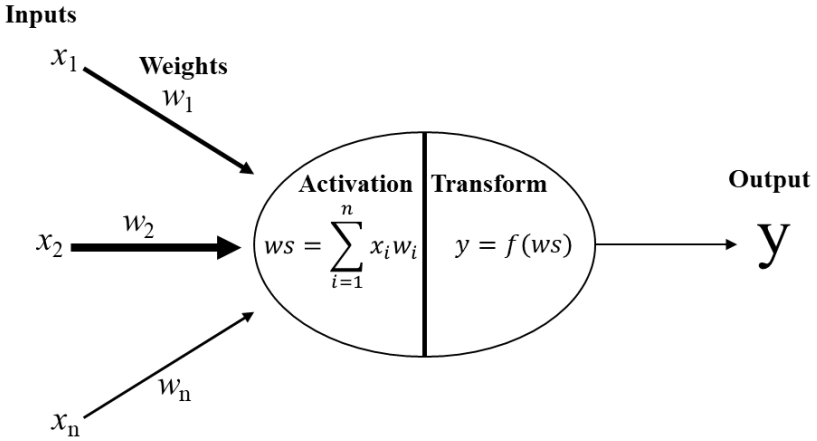


Figure 2.4.1 Schematic of an artificial neural network neuron. x_i are the inputs, w_i are the weights, ws is the weighted sum of the activations, f is the transfer function and y is the output.

There are many different methods to optimize an ANN, one of which is gradient descent, which was discussed previously. Gradient descent tries to decrease the error (difference between the desired output and the predicted output) by manipulating the weights based on the slope of the error gradient (Bishop, 2006). This can be mathematically depicted as:

$$w_{i+1} = w_i - \eta \nabla E(w_i) \tag{14}$$

where *w* is the connection weight, η is the step size (learning rate), and $\nabla E(w_i)$ is a vector that points in the direction of the greatest rate of increase of the error function (Bishop, 2006). Gradient descent is iterative and after each iteration the gradient is re-evaluated for the new weight vector until a local minimum is reached (Bishop, 2006). The numerous options to represent, evaluate and optimize neural networks allow for the development of very strong models, however, there is a negative relationship between flexibility and ease of developing a model. In addition, neural networks require large amounts of data to perform well.

There are many types of ANNs such as single-layer perceptron, multilayer perceptron, radial basis function (RBF) net, competitive network, Kohonen's self-organizing map, Hopfield network, recursive neural network (RNN), and convolutional neural network. However, for the purpose of this doctoral thesis, only RNNs will be discussed further. As mentioned previously, time-series data pose as one of the most challenging problems for machine learning as they tend to be complex, high-dimensional, highly-correlated, inherently noisy, and prone to overfitting (Esling and Agon, 2012; Yang and Wu, 2006). Conventional RNNs are a type of neural network where the connections between nodes form a directed graph along a temporal sequence, which allows the network to learn dynamic, temporally-varying behaviour (Hochreiter and Schmidhuber, 1997; Lukoševicius and Jaeger, 2009; Troje, 2002a; Weng and Shen, 2008). RNNs use state variables that allow for past information to be stored, creating a "memory" (Hochreiter and Schmidhuber, 1997). The unit can then use past information along with current input to calculate an output. RNNs work well when the gap between the current input and the relevant stored past information is small, however, if the gap is too large, the relevant information is lost (Hochreiter and Schmidhuber, 1997). Long short-term memory (LSTM) networks are a type of RNN that were developed to retain relevant information regardless of the sequential gap size (Hochreiter and Schmidhuber, 1997). This is achieved through the use of memory cells and gate units that allow for there to be a constant error flow through self-connected units. Gate units consist of an input gate, which is used to limit the data entering the unit to protect the information already within the unit from irrelevant data, and an output gate, which limits data exiting the unit in order to protect

other units from erroneous data (Hochreiter and Schmidhuber, 1997). Therefore, an LSTM can retain pertinent information regardless of gap size.

Another limitation of traditional RNNs is that backpropagation is computationally expensive as it backpropagates through time requiring large amounts of computing power and time to train the networks (Lukoševicius and Jaeger, 2009). To address this problem, reservoir computing (RC), also known as echo state networks, was developed. RC randomly generates the recurrent part (the reservoir) and then keeps it fixed with only the weights of the readout layer being trained, unlike the traditional RNN that trains the weights of each layer (Lukoševicius and Jaeger, 2009). Due to exploding or vanishing gradients, training the weights of the non-readout layers can lead to instability (exploding) or under training due to insufficient changes at the relevant layers (vanishing; Bengio et al., 1994; Hochreiter, 1998). Training only the read-out layer allows for RC to be more stable than other RNNs (Lukoševicius and Jaeger, 2009). Despite the simplification, the reservoir provides a robust pool of dynamic features, which has shown to be fit for solving numerous tasks (Bianchi et al., 2021).

When looking at the biomechanics literature, Clouthier et al. (2020) used an LSTM network in conjunction with convolutional neural networks to predict 13 different athletic movements with as few as three sensors with an accuracy of 85.9%. Another study looking at classifying six temporal phases of the tennis swing from optical motion capture data were able to classify the six phases using a singular RC network and an ensemble RC network with 85.3% and 95.3% accuracy, respectively, over 12 repetitions (Bačić, 2016).

2.4.8 Summary

Regression, support vector machines, *k*-nearest neighbors, linear discriminant analysis, Naïve Bayes, decision trees, and neural networks are just a few examples of algorithms for supervised learning techniques. For each machine learning problem, multiple algorithms can be used. Previously, research has compared and contrasted different types of algorithms to identify the algorithm that performs best based on a specific problem (Altun and Barshan, 2010; Lim and Loh, 2000). Although there are strengths and limitations to all supervised learning techniques, there is no one best algorithm and the optimal algorithm is situational. The optimal algorithm depends on multiple factors including: size and type of dataset, number of input features,

computational resources, and time (not only to run the model but also to develop and refine the model). Because there is no one best algorithm, the strengths, limitations and applicability of desired algorithms should be evaluated for all new supervised machine learning problems.

2.5 Motion and Shape Capture from Sparse Markers (MoSh)

Although marker-based motion capture has been widely accepted, and is often known as the gold standard within the research and entertainment industries; it is also widely criticized that motion capture creates lifeless and unnatural motions as a result of using an underlying skeleton that acts as a proxy for movement (Loper et al., 2014). Typically, for marker-based motion capture, markers are placed on bony, anatomical landmarks to define parts of the body. Markers are placed on parts of the body that are as rigid as possible to reduce soft-tissue motion, which is typically treated as noise. By placing markers on rigid body parts and processing soft-tissue motion as noise, subtle information about body motion is lost when moving from the captured motion to the rigid, skeletal representation of the movement (Loper et al., 2014). It is believed that the soft-tissue motion is not noise, but may hold important, subtle motions that are important to both research and animation (Loper et al., 2014). Loper et al. (2014) proposed a new method using minimal surface markers, named motion capture from sparse markers (MoSh), which instead of using an underlying skeleton, uses a 3D parametric body shape that simultaneously estimates the surface marker locations based on the 3D parametric body model, the shape of the subject, and the body pose. With MoSh, body shape is able to change over time, which allows for the technique to capture the non-rigid, soft-tissue motion. MoSh has previously been used in the field of psychology, more specifically vision and perception, to understand how perceived attractiveness and confidence relates to body-shape and walking motion and to better understand sex-based differences in gait styles (Thaler et al., 2020).

MoSh consists of five main components: 1) a parametric 3D body model that is able to realistically represent a large range of body shapes, poses, and allows for pose-dependent deformations to be captured, 2) the assumption that there is variation in how markers are placed between subjects, sessions, and researchers and therefore, assumes the exact marker location is unknown and instead optimizes the observed marker placement relative to the 3D parametric body model, 3) a 3D body shape that best explains the 3D mocap marker data, and 4) a 3D body pose. The first 4 steps are all incorporated into a single optimization function that is optimized for a

subset of the data. For component 5, following completing components 1-4, MoSh uses the previously calculated body shape and marker locations to calculate body pose for the remaining data. If multiple trials are collected during one session, marker locations and body shape from one trial can be used to calculate the body pose on the remaining trials (Loper et al., 2014).

There are three separate models: male, female and gender neutral. Each 3D parametric body model is a triangulated mesh consisting of 10,777 vertices and is parameterized by a global translation centre (γ), a vector of pose parameters (θ), a mean shape (μ), and a vector of shape parameters (β). Body-shape is approximated by the mean shape and a linear combination of shape basis vectors, where β is a vector of these 10 linear coefficients, which are calculated using PCA based on training body shapes (Loper et al., 2014). The body shape model was trained on 3803 CAESAR scans of people in an upright pose, while pose-dependent components of the model were trained on 1832 scans of 78 people in a wide range of poses (Loper et al., 2014).

Based on the assumption that marker locations are not placed precisely on the anatomical landmarks due to precise placement being difficult to achieve, the first step of MoSh solves for marker locations relative to a body mesh (Loper et al., 2014). A latent coordinate system is created that contains the markers and the body model in a neutral pose, so that the relationship between the body surface and the markers in a pose-independent, translation-independent fashion can be created, which is then transferred to the motion capture frames and the approximated default positions of the markers are calculated. A function is then defined that maps these default positions onto a particular shape, pose or location of the body, which are referred to as simulated markers. This is achieved by mapping a local surface geometry to the 3D marker positions, where each marker position is represented in an orthonormal basis defined by its nearest triangle in the latent coordinate system (Loper et al., 2014). The basis is defined by the triangle normal, one of the triangle's normalized edges, and the cross product between the previous two. Initially, the marker positions are transformed from the 3D latent-space position into a coordinate vector in the space of its local basis and then the coordinate vector is mapped onto a 3D observed-space position defined by the specific position or pose (Loper et al., 2014).

With the simulated markers approximated, latent markers, poses, body locations, and body-shape are estimated so that the simulated markers match the observed markers. To do this, multiple objective functions are defined (Loper et al., 2014). The first time is the data term, which is the

sum of squared distances between the simulated and the observed markers in centimetres. The second term, the surface distance energy term, optimizes the marker locations so that the known distance from the body surface is maintained. Another term penalizes the estimated latent markers if they deviate too far from the marker locations and adjusts the latent marker back towards the original position. Another term defines the pose and shape priors to regularize the estimation of body and shape, which are modelled as Gaussian and computed from the pose and shape of the training data. The priors are regularized by penalizing the squared Mahalanobis distance from the mean shape and pose. The last term is a velocity constancy term that helps to smooth noise. The final objective function is a weighted sum of the aforementioned terms (Loper et al., 2014). Optimization is a two-step process, where the time-independent parameters (body-shape and marker placements) are optimized initially, while the second step optimizes pose and body location (Loper et al., 2014). Throughout the optimization process, the transformation matrix from latent to observed coordinate systems are continuously re-estimated and the local basis and the body-shape coefficients are continuously refined to minimize the distance between the simulated and observed markers (Loper et al., 2014). Once optimized, based on the simulated marker positions, the body-shape coefficients and 3D pose of the individual are outputted for each individual and trial.

2.6 Research Questions and Hypotheses

Global Objective 1: Determine the inter- and intra-rater reliability of a movement screen without task-specific scoring criteria

Study 1:

Previous studies looking at the inter- and intra-rater reliability of movement screens had limited number of raters with the average being two; assessed the inter- and intra-rater reliability of a movement screen with scoring criteria, which has been criticized to not being sensitive enough to detect differences in athletes who are at high risk of injury; and have not investigated the effects of demographic information, such as body-shape, on reliability. To address these limitations and Global Objective 1, a custom online visualisation software using animations created using MoSh was developed to assess the inter- and intra-rater reliability of a movement screen without task-specific scoring criteria.

The study was specifically designed to answer the following three research questions:

S1Q1: Is there strong inter- and intra-rater reliability of movement competency scores between expert assessors for a movement screen without task scoring criteria?

S1Q2: How does intra-reliability change within and between sessions?

S1Q3: How does body-shape affect intra-rater reliability within the same session?

The following three hypotheses were then associated with the corresponding research question:

S1H1: There will be poor inter-rater reliability between expert assessors, but there will be good intra-rater reliability.

S1H2: Intra-rater reliability will be greater within session than between session.

S1H3: Due to personal biases, body-shape will have a negative effect on intra-rater reliability, therefore intra-rater reliability without body-shape modification will be greater than intra-rater reliability with body-shape modification.

Global Objective 2: Develop an objective movement screening tool that is able to objectively score an athlete's movement competency using optical motion capture data

Study 2:

Previous research has criticized subjective movement screens for having poor inter-rater reliability and not being sensitive enough to detect individuals at a higher risk of injury. Thus, to address these limitations and Global Objective 2, Study 2 looked at providing proof-of-principle for an objective movement screening tool.

The research questions addressed in Study 2 were:

S2Q1: Can differences in elite and novice athletes' movement patterns be detected and movement competency scores be calculated using pattern recognition and LDA?

It was hypothesized that:

S2H1: There will be detectable differences between elite and novice athletes; therefore, athletes will be able to be classified based on skill level using PCA and LDA. Based on the

linear discriminant function, movement competency scores will be able to be computed for each athlete.

Global Objective 3: Refine the object movement screening tool to allow for multiple data types and classifiers to be used and to determine the optimal machine learning algorithm(s).

Study 3 & 4:

To make the objective tool more robust and easier to implement within the field, more complex machine learning algorithms needed to be assessed as well as different types of input data. Therefore, Study 3 asked the following research questions:

S3Q1: Does the use of feature selection increase classification rates?

S3Q2: What is the optimal traditional linear or non-linear machine learning algorithm(s) for differentiating athletes based on skill level?

S3Q3: Can athletes be classified based on skill level using data that are able to be acquired by IMUs, such as linear acceleration and angular velocity?

It was hypothesized that:

S3H1: The use of feature selection will increase classification rates compared to using the first 35 principal components.

S3H2: The more complex traditional non-linear machine learning algorithms will have better classification rates than the previously used LDA.

S3H3: Athletes will be able to be classified based on data that can be collected by IMUs.

To determine whether demographic-specific models needed to be created and to test increasingly complex machine learning algorithms to determine the optimal combination of pattern recognition, feature selection, and machine learning techniques, Study 4 was designed with the following research questions in mind:

S4Q1: Can athletes be differentiated based on other classifiers such as sport played?

S4Q2: Do machine learning algorithms designed for time-series data perform better than the traditional classifiers for differentiating on both skill level and sport played?

For each research question, it was hypothesized that:

S4H1: Athletes competing in different sports would move differently; therefore, athletes would be able to be differentiated based on sport played.

S4H2: The RNNs would have higher classification rates compared to the linear and non-linear classifiers for both skill level and sport played.

Chapter 3: General Methodology

This thesis is comprised of studies utilizing a common, single data collection and subsequent post-processing and analyses. All data were collected by Motus Global (Rockville Centre, New York, USA), a for-profit company whose mission was to develop solutions that enhanced human performance, monitor rehabilitation, and provide insights to reduce the risk of injuries through motion-capture technology, between 2012-2016 as they conducted their proprietary screening process on 542 athletes. All athletes consented to the data being used for secondary use and data sharing and transfer agreements were agreed upon by the researchers and company prior to the researchers receiving the data. Ethics clearance for secondary use of the data was obtained for each study at each necessary institution (Appendix A).

3.1 Participants

Data were collected on 542 (479 males, 63 females) athletes competing in one of 12 sports (i.e., baseball, basketball, cricket, football, golf, lacrosse, rugby, soccer, squash, tennis, track and field, or volleyball) and ranging in skill level from youth to professional (e.g., NFL, NBA, MLB, FIFA). The average age, height, weight was 20.2 ±4.67 years, 183.3 ± 19.28 cm, and 83.1± 22.85 kgs, respectively. When broken down by sex, males had an average age, height and weight of 20.3 ± 4.5 years, 185.29 ± 19.39 cm, and 86.08 ± 22.25 kg, respectively and females had an average age, height, and weight of 18.97 ± 5.68 years, 167.96 ± 8.86 cm, and 61.14 ± 13.3 kg, respectively. Athletes were assigned as an elite or novice athlete based on their level of competition. Based on previous research that found that athletes who completed greater than 10,000 hours of deliberate practice are experts, athletes competing at the intercollegiate, semi-professional, and professional level were considered elite athletes, whereas all other lower levels (e.g., youth, recreational, high school, etcetera) of play were deemed novice athletes (Baker et al., 2003; Helsen et al., 1998). Prior to data collection, each athlete signed an informed consent form providing permission of their data to be used for future research. If the athlete was under the age of consent, permission was granted by their legal guardian/next to kin.

3.2 Equipment

In order to capture whole-body kinematics, athletes were outfitted with 45 passive, reflective markers (B&L Engineering, Santa Ana, CA, USA). 37 of the markers were placed on anatomical landmarks to define the head, trunk, upper arms, forearms, pelvis, thighs, shanks, and feet. These markers were placed on the: front, back, left and right side of the head; left and right clavicle, left and right acromion process; xyphoid process of the sternum; 2nd and 8th thoracic vertebrae; left and right anterior superior iliac spine (ASIS); left and right posterior superior iliac spine (PSIS); left and right medial and lateral epicondyles of the humerus; left and right ulnar and radial styloid processes; left and right greater trochanters; left and right medial and lateral condyles of the femur; left and right medial and lateral malleoli; left and right heel; and left and right head of the 2nd metatarsal. Eight additional tracking markers to assist with tracking segments and to help identify the left and right limbs were placed: bilaterally on the thighs, forearms, and biceps, and unilaterally on the right shank and scapula (Appendix D). All data were collected at 120Hz using an 8-camera Raptor-E motion capture system (Motion Analysis Corporation, Santa Rosa, CA, USA).

3.3 Protocol

Upon arriving at the Motus Global laboratory, athletes read and signed the informed consent form, had their height (with shoes) and weight taken, and provided details regarding their injury history from the previous ten years. Injury data included injury diagnosis, date of injury, time missed due the injury, and if surgery was necessary, the name of the surgeon, and city and state the surgery was performed. Athletes were then outfitted with the markers described in section 3.2. After outfitting the athlete with the markers, the athlete performed two calibration trials. The first calibration trial was a static calibration trial, where the athlete stood in the 'motorcycle' position with their feet shoulder-width apart, toes pointing straight forward, arms abducted 90° at the shoulder, elbows bent at 90°, and wrists in line with the forearm. The second calibration trial was a dynamic calibration trial where the initial pose was the same as the static calibration. The athlete was then instructed to internally and externally rotate their arms to their maximum range of motion (ROM), while maintaining their shoulders abducted and elbows bent at 90°.

The athlete then completed a 21-movement, proprietary protocol that was designed to assess the athletes' ROM at each joint, stability, power, and balance. For this thesis, only seven of

the movements were analyzed: bird-dog, drop jump, hop-down, L-hop, lunge, step-down, and T-balance. These seven tasks were chosen due to their dynamic nature and their ability to challenge the athletes' coordination, stability, and balance. The athletes continued to perform the movement until they believed they had completed the task to the best of their ability. Only their best trial was kept and analyzed.

3.3.1 Bird-Dog Right

The athlete started in a table-top position with one arm and the contralateral leg (e.g., left arm and right leg) fully extended in line with their trunk (Figurer 3.3.1A). The athlete then drew their elbow and knee together towards their transverse midline (Figure 3.3.1B) and then returned to the initial extended position (Figure 3.3.1C). The test was performed bilaterally.

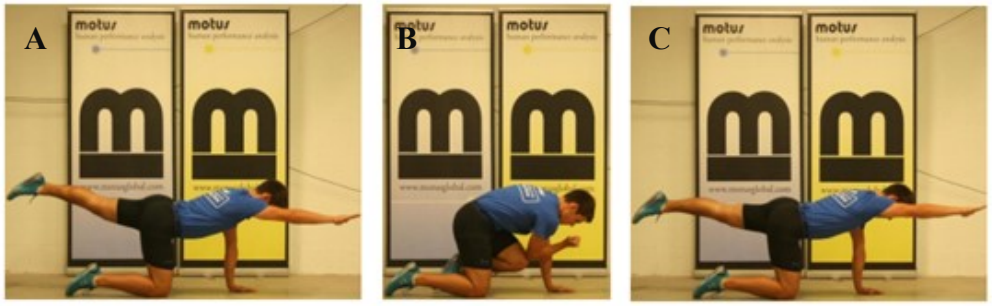


Figure 3.3.1. Bird-Dog.

3.3.2 Drop Jump

The athlete began by standing on a 30 cm tall platform (Figure 3.3.2A). The athlete then dropped down off the platform to the floor (Figure 3.3.2B) and upon landing immediately transitioned into a maximal vertical height jump (Figure 3.3.3C).

Figure 3.3.2. Drop Jump.

3.3.3 Hop-Down

Similar to the drop jump, the athlete started upon a 10 cm platform on a single leg (Figure 3.3.3A). The athlete then dropped down off the platform, landing on the same leg they initiated the task on (Figure 3.3.3B) and immediately transitioned into a single-leg maximal vertical height jump (Figure 3.3.3C). The movement was performed bilaterally.

Figure 3.3.3. Hop Down.

3.3.4 L-Hop

The athlete started standing on the floor on both legs (Figure 3.3.4A). The athlete then jumped forward as far as they were able to while landing on a single leg (Figure 3.3.4B). The athlete then immediately transitioned into a maximum lateral hop landing on the opposite leg as was initially landed on (Figure 3.3.4C). The movement was performed bilaterally.

Figure 3.3.4. L-Hop.

3.3.5 Lunge

The athlete started standing erect with their feet shoulder-width apart and their hands on their hips (Figure 3.3.5A). Keeping their hands on their hips for the duration of the trial, the athlete stepped forward the length of their leg (i.e., distance from floor to their greater trochanter) into a forward lunge position (Figure 3.3.5B). Using the same leg, they then transitioned into a backwards lunge (Figure 3.3.5D) with their back foot reaching the length of their leg past the initial position without touching their foot down during the transition (Figure 3.3.5C). The athlete then returned to their initial position (Figure 3.3.5A). The movement was performed bilaterally.

Figure 3.3.5. Lunge.

3.3.6 Step-Down

The athlete started standing on a single leg at the edge of a 20 cm platform so that the opposite foot was hanging off the platform (Figure 3.3.6A). The athlete then squatted down in a controlled manner so that the heel of their foot that was the non-weight bearing leg touched the ground (Figure 3.3.6B). The athlete then had to rise back to their initial position (Figure 3.3.6C). Throughout the movement, the athlete was to hold their hands in a prayer position at nipple height. The movement was performed bilaterally.

Figure 3.3.6. Step-Down.

3.3.7 T-Balance

The athlete started the movement by standing on a single leg with the non-weight bearing leg flexed 90° at the hip and knee and placing their hands in a prayer position at nipple height (Figure 3.3.7A). The initial position was held for three seconds. In one fluid motion, the athlete hinged at the hips as far as they were able while maintaining balance, so that their trunk became more parallel with the floor. At the same time as they were hinging forward, the athlete extended their non-weight bearing leg so that it was in line with their trunk and extended both of their arms out so the arms were perpendicular with their trunk, creating a T-position with their body (Figure 3.3.7B). Once this T-position was reached, the athlete reversed the motion returning to the starting position and holding the position for three seconds (Figure 3.3.7C). The movement was performed bilaterally.

Figure 3.3.7. T-Balance.

3.4 Data Analysis

3.4.1 Pre-processing

The data were labelled, cleaned and gap-filled in Cortex (Motion Analysis, Santa Rosa, CA, USA). Once labelled and clean, data were exported to Visual3D (C-Motion, Inc., Germantown, MD, USA), where a full-body model was developed based on the static calibration trial. The model was then applied to all motion trials. Joint centre triaxial positional data of the wrists, elbows, shoulders, distal end of the feet, ankles, knees, and hips; triaxial centre of gravity positional data were calculated for the trunk, head and pelvis; and triaxial positional raw data for the left and right heel, T2, T8, sternum, and the back, front, and sides of the head were calculated and extracted. The additional triaxial positional raw data of the select markers were chosen to more robustly model the feet, trunk, and head. All data were then exported to Matlab (The Mathworks,

Inc., Natick, MA, USA) or Python 3.0, where the data were filtered using a dual-pass, low-pass Butterworth filter with a cutoff of 15Hz, trimmed to specific start and stop criteria (Table B1), and time-normalized to 500 frames using a Piecewise Cubic Hermine Interpolating Polynomial (PCHIP) to control for differences in the absolute time taken to complete each movement.

Chapter 4: Study 1

ASSESSING INTER- AND INTRA-RATER RELIABILITY OF MOVEMENT SCORES AND THE EFFECTS OF BODY-SHAPE USING A CUSTOM VISUALISATION TOOL

Original Article

Gwyneth B. Ross[1], Nikolaus F. Troje[2], Steven L. Fischer[3], and Ryan B. Graham[1,3]

[1]School of Human Kinetics, Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada

[2]Centre of Vision Research & Department of Biology, York University, Toronto, Ontario, Canada

[3]Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada

Submitted to: *British Journal of Sports Medicine*

4.1 ABSTRACT

Objective: To assess inter- and intra-rater reliability of movement scores within and between sessions of expert assessors during a movement screen using a custom online visualisation software and to explore the effects of body-shape on reliability.

Methods: Kinematic data from 542 athletes performing seven movement tasks were used to create animations using motion and shape capture from sparse markers (MoSh). For each task, there were 90 animations, where 30 unique animations were repeated twice, making up 60 of the animations, to test intra-rater reliability and the remaining 30 animations consisted of 10 unique animations that were repeated three times with body-shape manipulated to be either underweight, normal, or overweight to test the effects of body-shape on intra-rater reliability. Using a custom developed visualisation tool, 10 expert assessors completed two identical sessions where they rated each movement on a scale of 1-10. The arithmetic mean of weighted Cohen's kappa for each task and day were calculated to test reliability.

Results: Across tasks, intra-rater reliability ranged from slight to fair agreement and inter-session reliability had slight to fair agreement. For intra-rater reliability, when looking at the average kappa values, intra-rater reliability within session with and without body manipulation and between sessions were 0.45, 0.37, and 0.35, respectively.

Conclusion: Reliability ranged from slight to moderate agreement, with a decrease in intra-rater reliability when body-shape was manipulated, suggesting that assessing movement competency based on visual appraisal during a movement screen, even with expert assessors, is not a reliable method. Therefore, objective methods and tools should be developed to reliably assess movement competency.

4.2 INTRODUCTION

Movement screens are used across a variety of settings including in ergonomic, clinical, and athletic settings to quantify 'movement quality' and identify movement patterns that are associated with an increased risk of injury and/or decreased performance (Cook et al., 2014; Donà et al., 2009; Kritz et al., 2009; McCall et al., 2014; McCunn et al., 2016; Padua et al., 2009). There are many different types of movement screens, with the Functional Movement Screen (FMS) being the most well-known (McCunn et al., 2016). Each assessment has its own unique battery of movements and scoring criteria, but each share one constant: they are all scored using visual appraisal which is a subjective approach (McCunn et al., 2016). In part due to the subjective nature, and in part to the relative recent adoption of movement screens by practitioners, there is relatively little scientific research on the use of movement screens (McCunn et al., 2016).

Within the small amount of scientific research available, there are conflicting results regarding inter- and intra-rater reliability, even within the same movement screen (McCunn et al., 2016). When looking across studies, this is most likely attributed to the small number of raters being compared within each study (with the majority of studies using only two raters), rater experience, real-time scoring versus scoring from videos, and the qualitative interpretation of reliability measures (McCunn et al., 2016). When looking within studies, the variability in inter- and intra-rater reliability is thought to be due to the dynamic nature of the movements, the rater's perspective, and/or rater bias (Onate et al., 2012; Smith et al., 2013).

For movement screens, the movements can be dynamic and fast-paced in nature involving multiple joints, making it difficult for the rater to evaluate all parts of the movement across all of the joints (Onate et al., 2012). In addition, the rater's perspective may have an influence on the score, as they may only see the performance from one vantage point, making it difficult for the rater to see scoring criteria that are either out of view or occluded by the athlete's body (Onate et al., 2012; Smith et al., 2013). Although the research is limited, previous research has found that scoring movements from video could increase the reliability (McCunn et al., 2016), since assessors are able to watch the movement multiple times. However, a limitation of video, is that the movement is reduced to one or a few vantage points, where important information may be out of view of the assessor. Another contributing factor is the rater's inherent bias which can influence the scores, especially bias towards the participants' body-shape. Research has consistently shown that there is pervasive implicit and explicit weight bias among clinicians, physical therapists,

physical education teachers, and strength and conditioning personnel, with males showing a larger bias (Panza, 2018). The use of customizable gaming software and animation techniques, such as motion and shape capture from sparse makers (MoSh), can provide an alternative assessment option that has the benefits of video assessment, while allowing raters to have 360° views of the ratee and for body-shape to be manipulated to better understand the effects of body-shape bias on the reliability of movement scores.

MoSh is an animation technique that translates 3D kinematic optical motion capture data into 3D animations visualizing both the kinematic movement patterns and the body-shape of the individual (Loper et al., 2015, 2014). With MoSh, there are three body-shape model templates (male, female, non-binary), that are manipulated by 10 beta values to fit a personalized body-shape to each individual (Loper et al., 2015). When using MoSh and a custom developed visualisation tool, a shape model is animated, much like an avatar in a video game, where the animation can be replayed multiple times, multiple raters can score the same movement, and the vantage point can be rotated to focus on specific points of interest. While 2-dimensional video can also be replayed and assessed by multiple raters, raters cannot rotate to see other vantage points unless additional synchronized video data were collected. In addition, shape models can be manipulated and rigged to different kinematic movement patterns which allows for the effect of body-shape on reliability to be studied by creating animations with identical movement patterns, but differing body-shapes.

Therefore, the purpose of this study was to assess the inter- and intra-rater reliability between two sessions and within the same session with and with-out body-shape modification between expert raters during a movement screen without a scoring-criteria using a custom developed visualisation tool with MoSh animations. It was hypothesized that the intra-rater reliability within the same session would have the best reliability, followed by the intra-rater between sessions without body-shape modification, and subsequently intra-rater between sessions with body-shape modification and inter-rater reliability.

4.3 METHODS

4.3.1 Participants

Ten expert movement assessors (6 males and 4 females) were recruited to take part in the study. The participants, here on referred to as the raters, all had formal training in movement assessment and conducted movement assessments as part of their daily job (e.g., orthopaedic

surgeons, physical therapists, strength and conditioning coaches, kinesiologists, and movement performance specialists) and had an average of 7.0 ± 3.3 years of experience in their current role.

4.3.2 Data Preparation

4.3.2.1 Animation Preparation

To create the animations, motion capture data from 542 athletes (473 males, 69 females) performing seven unique movement screening movements (i.e., bird-dog, drop-jump, hop-down, L-hop, lunge, step-down, and T-balance) were collected. At the time of collection, athletes competed in one of 12 sports (i.e., baseball, basketball, cricket, football, golf, lacrosse, rugby, soccer, squash, tennis, track and field, or volleyball) and ranged in skill level from youth to professional (e.g., NFL, NBA, MLB, FIFA). The average age, height, weight were 20.2 ±4.67 years, 183.3 ± 19.28 cm, and 83.1± 22.85 kgs, respectively. To collect whole body kinematics, 42 markers were placed on anatomical landmarks and captured using an 8-camera Raptor-E motion capture system (Motion Analysis, Santa Rosa, CA, USA). All data were labelled and gap-filled in Cortex (Motion Analysis, Santa Rosa, CA, USA). Once the data were clean, MoSh was applied to a single frame from a single movement in order to fit a body-shape (represented by 10 beta values) and rig an animation skin to each athlete's kinematic data (Loper et al., 2015, 2014). For MoSh, body-shape and kinematic data are coded so they can be manipulated independently from one

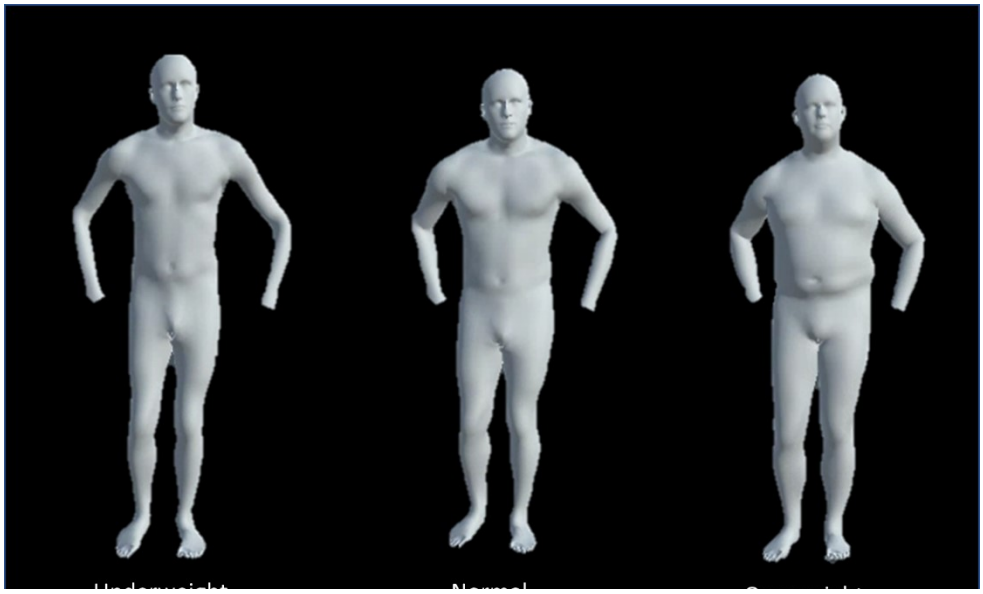


Figure 4.3.1. An example of the three different body-shapes (underweight, normal, and overweight) used for the intra-rater reliability within session with body-shape modification.

another. Body-shape is able to be manipulated by adjusting the 10 beta values that represent body-shape, whereas, kinematic data can be altered by changing the x,y,z positions of the desired joint(s) and frame(s) of data. The marker set used, although similar, was not identical to the ideal marker set, as suggested by Loper et al., 2014, with the main differences of missing markers placed on the breasts, buttocks, and hands. The breast and buttock markers were pertinent for fitting the female body-shape model, therefore, only male data were retained. The hand markers were necessary to create realistic hand movements, therefore, the hands were removed from the animations. For this study, the 5th, 50th, and 95th percentile body-mass indexes (BMI) of the dataset were calculated and used as the cut-offs for the three body-shape classes: underweight, normal, and overweight (Figure 4.3.1).

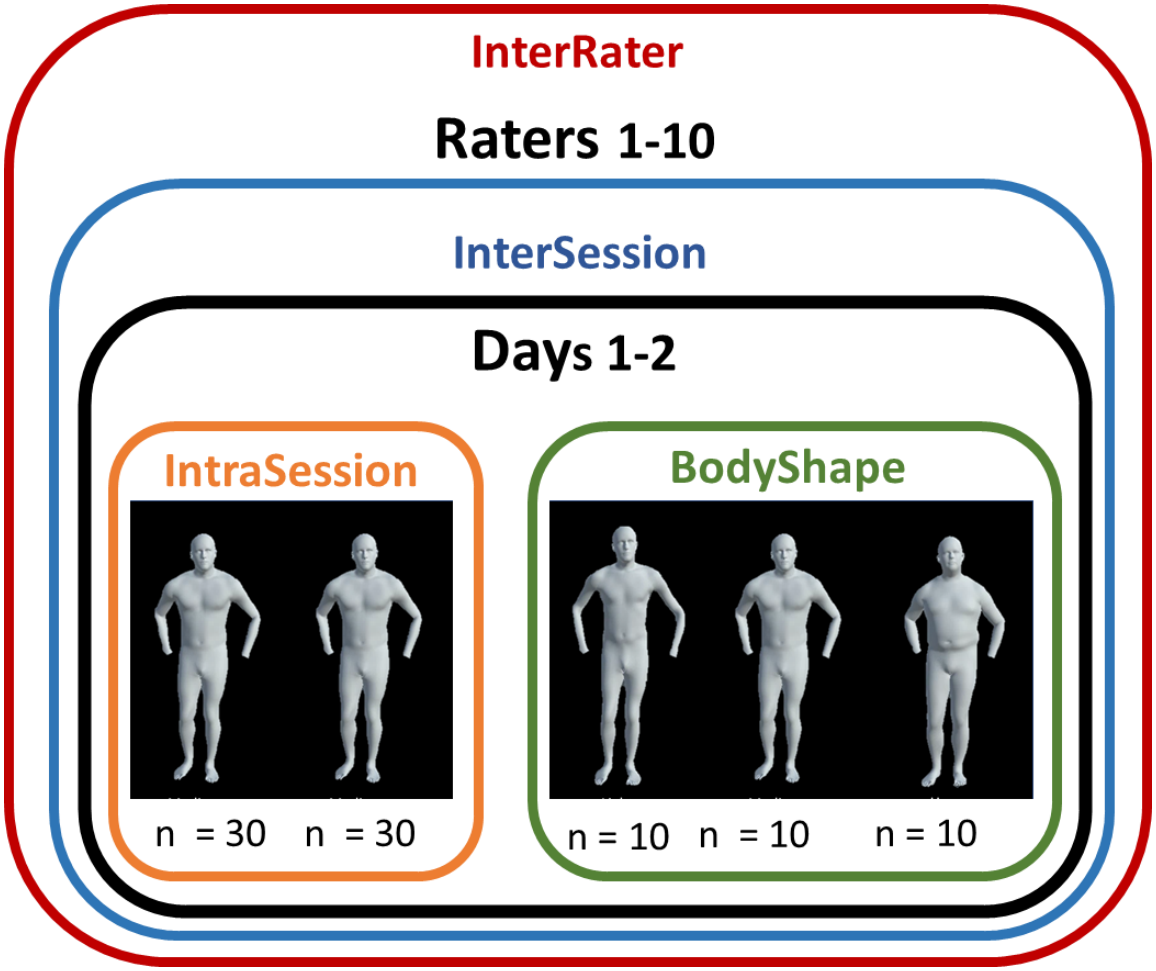


Figure 4.3.2. A visual depiction of the animations being compared to assess inter- and intra-rater reliability. InterRater = inter-rater reliability between raters. InterSession = intra-rater reliability between days. IntraSession = intra-rater reliability within session without body-shape modification. BodyShape = intra-rater reliability within session with body-shape modification.

A database of 630 animations were created consisting of 90 animations from each of the seven movements (7 movements x 90 animations = 630 animations). For within each of the seven movements, animations were selected to be able to test for intrasession reliability, intersession reliability, and weight bias (Figure 4.3.2.), as well as having a diversity of movement competency levels with approximated scores ranging from 1-10. To test intrasession reliability, 30 different movers were generated and duplicated, to create 60 of the 90 animations (Figure 4.3.2). Within the duplicated 30 animations, each movement pattern and body-shape were unique. To test weight-bias, for each approximated score, three animations were created with identical movement patterns but body- shape was manipulated so each of the three animations had a body-shape of a different class (e.g., underweight, normal, overweight), making up the remaining 30 animations (10 movement scores x 3 weight classes; Figure 4.3.2). If a task was performed bilaterally, only animations for the right-side were included.

4.3.2.2 Software Preparation

A custom-built, online, visualisation software was developed using the Unity game engine (Unity Software Inc., San Francisco, CA, USA), which was deployed on a Compute Canada server and linked to a common domain name. Within the software, there were three modules: Training, Day 1, and Day 2. Within each module, the raters were able to: zoom, rotate, and translate the animation for 360° views; play the animation; replay the animation; score the animation; move between the next and previous animation; view the control short-cut keys; and return to the main menu (Figure 4.3.3). For each animation, the score, date and time of score, time to score, and number of replays were recorded and stored in a MySQL database using phpMyAdmin.

4.3.3 Protocol

Before beginning the study, each rater read and signed a consent form as well as filled out an online form providing demographic information including age, gender, job title, years of experience, certifications, and average number of movement assessments performed per day, week, month or year. The consent form signed outlined the purpose of the study as to examine the inter-rater reliability of the used dataset. To try and obtain unbiased and/or natural reactions, the purposes of examining intra-rater reliability between sessions and within sessions were omitted. The Health Sciences Research Ethics Board at the University of Ottawa approved the study and the use of deception for research purposes (file no: H-10-19-4983).

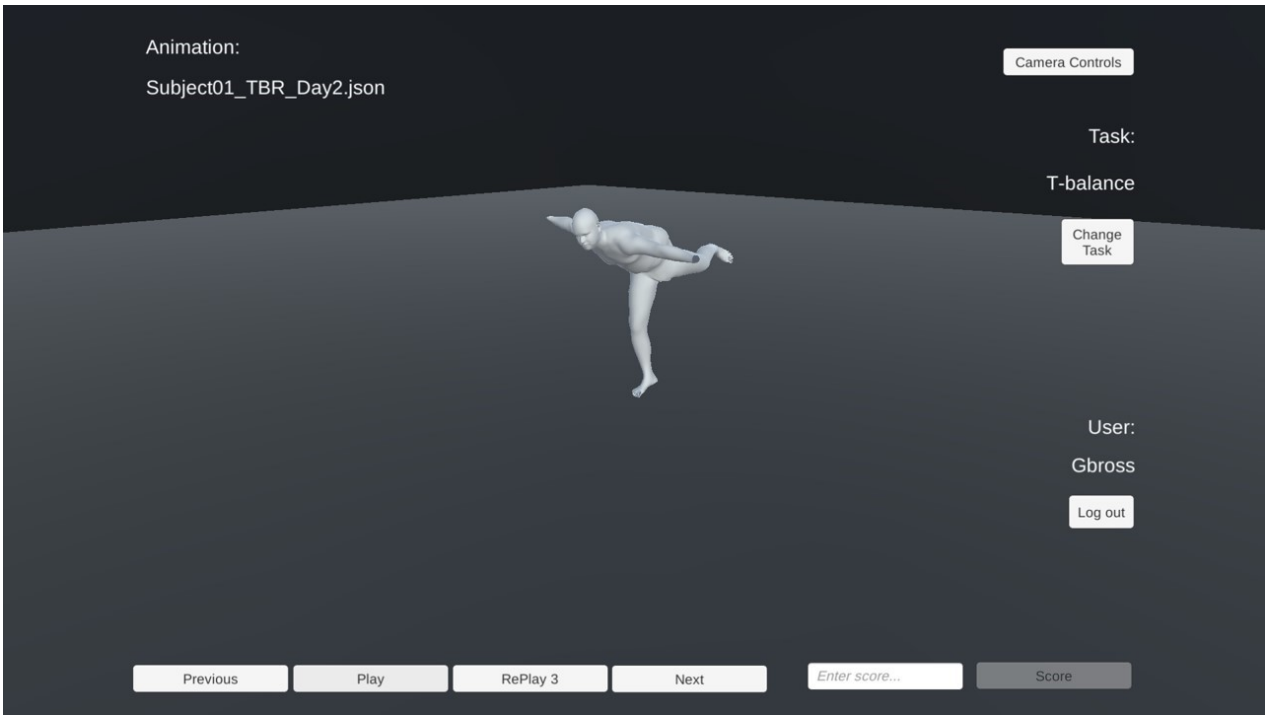


Figure 4.3.3. A screenshot of the custom visualisation tool user interface.

The study consisted of three modules: Training, Day 1 and Day 2. Before beginning to score movements, raters partook in the training module, where five animations for each movement were at their disposal to study. To select training module animations, two pilot raters completed Day 1 of the protocol and animations were chosen that had complete agreement between the two raters. Since the training module animations were part of the testing database, depending on the movement task, the training animations either had a score of {1, 3, 5, 7, or 9} or {2, 4, 6, 8, or 10}, to minimize the number of animations the raters were exposed to prior to the start of the study. In order to minimize bias, the pilot raters' scores were shown, but explanations for each score were not provided. Raters were asked to use their training and expertise to determine their own scoring criteria based on whole-body kinematics of the given training animations. The raters were able to return to the training module at any time during the study and were able to replay the animations as many times as they liked.

For the Day 1 and Day 2 module, raters scored each animation from 1-10 based on the animation's movement competency for each movement task. In order to decrease the risk of fatigue, raters did not have to complete all modules in one sitting but were able to complete them at their own pace. In addition, the raters were able to score the movements in whichever order they chose. The Day 1 and Day 2 modules had identical animations; however, the order in which the

animations were presented within each task were different between the two days. To decrease the risk of a learning effect, raters had to wait a minimum of 48 hours after completing the Day 1 module of the movement task before starting the Day 2 module of the same movement task. Raters were only able to replay each movement three times at real-time speed, but had the ability to zoom, translate, and rotate the vantage point during the movement. The limited number of replays was to decrease the risk of recall bias, especially since many of the movements were duplicates. If a rater submitted multiple scores for the same animation, the last score was taken. After completing Day 1 and Day 2 module, the true purposes of the study were disclosed, and the raters signed a post-study consent form that confirmed their acknowledgment and understanding of the use of deception in the study and permission to use their data.

4.3.4 Statistics

To test inter- and intra-rater reliability, the arithmetic means of weighted Cohen's kappa were used. For inter-rater reliability, comparisons between each rater and the mean of the 44 weighted Cohen's kappa values were calculated for each movement task. For the intra-rater reliability between sessions, weighted Cohen's kappa was calculated for each rater (except Rater 3 who only completed Day 1) between the exact same movements for Day 1 and Day 2. Both the individual and mean kappa values were retained for each movement task. For intra-rater reliability within session without body-shape manipulation, weighted Cohen's kappa was calculated between the 30 unique movements for each rater for each day, resulting in 19 kappa values (10 raters for Day 1 + 9 raters for Day 2) for each movement task. For looking at intra-rater reliability within session when body-shape was manipulated, weighted Cohen's kappa between each weight class for the 10 unique movements per day was calculated resulting in three kappa values (Overweight-Normal, Overweight-Underweight, Normal-Underweight) per rater per day. The kappa values were then averaged within raters. Weighted Cohen's kappa values were interpreted as no (≤ 0), slight (0.01-0.20), fair (0.21 -0.40), moderate (0.41-0.60), substantial (0.61-0.80), and almost perfect (0.81-1.00) agreement (McHugh, 2012).

4.4 RESULTS

For inter-rater reliability, there was slight to fair agreement with kappa values ranging from 0.09 (bird-dog) – 0.33 (lunge) across all tasks (Table 4.4.1). For intra-rater reliability between sessions, across raters, there was fair to moderate agreement with kappa values ranging from 0.27

(L-hop) – 0.46 (step-down; Table 4.4.2). For intra-rater reliability within session without body-shape manipulation, there was fair to moderate agreement with kappa values ranging from 0.33 (step-down) – 0.58 (T-balance) across all tasks (Table 4.4.3). For intra-rater reliability within session with body-shape manipulation, there was slight to moderate agreement with kappa values ranging from 0.17 (drop-jump) to 0.52 (Lunge) across all tasks (Table 4.4.4).

When looking at the individual rater level, for intra-rater reliability between sessions averaged across movement tasks, raters had fair to moderate agreement, with kappa values ranging from 0.22 (Rater 2) – 0.50 (Rater 5; Table 4.4.2). For intra-rater reliability within sessions without body-shape modification, reliability ranged from fair to moderate reliability with kappa values ranging from 0.31 (Rater 2, Day 1) – 0.57 (Rater 6, Day 1; Table 4.4.3). For intra-rater reliability within session with body-shape modification, reliability ranged from slight to moderate agreement with kappa values ranging from 0.2 (Rater 7, Day 2) - 0.54 (Rater 6, Day 2; Table 4.4.4).

4.5 DISCUSSION

The purpose of this study was to examine the inter-rater and intra-rater reliability of movement competency scores during a movement screen between and within sessions using a customized visualisation tool and to assess the effects of body-shape on reliability. It was hypothesized for that intra-rater reliability within the same session without body-shape manipulation would have been the highest, followed by intra-rater reliability between session and intra-rater reliability within the same session with body-shape manipulation, and inter-rater reliability would have the lowest agreement. The results support these hypotheses, where all forms of intra-rater ranged from slight to moderate agreement and inter-session reliability showed slight to fair agreement when looking across movement tasks and raters. For intra-rater reliability, when looking at the average kappa values, intra-rater reliability within session with and without body manipulation and between sessions were 0.45, 0.37, and 0.35, respectively.

Compared to other studies, the kappa values were on the lower end of the spectrum; however, the pattern of intra-rater reliability being better than inter-rater reliability was similar to previous results (Glaws et al., 2014; Onate et al., 2012; Smith et al., 2013). The lower scores could be due to the larger number of possible scores, the greater number of raters being compared, or the difference in scoring criteria. The FMS is scored between 0-3 for each task (Cook et al., 2014, 2006), whereas the movements for this study were scored between 1-10. With the greater number

Table 4.4.1. The arithmetic mean of weighted Cohen's kappa across all raters and days for inter-rater reliability, intra-rater reliability between sessions (InterSessions), intra-rater reliability within sessions without body-shape modification (IntraSession), and intra-rater reliability within session with body-shape modifications (BodyShape) for each task and the average and standard deviation across tasks. BD = bird-dog, DJ = drop-jump, HD = hop-down, LH = L-hop, LG = lunge, SD = step-down, and TB = T-balance.

	BD		**DJ**		**HD**		**LH**		**LG**		**SD**		**TB**		**Average**	**STD**
	Day 1	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*		
InterRater	0.09	0.10	0.14	0.14	0.13	0.12	0.14	0.11	0.31	0.33	0.21	0.25	0.22	0.18	**0.18**	**0.08**
Between Sessions	0.51		0.32		0.23		0.14		0.42		0.48		0.37		**0.35**	**0.13**
Within Session	0.47	0.46	0.43	0.48	0.41	0.35	0.36	0.46	0.48	0.51	0.33	0.39	0.54	0.58	**0.45**	**0.07**
BodyShape	0.41	0.41	0.17	0.26	0.29	0.29	0.39	0.50	0.52	0.46	0.34	0.35	0.40	0.35	**0.37**	**0.10**

Table 4.4.2. The weighted Cohen's kappa for intra-rater reliability between sessions for each rater for each task and the average and standard deviation (STD) across tasks and across raters. BD = bird-dog, DJ = drop-jump, HD = hop-down, LH = L-hop, LG = lunge, SD = step-down, and TB = T-balance.

Subject	**BD**	**DJ**	**HD**	**LH**	**LG**	**SD**	**TB**	**Average**	**STD**
1	0.35	0.29	0.23	0.25	0.32	0.43	0.44	**0.33**	**0.08**
2	0.03	0.27	0.23	0.16	0.31	0.20	0.36	**0.22**	**0.11**
3									
4	0.51	0.24	0.30	0.32	0.56	0.67	0.36	**0.43**	**0.16**
5	0.47	0.48	0.43	0.30	0.57	0.61	0.63	**0.50**	**0.12**
6	0.38	0.37	0.43	0.54	0.52	0.39	0.64	**0.47**	**0.10**
7	0.28	0.19	0.29	0.20	0.40	0.33	0.36	**0.29**	**0.08**
8	0.51	0.32	0.23	0.14	0.42	0.48	0.37	**0.35**	**0.13**
9	0.45	0.30	0.33	0.35	0.52	0.58	0.46	**0.43**	**0.10**
10	0.35	0.49	0.21	0.20	0.44	0.43	0.47	**0.37**	**0.12**
Average	**0.37**	**0.33**	**0.30**	**0.27**	**0.45**	**0.46**	**0.45**	**0.38**	
STD	**0.15**	**0.10**	**0.09**	**0.13**	**0.10**	**0.15**	**0.11**		**0.14**

Table 4.4.3. The arithmetic mean of weighted Cohen's kappa for intra-rater reliability within session without body-shape modification for each rater for each task and session day, as well as, the average and standard deviation (STD) across raters and tasks. BD = bird-dog, DJ = drop-jump, HD = hop-down, LH = L-hop, LG = lunge, SD = step-down, and TB = T-balance.

Rater	**BDR**		**DJ**		**HDR**		**LHR**		**LG**		**SDR**		**TBR**		**Average**		**STD**	
	Day 1	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	***Day 1***	***Day 2***	***Day 1***	***Day 2***
1	0.48	0.42	0.62	0.45	0.37	0.33	0.39	0.54	0.32	0.64	0.51	0.43	0.52	0.60	**0.46**	**0.49**	**0.11**	**0.11**
2	0.49	0.27	0.26	0.31	0.30	0.12	0.08	0.40	0.36	0.28	0.22	0.48	0.45	0.41	**0.31**	**0.32**	**0.14**	**0.12**
3	0.41		0.60		0.26		0.34		0.53		0.25		0.47		**0.41**		**0.13**	
4	0.51	0.46	0.36	0.51	0.59	0.41	0.32	0.60	0.45	0.47	0.43	0.31	0.50	0.56	**0.45**	**0.48**	**0.09**	**0.10**
5	0.34	0.52	0.47	0.58	0.44	0.43	0.37	0.50	0.60	0.52	0.39	0.43	0.59	0.66	**0.46**	**0.52**	**0.10**	**0.08**
6	0.52	0.54	0.61	0.42	0.56	0.43	0.55	0.68	0.69	0.49	0.32	0.60	0.74	0.72	**0.57**	**0.55**	**0.14**	**0.12**
7	0.50	0.42	0.33	0.50	0.32	0.34	0.21	0.40	0.34	0.35	0.02	0.30	0.58	0.47	**0.33**	**0.40**	**0.18**	**0.07**
8	0.55	0.49	0.17	0.51	0.40	0.32	0.36	0.09	0.43	0.59	0.35	0.34	0.46	0.58	**0.39**	**0.42**	**0.12**	**0.18**
9	0.40	0.49	0.37	0.47	0.35	0.39	0.50	0.41	0.53	0.72	0.50	0.41	0.51	0.54	**0.45**	**0.49**	**0.07**	**0.11**
10	0.49	0.57	0.54	0.56	0.52	0.42	0.49	0.53	0.52	0.50	0.33	0.18	0.52	0.65	**0.49**	**0.49**	**0.07**	**0.15**
Average	**0.47**	**0.46**	**0.43**	**0.48**	**0.41**	**0.35**	**0.36**	**0.46**	**0.48**	**0.51**	**0.33**	**0.39**	**0.54**	**0.58**	**0.45**			
STD	**0.06**	**0.09**	**0.16**	**0.08**	**0.11**	**0.10**	**0.14**	**0.17**	**0.12**	**0.14**	**0.14**	**0.12**	**0.09**	**0.10**			**0.13**	

Table 4.4.4. The arithmetic mean of weighted Cohen's kappa for intra-rater reliability within session with body-shape modification for each rater for each task and session day, as well as, the average and standard deviation (STD) across raters and tasks. BD = bird-dog, DJ = drop-jump, HD = hop-down, LH = L-hop, LG = lunge, SD = step-down, and TB = T-balance.

Rater	**BDR**		**DJ**		**HDR**		**LHR**		**LG**		**SDR**		**TBR**		**Average**		**STD**	
	Day 1	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*	*Day 1*	*Day 2*
1	0.51	0.38	0.69	0.23	0.46	0.31	0.49	0.65	0.47	0.37	0.48	0.52	0.43	0.30	**0.50**	**0.40**	**0.09**	**0.14**
2	0.27	0.24	0.12	0.15	0.15	0.37	0.05	0.42	0.56	0.42	0.07	0.07	0.34	0.30	**0.22**	**0.28**	**0.18**	**0.14**
3	0.48		0.00		0.04		0.57		0.34		0.25		0.30		**0.28**		**0.21**	
4	0.70	0.69	0.12	0.27	0.50	0.60	0.42	0.47	0.72	0.69	0.73	0.50	0.42	-0.10	**0.52**	**0.45**	**0.22**	**0.28**
5	0.35	0.36	0.18	0.25	0.13	0.39	0.48	0.34	0.55	0.61	0.47	0.33	0.42	0.53	**0.37**	**0.40**	**0.16**	**0.12**
6	0.14	0.53	0.09	0.06	0.63	0.51	0.55	0.66	0.58	0.79	0.16	0.62	0.41	0.60	**0.37**	**0.54**	**0.23**	**0.23**
7	0.19	0.25	0.15	0.37	0.05	-0.02	0.47	0.45	0.44	0.00	0.17	0.03	0.22	0.33	**0.24**	**0.20**	**0.15**	**0.20**
8	0.63	0.50	-0.03	0.16	0.21	0.16	0.45	0.60	0.58	0.35	0.47	0.50	0.45	0.28	**0.40**	**0.36**	**0.23**	**0.18**
9	0.46	0.39	0.13	0.20	0.38	0.23	0.28	0.26	0.43	0.37	0.16	0.10	0.38	0.45	**0.32**	**0.29**	**0.13**	**0.12**
10	0.36	0.33	0.20	0.64	0.33	0.07	0.15	0.65	0.56	0.54	0.42	0.51	0.63	0.45	**0.38**	**0.46**	**0.18**	**0.20**
Average	**0.41**	**0.41**	**0.17**	**0.26**	**0.29**	**0.29**	**0.39**	**0.50**	**0.52**	**0.46**	**0.34**	**0.35**	**0.40**	**0.35**	**0.37**			
STD	**0.18**	**0.15**	**0.20**	**0.17**	**0.20**	**0.20**	**0.18**	**0.15**	**0.10**	**0.23**	**0.21**	**0.23**	**0.11**	**0.20**			**0.20**	

of possible scores, there is greater sensitivity; however, the probability of raters selecting the same score is decreased. In addition, the sensitivity may be greater, but the human eye may not be able to distinguish the differences. Previous studies compared 2 (Onate et al., 2012), 3 (Glaws et al., 2014) and 4 (Smith et al., 2013) raters, whereas this study compared 10 raters. The increase in number of raters, due to needing to align more raters, may also contribute to the lower kappa values. Although the greater range in scores and number of raters likely contributed to the lower kappa values, the main reason was likely due to the scoring criteria.

For the movement screens that previously assessed inter-rater and intra-rater reliability, given task-specific scoring criteria were used to assess movement competency (Glaws et al., 2014; Onate et al., 2012; Smith et al., 2013), whereas for this study, the raters were asked to use their expertise to establish their own whole-body scoring criteria. Previous research has criticized the FMS for having poor criterion validity, which was attributed to the vagueness of the scoring criteria (Whiteside et al., 2014). In addition, many of the FMS task-specific scoring criteria are not linked (epidemiologically or biomechanically) to injury mechanisms or risk factors (Frost et al., 2015b) and individuals were able to increase their scores when made aware of the scoring criteria (Frost et al., 2015a). Furthermore, due to the large amount of movement variability between athletes, the FMS scoring criteria may be insensitive to potentially risky movement behavior, with previous research recommending that whole-body segment and joint kinematics should be incorporated when administering movement screens (Frost et al., 2015b). Therefore, for this study we opted to test the reliability of movement competency scores during the movement screen without task-specific scoring criteria, which likely led to the lower reliability scores compared to previous research.

4.5.1 Intra-Rater Reliability

For intra-rater reliability, as hypothesized, the within session without body-shape modification had the highest reliability compared to within session with body-shape modification and between session reliability. The within session without body-shape modification were identical movements and avatars, which the raters would have seen sometimes only two animations previously, therefore since there was a shorter duration between rescoring the two movements compared to the between session, which had a minimum of 48 hours between rescoring the animations, it was expected that the within session intra-rater reliability would be higher than

the between session reliability. Similar results have been reported when comparing intra-rater reliability within and between session for the Soccer Injury Movement Screen (McCunn et al., 2017). The poor intra-rater reliability of movement scores when scoring identical animations suggests that visual observation is not precise when observing whole-body kinematics. This may be due to the large amount of information the rater needs to observe, process, and analyze the movements in a short amount of time (Onate et al., 2012). In addition, due to the raters scoring multiple animations at once, the low intra-rater reliability may be because of a bias due to the influence of previously seen animations, where the bias may change with every new animation seen.

For the intra-rater reliability within session with body-shape modification, for average results within tasks across raters, the reliability was worse than the intra-rater reliability within session without body-shape modification. The without body-shape modification animations had identical movements and avatars, whereas, for the with body-shape modification the avatars looked different, which may contribute to the lower reliability. In addition, research has consistently shown that there is pervasive implicit and explicit weight bias among clinicians, physical therapists, physical education teachers, and strength and conditioning personnel (Panza, 2018).

When looking across raters, differences in average kappa values between without body-shape modification and with body-shape modification ranged from -0.02 to 0.17, with a negative value indicating better agreement with body-shape manipulation. The single rater who had a slight increase in reliability with body-shape manipulation also had just over double the amount of variability in scores across tasks compared to the without body-shape modification condition, suggesting that the observed differences were most likely attributed to the large amount of variability seen across all conditions. The range in differences in kappa values between the two conditions suggests that some raters were more affected by body-shape than others. These differences are likely due to rater bias, which has been well documented with research suggesting that the rater bias can account for just as much variance of scores as differences in the examinee's ability (Lumley and McNamara, 1993). Biases can be conscious or unconscious with common types of biases including: leniency bias (inflating scores due to feeling sympathetic towards the ratee), contrast bias (evaluating by comparing to previous person), central tendency bias (preferring to give an average, middle rating despite performance), similar to me bias (inflating

scores based on rater feeling similar to ratee), personal bias (scoring based on personal beliefs and ideologies), and halo effect (rating based only on one good aspect, despite the rest of the performance). Therefore, the variance in effect of body-shape on reliability between raters is likely due to differences in the type and severity of biases in effect.

4.5.2 Implications

On average across raters, body-shape had a negative effect on reliability compared to without body-shape modification with differences likely due to rater bias. With MoSh, one can manipulate personal characteristics of the animation, while maintaining movement patterns. Therefore, MoSh in combination with the customized visualisation software can provide a tool to minimize personal and similar to me bias by being able to standardize personal characteristics which may be biasing raters' scores such as: body-shape, facial expressions, gender expression, and race. In addition, the developed visualisation software can be modified to accommodate other virtual characters/environments or other animation/media file types (e.g., video) for future research/collaborations.

For both inter- and intra-rater reliability, at best for this study, there is fair reliability, which suggests that assessing movement competency via subjective assessment without specific scoring criteria is not a reliable method. However, as mentioned previously, the inclusion of scoring criteria has its own limitations, such as lack of sensitivity (Frost et al., 2015b). Previous research has discussed the importance of developing objective methods, tools, and thresholds to assess movement competency, which this study reinforces (Frost et al., 2015b; Whiteside et al., 2014).

4.5.3 Limitations

Two limitations of this study were the use of an identical dataset between Day 1 and Day 2 and the use of new software. The use of two identical datasets between Day 1 and Day 2 may have led to some learning effects. To try to combat the learning effects, raters had to wait a minimum of 48 hours between finishing Day 1 and starting Day 2. When looking between days for the inter-rater reliability, on average, there was no difference in kappa values between Day 1 and Day 2; whereas there was an average decrease in kappa values from Day 1 and 2 for both intra-session reliability without body-shape manipulation and intra-session reliability with body-shape manipulation of 0.02. Based on these results, a learning effect does not appear to be

influencing the results and differences seen at the individual or task level are more likely due to the large variability seen across all conditions. In addition, the use of a new software and the use of avatars rather than 2D video of human participants may have been influencing their scoring abilities. However, the software was very intuitive to use, each rater had the ability to familiarize themselves with the software and controls in the training session before starting the testing sessions, raters could access the control descriptions during any point in testing, and anecdotally, no raters mentioned any difficulty of using the program. The researchers believe that the benefits that the program provided such as the ability to have 360° views of the athletes and the ability to modify body-shape outweighed the potential of the use of the software and avatars potentially minimally affecting the rater's scores.

4.5.4 Conclusion

In conclusion, inter- and intra-rater reliability were low, with agreement ranging from slight to fair, suggesting that assessing movement competency via subjective assessment is not reliable when there are not task-specific scoring criteria. This is further compounded when athletes with different body-shape types are being assessed, with reliability decreasing on average across raters when body-shape was manipulated. This study supports previous literature which argues for the need for the development of objective methods, tools, and thresholds to assess movement competency (Frost et al., 2015b; Whiteside et al., 2014).

4.6 SUMMARY

- To test inter- and intra-rater reliability, a customized, online, visualisation software was used, which allowed for 360° views of the raters and for the effects of body-shape on reliability to be studied.
- Inter-rater and intra-rater reliability for movement screens without scoring criteria range from slight to fair, with body-shape having a negative effect on reliability.
- Objective methods, tools, and thresholds should be developed to reliably assess movement competency.

4.7 ACKNOWLEDGEMENTS

The authors would like to thank Motus Global for use of the dataset, all of the athletes for their participation, Reza Hajari for his help with developing the custom software, and Alexandre Mir-Orefice for his assistance with the development of the animation database.

4.8 COMPETING INTERESTS

The authors declare that there are no competing interests and that they results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

4.9 FUNDING

This research was funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada (PGSD3 - 504132 – 2017; Gwyneth Ross).

4.10 REFERENCES

Anderlucci, L., Lubisco, A., Mignani, S., 2021. Investigating the judges performance in a national competition of sport dance. Soc. Indic. Res. 156, 783–799.

Cook, G., Burton, L., Hoogenboom, B., 2006. Pre-participation screening: the use of fundamental movements as an assessment of function–Part 1. North Am. J. Sport. Phys. Ther. NAJSPT 1, 62. doi:10.1055/s-0034-1382055

Cook, G., Burton, L., Hoogenboom, B.J., 2014. Functional movement screening : The use of fundamental movements as an assssment of function- Part 2. Int. J. Sports Phys. Ther. 9, 549–563. doi:10.1111/j.1600-0838.2010.01267.x

Donà, G., Preatoni, E., Cobelli, C., Rodano, R., Harrison, A.J., 2009. Application of functional principal component analysis in race walking: An emerging methodology. Sport. Biomech. 8, 284–301. doi:10.1080/14763140903414425

Frost, D.M., Beach, T.A., Callaghan, J.P., McGill, S.M., 2015a. FMS^{TM} scores change with performers' knowledge of the grading criteria - Are general whole-body movement screens capturing "dysfunction"? J. Strength Cond. Res. 29, 3037–3044. doi:10.1519/JSC.0b013e3182a95343

Frost, D.M., Beach, T.A.C., Campbell, T.L., Callaghan, J.P., Mcgill, S.M., 2015b. An appraisal of the Functional Movement Screen^{TM} grading criteria - Is the composite score sensitive to risky movement behavior? Phys. Ther. Sport 16, 324–330. doi:10.1016/j.ptsp.2015.02.001

Glaws, K.R., Juneau, C.M., Becker, L.C., Di Stasi, S.L., Hewett, T.E., 2014. Intra- and inter-rater reliability of the Selective Functional Movement Assessment (SFMA). Int. J. Sports Phys. Ther. 9, 195–207.

Kritz, M., Cronin, J., Hume, P., 2009. The bodyweight squat: A movement screen for the squat pattern. Strength Cond. J. 31, 76–85. doi:10.1519/SSC.0b013e318195eb2f

Loper, M., Mahmood, N., Black, M.J., 2014. MoSh : Motion and shape capture from sparse markers. ACM Trans. Graph. 33, 1–13.

Loper, M., Mahmood, N., Romero, D., Pons-Moll, G., Black, M.J., 2015. SMPL: A skinned multi-person linear model. ACM Trans. Graph. 34, 1–16.

Lumley, T., McNamara, T.F., 1993. Rater characteristics and rater bias: Implications for training. Lang. Test. 12, 54–71.

McCall, A., Carling, C., Nedelec, M., Davison, M., Le Gall, F., Berthoin, S., Dupont, G., 2014. Risk factors, testing and preventative strategies for non-contact injuries in professional football: Current perceptions and practices of 44 teams from various premier leagues. Br. J. Sports Med. 48, 1352–7. doi:10.1136/bjsports-2014-093439

McCunn, R., aus der Fünten, K., Fullagar, H.H.K., McKeown, I., Meyer, T., 2016. Reliability and association with injury of movement screens: A critical review. Sport. Med. 46, 763–781. doi:10.1007/s40279-015-0453-1

McCunn, R., Fünten, K. Der, Govus, A., Julian, R., Schimpchen, J., Meyer, T., 2017. The intra- and inter-rater reliability of the Soccer Injury Movement Screen (SIMS). Int. J. Sport Phys. Ther. 12, 53–66.

McHugh, M.L., 2012. Interrater reliability: The kappa statistic. Biochem. Medica 22, 276–282.

Onate, J.A., Dewey, T., Kollock, R.O., Thomas, K.S., Van Lunen, B.L., DeMaio, M., Ringleb, S.I., 2012. Real-time intersession and interrater reliability of the Functional Movement Screen. J. Strength Cond. Res. 26, 408–415. doi:10.1519/JSC.0b013e318220e6fa

Padua, D.A., Marshall, S.W., Boling, M.C., Thigpen, C.A., Garrett, W.E., Beutler, A.I., 2009. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. Am. J. Sports Med. 37, 1996–2002. doi:10.1177/0363546509343200

Panza, G.A., 2018. Weight bias among exercise and nutrition professionals: A systematic review. Obeisity Rev. 19, 1492–1503. doi:10.1111/obr.12743

Smith, C.A., Chimera, N.J., Wright, N.J., Warren, M., 2013. Interrater and intrarater reliability of the Functional Movement Screen. J. Strength Cond. Res. 27, 982–987.

Whiteside, D., Deneweth, J.M., Pohorence, M.A., Sandoval, B., Russell, J.R., McLean, S.G., Zernicke, R.F., Goulet, G.C., 2014. Grading the Function Movement Screen: A comparison of manual (real-time) and objective methods. J. Strength Cond. Res. 30, 924–933.

Chapter 5: Study 2

OBJECTIVELY DIFFERENTIATING MOVEMENT PATTERNS BETWEEN ELITE AND NOVICE ATHLETES

Original Article

Gwyneth B. Ross[1], Brittany Dowling[2], Nikolaus F. Troje[3], Steven L. Fischer[4], and Ryan B. Graham[1,4]

[1]School of Human Kinetics, Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada

[2]Motus Global, Rockville Centre, New York, USA

[3]Centre of Vision Research & Department of Biology, York University, Toronto, Ontario, Canada

[4]Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada

Published by: Medicine & Science in Sports & Exercise (2018, Vol. 50, Issue 7, Pgs. 1457-1464)

5.1 ABSTRACT

Introduction: Movement screens are frequently used to identify abnormal movement patterns that may increase risk of injury or hinder performance. Abnormal patterns are often detected visually based on the observations of a coach or clinician. Quantitative, or data-driven methods can increase objectivity, remove issues related to inter-rater reliability and offer the potential to detect new and important features that may not be observable by the human eye. Applying principal components analysis (PCA) to whole-body motion data may provide an objective data-driven method to identify unique and statistically important movement patterns, an important first step to objectively characterize optimal patterns or identify abnormalities. Therefore, the primary purpose of this study was to determine if PCA could detect meaningful differences in athletes' movement patterns when performing a non-sport-specific movement screen. As a proof of concept, athlete skill level was selected *a priori* as a factor likely to affect movement performance.

Methods: Motion capture data from 542 athletes performing seven dynamic screening movements (i.e., bird-dog, drop jump, T-balance, step-down, L-hop, hop-down, and lunge) were analyzed. A PCA-based pattern recognition technique and linear discriminant analysis with cross-validation were used to determine if skill level could be predicted objectively using whole-body motion data.

Results: Depending on the movement, the validated linear discriminant analysis models accurately classified 70.66-82.91% of athletes as either elite or novice.

Conclusion: We have provided proof that an objective data-driven method can detect meaningful movement pattern differences during a movement screening battery based on a binary classifier (i.e., skill level in this case). Improving this method can enhance screening, assessment and rehabilitation in sport, ergonomics and medicine.

5.2 INTRODUCTION

Movement screens are used to identify aberrant movement patterns believed to increase risk of injury and/or impede performance. Individuals' kinetic and/or kinematic data can be recorded as they perform standardized motions, where those data can be evaluated to determine the correctness and/or proficiency of movement (Cook et al., 2014; Donà et al., 2009; Kritz et al., 2009; McCall et al., 2014; McCunn et al., 2016; Padua et al., 2009). However, movement screening is most commonly applied as a visual appraisal of movement (McCunn et al., 2016). While visual appraisal of movement is useful in terms of coaching, data-driven, quantitative methods could enhance movement screening by increasing objectivity (Federolf et al., 2014; McCunn et al., 2016) and reducing error associated with visual-based appraisal. Data-driven methods may also create the potential to detect new and important movement features that may not be easily visible to the human eye (Federolf et al., 2014). This study explores the viability of a principal component analysis (PCA)-based movement pattern recognition technique as a data-driven approach to objectively characterize movement during a movement screen.

Movement screens are growing in popularity amongst practitioners, coaches and athletes to identify abnormal movements (Cook et al., 2006; Frost et al., 2012; Harris-Hayes and Van Dillen, 2009; Kritz et al., 2009; McCunn et al., 2016). The Functional Movement Screen (FMS™; (Functional Movement Systems, Chatham, VA, USA) offers one example of a movement screen whose use has recently increased in popularity amongst coaches and trainers and where a number of studies have identified prospective challenges with inter- and intra-rater reliability (Gribble et al., 2013; Gulgin and Hoogenboom, 2014; Minick et al., 2010; Onate et al., 2012; Smith et al., 2013) . The FMS™ is a quantitative movement screen in that movements receive a numerical score; however, it is a subjective, quantitative measure because that score is derived based on visual observations by a rater (Cook et al., 2014, 2006). Consistent with typical movement screens and the movement screen used in this study, the FMS™ is comprised of movements believed to challenge mobility and stability by requiring participants to move into and out of extreme positions using a self-controlled pace. There is agreement within the literature, that when looking at the overall scores, there is strong inter-rater and intra-rater reliability (Gribble et al., 2013; Gulgin and Hoogenboom, 2014; Minick et al., 2010; Onate et al., 2012; Smith et al., 2013) for both novice and experienced raters. However, inter-rater reliability is poor for some individual movements

such as the hurdle step (Onate et al., 2012; Smith et al., 2013), as well, the hurdle step had poor intersession reliability (participants were tested on two days, separated by a week; Onate et al., 2012). In addition, when comparing multiple pairings of novice and experienced raters, only half of the movements (n = 6) had perfect agreement, a quarter of the movements had moderate agreement (n = 3), and a quarter of the movements had slight agreement (n = 3; Gulgin and Hoogenboom, 2014). The poor inter-rater reliability for some movements is thought to be due to the dynamic nature of the movement(s) (Onate et al., 2012) and the rater's perspective, where they may only see the performance from one vantage point, making it difficult for the rater to see scoring criteria that are either out of view or occluded by the athlete's body (Onate et al., 2012; Smith et al., 2013). A possible solution to address the aforementioned limitations is the use of motion capture.

Motion capture systems are able to objectively track 3D motions. Motion capture data can be collected to an accuracy of 0.2 mm during static movements and 2 mm during dynamic movements (however error continues to decrease as technology improves), likely offering greater resolution to detect movement differences compared to the visual observation (Eichelberger et al., 2016; Merriaux et al., 2017). Furthermore, 3D motion capture enables the movement to be assessed from multiple vantage points, and can be used to collect data in order to calculate measures that can provide athletes and coaches with quantitative, objective feedback of movement performance (Federolf et al., 2014). However, whole-body 3D motion capture data are extremely high-dimensional, which can be difficult to analyze and interpret. Reducing data to discrete parameters or by applying other data reduction methods, such as PCA can help in analyzing and interpreting motion capture data. The use of PCA for motion analysis, enabling the extraction of fundamental movement patterns, has dramatically increased (Federolf et al., 2014; Masoud et al., 2017; Pion et al., 2014; Troje, 2002a).

PCA is a multivariate statistical technique that can be applied to reduce the dimensionality of high-dimensional data sets. PCA can reduce dimensionality by identifying redundancies in the data, where a subset of the data will often explain the majority of the variance (Troje, 2002a). While PCA has been applied to joint waveforms (e.g., angles and moments; Deluzio and Astephen, 2007; Wrigley et al., 2006), it can also be applied directly to 3D motion capture marker data (Troje, 2002a). This allows for whole-body kinematics to be analyzed instead of individual joints, for the

data to be reconstructed in a visually meaningful way and to account for body-size. As an example, walking can be characterized using motion capture data by describing the 3D positions of multiple reflective markers at each point in time. Applying PCA, an individual's walking pattern can be more efficiently described using the first four principal components, which account for 98% of the total variance in the overarching 3D positional data (Troje, 2002a). Using PCA to examine 3D motion capture data, (i.e., all marker trajectories) can help accurately identify embedded patterns of complex movements.

PCA-based movement pattern recognition techniques have been effectively used to detect and explain differences in whole-body movements. For example differences in gait patterns were detected and explained based on factors including age, body mass index, sex, and feelings (happy/sad, nervous/relaxed; Baydal-Bertomeu et al., 2016; Troje, 2008, 2002a, 2002b). When applied to ergonomics, egress/ingress motion patterns were detected and explained based on vehicle design (Masoud et al., 2017). In a sport example, PCA was applied objectively to analyze down-hill (Federolf et al., 2014) and cross-country skiing (Gløersen et al., 2017) technique, and to develop an objective judging tool for competitive diving (Young and Reinkensmeyer, 2014). An emerging benefit of this technique is its ability to provide objective biomechanical feedback to coaches and athletes to use when planning the training of athletes (Federolf et al., 2014). For example, a golf coach can show how a golfer's full body movement patterns and timing of movements during their golf swing compare to that of a professional golfer. However, a less obvious, secondary benefit of the technique is that it can be used to support the development of instructors' or coaches' skills to assess an athlete's movements (Federolf et al., 2014). For example, PCA can identify principal positions/patterns that can be highlighted to coaches, who can then ensure their athletes attain those positions/patterns during practice. PCA-based movement pattern recognition techniques also allowed researchers to characterize movement features during diving, where dive scores were calculated based on how a given feature pattern related to the mean (Young and Reinkensmeyer, 2014), reducing the inherent variability in traditional observational-based judging. PCA applied to whole-body 3D motion capture data offers researchers, coaches, and judges with a method to objectively assess and even score movement patterns of athletes as they perform their specific movement skills, removing subjectivity/bias from traditional observational-based assessment.

PCA-based movement pattern recognition shows promise but has not been applied to a movement screen. Objectively classifying athletes' movement patterns based on certain demographics (e.g., height, weight, skill level, sport played, and injury history) could provide a unique data-driven approach to enhance movement screening, training, and rehabilitation. Therefore, the purpose of the current study was to assess the application of a PCA-based pattern recognition technique as a method to differentiate whole-body movement patterns of athletes using skill level as a dichotomous factor. We chose this factor due to its likelihood to influence movement quality and performance. Further, an example is provided discussing how this data-driven approach could be applied as a coaching tool to inform corrective strategies. Based on previous research, it was hypothesized that a PCA-driven movement pattern recognition technique would be able to effectively differentiate between the movements of athletes of different skill levels while performing movements that are not specific to the athletes' particular discipline.

5.3 METHODS

5.3.1 Participants

Motion capture data were collected from 542 athletes by Motus Global (Rockville Centre, NY, USA). The sample included athletes competing in baseball, basketball, soccer, golf, tennis, track and field, squash, cricket, lacrosse, football, or volleyball. Athletes' skill level varied from recreational to professional (e.g., NBA, MLB, NFL, PGA, FIFA). Athletes were dichotomized based on skill level (-1 = novice, 1 = elite) based on the cut-off that elite athletes should have accumulated at least 10,000 hours of deliberate practice (Baker et al., 2003; Helsen et al., 1998). Thus, elite athletes were considered as those competing at the collegiate, semi-professional, and professional levels and novice athletes were considered as those competing at all other levels (e.g., high-school, recreational, etc.). Table 5.3.1 summarizes the demographics and skill level of participants. Prior to data collection, each participant read and signed an informed consent form permitting Motus Global to use the data for future research. The Health Sciences Research Ethics Board at Queen's University approved the secondary use of the data for research purposes (File No: 6017208).

5.3.2 Protocol

Upon arrival at the Motus Global laboratory, participants read and signed the consent form. Once the consent form was signed, participants' height (with shoes on) and weight were measured. Athletes were outfitted with 45 passive, reflective markers (B&L Engineering, Santa Ana, CA, USA) placed on specific body locations as required to characterize whole-body motion. Markers (n=37) were placed on anatomical landmarks including: the spinal process of the second thoracic vertebrae (T_2); the spinal process of the eighth thoracic vertebrae (T_8); xyphoid process; the front, left, right, and back of the head; and, bilaterally on the acromioclavicular joint, sternoclavicular joint, medial and lateral epicondyle of the humerus, ulnar styloid process, radial styloid process, anterior and posterior superior iliac spine, greater

Table 5.3.1. Athletes' mean age, height, and weight broken down by sex and skill level. Standard deviations are in brackets.

		Elite	**Novice**	**Elite & Novice**
Male	n	306	174	480
	Age	22.5 (3.5)	16.8 (4.2)	20.05(4.7)
	Height	191.5 (19.8)	174.7 (13.0)	185.3 (19.4)
	Weight	95.7 (18.7)	69.7 (18.0)	86.1 (22.3)
Female	n	12	50	62
	Age	23.3 (4.0)	18.0 (5.6)	19.00 (5.7)
	Height	170.4 (7.2)	167.0 (8.7)	168.00 (8.9)
	Weight	64.3 (8.8)	60.0 (13.7)	61.1 (13.3)
Male & Female	n	318	224	542
	Age	22.6 (3.6)	17.1 (4.6)	20.2 (4.7)
	Height	190.7 (19.9)	172.9 (12.6)	183.3 (19.3)
	Weight	94.5 (19.4)	67.5 (17.6)	83.1 (22.9)

trochanter, medial and lateral epicondyle of the femur, medial and lateral malleolus, distal end of the third metatarsal, and calcaneus, as required to describe major body segments (McPherson et al., 2016). Additional tracking markers (n=8; one marker placed on each thigh, each forearm, each bicep, the right shank, and the right scapula) were applied to assist with tracking the segments and also to aid in differentiating between the left and right sides. Once outfitted with the markers, athletes performed static and dynamic calibration trials. During the static calibration trial, the athlete stood with their feet shoulder width apart and toes pointing straight forward. The arms were

abducted 90° with a 90° bend in the elbow. The static calibration was used to align the athlete to the laboratory global coordinate system as well as to define local coordinate systems specific to each athlete. For the dynamic calibration trial, the athlete held the same position as the static calibration trial and then rotated his or her arms 90° both internally and externally.

Each athlete then completed a series of movement tests. The battery consisted of 21 unique movements; however, only seven were used in this analysis including the: bird-dog, drop jump, T-balance, step-down, L-hop, hop-down, and lunge movement (see Appendix B2, which visually shows and provides detailed explanations for each movement). These seven movements were selected as those most likely to challenge mobility and stability across the shoulder, spine, hip, knee, and ankle joints. With the exception of the bilateral drop jump, all movements were performed on both the right and left sides for a total of 13 motion profiles per athlete. Full-body motion data were captured at 120 Hz using an 8-camera Raptor-E (Motion Analysis, Santa Rosa, CA, USA) motion capture system.

5.3.3 Data Analysis

5.3.3.1 Pre-Processing

Motion capture data were collected, labelled, and gap-filled using Cortex (Motion Analysis, Santa Rosa, CA, USA). Data from the anatomical landmarks and the tracking markers during the calibration trial were used to develop a 3D, whole-body kinematic model in Visual3D (C-Motion, Inc., Germantown, MD, USA). The model was then applied to all motion trials outputting three types of positional data: joint centers bilaterally for the wrist, elbow, shoulder, foot, ankle, knee, and hip; centers of gravity for the trunk, head, and pelvis; and marker positional data for the left and right heel, T_2, T_8, sternum, and the back, front and sides of the head to model the feet, trunk and head more robustly when reconstructing the data. Data were exported to Matlab (The MathWorks, Natick, MA, USA) for further analysis. In Matlab, all trials were clipped to specific start and end-point criteria (see Appendix B, Table B1, which outlines the specific start and end criteria for each trial), filtered using a dual-pass, low-pass Butterworth filter with a cut-off of 15 Hz, and time normalized to 500 frames to control for differences in the absolute movement time.

5.3.3.2 Application of PCA as a Movement Pattern Recognition Technique

A PCA-based movement pattern recognition technique was used to analyze the data. Using the motion data from each movement, an *n x m* matrix was created for each movement; where *n* represented the number of athletes contributing data and *m* represented the time-series motion data for each participant. Noted in Table 5.4.1, the number of athletes contributing data to each movement varied as trials were removed from analysis where athletes did not perform a particular movement or where marker occlusions did not permit a whole-body representation for the duration of the movement. Time-series motion data were expressed as the x, y and z positional data for each aforementioned joint center, segment center of gravity and retained marker (26 positions x 3 axes) for all 500 time points, such that m = 39,000 (Troje, 2002b). Separate PCAs were applied to the movement data specific to each movement. On average, the sample size for the movement tasks was 365. Using the suggested methodology of 10 participants per predictor in the linear discriminant model (Field, 2009), the first 35 principal components (PC) and associated scores for every participant were retained. Since elites were statistically taller than novices ($F= 138.25$, $p<0.001$), all x,y,z positional data for each movement were normalized by each athlete's individual height by dividing each raw data point by their own height. This ensured that differences in PC scores between groups were not strictly due to variation in body size.

5.3.4 Statistical Analysis

Linear discriminant analyses (LDA) with leave-one-out cross-validation and without validation were used to determine the ability to differentiate athletes based on an athlete's skill level (elite or novice) for each movement individually. For each movement, all 35 PC scores retained from the PCA were used as predictors for the model. Lacking a whole new set of data to test the linear classifier, we ran a leave-out-one validation procedure, where one of the athletes' data were taken out (test athlete) and a linear classifier was computed on the remaining athletes (training athletes). After having done so, the test athlete was projected first onto the principal components derived from the training athletes and then the resulting score vector was then projected onto the discriminant function in the subspace spanned by the first 35 components (Troje, 2002a). Data were reconstructed such as presented in Figure 5.4.3, using the linear discriminant functions for each movement (Troje, 2002a). The reconstructed data were calculated using the equation:

$$r = m + \alpha d \quad (15)$$

where r is the reconstructed movement data, m is the mean movement data across all athletes, and d is the linear discriminant function. The scalar function α is scaled in terms of standard deviations (Troje, 2002a), where positive values emphasize features characterizing elites and negative values characterizing novices. As α changes from a negative integer to a positive integer, the reconstructed data is more representative of an elite athlete (Troje, 2002a). For visualization purposes (e.g., Figures 5.4.3 and Appendix B, which show the differences in movement patterns between elite and novice athletes for each movement), α was set at 1 to best show differences between the two groups.

5.3.5 Application of the Movement Pattern Recognition Technique

For a practical application of the movement pattern recognition technique, two athletes (one novice and one elite) from the existing database were extracted at random to classify how they scored along the linear discriminant function in terms of Z-scores on either side of the mean. Z-scores were be calculated by:

$$Z = \frac{X - \mu}{\sigma} \quad (16)$$

where Z is the athlete's Z-score, X is the athlete's testing projection (the distance between the athlete's score and the linear discriminant function), μ is the mean testing projection across all athletes, and σ is the standard deviation of the mean testing projection across all athletes. From the Z-score, using a normal confidence distribution function with μ equaling 0 and σ equaling 1, the likelihood of being an elite athlete for each movement was calculated. A likelihood of <50% represents more novice-like movements, whereas a likelihood of >50% represents more elite-like movement. An individualized movement report describing the likelihood (percentage) that the athlete was elite for each movement was created.

5.4 RESULTS

When 35 PCs were retained, the explained variance for each movement was greater than 99.8% (Table 5.4.1). The PCA-pattern recognition technique was able to accurately classify between 77.89% (bird-dog left) and 88.36% (drop-jump) of athletes as either elite or novice depending on the movement when not using any validation (Table 5.4.1). When leave-one-out

validation was used, the PCA-pattern recognition technique was able to accurately classify between 70.66% (T-balance left) and 82.91% (drop-jump; Figure 5.4.4) of athletes depending on the movement (Table 5.4.1).

Table 5.4.1. The number of athletes, percent explained variance and percentage of correctly classified athletes as either elite or novice using a linear discriminant analysis model with and without leave-one-out validation for each movement.

		Male		**Female**			**% of Correctly Classified Athletes**	
Movement	*n*	*Elite*	*Novice*	*Elite*	*Novice*	*PEV (%)*	*No Val. (%)*	*LOO Val. (%)*
Bird-Dog Left	380	242	83	12	43	99.95	77.89	72.63
Bird-Dog Right	387	244	88	11	44	99.95	80.62	75.97
Drop-Jump	275	168	64	7	36	99.91	88.36	82.91
Hop-Down Left	396	242	99	10	45	99.94	81.31	76.77
Hop-Down Right	396	242	97	11	46	99.93	80.30	76.26
L-Hop Left	266	159	67	6	34	99.91	80.83	76.69
L-Hop Right	267	160	67	6	34	99.91	87.27	80.15
Lunge Left	399	246	97	12	44	99.87	81.45	74.69
Lunge Right	401	248	97	12	44	99.88	79.55	76.31
Step-Down Left	399	246	98	12	43	99.95	82.21	78.20
Step-Down Right	399	247	96	11	45	99.95	81.20	75.94
T-Balance Left	392	244	92	11	45	99.92	78.57	70.66
T-Balance Right	395	244	94	12	45	99.93	79.49	73.67

PEV = Percentage of explained variance; Val.= Validation; LOO = Leave-One-Out

For all tasks, the classification rate (Figure 5.4.1), sensitivity (percent likelihood of an elite being classified as an elite) and specificity (percent likelihood of a novice athlete being classified as a novice; Figure 5.4.2) increase as more PCs are retained. Although some tasks level off after retaining 35 PCs, some tasks (bird-dog right, drop-jump, hop-down left and right, L-hop left and right, and T-balance left) continue to increase with the more PCs retained, meaning that some of the classification rates, sensitivity and specificity could improve.

Reconstructed motion data (using LDFs) allowed visualization of the differences between the elite and novice athletes' movement patterns. For example, during the T-balance right, elite athletes demonstrated greater hip flexion at the beginning of the movement (Figure 5.4.3A), greater forward hip rotation such that their trunk and left leg was more parallel with the ground at the mid-point of the movement (5.4.3B), and greater hip flexion again at the end of the movement (Figure 5.4.3C; see Appendix B, which visually shows and provides detailed descriptions of

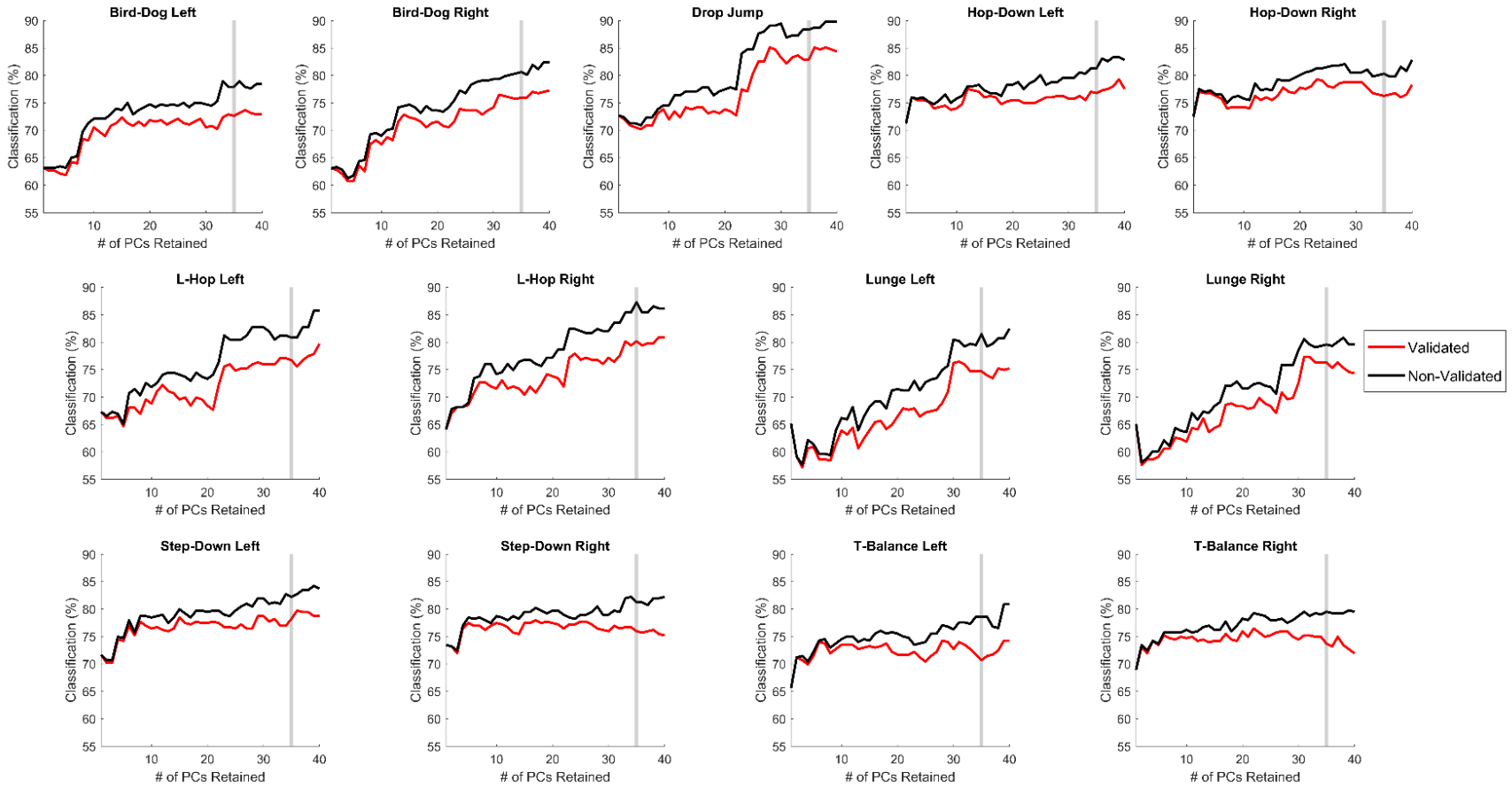


Figure 5.4.1. The percent of correctly classified athletes for when 1 to 40 PC scores were retained for the linear discriminant analysis models with leave-one-out validations (red) and no validation (black) for each movement task. The vertical gray line represents the number of PCs retained for this study (35 PCs).

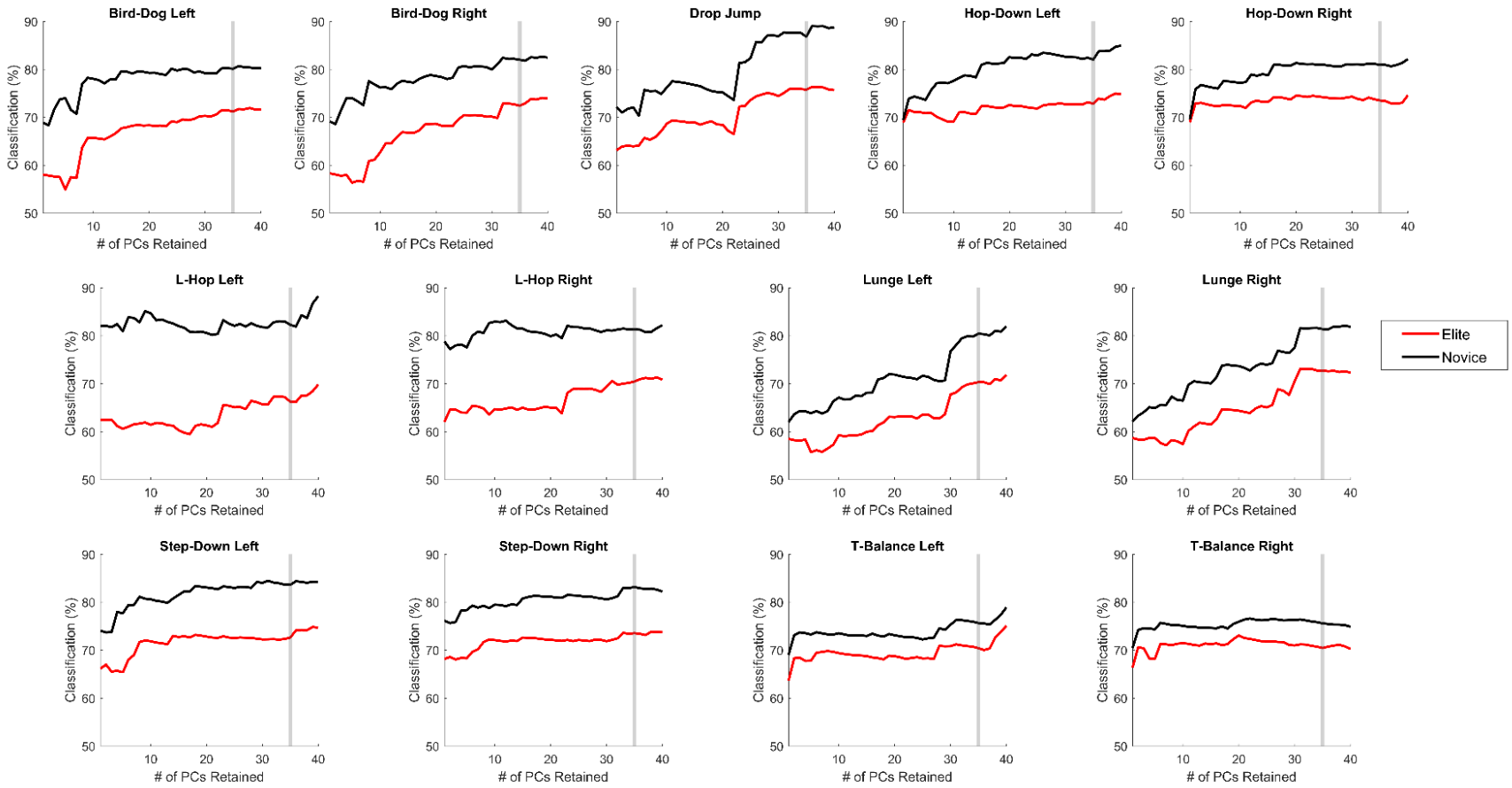


Figure 5.4.2. The sensitivity (percent likelihood of correctly classifying an elite athlete; red line) and the specificity (percent likelihood of correctly classifying a novice athlete; black line) for when 1 to 40 PCs scores were retained for each movement task. The vertical gray line represents the number of PCs retained for this study (35 PCs).

differences between skill levels for all movements). Although comparisons rely on visual observation, the differences observed are detected by the data-driven approach. Visualizations provide insight to what the differences are that discriminate between elite and novice athletes. They can also be used to provide feedback to athletes who aim to improve their performance.

The model accurately classified athletes as either elite or novice, so a practical application of the technique was explored. When interpreting the reports, assuming the criterion is to move like an elite athlete, a lower percentage represents poorer movement performance (more novice-like), whereas a higher percentage represents superior movement performance (more elite-like).

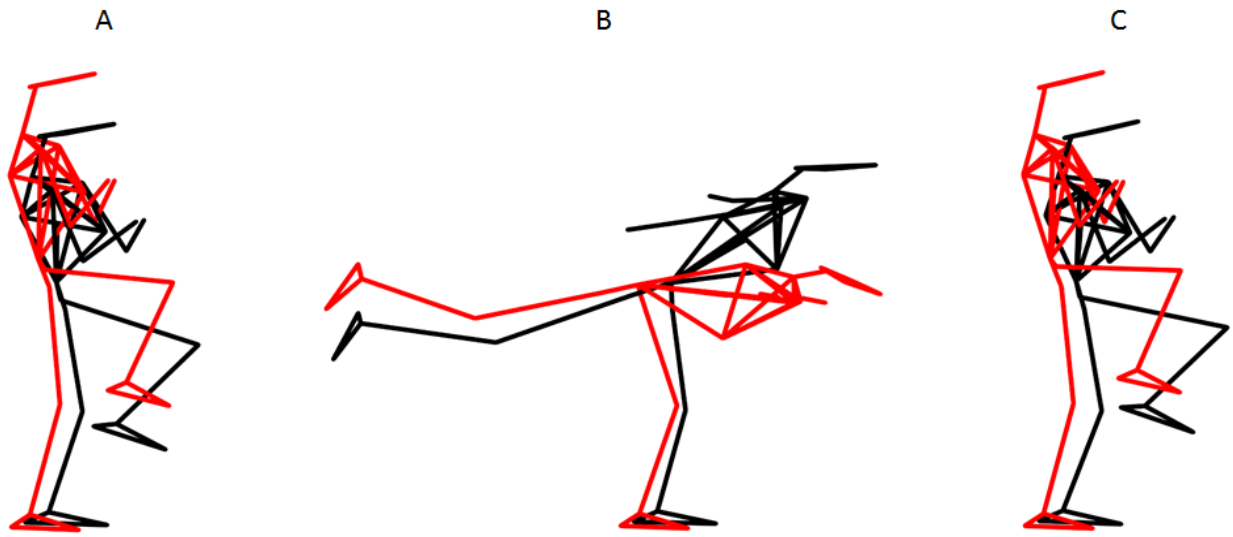


*Figure 5.4.3. Reconstruction of the linear discriminant function differentiating elite and novice athletes during the T-balance right movement at 0, 50 and 100% of the movement. Data are reconstructed using the equation: reconstructed data = mean data ± 1*linear discriminant function. Red represents elite athletes, black represents novice athletes.*

The two randomly selected athletes were an elite basketball player and a novice golfer. Depending on the task, the elite basketball player percent likelihood of being an elite athlete ranged from 63.61% (drop-jump) to 99.10% (lunge left; Figure 5.4.4). For the novice golfer, their percent likelihood of being an elite athlete ranged from 3.42% (drop-jump) to 82.15% (lunge left).

5.5 DISCUSSION

The purpose of this study was to assess the ability of a PCA-based pattern recognition technique to differentiate whole-body movement patterns between novice and elite athletes when

performing non-sport-specific movement screening movements. The two-staged approach applied PCA to reduce the dimensionality of the complex 3D motion capture data where PC scores were used to develop linear discriminant analysis models for classification. The approach yielded an objective, data-driven model to classify athletes' performance during individual movement screening movements, where the results supported the hypothesis that movement patterns could be differentiated on the basis of athletes' skill level.

The ability to classify athletes accurately using linear discriminant analysis was movement dependent. The movement dependency of model classification accuracy may be explained by considering the challenge, constraints and strength needed with each specific movement. On average, the movement tasks that involved jumping (drop-jump, L-hop, and hop-down) had the highest percentage of correctly classified athletes. When looking at the reconstructed data, there are not large differences in distances jumped compared to height. However, there are differences in technique. For the drop-jump (see Appendix B, Figure B2, which visually shows differences between skill levels during the drop-jump), the elite athletes have a greater range of motion of their arm swing which may result in a greater take-off velocity (Lees et al., 2004). For the hop-down

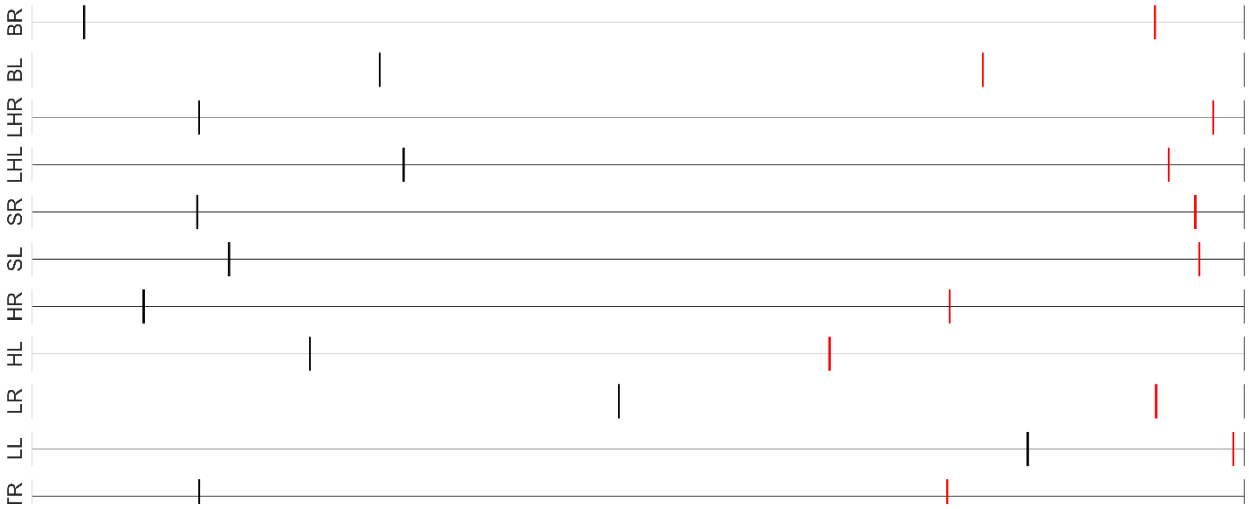


Figure 5.4.4. An example movement report created for a novice golfer (black) and professional basketball player (red) based on the linear discriminant analysis model and individual PC scores. The scores describe percent likelihood of the athlete based on the athlete's performance for each movement. Each horizontal bar represents a 0-100% likelihood for a movement. BR = bird-dog right; BL = bird-dog left; LHR = L-hop right; LHL = L-hop left; SR = step-down right; SL = step-down left; HR = hop-down right; HL = hop-down left; LR = lunge right; LL = lunge left; TR = T-balance right; TL = T-balance left; DJ = drop jump

(see Appendix B, Figure B3, which visually shows the difference between skill levels during the hop-down), elite athletes have greater range of motion of their non-stance leg during the transition from landing to take-off than novice athletes. The non-stance leg is thought to work similarly to the arms during a max vertical jump, where an increase in swing increases upwards momentum. For the L-hop (see Appendix B, Figure B4, which visually shows the differences between skill levels during the L-hop), elite athletes achieve greater jump distance when performing the lateral portion of the jump than the novice athletes. This could be due to the elite athletes achieving a deeper squat when landing from the forward jump, allowing for elite athletes to achieve greater distance during the lateral jump (Domire and Challis, 2017). In addition, elite athletes better minimized vertical jump height when jumping for maximal horizontal difference than novice athletes, demonstrating a more efficient jump.

In addition to the jumping tasks, the step-down movements on average had high classification rates. During data collection, athletes often commented on the step-down being the most difficult. It was also the movement that was most often re-collected due to a failure to correctly complete the movement. This suggests that the step-down challenged athletes' mobility and stability more than the other movements. When observing the reconstructed movement patterns, clear differences between elite and novice athletes can be seen. During the step-down movement, elite athletes employed a different movement strategy, moving their center of mass posteriorly to complete the movement.

Since the model accurately classified athletes as either elite or novice, a practical application of the technique was explored. The elite basketball player's movement report is highlighted as the red lines in Figure 5.4.4. The elite basketball player had a greater than 50% likelihood of being an elite athlete based on their performance in every movement, however they performed the poorest on the drop-jump and hop-down left. The poor drop-jump and hop-down left performance may be indicative of underlying dysfunction or poor movement, which could prompt the coach or trainer to probe for possible reasons causing the deficiency (e.g., lack of stability, muscle imbalances, poor range of motion). In contrast, the novice golfer generally scored poorly across all movements except for the lunges (Figure 5.4.4), indicative of the golfer's novice status. However, the golfer performed poorest on the drop-jump (3.42% chance of being an elite athlete), which may prompt the coach or trainer to focus on movements related to drop-jump

performance. This application of the technique demonstrates how an objective, data-driven method can be used to detect relevant performance differences within a movement screen, in this case using a scoring spectrum related to athlete skill level.

The model developed here improves upon the limitations associated with previous subjective movement screens, such as scores being at the discretion of the test administrator, only observing from one vantage point (Onate et al., 2012; Smith et al., 2013), and the test administrator needing to be able to assess multiple scoring criteria at once during quick, dynamic movements (Onate et al., 2012). The outcome scoring variables proposed in this study are purely objective (although the assessment of differences between the two groups is subjective), scores are not reliant on the discretion of the test administrator and data are collected from multiple cameras at different vantage points allowing for a 360º view of the athlete throughout the movement.

A limitation on the application of this approach for athlete training is the use of movement patterns of elite athletes as the gold standard. Although this is a limitation, because the optimal movement patterns for these movements are unknown, the use of elite athletes as the gold standard is an appropriate option due to their extensive athletic training. To identify optimal movement patterns, future research should focus on using inverse and forward dynamics and optimal-control models to try and identify movement patterns (using the PC scores) that are optimal based on reduced joint loading and minimizing cost functions. Based on results from the models, individuals' movement patterns can be identified on a scale that denotes risk of injury. The identification of optimal movements could lead to better training programs to help decrease injury and improve performance. In addition, because the current methods require an expensive motion capture system, we are currently investigating the use of inexpensive motion capture systems (i.e., inertial movement sensors, inexpensive cameras) to make the use of these methods more accessible to athletes, coaches, and clinicians.

In conclusion, a novel pattern recognition technique using PCA was able to accurately classify athletes based on level of expertise during a movement battery. This technique can potentially be utilized for the training and rehabilitation of athletes and in other fields including ergonomics. Future research should examine other classifiers (i.e., sport played, injury history) and validate the use of inexpensive motion capture systems. In addition, the use of inverse and

forward dynamics and optimal-control models to try to identify common strategies and movements to reduce joint loading and minimize cost functions should be explored.

5.6 ACKNOWLEDGEMENTS

We would like to thank our funding sources: The Natural Sciences and Engineering Research Council (NSERC) of Canada, the Queen Elizabeth II Graduate Scholarships in Science and Technology (QEII-GSST), and the University of Ottawa School of Graduate Studies. In addition, we would like to thank the athletes who partook in the study.

5.7 CONFLICT OF INTEREST

The results of the present study do not constitute endorsement by ACSM. In addition, the authors declare that there are no conflicts of interest. Although Motus is a for-profit company and the Director of Research at Motus, Brittany Dowling, is an author on the paper, the approach used in this paper is an objective, data-driven approach and is not only applicable to this data set. The results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation

5.8 REFERENCES

Baker, J., Cote, J., Abernethy, B., 2003. Sport-specific practice and the development of expert decision-making in team ball sports. J. Appl. Sport Psychol. 15, 12–25. doi:10.1080/10413200305400

Baydal-Bertomeu, J.M., Durá-Gil, J.V., Piérola-Orcero, A., Parrilla Bernabé, E., Ballester, A., Alemany-Munt, S., 2016. A PCA-based bio-motion generator to synthesize new patterns of human running. PeerJ Comput. Sci. 2, e102. doi:10.7717/peerj-cs.102

Cook, G., Burton, L., Hoogenboom, B., 2006. Pre-participation screening: the use of fundamental movements as an assessment of function–Part 1. North Am. J. Sport. Phys. Ther. NAJSPT 1, 62. doi:10.1055/s-0034-1382055

Cook, G., Burton, L., Hoogenboom, B.J., 2014. Functional movement screening : The use of fundamental movements as an assssment of function- Part 2. Int. J. Sports Phys. Ther. 9, 549–563. doi:10.1111/j.1600-0838.2010.01267.x

Deluzio, K.J., Astephen, J.L., 2007. Biomechanical features of gait waveform data associated with knee osteoarthritis. An application of principal component analysis. Gait Posture 25, 86–93. doi:10.1016/j.gaitpost.2006.01.007

Domire, Z.J., Challis, J.H., 2017. The influence of squat depth on maximal vertical jump performance 414. doi:10.1080/02640410600630647

Donà, G., Preatoni, E., Cobelli, C., Rodano, R., Harrison, A.J., 2009. Application of functional principal component analysis in race walking: An emerging methodology. Sport. Biomech. 8, 284–301. doi:10.1080/14763140903414425

Eichelberger, P., Ferraro, M., Minder, U., Denton, T., Blasimann, A., Krause, F., Baur, H., 2016. Analysis of accuracy in optical motion capture – A protocol for laboratory setup evaluation. J. Biomech. 49, 2085–2088. doi:10.1016/j.jbiomech.2016.05.007

Federolf, P., Reid, R., Gilgien, M., Haugen, P., Smith, G., 2014. The application of principal component analysis to quantify technique in sports. Scand. J. Med. Sci. Sport. 24, 491–499. doi:10.1111/j.1600-0838.2012.01455.x

Field, A., 2009. Discovering Statistics Using SPSS, 3rd ed. SAGE Publications Ltd., London, UK.

Frost, D., Andersen, J., Lam, T., Finlay, T., Darby, K., McGill, S., 2012. The relationship between general measures of fitness, passive range of motion and whole-body movement quality. Ergonomics 139, 1–16. doi:10.1080/00140139.2011.620177

Gløersen, Ø., Myklebust, H., Hallén, J., Federolf, P., 2017. Technique analysis in elite athletes using principal component analysis. J. Sports Sci. 0, 1–9. doi:10.1080/02640414.2017.1298826

Gribble, P.A., Brigle, J., Piertrosimone, B.G., Pfile, K.R., Webster, K.A., 2013. Intrarater reliability of the function movement screen. J. Strength Cond. Res. 27, 978–981.

Gulgin, H., Hoogenboom, B., 2014. The Functional Movement Screening (FMS)^{TM}: An inter-rater reliability study Between raters of varied experience. Int. J. Sports Phys. Ther. 9, 14–20.

Harris-Hayes, M., Van Dillen, L.R., 2009. The inter-tester reliability of physical therapists classifying low back pain problems based on the Movement System Impairment Classification System. PM R 1, 117–126. doi:10.1016/j.pmrj.2008.08.001

Helsen, W.F., Starkes, J.L., Hodges, N.J., 1998. Team sports and the theory of deliberate practice. J. Sport Exerc. Psychol. 20, 12–34.

Kritz, M., Cronin, J., Hume, P., 2009. The bodyweight squat: A movement screen for the squat pattern. Strength Cond. J. 31, 76–85. doi:10.1519/SSC.0b013e318195eb2f

Lees, A., Vanrenterghem, J., De Clercq, D., 2004. Understanding how an arm swing enhances performance in the vertical jump 37, 1929–1940. doi:10.1016/j.jbiomech.2004.02.021

Masoud, H.I., Zerehsaz, Y., Jin, J., 2017. Analysis of human motion variation patterns using UMPCA. Appl. Ergon. 59, 401–409. doi:10.1016/j.apergo.2016.09.016

McCall, A., Carling, C., Nedelec, M., Davison, M., Le Gall, F., Berthoin, S., Dupont, G., 2014. Risk factors, testing and preventative strategies for non-contact injuries in professional football: Current perceptions and practices of 44 teams from various premier leagues. Br. J. Sports Med. 48, 1352–7. doi:10.1136/bjsports-2014-093439

McCunn, R., aus der Fünten, K., Fullagar, H.H.K., McKeown, I., Meyer, T., 2016. Reliability and

association with injury of movement screens: A critical review. Sport. Med. 46, 763–781. doi:10.1007/s40279-015-0453-1

McPherson, A.L., Dowling, B., Tubbs, T.G., Paci, J.M., 2016. Sagittal plane kinematic differences between dominant and non-dominant legs in unilateral and bilateral jump landings. Phys. Ther. Sport 22, 54–60.

Merriaux, P., Dupuis, Y., Boutteau, R., Vasseur, P., Savatier, X., 2017. A study of Vicon system positioning performance. Sensors 17, 1591–1608. doi:10.3390/s17071591

Minick, K.I., Kiesel, K.B., Burton, L., Taylor, A., Plisky, P., Butler, R.J., 2010. Interrater Reliability of the Functional Movement Screen. Strength Cond. 24, 479–486.

Onate, J.A., Dewey, T., Kollock, R.O., Thomas, K.S., Van Lunen, B.L., DeMaio, M., Ringleb, S.I., 2012. Real-time intersession and interrater reliability of the Functional Movement Screen. J. Strength Cond. Res. 26, 408–415. doi:10.1519/JSC.0b013e318220e6fa

Padua, D.A., Marshall, S.W., Boling, M.C., Thigpen, C.A., Garrett, W.E., Beutler, A.I., 2009. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. Am. J. Sports Med. 37, 1996–2002. doi:10.1177/0363546509343200

Pion, J., Fransen, J., Lenoir, M., Segers, V., 2014. The value of non-sport-specific characteristics for talent orientation in young male judo, karate and taekwondo athletes. Arch. Budo 10, 147–154.

Smith, C.A., Chimera, N.J., Wright, N.J., Warren, M., 2013. Interrater and intrarater reliability of the Functional Movement Screen. J. Strength Cond. Res. 27, 982–987.

Troje, N.F., 2008. Retrieving information from human movement patterns, in: Thomas, F., Zacks, J. (Eds.), Understanding Events: From Perception to Action. Oxford University Press, Oxford, pp. 308–335.

Troje, N.F., 2002a. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. J. Vis. 2, 371–387. doi:10.1167/2.5.2

Troje, N.F., 2002b. The little difference: Fourier based synthesis of gender-specific biological

motion, in: Würtz, R., Lappe, M. (Eds.), Dynamic Perception. Aka Press, Berlin, pp. 115–120.

Wrigley, A.T., Albert, W.J., Deluzio, K.J., Stevenson, J.M., 2006. Principal component analysis of lifting waveforms. Clin. Biomech. 21, 567–578. doi:10.1016/j.clinbiomech.2006.01.004

Young, C., Reinkensmeyer, D.J., 2014. Judging complex movement performances for excellence: A principal components analysis-based technique applied to competitive diving. Hum. Mov. Sci. 36, 107–122. doi:10.1016/j.humov.2014.05.009

Chapter 6: Study 3

CLASSIFYING ELITE FROM NOVICE ATHLETES USING SIMULATED WEARABLE SENSOR DATA

Original Article

Gwyneth B. Ross[1], Brittany Dowling[2], Nikolaus F. Troje[3], Steven L. Fischer[4], and Ryan B. Graham[1,4]

[1]School of Human Kinetics, Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada

[2]Motus Global, Rockville Centre, New York, USA

[3]Centre of Vision Research & Department of Biology, York University, Toronto, Ontario, Canada

[4]Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada

Published by: Frontiers in Bioengineering and Biotechnology (2020, Volume 8, Article 814)

6.1 ABSTRACT

Introduction: Movement screens are frequently used to identify differences in movement patterns such as pathological abnormalities or skill related differences in sport; however, abnormalities are often visually detected by a human assessor resulting in poor reliability. Therefore, our previous research has focused on the development of an objective movement assessment tool to classify elite and novice athletes' kinematic data using machine learning algorithms. Classifying elite and novice athletes can be beneficial to objectively detect differences in movement patterns between the athletes, which can then be used to provide higher quality feedback to athletes and their coaches. Currently, the method requires optical motion capture, which is expensive and time-consuming to use, creating a barrier for adoption within industry. Therefore, the purpose of this study was to assess whether machine learning could classify athletes as elite or novice using data that can be collected easily and inexpensively in the field using inertial measurement units (IMUs). A secondary purpose of this study was to refine the architecture of the tool to optimize classification rates.

Methods: Motion capture data from 542 athletes performing seven dynamic screening movements were analyzed. A principal component analysis (PCA)-based pattern recognition technique and machine learning algorithms with the Euclidean norm of the segment linear accelerations and angular velocities as inputs were used to classify athletes based on skill level.

Results: Depending on the movement, using metrics achievable with IMUs and a linear discriminant analysis (LDA), 75.1-84.7% of athletes were accurately classified as elite or novice.

Conclusion: We have provided evidence that suggests our objective, data-driven method can detect meaningful differences during a movement screening battery when using data that can be collected using IMUs, thus providing a large methodological advance as these can be collected in the field using sensors. This method offers an objective, inexpensive tool that can be easily implemented in the field to potentially enhance screening, assessment, and rehabilitation in sport and clinical settings.

6.2 INTRODUCTION

Movement screens are widely used across many disciplines including in ergonomic, clinical, and athletic settings to identify aberrant movement patterns in hopes of decreasing risk of injury and/or improving performance (Cook et al., 2014; Donà et al., 2009; Kritz et al., 2009; McCall et al., 2014; McCunn et al., 2016; Padua et al., 2009). Most commonly, during a movement screen, an individual's movement is evaluated based on visual appraisal (McCunn et al., 2016); however, there is agreement within the literature that inter-rater and inter-session (participants tested during two separate sessions) reliability of these subjective movement screens are poor (Gulgin and Hoogenboom, 2014; Onate et al., 2012; Smith et al., 2013). Therefore, our previous research focused on the development and application of an objective framework as a data-driven alternative to objectively classify movement strategies and quality during a movement screen (Ross et al., 2018), known as the Objective Movement Assessment Tool (OMAT).

The previously published technique with optical motion capture, herein referred to as OMAT-OPT, uses principal component analysis (PCA; Federolf et al., 2014; Troje, 2002; Young and Reinkensmeyer, 2014) in conjunction with linear discriminant analysis (LDA) to objectively differentiate and score whole-body movement patterns between desired binary classifiers (Ross et al., 2018). For OMAT-OPT, the data input into the PCA are time-series trajectories of joint centers and select anatomical markers, representing the whole-body, captured using an optical motion capture system. During a non-sport-specific movement screening battery consisting of seven unique dynamic movements that challenge stability and mobility across all major joints, between 70.7% and 82.9% of athletes were appropriately classified as either elite or novice depending on the movement (Ross et al., 2018). Although OMAT-OPT provides an objective, data-driven method that can detect meaningful movement pattern differences during a movement screening battery for binary classification, it requires optical motion capture technology, which is expensive and time-consuming to set up, capture and post-process data, reducing the accessibility and feasibility of the current technique in clinical, ergonomic, and sport settings (Hadjidj et al., 2013).

The use of wearable systems are increasing in popularity in clinical, sport, and ergonomic settings (Hadjidj et al., 2013; Patel et al., 2012), offering an inexpensive alternative to optical motion capture systems. The wearable systems are easily transportable, require minimal post-processing, are able to collect data in larger capture volumes compared to optical systems, and are immune to problems associated with optical systems such as occlusion and line-of-sight problems

(Zhou and Hu, 2008). A common type of sensor used is the inertial measurement unit (IMU). IMUs contain an accelerometer, gyroscope, and magnetometer, allowing measurement of linear accelerations and angular velocities in three axes and the triaxial magnetic fields of the earth. IMUs are susceptible to drift, especially when close to metal, although more robust algorithms are continuously being developed to mitigate these effects (Madgwick et al., 2011; Wittmann et al., 2019), making them more suitable for use in the field.

IMU data have been used to objectively classify movement based on different classifiers during non-sport specific tasks (Johnston et al., 2019, 2016; Francesco Sgro et al., 2017; Zago et al., 2019). Machine learning with IMU data as the input has been able to objectively identify children of different motor development levels during a standing long jump (Francesco Sgro et al., 2017), rugby players at a higher risk of a sport-related concussion based on a Y-balance test (Johnston et al., 2019), Australian football players at different levels of fatigue during a Y-balance test (Johnston et al., 2016), and to predict change of direction, speed, and mechanical work during cutting maneuvers (Zago et al., 2019), to name a few. Although these studies only looked at a single IMU placed on the low-back of the participant, these findings suggest that IMUs can be used as an inexpensive alternative to optical motion capture to characterize and classify motion.

Although research using machine learning to classify elite and novice athletes is limited, discriminant analysis has been previously used to classify novice, good, and elite rowers during ergometer testing (Smith and Spinks, 1995). The ability to objectively differentiate movement patterns between novice and elite athletes is useful to highlight emergent differences in movement performance. Guided by those differences, coaches can improve quality of feedback to their athletes (Smith and Spinks, 1995). We chose skill level as the dichotomous factor to initially assess due to its likelihood to influence movement quality and performance, with the intention of in the future expanding to sex, sport played, and injury history or risk.

Feature selection approaches and machine learning algorithms may also influence the accuracy of classification between elite and novice athletes using IMU data and are therefore important secondary considerations. Previously, the OMAT-OPT used the first 35 principal component (PC) scores as the input data for the LDA; however, alternative feature selection approaches could provide an objective method to best decide which PC scores to use as input data to maximize classification. Ensemble feature selection, which is based on the same ideology of ensemble supervised classifiers, is a useful approach to evaluate. Ensemble feature selection

includes the use of multiple feature selection algorithms to select features and has been found to have greater stability (i.e., less likelihood of features changing if data are added or removed) and better generalizability than using a single feature selection technique (Saeys et al., 2008). In addition, the OMAT currently uses LDA, which was selected due to superior performance during testing. However, it is unknown whether LDA would still garner the highest classification rates when using PC scores selected by an ensemble feature selection approach, rather than the first 35 PC scores and/or when using IMU data. Alternative machine learning algorithms including binary logistic regression (BLR), decision trees (DT), K-nearest neighbors (kNN), naïve Bayes (NB), support vector machine with a linear kernel (SVM), and support vector machine with a radial basis function kernel (RBF) may strengthen classification accuracy relative to our existing LDA approach. As a result, while investigating the utility of IMUs to classify movements between novice and elite athletes, it remains important to concurrently evaluate the underlying machine learning model architecture required to generate the best possible classification.

Therefore, the purpose of this study was to assess the ability of the previously developed framework to differentiate whole-body movement patterns between novice and elite athletes performing a non-sport-specific movement screening battery when using data extractable from an IMU (i.e., simulated IMU data; OMAT-sIMU), which can be collected easily and inexpensively in the field. Although data in the current study are simulated IMU data based on optical motion capture, this study provides proof-of-concept that IMU-based data can provide enough information to successfully classify athletes' movement patterns based on skill level. A secondary purpose of this study was to refine the architecture of the OMAT to optimize classification rates by incorporating feature selection and multiple machine learning algorithms (i.e., BLR, DT, KNN, LDA, NB, SVM, and RBF) for both the OMAT-OPT and OMAT-sIMU.

6.3 METHODS

6.3.1 Participants

Kinematic data were collected on 542 athletes by Motus Global (Rockville Centre, NY, USA). The sample included athletes competing in 11 different sports (i.e., baseball, basketball, soccer, golf, tennis, track and field, squash, cricket, lacrosse, football, or volleyball) and ranging in skill level from recreational to professional (e.g., NBA, MLB, NFL, PGA, FIFA). c Before data collection, each athlete read and signed an informed consent form permitting Motus Global to use

the data for future research. The Health Sciences Research Ethics Board at the University of Ottawa approved the secondary use of the data for research purposes (file no: H-08-18-1085).

6.3.2 Protocol

Upon arrival to the Motus Global laboratory, each athlete read and signed an informed consent form, provided information on injury history for the previous 10 years, and had their height (with shoes on) and weight recorded. The athlete was then outfitted with 45 passive, reflective markers (B&L Engineering, Santa Ana, CA) to capture whole-body motion (McPherson et al., 2016; Ross et al., 2018). After being outfitted with the markers, the athlete completed a static and dynamic calibration trial (Ross et al., 2018). The static calibration trial was used to develop a whole-body biomechanical model for each athlete.

After the calibration trials, each athlete completed a movement battery consisting of 21 unique movements testing athletes' range of motion at each joint, stability, power and balance. However, only seven movements were used in the analysis due to their dynamic nature and ability to challenge the athletes' coordination, stability, and mobility across all major joints. The seven tasks included: bird-dog, drop jump, hop-down, L-hop, lunge, step-down, and T-balance (Figure 6.3.1). Each movement was performed bilaterally on the left and right side except for the drop jump which was performed symmetrically, resulting in a total of 13 movement trials (Ross et al., 2018). The athlete performed each task until they believed they did it to the best of their ability with only the trial that was deemed the best being retained for each athlete. Full-body motion data were captured at 120 Hz using an 8-camera Raptor-E (Motion Analysis Corporation, Santa Rosa, CA) motion capture system.

6.3.3 Data Analysis

6.3.3.1 Pre-Processing

Motion capture data were collected, labeled, and gap-filled using Cortex (Motion Analysis Corporation, Santa Rosa, CA). Data from anatomical landmarks and the tracking markers during the calibration trial were used to develop a whole-body 3D kinematic model in Visual3D v6 (C-Motion, Inc., Germantown, MD). The model was then applied to all motion trials outputting joint centers bilaterally for the wrist, elbow, shoulder, foot, ankle, knee, and hip; centers of gravity for

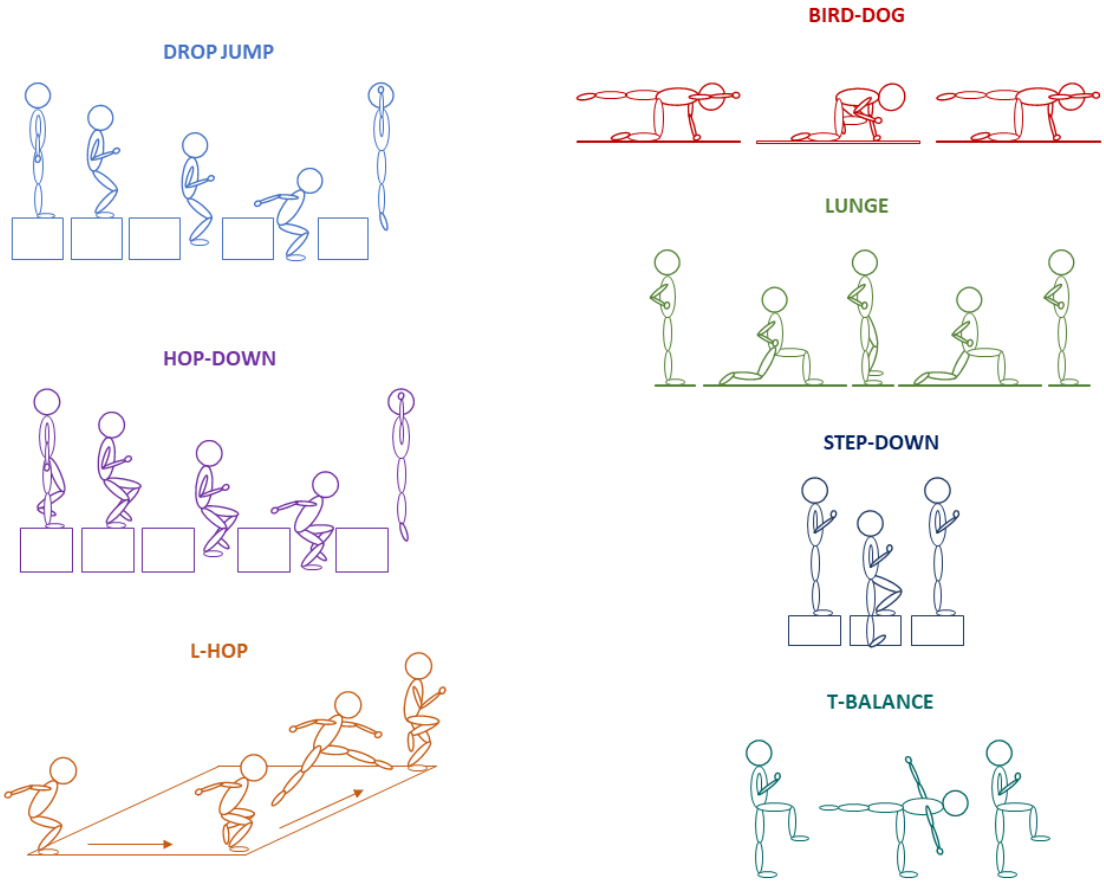


Figure 6.3.1. Schematic drawings of the seven unique movements performed by the athletes: drop-jump, hop-down, L-hop, bird-dog, lunge, step-down, T-balance.

the trunk, head, and pelvis; marker positional data for the left and right heel, T_2, T_8, sternum, and the back, front and sides of the head for the OMAT-OPT model and segment angular velocities

and center of gravity (CoG) linear velocities of the head, trunk, pelvis, upper arms, forearms, thighs, shanks, and feet for the OMAT-sIMU model. Data were then exported and analyzed using Python 3.0. All trials were trimmed to specific start and end-point criteria (Ross et al., 2018), and filtered using a dual-pass, low-pass Butterworth filter with a cutoff of 15 Hz. Since elite athletes were significantly taller than novices ($F= 138.25$, $p <0.001$), all data for each movement were normalized by each athlete's individual height by dividing each raw data point by their own height. Normalization ensured that differences in PC scores between groups were not strictly due to variation in size.

6.3.3.1.1 OMAT-OPT Data

The 3D positional data of the joint centers and markers retained in the OPT model for each participant were rotated so that the local coordinate system of the trunk was aligned with the global coordinate system. The data were then translated so that the midpoint between the left and right hip of the first frame of data was aligned with the global origin (i.e., midpoint of left and right hip equaled 0,0,0 for x, y, and z coordinates, respectively). The rotated 3D data were then time normalized to 500 frames using Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation to control for differences in absolute movement time for each participant. An [n x 39,000] matrix for each movement was then constructed, where n was the number of subjects and 39,000 was the time-normalized x, y, and z data for each joint center, center of gravity, and retained markers mentioned above (26 positions x 3 axes x 500 time points). Due to marker occlusion and some athletes not performing all tasks, n was dependent on the movement task (Table 6.4.1).

6.3.3.1.2 OMAT-sIMU Data

In order to simulate IMU accelerometer data, we extracted the segment CoG linear velocities of each segment retained in the model and then differentiated the data once to calculate segment linear accelerations, and to simulate the IMU gyroscope data, we extracted the segment angular velocities of the same segments. Once all data were extracted and calculated, the data were time normalized to 500 frames as per the OMAT-OPT. In anticipation of the future implementation of the method in clinic or industry, the Euclidean norm (i.e., square root of the sum of squares) of the x, y, and z axes of the linear segment accelerations and segment angular velocities were taken to minimize the effect of sensor brand or orientation (Clouthier et al., 2020) and to reduce the dimensionality of the data (Bergmann et al., 2014).

A matrix for each movement was then constructed with the Euclidean norm of the linear segment accelerations and segment angular velocities for each segment and each participant. Segment linear accelerations and angular velocities were chosen to mimic outputs collected via IMUs. Each matrix was n (number of participants; Table 6.4.2) x 13000 (Euclidean norm x 2 data features x 13 body segments x 500 time points). Because the units were different between the data features (i.e., linear accelerations in m/s^2, angular velocities in rad/s), the scale of the data between the two data features varied widely, which would lead to classification being driven primarily by the data feature with the larger scale. Therefore, the data were feature scaled to be between 0 and

1 for each movement using scikit-learn Robust Scaler (Pedregosa et al., 2011), which removes the median from each feature and scales the data according to the 1st and 3rd quartile of the data, mitigating the effect of outliers during scaling.

6.3.3.2 Feature Selection

For both OMAT-sIMU and OMAT-OPT, PCA was applied to each matrix, resulting in a unique model per task per data type (i.e., OMAT-sIMU, OMAT-OPT). Using the PC scores as features, ensemble feature selection, consisting of six common feature selection techniques (Pearson correlation, chi-squared, recursive feature elimination, lasso, random forest, and LightGBM), was used to rank the PCs based on contribution to the model for each movement task and data type. Ensemble feature selection has been found to improve the robustness of feature ranking and feature subset selection as well as increase the generalizability of the features selected (Saeys et al., 2008). The scikit-learn library was used for the chi-squared, recursive feature elimination, lasso, and random forest (Pedregosa et al., 2011). The top 25 features per data type were retained for each technique. The features were then sorted based on the number of techniques where they ranked in the top 25 features. PC scores that ranked in the top 25 for at least 50% of the techniques (i.e., 3) were retained for the classifier (Tables 6.4.1; OMAT-OPT & 6.4.2; OMAT-sIMU). To minimize overfitting of the models, the maximum number of features retained was the square root of the number of samples for each movement task (Hua et al., 2005; e.g., lunge right had 401 samples, therefore a maximum of 20 PC scores could be retained for that task).

6.3.3.3 Classification

To refine the architecture of the OMAT, seven different kinds of classifiers were used: BLR, DT, kNN, LDA, NB, SVM and RBF to classify athletes based on skill level (elite versus novice). All classifiers were employed using the scikit-learn library (Pedregosa et al., 2011). For all classifiers, PC scores retained from feature selection were used as predictors and leave-one-out cross-validation was used for validation. Each model was rerun to use between 1 and the total number of PCs retained to determine the optimal number of PCs to retain for each classifier for each movement task. The model with the highest classification rate was deemed the optimal model. Due to a lack of a testing dataset, leave-one-out validation was used where one of the athletes' data were taken out (test athlete) and the PCA, feature selection, and classifier models were computed on the remaining athletes (training athletes). After computing the new PCA, feature selection, and

classifier models, the test athlete was projected into the PCA, feature selection, and classifier model spaces computed on the training athletes. The procedure was repeated until all athletes had been left out and projected back into the PCA, features selection, and classifier models (Ross et al., 2018; Troje, 2002).

6.3.3.4 Signal Detection Theory

For the best classifier for each data type, to test the separation between the signal and the noise and to determine the strategy used by the frameworks, a signal detection theory (SDT) model was used for each optimal model retained. In SDT, there are four types of classification: 1.) Hit, 2.) Miss, 3.) False alarm (FA), and 4.) Correct rejection (CR; Abdi, 2007). For this study, a hit was when an elite athlete was correctly classified as an elite (equivalent to sensitivity), a miss was when an elite athlete was misclassified as a novice, a FA was when a novice athlete was misclassified as an elite, and a CR was when a novice athlete was correctly classified as a novice (equivalent to specificity). Parameter D' is calculated subtracting the probability (z-score) of a false alarm from the probability (z-score) of a hit and tells the distance between the two peaks (e.g., elite and novice) in standard deviations; the higher the score the more separable the two groups are with a score of 0 representing chance (Abdi, 2007). Parameter C is calculated by taking the average probability (z-score) of a hit and false alarm and represents the strategy used by the framework. A positive value represents the framework being conservative (e.g., more likely to classify an athlete as novice), where as a negative value represents the framework being liberal (e.g., more likely to classify an athlete as elite; Abdi, 2007). The closer the value is to 0, the closer the framework is to being the ideal observer (e.g., not more likely to classify as either elite or novice; Abdi, 2007).

6.4 RESULTS

6.4.1 OMAT-OPT

For all tasks the linear classifiers (i.e., BLR, LDA and SVM) outperformed DT, kNN, and RBF, except RBF performed as well as the linear classifiers for the lunge left and step-down left (Figure 6.4.1). For the drop-jump, hop-down left, L-hop left, and lunge left, NB performed as well as the linear classifiers, however, for all other tasks, they performed in between the linear classifiers and DT, kNN, and RBF. Since there were minimal differences (< 0.5%) on the average

classification rates for all tasks between BLR, LDA, and SVM, and to be able to compare the current results to previous results (Ross et al., 2018), LDA was selected for further analysis. When using LDA, the optimal number of PCs retained ranged from 9 (hop-down right) to 18 (bird-dog left, T-balance right) with an average of 14.15 ± 3.02 PCs retained (Table 6.4.1). The OMAT-OPT accurately classified between 73.1% (T-balance right) to 83.8% (L-hop left) of athletes as either elite or novice (Table 6.4.1). The average classification rate across all tasks was 78.1% ± 3.26%. For SDT, on average, OMAT-OPT had a hit, miss, FA, and CR rate of 0.87 ± 0.03, 0.13 ± 0.03, 0.38 ± 0.06, and 0.62 ± 0.06, respectively (Table 6.4.1). The average D' was 1.44 ± 0.25 and the average C was -0.41 ± 0.07 (Table 6.4.1).

6.4.2 OMAT-sIMU

Similar to OMAT-OPT, for all tasks the linear classifiers (i.e., BLR, LDA, and SVM) outperformed all the other classifiers (i.e., DT, kNN, NB, and RBF), except KNN performed as well as the linear classifiers in the bird-dog left and right, hop-down left, and step-down left (Figure 6.4.2). Since there were again minimal differences between the average classification rates for all movement tasks between BLR, LDA, and SVM, LDA was selected for further analysis. When using segment linear accelerations and angular velocities, data available from an IMU system, the optimal number of PCs retained ranged from 6 (bird-dog left, step-down right) to 18 (lunge left and right) with an average of 11.92 ± 4.03 PCs retained (Table 6.4.2). OMAT-sIMU accurately classified between 75.1% (T-balance right) to 84.7% (drop-jump) of athletes as either elite or novice (Table 6.4.2). The average classification rate across all tasks was 80.0% ± 3.19%. For SDT, on average, OMAT-sIMU had a hit, miss, FA, and CR rate of 0.89 ± 0.02, 0.11 ± 0.02, 0.43 ± 0.1, and 0.57 ± 0.1, respectively (Table 6.4.2). The average D' was 1.44 ± 0.25 and the average C was -0.53 ± 0.15 (Table 6.4.2).

When comparing the OMAT-OPT and OMAT-sIMU classification rates on average, OMAT-sIMU had higher classification rates than OMAT-OPT by 1.92%. OMAT-sIMU outperformed OMAT-OPT in the bird-dog right (1.1%), drop-jump (4.31%), hop-down left (4.48%) and right (5.46%), L-hop right (3.68%), lunge left (2.50%) and right (2.95%), step-down right (2.18%), and T-balance right (1.97%), whereas OMAT-OPT had a higher classification rate than OMAT-sIMU for bird-dog left (1.41%; Figure 6.4.3). The two models performed relatively the same for the L-hop left (0.68%), step-down left (0.75%), and T-balance left (0.83%).

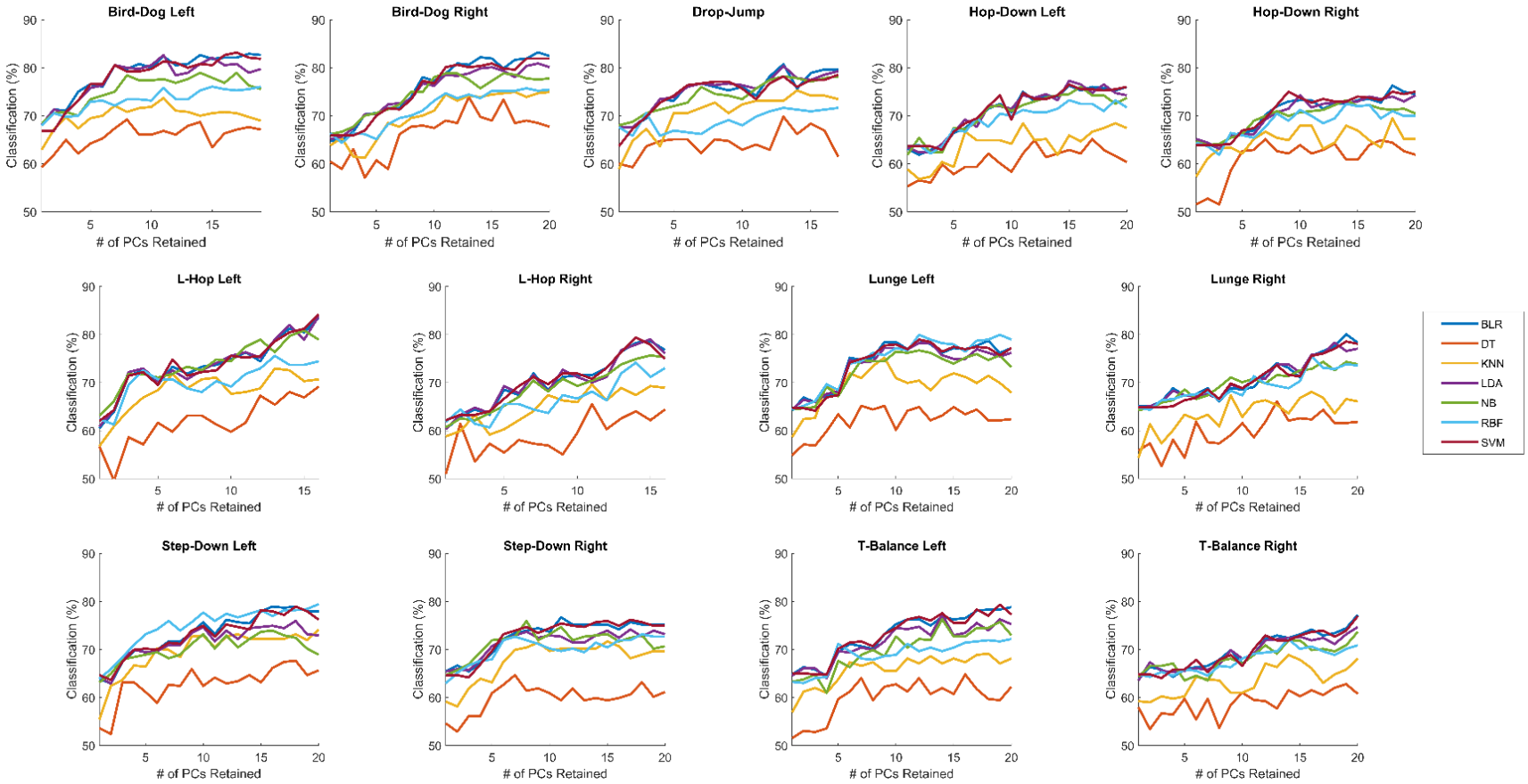


Figure 6.4.1. OPT: The percent of correctly classified athletes as either elite or novice for when 1 to the total number of PCs retained were retained for binary logistic regression (BLR), decision tree (DT), linear discriminant analysis (LDA), k-nearest neighbors (kNN), naïve bayes (NB), and support vector machine with a linear kernel (SVM) and a radial basis function kernel (RBF) with leave-one-out validation for OPT. See number of PCs retained in Table 6.4.1.

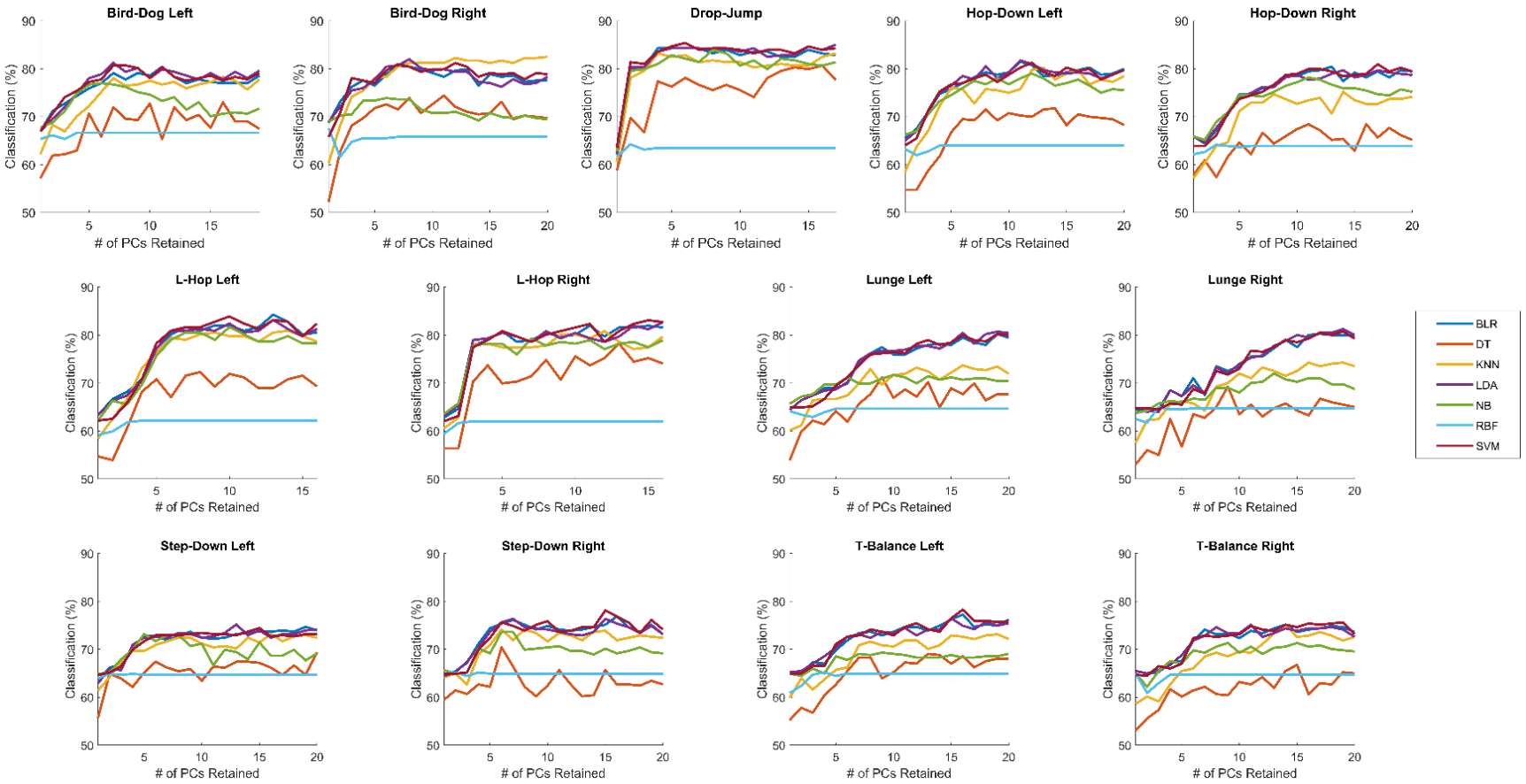


Figure 6.4.2. sIMU: The percent of correctly classified athletes as either elite or novice for when 1 to the total number of PCs retained were retained for binary logistic regression (BLR), decision tree (DT), linear discriminant analysis (LDA), k-nearest neighbors (kNN), naïve bayes (NB), and support vector machine with a linear kernel (SVM) and a radial basis function kernel (RBF) with leave-one-out validation for OPT. See number of PCs retained in Table 6.4.2.

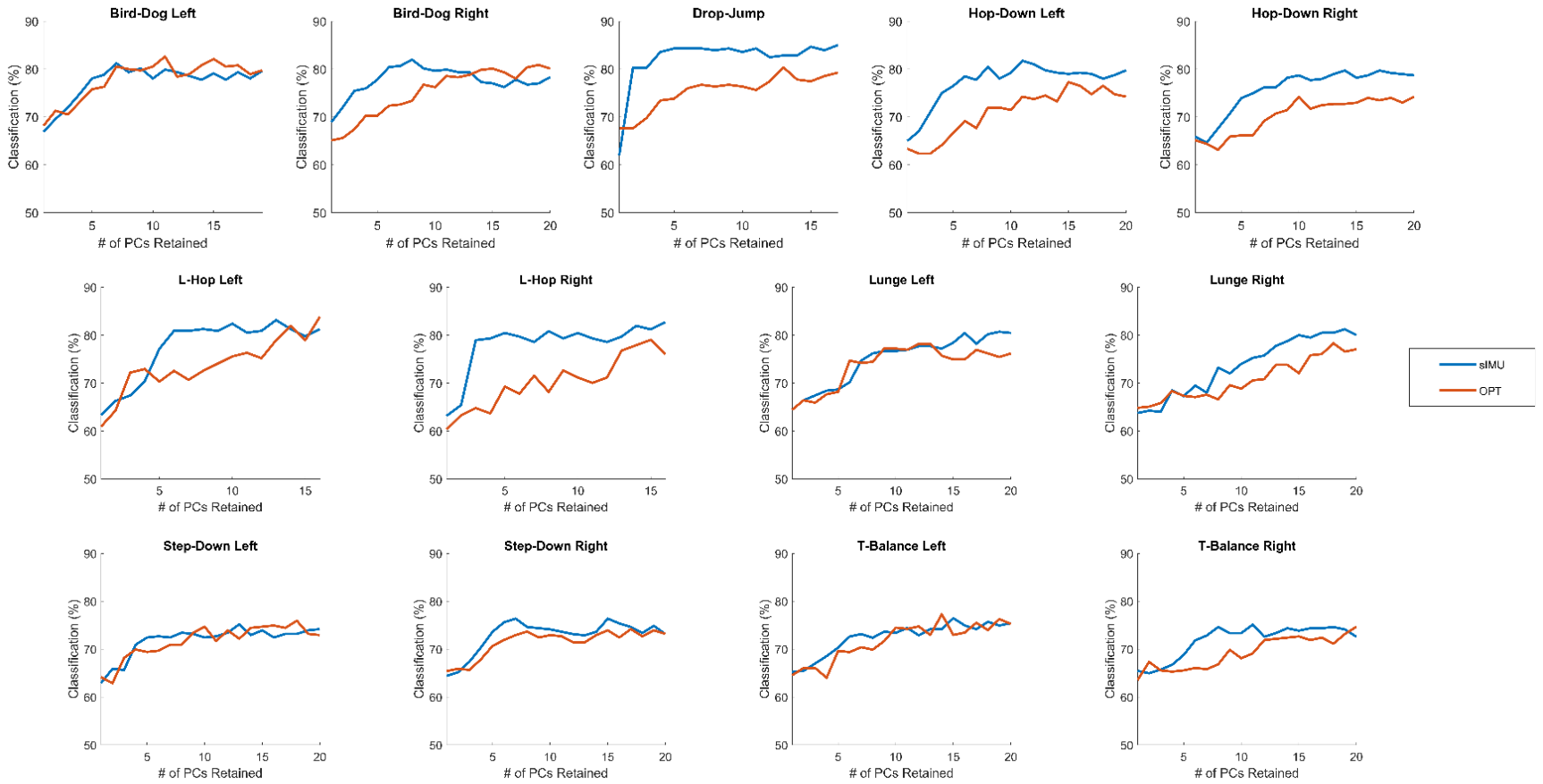


Figure 6.4.3. The percent of correctly classified athletes as either elite or novice for when 1 to the total number of PCs retained were retained for the LDA with leave-one-out validation when using OPT (orange) and sIMU (blue) data.

Table 6.4.1. OMAT-OPT: The number of athletes broken down by sex and skill level and the overall classification accuracy, hit rate, miss rate, false alarm (FA) rate, correct rejection (CR) rate, d' and c when the optimal number of PCs were retained for each movement task and the average and standard deviation across all movement tasks.

		Male		**Female**									
Movement	n	Elite	Novice	Elite	Novice	# of PCs	Acc. (%)	Hit	Miss	FA	CR	D'	C
Bird-Dog Left	380	242	83	12	43	10	82.63	0.91	0.09	0.34	0.66	1.75	-0.46
Bird-Dog Right	387	244	88	11	44	18	80.88	0.90	0.10	0.36	0.64	1.62	-0.46
Drop Jump	275	168	64	7	36	12	80.36	0.88	0.12	0.33	0.67	1.61	-0.37
Hop-Down Left	396	242	99	10	45	14	77.27	0.87	0.13	0.40	0.60	1.39	-0.45
Hop-Down Right	396	242	97	11	46	9	74.24	0.84	0.16	0.43	0.57	1.17	-0.40
L-Hop Left	266	159	67	6	34	15	83.83	0.89	0.11	0.25	0.75	1.91	-0.27
L-Hop Right	267	160	67	6	34	14	79.03	0.87	0.13	0.34	0.66	1.54	-0.35
Lunge Left	399	246	97	12	44	11	78.20	0.87	0.13	0.38	0.62	1.43	-0.40
Lunge Right	401	248	97	12	44	17	78.30	0.88	0.12	0.39	0.61	1.44	-0.44
Step-Down Left	399	246	98	12	43	17	75.94	0.84	0.16	0.40	0.60	1.28	-0.38
Step-Down Right	399	247	96	11	45	16	74.19	0.83	0.17	0.42	0.58	1.16	-0.37
T-Balance Left	392	244	92	11	45	13	77.30	0.89	0.11	0.45	0.55	1.37	-0.56
T-Balance Right	395	244	94	12	45	18	73.16	0.83	0.17	0.45	0.55	1.08	-0.41
Average	**365.54**	**225.54**	**87.62**	**10.23**	**42.15**	**14.15**	**78.10**	**0.87**	**0.13**	**0.38**	**0.62**	**1.44**	**-0.41**
STD	**55.17**	**36.14**	**13.11**	**2.31**	**4.38**	**3.02**	**3.26**	**0.03**	**0.03**	**0.06**	**0.06**	**0.25**	**0.07**

Table 6.4.2. OMAT-sIMU: The number of athletes broken down by sex and skill level and overall classification accuracy, hit rate, miss rate, false alarm rate, correct rejection rate, d' and c when the optimal number of PCs were retained for each movement task and the average and standard deviation across all movement tasks.

		Male		**Female**									
Movement	n	Elite	Novice	Elite	Novice	# of PCs	Acc. (%)	Hit	Miss	FA	CR	D'	C
Bird-Dog Left	380	242	83	12	43	6	81.22	0.92	0.08	0.45	0.55	1.56	-0.66
Bird-Dog Right	387	244	88	11	44	7	81.98	0.90	0.10	0.49	0.51	1.34	-0.64
Drop Jump	275	168	64	7	36	14	84.67	0.89	0.11	0.27	0.73	1.81	-0.29
Hop-Down Left	396	242	99	10	45	10	81.75	0.88	0.13	0.36	0.64	1.51	-0.40
Hop-Down Right	396	242	97	11	46	13	79.70	0.89	0.11	0.40	0.60	1.49	-0.50
L-Hop Left	266	159	67	6	34	12	83.15	0.87	0.13	0.29	0.71	1.70	-0.29
L-Hop Right	267	160	67	6	34	15	82.71	0.90	0.10	0.33	0.67	1.71	-0.41
Lunge Left	399	246	97	12	44	18	80.70	0.93	0.07	0.50	0.50	1.47	-0.74
Lunge Right	401	248	97	12	44	18	81.25	0.91	0.09	0.38	0.62	1.62	-0.51
Step-Down Left	399	246	98	12	43	12	75.19	0.87	0.13	0.52	0.48	1.07	-0.60
Step-Down Right	399	247	96	11	45	6	76.37	0.88	0.12	0.50	0.50	1.17	-0.60
T-Balance Left	392	244	92	11	45	14	76.47	0.90	0.10	0.55	0.45	1.15	-0.71
T-Balance Right	395	244	94	12	45	10	75.13	0.87	0.13	0.50	0.50	1.14	-0.58
Average	**365.54**	**225.54**	**87.62**	**10.23**	**42.15**	**11.92**	**80.02**	0.89	0.11	0.43	0.57	1.44	-0.53
STD	**55.17**	**36.14**	**13.11**	**2.31**	**4.38**	**4.03**	**3.19**	0.02	0.02	0.10	0.10	0.25	0.15

6.5 DISCUSSION

The primary purpose of this study was to assess the ability of the OMAT to differentiate whole-body movement patterns between novice and elite athletes performing a non-sport-specific movement screening battery using data able to be collected via an IMU. The secondary purpose of this study was to refine the architecture of the OMAT by incorporating feature selection and testing multiple classifiers. For both the OMAT-OPT and OMAT-sIMU, BLR, LDA, and SVM, on average, outperformed all other classifiers tested. These findings suggest that the data can be separated using a linear plane; and therefore, the use of more complicated, computationally expensive non-linear classifiers is not only not required, but can be detrimental. There were minimal differences between BLR, LDA, and SVM, so therefore, in order to easily compare the current results with previous work, LDA was chosen as the classifier to report the results.

OMAT-OPT with feature selection outperformed the previously published results on 7 of the 13 tasks (i.e., bird-dog left and right, hop-down left, L-hop left, lunge left and right, and T-balance left) and OMAT-sIMU was able to outperform the previously published results of the OMAT-OPT in all tasks except the step-down left (Ross et al., 2018). This in part is due to the introduction of feature selection into the methodology, reinforcing the value of this approach for future work aiming to objectively classify movement patterns. Compared to the previous study, where PCs 1-35 were retained (Ross et al., 2018), using feature selection, we are now able to have greater classification rates using fewer PCs, which requires fewer computational resources and decreases the risk of overfitting for 7 of the 13 tasks. OMAT-sIMU outperformed or performed equally to OMAT-OPT in all movements except the bird-dog left. These findings suggest that the OMAT-sIMU approach better captures movement pattern differences between novice and elite athletes compared to OMAT-OPT data. This is thought to be due to the different types of data analyzed for OMAT-OPT and OMAT-sIMU. OMAT-OPT uses joint center trajectories, which due to the constrained nature of the tasks, may be capturing more gross motor patterns that are unrelated to skill. In contrast, the OMAT-sIMU uses linear acceleration and angular velocity that are more likely to capture the smoothness of the movement, which may be a better indicator of skill level than gross motor patterns. However, for both the OMAT-OPT and OMAT-sIMU, when looking at trends in individual athlete data across tasks, if there were differences in how the athlete was classified between tasks, athletes tended to be classified the same on all tasks that were

targeting the same skill set (e.g., trunk stability, jumping, balance) and if there were discrepancies on how the left/right tasks were classified, the dominant side was usually classified as elite. This suggests that relevant differences between elite and novice-like movement patterns can be detected using both data types. A combined approach of using both sIMU and OPT data may provide even better classification rates than using sIMU or OPT alone due to the two types of data potentially capturing different movement features.

Previously, in order to assess how well the framework was classifying elite and novice athletes on a group basis, the percent of correctly classified elites and novices were calculated (Ross et al., 2018). SDT was chosen for this current study because it provides classification rates for each group (e.g., hit and correct rejection) as well as the additional information of response bias. For all tasks, both OMAT-OPT and OMAT-sIMU had higher rates of correctly classifying elite athletes (depicted by the increased hit and decreased miss rates) compared to novice athletes (depicted by the decreased CR and increased FA rates). For all tasks, D' was greater than 1.08 and 1.07 for OMAT-OPT and OMAT-sIMU, respectively, suggesting that elite and novice athletes are separable when using both OMAT-OPT and OMAT-sIMU. However, on average, the data are more robustly classified when using OMAT-sIMU data compared to OMAT-OPT data. Lastly, for all tasks, for both OMAT-OPT and OMAT-sIMU, the framework was more likely to classify the athlete as elite than the ideal observer. A potential reason for this could be that some of the novice athletes were attending an elite youth sports academy, which boasts a high percentage of students continuing to compete at the collegiate and professional levels. Therefore, some of the novice athletes were on track to become elite athletes at the time of testing. On average, OMAT-OPT acted more closely to the ideal observer than OMAT-sIMU, based on our definition of a hit and correct rejection; this is represented by the smaller C value.

Although on average the models using the OMAT-sIMU data as the inputs, had higher classification rates than OMAT-OPT, a limitation of the OMAT-sIMU data is that it is more difficult to interpret differences between elites and novices compared to OMAT-OPT, making it harder to train individuals to improve their movement patterns. With only linear accelerations and angular velocities, and no video data, it is hard to discern exactly how the athlete is moving within space to obtain a score more representative of an elite or novice athlete. IMUs may offer an inexpensive measurement device to objectively screen movement abilities, where those

individuals identified with weaknesses can then be tested more in depth with optical motion capture to inform targeted corrective exercise approaches.

A limitation of this study is the use of camera-based motion capture to calculate the linear accelerations and angular velocities for each segment and not raw data collected from IMUs. This technique has been used in previous research when the desired database does not contain IMU data (Young, 2010). Although linear accelerations and angular velocities would change when using IMUs, due to the inability to place IMUs at the CoG of segments, previous research has found strong agreement between IMU outputs and optical motion capture outputs (Bolink et al., 2018; Mcginnis et al., 2014). Even though the data were not raw data from an IMU, we purposefully took the Euclidean norm of the data to increase ecological validity and to remove the effect of different local coordinate system orientations of the linear acceleration and angular velocities within the global coordinate system. In addition, we differentiated positional data, which introduces noise to the data that would not be present when collecting data via IMU and were still able to get high classification rates. We are confident that these classification results are representative and may be lower than that of what would be achieved using sensors themselves, which we are in the process of testing. A second limitation of the study is the assumption athletes at the collegiate and professional level completed 10,000 hours of deliberate practice. However, athletes competing at the professional and inter-collegiate levels would be in the higher echelon of athletes in their sport even if not completing 10,000 hours. Nonetheless, this paper provides proof-of-concept that the OMAT is able to accurately classify athletes as novice or elite with consistent or improved accuracy when using data available from IMUs, relative to whole-body marker data. Future research should investigate the ability to classify athletes using the OMAT using segment linear acceleration and angular velocity data collected using IMUs, fine-tuning algorithms to increase classification rates, and exploring other classifiers such as sport played, injury risk, and sex.

In conclusion, the introduction of feature selection increased the classification rates compared to using the first 35 PC scores and BLR, LDA, and SVM produced the highest classification rates although there were minimal differences (< 0.5%) between the three. Segment linear acceleration and angular velocity data readily available from an IMU, could differentiate athletes' movement performance based on skill level when using a novel machine learning

approach (Ross et al., 2018) with a level of accuracy consistent with the use of whole-body motion capture data. These data suggest that IMUs, in conjunction with OMAT, may provide an inexpensive and timely way to objectively characterize and classify movement performance in the field, providing a feasible method for coaches and clinicians to objectively measure performance.

6.6 CONFLICT OF INTEREST

Britney Dowling is the Director of Research at Motus. Dowling was involved in the data collection, pre-processing of the data, and editing of the manuscript and was not involved in the implementation of the OMAT or analysis of the results. In addition, the OMAT used in this paper is an objective, data-driven approach with no human interaction and does not only apply to this data set. All other authors declare no potential conflict of interest.

6.7 AUTHOR CONTRIBUTIONS

GR, RG, NT and SF conceived the study and interpreted the results. GR and BD collected the data and performed the pre-processing of the data, GR implemented the OMAT, analysed the results and prepared the manuscript. All authors revised the manuscript.

6.8 FUNDING

This work was funded by Natural Sciences and Engineering Research Council (NSERC) of Canada and the Ontario Early Researcher Awards Program.

6.9 ACKNOWLEDGEMENTS

The authors would like to thank the athletes who partook in the study.

6.10 DATA AVAILABLITY STATEMENT

Sample code and data are available at https://doi.org/10.5281/zenodo.3575075.

6.11 REFERENCES

Abdi, H., 2007. Signal Detection Theory (SDT), in: Salkini, N., Rasmussen, K. (Eds.), Encyclopedia of Measurement and Statistics. SAGE Publications Ltd., Thouse Oaks, pp. 886–889.

Baker, J., Cote, J., Abernethy, B., 2003. Sport-specific practice and the development of expert decision-making in team ball sports. J. Appl. Sport Psychol. 15, 12–25. doi:10.1080/10413200305400

Bergmann, J.H.M., Langdon, P.M., Mayagoitia, R.E., Howard, N., 2014. Exploring the use of sensors to measure behavioral interactions: An experimental evaluation of using hand trajectories. PLoS One 9, 1–10. doi:10.1371/journal.pone.0088080

Bolink, S.A.A.N., Naisas, H., Senden, R., Essers, H., Heyligers, I.C., Meijer, K., Grimm, B., 2018. Validity of an inertial measurement unit to assess pelvic orientation angles during gait, sit-stand transfers and step-up transfers: Comparison with an optoelectronic motion capture system. Med. Eng. Phys. 13, 1–7. doi:10.1016/j.medengphy.2015.11.009

Clouthier, A.L., Ross, G.B., Graham, R.B., Graham, R.B., 2020. Sensor data required for automatic recognition of athletic tasks using deep neural networks. Front. Bioeng. Biotechnol. 7, 473. doi:10.3389/fbioe.2019.00473

Cook, G., Burton, L., Hoogenboom, B.J., 2014. Functional movement screening : The use of fundamental movements as an assssment of function- Part 2. Int. J. Sports Phys. Ther. 9, 549–563. doi:10.1111/j.1600-0838.2010.01267.x

Donà, G., Preatoni, E., Cobelli, C., Rodano, R., Harrison, A.J., 2009. Application of functional principal component analysis in race walking: An emerging methodology. Sport. Biomech. 8, 284–301. doi:10.1080/14763140903414425

Federolf, P., Reid, R., Gilgien, M., Haugen, P., Smith, G., 2014. The application of principal component analysis to quantify technique in sports. Scand. J. Med. Sci. Sport. 24, 491–499. doi:10.1111/j.1600-0838.2012.01455.x

Gulgin, H., Hoogenboom, B., 2014. The Functional Movement Screening (FMS)^{TM}: An inter-rater reliability study Between raters of varied experience. Int. J. Sports Phys. Ther. 9, 14–20.

Hadjidj, A., Souil, M., Bouabdallah, A., Challal, Y., Owen, H., 2013. Wireless sensor networks for rehabilitation applications: Challenges and opportunities. J. Netw. Comput. Appl. 36, 1–15. doi:10.1016/j.jnca.2012.10.002

Helsen, W.F., Starkes, J.L., Hodges, N.J., 1998. Team sports and the theory of deliberate practice. J. Sport Exerc. Psychol. 20, 12–34.

Hua, J., Xiong, Z., Lowey, J., Suh, E., Dougherty, E.R., 2005. Optimal number of features as a function of sample size for various classification rules. Bioinformatics 21, 1509–1515. doi:10.1093/bioinformatics/bti171

Johnston, W., O'Reilly, M., Dolan, K., Reid, N., Coughlan, G., Caulfield, B., 2016. Objective classification of dynamic balance using a single wearable sensor, in: 4th International Congress on Sports Sciences Research and Technology Support 2016. Porto, Portugaul, pp. 15–24.

Johnston, W., Reilly, M.O., Duignan, C., Liston, M., Mcloughlin, R., Coughlan, G.F., Caulfield, B., 2019. Association of dynamic balance with sports-related concussion: A prospective cohort study. Am. J. Sports Med. 47, 197–205. doi:10.1177/0363546518812820

Kritz, M., Cronin, J., Hume, P., 2009. The bodyweight squat: A movement screen for the squat pattern. Strength Cond. J. 31, 76–85. doi:10.1519/SSC.0b013e318195eb2f

Madgwick, S.O.H., Harrison, A.J.L., Vaidyanathan, R., 2011. Estimation of IMU and MARG orientation using a gradient descent algorithm, in: IEEE International Conference on Rehabilitation Robotics (ICORR). Zurich, Switzerland, pp. 1–7.

McCall, A., Carling, C., Nedelec, M., Davison, M., Le Gall, F., Berthoin, S., Dupont, G., 2014. Risk factors, testing and preventative strategies for non-contact injuries in professional football: Current perceptions and practices of 44 teams from various premier leagues. Br. J. Sports Med. 48, 1352–7. doi:10.1136/bjsports-2014-093439

McCunn, R., aus der Fünten, K., Fullagar, H.H.K., McKeown, I., Meyer, T., 2016. Reliability and association with injury of movement screens: A critical review. Sport. Med. 46, 763–781. doi:10.1007/s40279-015-0453-1

Mcginnis, R.S., Cain, S.M., Davidson, S.P., Vitali, R. V, Mclean, S.G., Perkins, N.C., 2014. Validation of complementary filter based IMU data fusion for tracking torso angle and rifle orientation, in: ASME 2014 International Mechanical Engineering Congress and Exposition IMECE. Montreal, Quebec. doi:10.1115/IMECE201436909

McPherson, A.L., Dowling, B., Tubbs, T.G., Paci, J.M., 2016. Sagittal plane kinematic differences between dominant and non-dominant legs in unilateral and bilateral jump landings. Phys. Ther. Sport 22, 54–60.

Onate, J.A., Dewey, T., Kollock, R.O., Thomas, K.S., Van Lunen, B.L., DeMaio, M., Ringleb, S.I., 2012. Real-time intersession and interrater reliability of the Functional Movement Screen. J. Strength Cond. Res. 26, 408–415. doi:10.1519/JSC.0b013e318220e6fa

Padua, D.A., Marshall, S.W., Boling, M.C., Thigpen, C.A., Garrett, W.E., Beutler, A.I., 2009. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. Am. J. Sports Med. 37, 1996–2002. doi:10.1177/0363546509343200

Patel, S., Park, H., Bonato, P., Chan, L., Rodgers, M., 2012. A review of wearable sensors and systems with application in rehabilitation. J. Neuroeng. Rehabil. 9, 1–17.

Pedregosa, F., Weiss, R., Brucher, M., 2011. Scikit-learn : Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Ross, G.B., Dowling, B., Troje, N.F., Fischer, S.L., Graham, R.B., 2018. Objectively differentiating movement patterns between elite and novice athletes. Med. Sci. Sport. Exerc. 50, 1457–1464. doi:10.1249/MSS.0000000000001571

Saeys, Y., Abeel, T., Van de Peer, Y. Van, 2008. Robust feature selection using ensemble feature selection techniques, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidleberg, pp. 313–325.

Sgro, F., Mango, P., Pignato, S., Schembri, R., Licari, D., Lipoma, M., 2017. Assessing standing long jump developmental levels using an inertial measurement unit. Percept. Mot. Skills 124, 21–38. doi:10.1177/0031512516682649

Smith, C.A., Chimera, N.J., Wright, N.J., Warren, M., 2013. Interrater and intrarater reliability of the Functional Movement Screen. J. Strength Cond. Res. 27, 982–987.

Smith, R.M., Spinks, W.L., 1995. Discriminant analysis of biomechanical differences between novice , good and elite rowers. J. Sport Sci. 13, 377–385. doi:10.1080/02640419508732253

Troje, N.F., 2002. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. J. Vis. 2, 371–387. doi:10.1167/2.5.2

Wittmann, F., Lambercy, O., Gassert, R., 2019. Magnetometer-based drift correction during rest in IMU arm motion tracking. Sensors 19, 1312–1330. doi:10.3390/s19061312

Young, A.D., 2010. From posture to motion: The challenge for real time wireless inertial motion capture, in: Proceedings of the Fifth International Conference on Body Area Networks. Corfu, Greece, pp. 131–137.

Young, C., Reinkensmeyer, D.J., 2014. Judging complex movement performances for excellence: A principal components analysis-based technique applied to competitive diving. Hum. Mov. Sci. 36, 107–122. doi:10.1016/j.humov.2014.05.009

Zago, M., Sforza, C., Dolci, C., Tarabini, M., Galli, M., 2019. Use of machine learning and wearable sensors to predict energetics and kinematics of cutting maneuvers. Sensors 19, 3094–3105.

Zhou, H., Hu, H., 2008. Human motion tracking for rehabilitation — A survey. Biomed. Signal Process. Control 3, 1–18. doi:10.1016/j.bspc.2007.09.001

Chapter 7: Study 4

COMPARISON OF MACHINE LEARNING CLASSIFIERS FOR DIFFERENTIATING LEVEL AND SPORT USING MOVEMENT DATA

Original Article

Gwyneth B. Ross[1], Allison Clouthier[1], Alistair Boyle[2], Steven L. Fischer[3], and Ryan B. Graham[1,3]

[1]School of Human Kinetics, Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada

[2] Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada

[3]Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada

Submitted to: *Journal of Biomechanics*

7.1 ABSTRACT

Purpose: The purpose of this study was to determine if recurrent neural networks (RNNs) designed for multivariate, time-series analyses (i.e., long short-term memory (LSTM) and reservoir computing (RC)) outperform traditional linear and non-linear machine learning classifiers when classifying athletes based on competition level and sport played.

Methods: Optical-based kinematic data from 542 athletes were used as input data for the machine learning classifiers to classify athletes based on competition level and sport played. For the traditional machine learning classifiers, principal component analysis (PCA) and feature selection were used to reduce the data dimensionality to help prevent overfitting and to determine the best principal component scores to retain as the input data. Machine learning classifiers used were RC, LSTM, linear discriminant analyses, logistic regression, support vector machines with a linear and radial basis function kernel, *k*-nearest neighbors, decision trees, and naïve bayes.

Results: Across tasks, the RNNs and linear machine learning classifiers tended to outperform the non-linear machine learning classifiers and all classifiers except for *k*-nearest neighbors and decision trees had significantly better classification rates than the naïve classification rates. For all tasks, RC took the least amount of time to train.

Conclusion: Across tasks, RC had one of the highest classification rates and took the least amount of time to train, therefore going forward, for these types of analyses we recommend using RC. In addition, athletes were successfully classified based on sport suggesting that athletes competing in different sports move differently during non-sport specific movements. Therefore, movement assessment screens should incorporate sport-specific scoring criteria.

7.2 INTRODUCTION

Movement screens are used to identify aberrant movement patterns believed to increase risk of injury and/or impede performance. Individuals' kinetic and/or kinematic data can be recorded as they perform standardized motions, where those data can be evaluated to determine the correctness and/or proficiency of movement (Cook et al., 2014; Donà et al., 2009; Kritz et al., 2009; McCall et al., 2014; McCunn et al., 2016; Padua et al., 2009). A criticism of movement screens is that they are not able to predict injury risk (McCunn et al., 2016), which is thought to be due to a lack of sensitivity within the scoring criteria (Frost et al., 2015b). The scoring criteria does not account for natural variability between athletes (Frost et al., 2015b), which may be due to skill level or sport played. Previous research has found that during running, using a single accelerometer placed on the hip (Kobsar et al., 2014), runners can be differentiated from soccer players, however, it is unknown whether athletes of different sports move differently during non-sport specific movement screens. Another limitation of movement screens is the poor inter- and intra-rater reliability due to scores being based on visual appraisal (Gulgin and Hoogenboom, 2014; McCunn et al., 2016; Onate et al., 2012; Smith et al., 2013).

Therefore, our previous research has focused on the development of an objective movement screening tool that uses three-dimensional (3D) motion captured data, principal component analysis (PCA), ensemble feature selection, and traditional machine learning to differentiate between elite and novice athletes (Ross et al., 2020, 2018). Our previous research found that for differentiating elite and novice athletes, linear traditional machine learning classifiers outperformed other non-linear traditional machine learning classifiers both when using optical motion capture data and simulated IMU data (Ross et al., 2020). While past research highlights the utility of linear classifiers in the context of athlete movement screening, classification rates ranged from 78.1-84.7% depending on the movement task (Ross et al., 2020). Emerging methods that better consider the time-series nature of movement, may permit better classification

Traditionally, time-series data has been considered one of the most challenging problems for machine learning as it tends to be complex, high-dimensional, highly-correlated (Esling and Agon, 2012), inherently noisy (Yang and Wu, 2006), and prone to overfitting (Esling and Agon,

2012). Previous research has focused on the development of techniques, such as principal component analysis (PCA) and recurrent neural networks (RNNs) to combat traditional issues with time-series data. In RNNs, the connections between nodes form a directed graph along a temporal sequence, which allows the network to learn dynamic, temporally-varying behavior (Hochreiter and Schmidhuber, 1997; Lukoševicius and Jaeger, 2009; Troje, 2002a; Weng and Shen, 2008).

RNNs can then use past information along with current input to calculate an output, which. works well when the gap between the current input and the relevant stored past information is small, however, if the gap is too large, the relevant information is lost (Hochreiter and Schmidhuber, 1997). LSTMs are a type of RNN that were developed to retain relevant information regardless of the sequential gap size (Hochreiter and Schmidhuber, 1997) by incorporating memory cells and gate units. Another limitation of traditional RNNs is that backpropagation is computationally expensive as it backpropagates through time requiring large amounts of computing power and time to train the networks (Lukoševicius and Jaeger, 2009). To address this problem, RC was developed. RC randomly generates the recurrent part (the reservoir) and then keeps it fixed with only the weights of the readout layer being trained, unlike the traditional RNN that trains the weights of each layer (Lukoševicius and Jaeger, 2009), decreasing the computational needs. However, it is unknown if RNNs will outperform the previously used traditional machine learning classifiers.

Therefore, the purpose of this study was two-fold: 1) to determine if RNN models designed for time-series analyses (i.e., LSTM and RC) can outperform the previously used linear and non-linear classifiers at classifying athlete skill level, and 2) if athletes can be differentiated based on the sport played, using the same technique, and if so, to identify which machine learning algorithm(s) discussed previously performs the best. To the best of the authors' knowledge, this is the first study to attempt to classify athletes based on sport played using movement patterns as input data.

7.3 METHODS

7.3.1 Participants

Kinematic data were collected from 542 athletes ranging in competition level from youth to professional (e.g., NBA, NFL, FIFA, MLB, etc.) and competing in 11 different sports (baseball, basketball, soccer, golf, tennis, track and field, squash, cricket, lacrosse, football, or volleyball). However, due to collection error, obstructed data, and not all athletes performing all tasks, the number of athletes per task varied (see tables S1-S4 in the supplementary material). Before data collection, each athlete read and signed an informed consent form permitting the data to be used for secondary analyses. The Health Sciences Research Ethics Board at the University of Ottawa approved the secondary use of the data for research purposes (file no: H-08-18-1085). Athletes competing at the inter-collegiate, semi-professional, or professional level were considered elite athletes, whereas athletes competing at all other levels were considered novice. For classifying sports, only elite athletes competing in football (n = 128), baseball (n = 96), basketball (n = 37), and soccer (n = 33) were analyzed, as these were the sports with the greatest number of athletes. For the classification of sport, after constraining to elite athletes from these four sports all retained athletes were male (see Table S2 in the supplementary material).

7.3.2 Protocol

Before data collection began, each athlete provided an injury history from the past 10 years, had their height (with shoes), and weight taken. For the data collection, the athletes were outfitted with 45 passive, reflective markers (B&L Engineering, Santa Ana, CA) to capture whole-body kinematics (McPherson et al., 2016; Ross et al., 2020, 2018). Once outfitted with the markers, the athlete completed both a static and dynamic calibration trial.

After the calibration trials, each athlete completed a movement screening battery consisting of 21 unique movements which were selected to test each athletes' range of motion at each joint and their global stability, power, and balance. For this study, only seven movements were used in the analysis. The seven tasks retained were the bird-dog, drop-jump, hop-down, L-hop, lunge, step-down, and T-balance, with each movement, except the drop-jump, being performed bilaterally, resulting in a total of 13 movement trials (Ross et al., 2020, 2018). Only one trial for each task was retained, with the athlete performing the task until they believed that they had performed it to the

best of their ability. Kinematic data were captured using an 8-camera Raptor-E (Motion Analysis Corporation, Santa Rosa, CA) motion capture system at a rate of 120 Hz.

7.3.3. Data Analysis

7.3.3.1 Pre-processing

All motion capture data were collected, labeled, and gap-filled using Cortex (Motion Analysis Corporation, Santa Rosa, CA). The data were then exported to (Visual3D C-Motion, Inc., Germantown, MD) and a 3D model was developed using the static calibration trial and applied to the motion capture data to calculate joint centers bilaterally for the wrist, elbow, shoulder, foot, ankle, knee, and hip; centres of gravity for the trunk, head, and pelvis; and marker positional data for the left and right heel, T2, T8, sternum, and the back, front, and sides of the head. Data were then exported to Python and all trials were trimmed to specific start and end-point criteria (Ross et al., 2018), and filtered using a zero-phase, fourth-order, low-pass Butterworth filter with a cut-off frequency of 15 Hz. Due to there being significant differences in height between groups (level: $F = 138.25$, $p < 0.001$; sport: $F = 12.32$, $p < 0.001$), the data for each movement were normalized to each athlete's height. The 3D positional data were then rotated and translated so that the local coordinate system of the trunk was aligned with the global coordinate system with the origin of the coordinate system aligning with the midpoint between the left and right hip of the first frame of the data. The rotated data were then time-normalized to 500 frames.

7.3.3.2 Traditional Machine Learning

The time-normalized data were then placed in an [n x 39,000] matrix for each movement, where n was the number of subjects and 39,000 was the time-normalized data for the x, y, and z axes for each joint centre, centre of gravity, and retained markers (500 timepoints x 3 axes x 26 positions). PCA was then applied to each matrix, resulting in a unique model per task per classifier. Using the principal components (PC) scores as features to classify skill level and sport, the Python scikit-learn library (Pedregosa et al., 2011) ensemble feature selection with Pearson Correlation, chi-squared, recursive feature elimination, lasso, random forest, and LightGBM was used to rank the PC scores based on contribution for each movement task and classifier. The PC scores that ranked in the top 25 for at least 50% of the techniques were retained for that task and classifier.

To minimize overfitting, the maximum number of PC scores retained was less than or equal to the square root of the number of subjects for that task and classifier.

To classify the data, PC scores retained from the ensemble feature selection were used as inputs and either sport (basketball/football/baseball/soccer) or level (elite/novice) were used as the class. Seven classifiers were used: binary (level) or multiple (sport) logistic regression (MLR), a decision tree classifier (DT), *k*-Nearest Neighbors (KNN), naïve bayes (NB), linear discriminant analysis (LDA), and support vector machines with a linear (SVM) and radial basis function (RBF) kernel. All classifiers were programmed using the scikit-learn library (Pedregosa et al., 2011).

7.3.3.3 RNNs

For the time-series-specific classifiers (RC and LSTM), for each task, a 3D matrix was constructed that was [n x 500 x 78], where n was the number of subjects, 500 was the number of frames, and 78 was the number of positions times the number of axes (26 x 3). The class labels were then encoded into one-hot arrays. For both RC (Bianchi et al., 2021) and LSTM (Clouthier et al., 2021), the hyperparameters were tuned using grid search (see Appendix C, Tables C5 and C6 for hyperparameters tuned and ranges).

7.3.3.4 Validation

All classifiers were validated using 10-fold cross-validation with 80% of the data being used for training and 20% for testing. Due to unequal group sizes for both level and sport, the training and testing sets were assigned the same ratio of elite and novice and basketball, football, baseball, and soccer players as the overall sample. Lastly, a naïve algorithm was used that predicted all athletes in the testing set as the majority group for that training set. For all classifiers, accuracy and time to complete each cross-fold was recorded and the mean and standard deviation across all tasks were calculated for each classifier. In addition, the maximum random access memory (RAM) required across all tasks for each classifier were recorded.

7.3.3.5 Statistics

In SPSS 20.0 (IBM, Armonk, New York, USA), Shapiro-Wilks tests were run to test for normality ($p > 0.05$). With the data being normally distributed, one-way ANOVAs with Tukey post-hoc tests were run in SPSS for each class and movement task between the ten different classifiers to test for significant differences ($p < 0.05$) in accuracy.

7.4 RESULTS

7.4.1 Time

For both level and sport, RC took the least amount of time to train with an average time of 0.16 ± 0.08 seconds and 0.12 ± 0.06 seconds, respectively (Table 7.4.1). The traditional classifiers all had similar training times with average times of 25.6 ± 11.7 seconds (level) and 81.6 ± 13.7 seconds (sport; Table 7.4.1). Due to the LSTM being computationally expensive and in order to reduce training times, all LSTM models were trained using an NVIDIA Titan RTX GPU and had average times of 76.6 ± 52.0 seconds and 23.4 ± 18.9 seconds for level and sport, respectively, where 69.1 GB (level) and 88.4 GB (sport) of RAM were required (Table 7.4.1). All other classifiers were less computationally expensive and were able to be trained in a timely fashion using an Intel® Xeon® Gold 6248 CPU and 384GB of ECC RAM. In terms of memory, the traditional classifiers all required the same amount (level: 2.4 GB, sport: 15.8 GB). RC required 12.6 GB and 48.0 GB for level and sport, respectively (Table 7.4.1).

Table 7.4.1. The computer architecture, the average training time in seconds per cross-fold validation and standard deviation, and the maximum memory needed across all tasks for level and sport.

		Level		**Sport**	
ML Algorithm	*Computing Architecture*	*Time (s)*	*Memory (GB)*	*Time (s)*	*Memory (GB)*
RC	CPU	0.16 (0.08)	12.6	0.12 (0.06)	48.0
LSTM	GPU	76.7 (52.0)	69.1	23.4 (18.9)	88.4
LDA	CPU	25.6 (11.7)	2.4	81.6 (13.7)	15.8
MLR	CPU	25.6 (11.7)	2.4	81.6 (13.7)	15.8
SVM	CPU	25.6 (11.7)	2.4	81.6 (13.7)	15.8
KNN	CPU	25.6 (11.7)	2.4	81.6 (13.7)	15.8
DT	CPU	25.6 (11.7)	2.4	81.6 (13.7)	15.8
NB	CPU	25.6 (11.7)	2.4	81.5 (13.7)	15.8
RBF	CPU	25.6 (11.7)	2.4	81.6 (13.7)	15.8

7.4.2 Level

Overall, the RNNs (LSTM and RC) and the linear classifiers (LDA, MLR, and SVM) outperformed the non-linear classifiers (KNN, DT, NB, and RBF) and the naïve classifier (Figure 7.4.1). Within the non-linear classifiers, RBF and NB outperformed KNN and DT. There was a significant main effect between classifiers for all tasks ($p < 0.001$; Table 7.4.2). Since similar results were found between tasks, only the results of the post-hoc tests for the lunge right will be

presented in text (see Appendix C for the results of all other post-hoc tests). Looking at just the average classification rates for the lunge right, RC (77.9% ± 4.7%) , LSTM (79.0% ± 3.8%), LDA (76.9% ± 3.9), MLR(77.0% ± 3.8%), and SVM (76.3% ± 3.1%) all had significantly greater classification rates than KNN (66.0% ± 4.3%; $p < 0.001$), DT (62.6% ± 4.6%; $p < 0.001$), and naïve (65.4%; $p < 0.001$), however, there were no significant differences between RC, LSTM, LDA, MLR, or SVM ($p > 0.001$; Table 7.4.3). RBF (73.6% ± 4.1%) had significantly greater classification rates than KNN ($p = 0.001$), NB (72.6% ± 3.9% ; $p < 0.001$), and naïve ($p < 0.001$; Table 7.4.2) and NB had significantly greater classification rates than DT ($p < 0.001$; Table 7.4.3).

Table 7.4.2. The overall results of the one-way ANOVAs for each movement task for level and sport. Main effect is denoted by F and significance level is denoted as p.

	Level		**Sport**	
	F	*p*	*F*	*p*
Bird-Dog Left	12.01	<0.001	11.851	<0.001
Bird-Dog Right	11.427	<0.001	22.948	<0.001
Drop-Jump	17.352	<0.001	15.131	<0.001
Hop-Down Left	10.123	<0.001	16.96	<0.001
Hop-Down Right	14.733	<0.001	14.821	<0.001
L-Hop Left	15.921	<0.001	11.282	<0.001
L-Hop Right	14.74	<0.001	8.811	<0.001
Lunge Left	24.558	<0.001	21.324	<0.001
Lunge Right	23.761	<0.001	20.994	<0.001
Step-Down Left	16.063	<0.001	15.94	<0.001
Step-Down Right	14.964	<0.001	20.939	<0.001
T-Balance Left	5.453	<0.001	26.053	<0.001
T-Balance Right	12.561	<0.001	20.599	<0.001

Table 7.4.3. The results of the Tukey post-hoc test for the lunge for level. The orange cells denote Lunge Left, whereas the blue cells denote Lunge Right. Red font denotes significance.

	Lunge									
	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		1	1	1	0.995	<0.001	<0.001	0.076	0.276	<0.001
LSTM	1		0.967	0.978	0.855	<0.001	<0.001	0.012	0.063	<0.001
LDA	1	1		1	1	<0.001	<0.001	0.276	0.644	<0.001
MLR	1	1	1		1	<0.001	<0.001	0.241	0.595	<0.001
SVM	1	1	1	1		<0.001	<0.001	0.496	0.855	<0.001
KNN	<0.001	<0.001	<0.001	<0.001	<0.001		0.595	0.009	0.001	1
DT	<0.001	<0.001	<0.001	<0.001	<0.001	0.045		<0.001	<0.001	0.819
NB	1	0.992	0.986	0.986	1	0.001	<0.001		1	0.03
RBF	0.999	0.978	0.967	0.967	0.998	0.002	<0.001	1		<0.001
Naïve	<0.001	<0.001	<0.001	<0.001	<0.001	0.234	1	<0.001	<0.001	

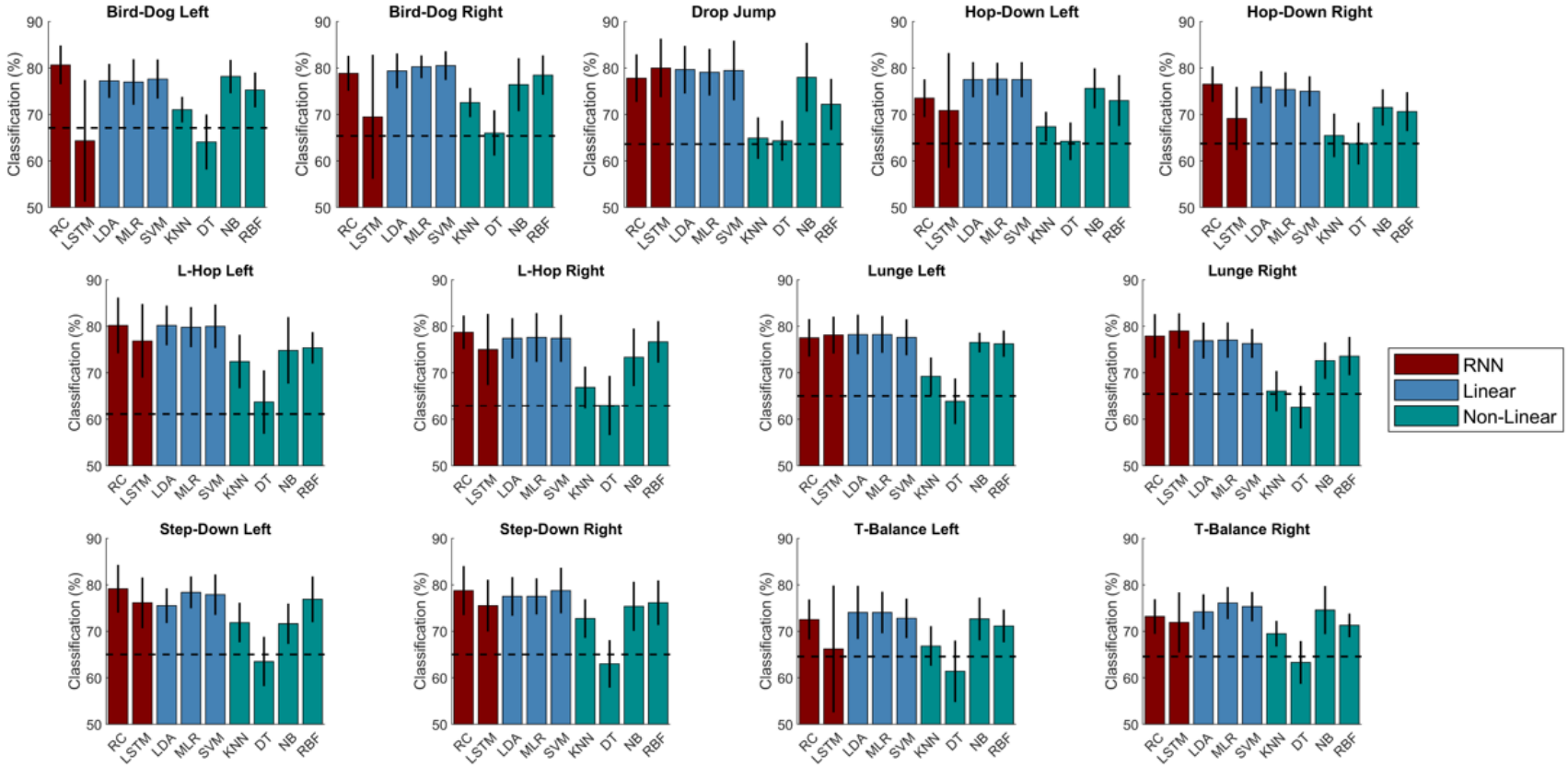


Figure 7.4.1. The mean classification rates and standard deviations when classifying for level for all movement tasks and all machine learning classifiers. Maroon bars represent recurrent neural network classifiers, blue bars represent linear machine learning classifiers and green bars represent non-linear machine learning classifiers. The dotted black line denotes the naïve classification rates.

7.4.3 Sport

For classifying sport, the results closely resembled classification of athlete's level. Overall, the RNNs and linear classifiers outperformed the non-linear classifiers and the naïve classifier, and within the non-linear classifiers, RBF and NB outperformed DT and KNN (Figure 7.4.2). There was a significant main effect between the different classifiers for all tasks ($p < 0.001$; Table 7.4.2). Similar to level, similar results were found between tasks, therefore only the results of the post-hoc test for the lunge right will be presented in text (see Appendix C for the results of all other post-hoc tests). Looking at just mean accuracy rates for the lunge right, RC (72.9% ± 6.1%) and LSTM (72.5% ± 6.5%) had a significantly better classification rates than KNN (58.1% ± 5.2%; $p < 0.001$), DT (47.5% ± 9.0%; $p < 0.001$), and naïve (47.9%; $p < 0.001$), however, there were no significant differences between RC and LSTM ($p = 1.00$; Table 7.4.4). LDA (66.0% ± 6.9%), MLR (65.8% ± 7.2%), SVM (65.6% ± 5.6%), NB (64.4% ± 7.0%), and RBF (63.8% ± 2.0%) had significantly greater classification rates than DT ($p < 0.001$) and naïve ($p < 0.001$), however there were no significant differences between LDA, MLR, SVM, NB and RBF ($p > 0.001$; Table 7.4.4). KNN ($p = 0.011$) and DT ($p = 1.00$) were not significantly different than naïve (Table 7.4.4).

Table 7.4.4. The results of the Tukey post-hoc test for the lunge for sport. The orange cells denote lunge left, whereas the blue cells denote lunge right. Red font denotes significance.

	Lunge									
	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		1	0.273	0.235	0.201	<0.001	<0.001	0.068	0.036	<0.001
LSTM	1		0.358	0.314	0.273	<0.001	<0.001	0.1	0.056	<0.001
LDA	0.024	0.086		1	1	0.121	<0.001	1	0.998	<0.001
MLR	0.018	0.067	1		1	0.144	<0.001	1	0.999	<0.001
SVM	0.006	0.024	1	1		0.171	<0.001	1	1	<0.001
KNN	<0.001	<0.001	<0.001	<0.001	0.002		0.171	0.406	0.559	0.011
DT	<0.001	<0.001	0.001	0.001	0.003	1		<0.001	<0.001	1
NB	<0.001	0.002	0.971	0.983	0.999	0.018	0.031		1	<0.001
RBF	0.031	0.108	1	1	1	<0.001	<0.001	0.953		<0.001
Naïve	<0.001	<0.001	<0.001	<.001	<.001	0.86	0.763	<0.001	<0.001	

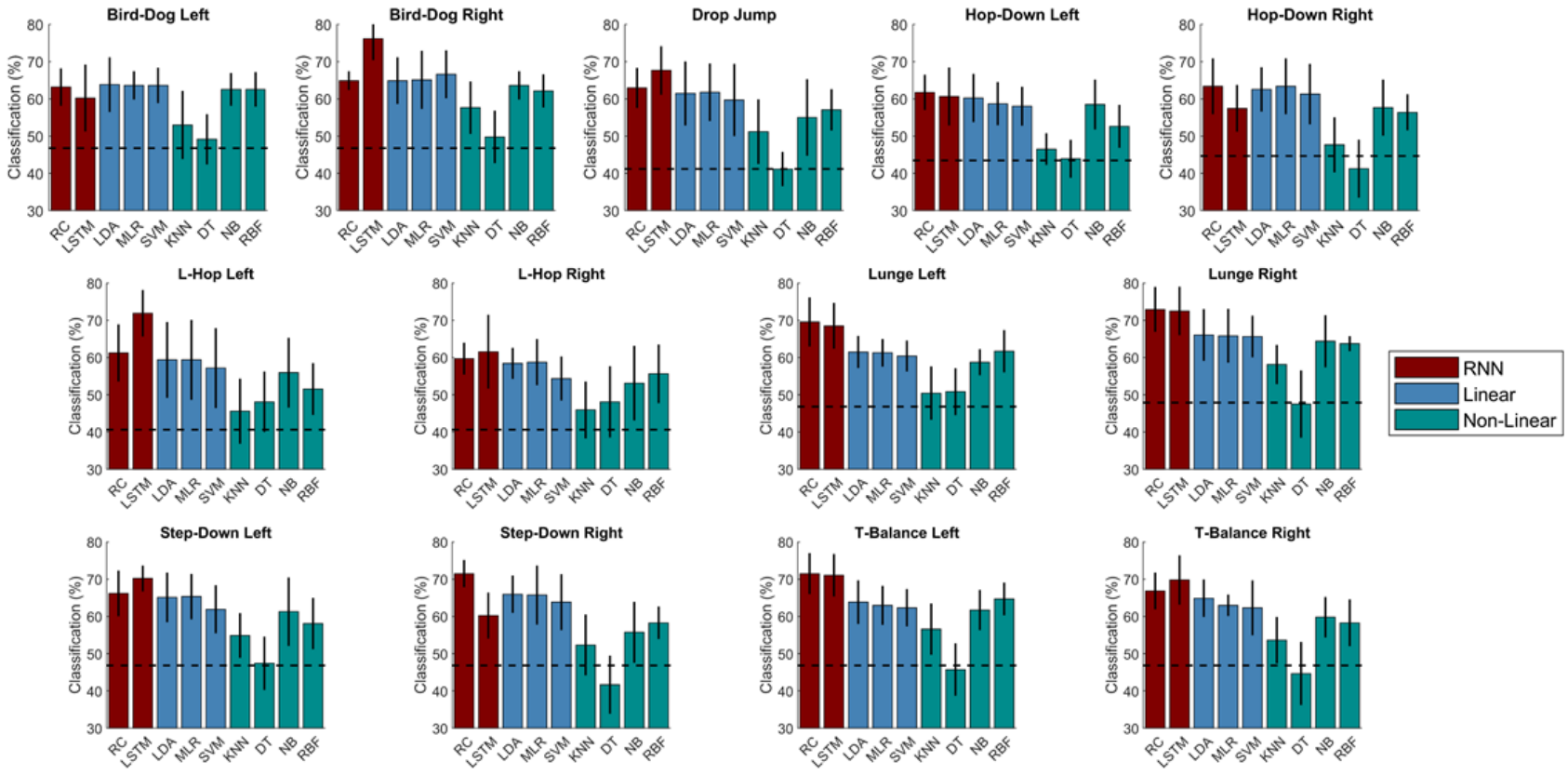


Figure 7.4.2. The mean classification rates and standard deviations when classifying for sport for all movement tasks and all machine learning classifiers. Maroon bars represent recurrent neural network classifiers, blue bars represent linear machine learning classifiers and green bars represent non-linear machine learning classifiers. The dotted black line denotes the naïve classification rates.

7.5 DISCUSSION

The purpose of this study was two-fold: 1) to determine if RC and LSTM were better at predicting athletic level than the machine learning classifiers previously used, and 2) to determine if athletes could be classified based on sport. Overall, for both level and sport, the RNNs performed better than non-linear classifiers and similarly to the linear classifiers. Both the RNNs and linear classifiers performed better than the naïve classifier.

For level, RC and LSTM performed similarly for most tasks except for the bird-dog left and right and the T-balance, where RC performed better than LSTM, with only the bird-dog left being significantly different. In addition, for the bird-dog left and right, hop-down left and right, L-hop left and right, and the T-balance left and right, LSTM had higher variability than RC. This was expected due to the nature of the classifiers. RNNs in general are known for being unstable due to the exploding/vanishing gradient effect due to back propagation through all layers (Bengio et al., 1994; Hochreiter, 1998). RC is designed to be more stable due to only manipulating the weights of the read-out layer (Lukoševicius and Jaeger, 2009). For the linear and non-linear classifiers, similar results were found to previous research (Ross et al., 2020). The linear classifiers within themselves all performed similarly. For the non-linear classifiers, KNN and DT tended to perform significantly worse than NB and RBF.

For sport, the RNNs performed similarly to each other, except for the bird-dog left where LSTM had significantly greater classification rates ($p = 0.001$). The instability of the LSTM seen with the level classifier was not as present for the sport classifier, which may be due to the smaller sample sizes and to larger differences in group sizes rather than the models being more stable. Again, similar to previous research (Ross et al., 2020) and the level classifier, the linear classifiers within themselves performed similarly and for the non-linear classifiers, KNN and DT tended to perform significantly worse than NB and RBF.

All classifiers were trained on an Intel® Xeon® Gold 6248 CPU and 384GB of ECC RAM, except for LSTM which was trained on an NVIDIA Titan RTX GPU. For both sport and level, RC took the least amount of time. These results agree with previous research that has found that the LSTM and RC models perform similarly, but that the LSTM takes a significantly longer time to train due to backpropagating through each layer (Jirak et al., 2020). The traditional classifiers all

performed similarly, with the feature selection portion requiring the most amount of time, and took longer to train than RC.

For all tasks except for the T-balance left for level, the RC and the linear classifiers had significantly better classification rates than the naïve classifier, suggesting that the classifiers were classifying based on actual differences between classes and not noise or by chance. Due to having one of the highest classification rates and taking the least amount of time to train for all tasks, going forward, it is suggested to use RC for these types of analyses. In addition, it was possible to classify athletes based on the sport, which suggests that athletes move differently based on the sport they play. Currently, popular movement screens do not consider athlete-specific demographics such as competition level, sport, age, or sex. Based on athletes moving differently depending on their level of play and sport, it argues that there should be sport- and level- specific scoring criteria for movement competency assessments, which may increase their ability to predict injury risk.

The dataset for this study contained unequal group sizes especially when classifying sport which may limit the generalizability of these results. To help combat this limitation, the ratio of athletes within groups for the entire dataset was maintained within each cross-fold. A further limitation for classifying sport was the absence of female athletes due to the selection criteria being restricted to elite athletes in baseball, basketball, football and soccer.

In conclusion, across all tasks, RC had one of the highest classification rates and took the least amount of time to train, therefore going forward, for these types of analyses we recommend using RC. Both the RNNs and linear classifiers performed better than the non-linear and naïve classifiers for level of play and sport played, suggesting the classifiers were classifying based on differences rather than chance. In addition, athletes were successfully classified based on sport suggesting that athletes competing in different sports move differently during non-sport-specific movements. Therefore, movement assessment screens should incorporate sport-specific scoring criteria.

7.6 ACKNOWLEDGEMENTS

We would like to thank our funding sources: the Natural Sciences and Engineering Research Council (NSERC) of Canada (PGSD3-504132-2017) and the Ontario Graduate

Scholarship. In addition, we would like to thank Motus Global, Inc. for the data and the athletes who partook in the study.

7.7 CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest and results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

7.8 REFERENCES

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Networks 5, 157–166.

Bianchi, F.M., Scardapane, S., Lokse, S., Jenssen, R., 2021. Reservoir computing approaches for representation and classification of multivariate time series. IEEE Trans. Neural Networks Learn. Syst. 32, 2169–2179.

Clouthier, A.L., Ross, G.B., Mavor, M.P., Coll, I., Boyle, A., Graham, R.B., 2021. Development and validation of a deep learning algorithm and open-source platform for the automatic labelling of motion capture markers. IEEE Access 9, 36444–36454.

Cook, G., Burton, L., Hoogenboom, B.J., 2014. Functional movement screening : The use of fundamental movements as an assssment of function- Part 2. Int. J. Sports Phys. Ther. 9, 549–563. doi:10.1111/j.1600-0838.2010.01267.x

Donà, G., Preatoni, E., Cobelli, C., Rodano, R., Harrison, A.J., 2009. Application of functional principal component analysis in race walking: An emerging methodology. Sport. Biomech. 8, 284–301. doi:10.1080/14763140903414425

Esling, P., Agon, C., 2012. Time-series data mining. ACM Comput. Surv. 45, Article 12. doi:10.1145/2379776.2379788

Frost, D.M., Beach, T.A.C., Campbell, T.L., Callaghan, J.P., Mcgill, S.M., 2015. An appraisal of the Functional Movement Screen^{TM} grading criteria - Is the composite score sensitive to risky movement behavior? Phys. Ther. Sport 16, 324–330. doi:10.1016/j.ptsp.2015.02.001

Gulgin, H., Hoogenboom, B., 2014. The Functional Movement Screening (FMS)^{TM}: An inter-rater reliability study Between raters of varied experience. Int. J. Sports Phys. Ther. 9, 14–20.

Hochreiter, S., 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. Int. J. Uncertainty, Fuzziness, Knowledge-Based Syst. 6, 107–116.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.

Jirak, D., Tietz, S., Ali, H., Wermter, S., 2020. Echo state networks and long short-term memory for continuous gesture recognition: A comparative study. Cognit. Comput. 1–13.

Kobsar, D., Osis, S.T., Hettinga, B.A., Ferber, R., 2014. Classification accuracy of a single tri-axial accelerometer for training background and experiece level in runners. J Biomech. 47, 2508-2511.

Kritz, M., Cronin, J., Hume, P., 2009. The bodyweight squat: A movement screen for the squat pattern. Strength Cond. J. 31, 76–85. doi:10.1519/SSC.0b013e318195eb2f

Lukoševicius, M., Jaeger, H., 2009. Reservoir computing approaches to recurrent neural network training. Comput. Sci. Rev. 3, 127–149. doi:10.1016/j.cosrev.2009.03.005

McCall, A., Carling, C., Nedelec, M., Davison, M., Le Gall, F., Berthoin, S., Dupont, G., 2014. Risk factors, testing and preventative strategies for non-contact injuries in professional football: Current perceptions and practices of 44 teams from various premier leagues. Br. J. Sports Med. 48, 1352–7. doi:10.1136/bjsports-2014-093439

McCunn, R., aus der Fünten, K., Fullagar, H.H.K., McKeown, I., Meyer, T., 2016. Reliability and association with injury of movement screens: A critical review. Sport. Med. 46, 763–781. doi:10.1007/s40279-015-0453-1

McPherson, A.L., Dowling, B., Tubbs, T.G., Paci, J.M., 2016. Sagittal plane kinematic differences between dominant and non-dominant legs in unilateral and bilateral jump landings. Phys. Ther. Sport 22, 54–60.

Onate, J.A., Dewey, T., Kollock, R.O., Thomas, K.S., Van Lunen, B.L., DeMaio, M., Ringleb, S.I., 2012. Real-time intersession and interrater reliability of the Functional Movement Screen. J. Strength Cond. Res. 26, 408–415. doi:10.1519/JSC.0b013e318220e6fa

Padua, D.A., Marshall, S.W., Boling, M.C., Thigpen, C.A., Garrett, W.E., Beutler, A.I., 2009. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. Am. J. Sports Med. 37, 1996–2002. doi:10.1177/0363546509343200

Pedregosa, F., Weiss, R., Brucher, M., 2011. Scikit-learn : Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Ross, G.B., Dowling, B., Troje, N.F., Fischer, S.L., Graham, R.B., 2018. Objectively differentiating movement patterns between elite and novice athletes. Med. Sci. Sport. Exerc.

50, 1457–1464. doi:10.1249/MSS.0000000000001571

Ross, G.B., Dowling, B., Troje, N.F., Fischer, S.L., Graham, R.B., Graham, R.B., 2020. Classifying elite From novice athletes using simulated wearable sensor data. Front. Bioeng. Biotechnol. 8, 1–10. doi:10.3389/fbioe.2020.00814

Smith, C.A., Chimera, N.J., Wright, N.J., Warren, M., 2013. Interrater and intrarater reliability of the Functional Movement Screen. J. Strength Cond. Res. 27, 982–987.

Troje, N.F., 2002. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. J. Vis. 2, 371–387. doi:10.1167/2.5.2

Weng, X., Shen, J., 2008. Knowledge-based systems classification of multivariate time series using two-dimensional singular value decomposition. Knowledge-Based Syst. 21, 535–539. doi:10.1016/j.knosys.2008.03.014

Yang, Q., Wu, X., 2006. 10 challenging problems in data mining research. Int. J. Inf. Technol. Decis. Mak. 5, 597–604.

Chapter 8: General Discussion

The general overarching goal of this thesis was to develop an objective movement screening tool that utilizes whole-body kinematic data and machine learning. More specifically, the three global objectives were: 1) to determine the inter- and intra-rater reliability of a movement screen without task-specific scoring criteria, 2) to develop an objective movement screening tool that is capable of objectively scoring an athlete's movement competency, and 3) to refine the objective movement screening tool to score movement competency based on multiple types of data and classifiers. The first global objective was achieved through Study 1, which looked at the inter- and intra-rater reliability of the movement screen across 10 expert assessors. In addition, the study narrowed in on intra-rater reliability and looked at how intra-rater reliability changed between sessions, within the same session, and within the same session when body-shape was manipulated to represent someone who was underweight, normal, and overweight. The second global objective was achieved through Study 2, which developed a framework for the objective movement screening tool that used principal component analysis (PCA) and linear discriminant analysis (LDA) to classify and score movement competency of elite and novice athletes, providing proof-of-concept for the objective movement screening tool. The third global objective was achieved through Study 3 & 4. Study 3 explored using a wider breadth of increasingly complex machine learning algorithms such as support vector machines, binary logistic regression and decision trees, as well as, testing the objective tool's ability to classify elite and novice athletes using two different types of data (i.e., optical motion capture and simulated inertial measurement unit (IMU) data). Study 4 further refined the objective movement screening tool by exploring neural networks well-suited for time-series data and if the objective tool could classify athletes based on sport played. The results of this thesis show promise of an objective movement screening tool for future use within the field.

8.1 Rater Reliability

It was hypothesized for rater reliability that intra-rater within the same session without body-shape manipulation would have the highest agreement, followed by intra-rater between session and intra-rater within the same session with body-shape manipulation, and inter-rater

reliability would have the lowest agreement. The results of the first study supported these hypotheses, where intra-rater reliability within session without body modification and between sessions ranged from fair to moderate agreement and intra-rater reliability with body-shape modification had slight to moderate agreement when looking across tasks and subjects. When looking at the average weighted kappa values, inter-rater reliability within session with and without body manipulation and between sessions were 0.45, 0.37, and 0.35, respectively.

Although the weighted Cohen's kappa values are on the lower end of the spectrum when compared to other studies, the pattern of intra-rater reliability being better than inter-rater reliability is similar to previous results (Glaws et al., 2014; Onate et al., 2012; Smith et al., 2013). The lower scores could be due to the larger number of possible scores, the greater number of assessors being compared, the difference in scoring criteria, and/or use of virtual avatars. The Functional Movement Screen (FMS) is scored between 0-3 for each task (Cook et al., 2014, 2006), whereas the movements in Study 1 were scored between 1-10. Although with greater number of scores there is greater sensitivity, the probability of assessors selecting the same score is decreased. In addition, the sensitivity may be greater, but the human eye may not be able to distinguish the differences. Previous studies compared 2 (Onate et al., 2012), 3 (Glaws et al., 2014) and 4 (Smith et al., 2013) raters, whereas Study 1 compared 10 raters. The increase in number of raters may also contribute to the lower kappa values. Additionally, scoring virtual avatars rather than human participants may have lowered scores, however, it would not be possible to modify body-shape with human participants. Although the greater range in scores, number of raters, and/or avatars likely contributed to the lower kappa values, the main reason was likely due to the scoring criteria.

For the movement screens that previously assessed inter-rater and intra-rater reliability, given task-specific scoring criteria were used to assess movement competency (Glaws et al., 2014; Onate et al., 2012a; Smith et al., 2013), whereas for Study 1, the assessors were asked to use their expertise to establish their own whole-body scoring criteria. Previous research has criticized the FMS for having poor criterion validity, which was attributed to the vagueness of the scoring criteria (Whiteside et al., 2014). In addition, many of the FMS task-specific scoring criteria are not linked (epidemiologically or biomechanically) to injury mechanisms or risk factors (Frost et al., 2015b). Furthermore, due to the large amount of movement variability between athletes, the FMS scoring criteria may be insensitive to potentially risky movement behaviour, with previous research recommending that whole-body segment and joint kinematics should be incorporated when

administering movement screens (Frost et al., 2015b). Therefore, Study 1 opted to test the reliability of the movement screen without task-specific scoring criteria, which likely led to the lower reliability scores compared to previous research.

In terms of comparing intra-rater within session with and without body-shape modification, to the best of the authors knowledge, this was the first study of its kind that aimed to study the effects of body-shape on intra-rater reliability. The results of Study 1 supported our hypothesis of without body-shape manipulation having better agreement than with body-shape manipulation when compared across tasks. The without body-shape modification animations had identical movements and avatars, whereas, for the with body-shape modification the avatars looked different. In addition, research has consistently shown that there is pervasive implicit and explicit weight bias among clinicians, physical therapists, physical education teachers, and strength and conditioning personnel (Panza, 2018). When looking across subjects, differences in average kappa values between without body-shape modification and with body-shape modification ranged from -0.02 to 0.17, with a negative value indicating better agreement with body-shape manipulation. The range in differences in weighted kappa values between the two conditions suggests that some assessors were more effected by body-shape than others. The single subject who had a slight increase in reliability with body-shape manipulation also had just over double the amount of variability in scores across tasks compared to the without body-shape modification condition, suggesting that the observed differences were most likely attributed to the large amount of variability seen across all conditions. For the majority of assessors, their biases influenced their scoring, which further shows that movement scores are not based solely on the athlete's movement competency, but influenced by extraneous variables, such as biases.

When looking between days for the inter-rater reliability, on average, there was no difference in weighted kappa values between Day 1 and Day 2; whereas there was an average decrease in weighted kappa values from Day 1 and 2 for both intra-session reliability without body-shape manipulation and intra-session reliability with body-shape manipulation of 0.02. Based on these results, a learning effect does not appear to be influencing the results and differences seen at the individual or task level are more likely due to the large variability seen across all conditions.

Overall, inter- and intra-rater reliability were low, with agreement ranging from slight to fair, suggesting that assessing movement competency via subjective assessment is not reliable, especially when there are not task-specific scoring criteria. This is further compounded when athletes with different body-shape types are being assessed, with reliability decreasing on average across participants when body-shape was manipulated. Movement scores are not based solely on the athlete's movement competency, but influenced by extraneous variables, such as the rater's personal biases. Study 1 supports the need for the development of objective methods, tools, and thresholds to assess movement competency which has been argued by previous literature (Frost et al., 2015b; Whiteside et al., 2014) .

8.2 Machine Learning Algorithms

Traditionally, time-series data have been considered one of the most challenging problems for machine learning as it tends to be complex, high-dimensional, highly-correlated, inherently noisy, and prone to overfitting (Esling and Agon, 2012; Yang and Wu, 2006). Previous research has focused on the development of techniques and classifiers to combat traditional issues with time-series data, for example: using data-reduction techniques such as singular value decomposition (SVD) and PCA; and, the development of recurrent neural networks (RNNs; Hochreiter and Schmidhuber, 1997; Lukoševicius and Jaeger, 2009; N F Troje, 2002; Weng and Shen, 2008). Therefore, in a Study 2, an objective movement tool framework was developed using PCA and LDA to classify athletes based on level of play (e.g., elite versus novice) as proof-of-concept. Study 3 and 4 then refined the framework to determine the optimal machine learning algorithm to classify athletes.

Study 3 explored using PCA with both linear and non-linear traditional machine learning classifiers including: linear discriminant analysis, binary logistic regression, support vector machine with a linear and radial basis function kernel, Naïve Bayes, *k*-nearest neighbours, and decision trees to determine an optimal machine learning algorithm. The results of Study 3 showed that the linear classifiers outperformed the non-linear classifiers, with their being little difference between linear classifiers. LDA was chosen to be further analyzed, since using the linear function from LDA, one is able to easily calculate a movement competency score and visualize differences in movement patterns between the two classes. In addition, Study 3 also explored the use of feature selection to select the principal component (PC) scores that differentiated the two classes rather

than using the 35 PC scores that described the greatest variance and found that the inclusion of feature selection within the framework increased classification rates. This was thought to be due to the tight constraints surrounding each movement. The athletes had to achieve certain start, mid, and end point criteria in order to have the movement be viable. Therefore, the first PCs, which represent the greatest amount of variability, were representing those gross movements, whereas the differences between elite and novice athletes were potentially those smaller movements that would be captured by later PC scores.

A limitation of PCA is since a subset of the data is taken, pertinent information for classifying athletes may be excluded, therefore, Study 4 assessed the use of RNNs. RNNs use connections between nodes from a directed graph along a temporal sequence, allowing for the networks to learn dynamic, temporally-varying behaviour, such as time-series data (Hochreiter and Schmidhuber, 1997; Lukoševicius and Jaeger, 2009). When compared to linear classifiers, RNNs performed similarly for both classifying level of play and sport played, however, reservoir computing required less time to train compared to the traditional linear classifiers. However, reservoir computing is not able to easily explain or visualize kinematic differences between the classes, as it is more of a 'black box' approach compared to the other linear classifiers. Therefore, depending on the intended purpose of the results of the objective tool, it may be beneficial to use a traditional linear classifier such as LDA, although they require more training time, to be able to score and visualize the differences between desired kinematic patterns.

8.4 Data Types

Study 3 examined the use of IMU data for differentiating athletes based on skill level. Previously, the framework was using optical motion capture data collected using an 8-camera high speed camera motion system, which is costly in both time and money, creating a barrier for adoption of use of the equipment in the field. Due to the dataset being collected for commercial purposes prior to the conception of Study 3 and it not being realistic nor feasible to recreate the dataset with both optical and IMU data collected on athletes, the data used for Study 3 were simulated IMU data calculated from the previously collected optical motion capture data. Results showed that athletes can be differentiated based on skill level using simulated IMU data and in fact, models using simulated IMU data, on average, outperformed models using optical motion capture data across all tasks. This is thought to do be due to the different types of data that are

being captured with the two different types of motion capture technology. The optical motion capture data were using joint centre trajectories, which due to the constrained nature of the tasks, may be capturing more gross motor patterns that are unrelated to skill. In contrast, the simulated IMU data comprised of linear acceleration and angular velocity that are more likely to capture the smoothness of the movement, which may be a better indicator of skill level than gross motor patterns.

For both data types, when looking at trends in individual athlete data across tasks, if there were differences in how the athlete was classified between tasks, athletes tended to be classified the same on all tasks that were targeting the same skill set (e.g., trunk stability, jumping, balance) and if there were disparities on how the left/right tasks were classified, the dominant side was usually classified as elite. This suggests that relevant differences between elite and novice-like movement patterns can be detected using both data types. However, similar to RNNs, depending on the application of the results of the objective tool, a possible limitation of using IMU data is that it is more difficult to interpret the differences between classes compared to optical motion capture. With only linear accelerations and angular velocities, and no video data, it is hard to discern how the athlete is moving within space. IMUs may offer an inexpensive measurement device to objectively screen movement abilities, where those individuals identified with weaknesses can then be tested more in depth with optical motion capture to inform targeted corrective exercise approaches.

8.5 Classes

Throughout this thesis, two different classes were assessed, level of play and sport played. For level of play, elite versus novice athletes were classified whereas for sport, the movement patterns of basketball, soccer, football, and baseball players were compared. The overall classification rates for both level of play and sport played were significantly greater than the naïve, where all athletes were labelled as the largest class in terms of numbers for all tasks. To the best of the author's knowledge, Study 4 was the first study aimed to determining whether differences exist between athletes' movement patterns during non-sport specific movement screening who compete in different sports. Since it was possible to classify athletes based on the sport, it can be determined that athletes move differently based on the sport they play. Currently, popular movement screens do not consider athlete-specific demographics such as competition level, sport,

age, or sex. Based on athletes moving differently depending on their level of play and sport, it argues that there should be sport- and level- specific scoring criteria for movement competency assessments, which may increase movement screens' ability to predict desired results such as likelihood of injury.

8.4 Limitations

Aside from the limitations that were discussed within each study, there are general limitations that should be considered.

8.4.1 Participants

There are many considerations and confounding variables that come into play when working with human participants, including age, height, sex, and athleticism. The data were collected by Motus Global, a for-profit sports biomechanics company aimed at serving world-class athletes. Motus Global was partnered with IMG Academy, an elite training facility that boasts a large percentage of their graduates continuing onto NCAA Division I schools and has been a training centre for many professional and national level teams. Therefore, the level of athleticism of the 'novice' athletes in the dataset compared to the general public, is likely higher. Therefore, the movement competency scores for the novice athletes may be under-valued compared to if a larger spectrum of athletic ability was included. In addition, due the novice athletes on average being younger than the elite athletes, there was a significant difference in height between the novice and elite groups. However, the effects of height were attempted to be mitigated by dividing out the individual height of each athlete from the kinematic data. Lastly, there were significantly more males within the dataset than females, with the majority of the females being novice athletes. Furthermore, in Study 4, due to restricting the inclusion criteria to elite athletes in basketball, baseball, football and soccer, no female athletes were included. Therefore, not only for Study 4, but all studies, the results cannot be generalized to female athletes. Similarly to males, it is hypothesized that female athletes would be able to be differentiated based on skill level and sport played; however, there would be differences in movement patterns between males and females (Anderson et al., 2015; Chimera et al., 2015).

8.4.2 Assumptions

For this thesis three assumptions were made: 1) the definition of elite and novice athletes, 2) all elite athletes had completed 10,000 hours of training, and 3) elite athletes have superior movement competency compared to novice athletes. Based on previous research, elites were defined as those with greater than 10,000 hours of deliberate practice, therefore, athletes competing at the intercollegiate, semi-professional, and professional level were considered elite athletes, whereas all other lower levels (e.g., youth, recreational, high school, etc.) of play were deemed novice athletes (Baker et al., 2003; Helsen et al., 1998). Although this assumes athletes at the collegiate and professional level completed 10,000 hours of deliberate practice, athletes competing at the professional and inter-collegiate levels would be in the higher echelon of athletes in their sport even if not completing 10,000 hours.

Within the literature, there is no agreement on the definition of good movement quality, with each movement screen consisting of its own unique scoring criteria, however, it generally encapsulates performing correct posture and joint alignment while maintaining balance through the movement battery (McCunn et al., 2016). Therefore, to objectively score the athletes, it was assumed that elite athletes had better movement competency than novice athletes. Although assumed, previous research has shown that elite athletes perform higher on subjective movement competency screens and adopt more mature, skill-specific neuromuscular patterns compared to novice athletes (Fox et al., 2014; Zaggelidis and Lazaridis, 2013).

8.4.2 Homogenous Data

The data of this thesis were collected using the same equipment, were for the majority collected and labelled by the same four researchers, and primarily collected in the same laboratory, which could make it hard to introduce a small amount of new data to the data set. When comparing multiple optical motion capture systems, differences of up to approximately 2 mm have been reported (Richards, 1999). Although this would likely be missed by the human eye, PCA and machine learning algorithms could possibly score movement based on systematic differences between motion capture systems, rather than actual differences in movement competency. In addition, the majority of the data were collected and labelled by the same researchers. Research has shown that there is good reliability when comparing kinematics collected using marker-based motion capture data; however, it was not perfect (Malus et al., 2021; Tsushima et al., 2003), creating small differences that may be able to be detected by machine learning algorithms. The

homogeneity of the data could lead to difficulties in the future when trying to use the objective tool to score or classify newly collected data that is not using the same motion capture equipment, researchers, or laboratory space.

8.5 Future Directions

As this thesis is part of a larger project to commercialize an objective movement screening tool for use in the field, there are multiple future studies that should be completed. Foremost, research should focus on validating the objective movement screening tool, making it more accessible for use in the field, and growing the dataset to better represent the general population.

8.5.1 Validation

The machine learning models have been validated within each study in this thesis with either leave-one-out validation or cross-fold validation; therefore, it is expected that there is good criterion validity. However, it is unknown if the objective screening tool has good external and construct validity. Therefore, future research should focus on testing both external and construct validity. To test for external validity, new data should be collected with different equipment and researchers to test whether the new athletes when projected into the objective screening tool receive realistic scores.

Since optimal movement patterns for these movements are unknown, there is not a measure to compare the objective movement screen directly to determine construct validity. Therefore, there are multiple ways to measure construct validity. Firstly, once external validity is determined, future research should collect new kinematic and kinetic data to be able to model joint loading and muscle activations and to develop optimal control models to identify movement patterns that are optimal based on reduced joint loading and minimizing cost functions. The objective movement screening tool can then be compared against these pre-determined optimal movements. A second method is if the tool is being used for injury prediction or talent identification, a longitudinal study can be used to determine whether there is a correlation between current movement competency and future injury or athletic success, respectively.

8.5.2 Data Types

Currently, the objective movement screening tool has been tested using both marker-based, optical motion capture and simulated IMU data. To better the accessibility of using the objective

movement screening tool in the field, future research should explore additional types of data such as IMU data that are collected using an IMU system and markerless motion capture. Study 3 showed that athletes can be differentiated based on data types that can be collected via an IMU (i.e., linear acceleration and angular velocity); however, those data were calculated from marker-based, optical motion capture data. Therefore, future research should investigate whether athletes are able to be differentiated using linear acceleration and angular velocity data captured using IMUs.

Although IMUs are portable, inexpensive, and have a large capture volume, interpreting the results of IMU data can be difficult since there are no visual representations of how the body is moving in space. Therefore, there has been an increased interest within the biomechanics community in markerless motion capture systems. Markerless motion capture systems use either one or multiple video-cameras and deep neural networks to recognize key points on the body (Mündermann et al., 2006). Markerless motion capture is inexpensive, requires minimal post-processing, and does not require special equipment to be placed on the individual making it very accessible to be used in the field. In addition, because it is video-based, it is more intuitive to interpret the results compared to IMUs. Therefore, future research should investigate the use of markerless motion capture data as the input data for the objective movement screening tool.

8.5.3 Assessing for Talent Identification/Risk of Injury

Movement screens are often used for identifying athletes at a higher risk of injury or for talent identification (Cook et al., 2014; Donà et al., 2009; Kritz et al., 2009; McCall et al., 2014; McCunn et al., 2016; Padua et al., 2009). Throughout studies 2-4, approximately 15-30% of athletes were misclassified based on the task being examined and/or the methodology being used. Future research should perform a longitudinal study to look further into the misclassified athletes and whether novice athletes misclassified as elite athletes have a higher likelihood of making it to an elite level of play or if elite athletes misclassified as novice athletes have a higher risk of injury. In addition, future research should look at the combination of movements, whether it be all movements together or grouping movements that target the same system (e.g., balance, power, trunk stability, etc), to determine whether cumulative scores have better predictability of talent and/or injury-risk than individual movements.

8.5.4 Growing the Dataset

Future research should focus on expanding the current dataset to better represent female athletes, more beginner athletes, and athletes competing in a greater variety of sports in order to better represent the athletic population. In addition, as the dataset grows, the number of classes can be expanded within each demographic data (e.g., instead of classifying as elite and novice, classifying as specific as professional, collegiate, high-school, youth, etc.). Tailored scoring systems should be designed based on the athlete's demographics to create a more sensitive screening tool.

8.6 Conclusions

This doctoral thesis consisted of four studies with the goal of providing foundational research for an objective movement screening tool. When looking at inter- and intra-rater reliability of expert assessors during subjective scoring of movements, there was slight to fair agreement, with inter-rater reliability having the worst agreement. Within intra-rater reliability, agreement was best within in session compared to between session; however, within session, if body-shape was modified, agreement worsened. For the objective movement screening tool, the tool was able to accurately differentiate based on marker-based optical motion capture data and simulated IMU data and was able to differentiate based on the athletes' skill level and sport played. Lastly, both RNNs and PCA in combination with a traditional linear classier were able to differentiate athletes better than naïve classifiers. This thesis is part of a larger project to bring an objective movement screening tool for field-use to market and provides solid foundational knowledge to use in the continued development of an objective movement screening tool.

Chapter 9: General References

Abdi, H., 2007. Signal Detection Theory (SDT), in: Salkini, N., Rasmussen, K. (Eds.), Encyclopedia of Measurement and Statistics. SAGE Publications Ltd., Thouse Oaks, pp. 886–889.

Abdi, H., Williams, L.J., 2010. Principal Component Analysis. Wiley Interdiscip. Rev. Comput. Stat. 2, 433–459. doi:10.1002/wics.101

Agatonovic-Kustrin, S., Beresford, R., 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J. Pharm. Biomed. Anal. 22, 717–727. doi:10.1016/S0731-7085(99)00272-1

Altun, K., Barshan, B., 2010. Human activity recognition using inertial/magnetic sensor units, in: Springer (Ed.), International Workshop on Human Behavior Understanding. Berlin, Heidelberg, pp. 38–51.

Anderlucci, L., Lubisco, A., Mignani, S., 2021. Investigating the judges performance in a national competition of sport dance. Soc. Indic. Res. 156, 783–799. doi:10.1007/s11205-019-02256-z

Anderson, B.E., Newmann, M.L., Huxel Bliven, K.C., 2015. Functional Movement Screen differences between male and female secondary school athletes. J. Strength Cond. Res. 29, 1098–1106.

Armand, S., Watelain, E., Roux, E., Mercier, M., Lepoutre, F.-X., 2007. Linking clinical measurements and kinematic gait patterns of toe-walking using fuzzy decision trees. Gait Posture 25, 475–484. doi:10.1016/j.gaitpost.2006.05.014

Ashouri, S., Abedi, M., Abdollahi, M., Dehghan, F., 2017. A novel approach to spinal 3-D kinematic assessment using inertial sensors: Towards effective quantitative evaluation of low back pain in clinical settings. Comput. Biol. Med. 89, 144–149. doi:10.1016/j.compbiomed.2017.08.002

Bačić, B., 2016. Echo state network ensemble for human motion data temporal phasing: A case study on tennis forehands, in: Hirose, A., Ozawa, S., Doya, K., Ikeda, K., Lee, M., Liu, D.

(Eds.), Neural Information Processing. Springer International Publishing, Cham, pp. 11–18.

Baker, J., Cote, J., Abernethy, B., 2003. Sport-specific practice and the development of expert decision-making in team ball sports. J. Appl. Sport Psychol. 15, 12–25. doi:10.1080/10413200305400

Bardenett, S.M., Micca, J.J., DeNoyelles, J.T., Miller, S.D., Jenik, D.T., Brooks, G.S., 2015. Functional Movement Screen normative values and validity in high school athletes: Can the FMS be used as a predictor of injury? Int. J. Sports Phys. Ther. 10, 303–308.

Baydal-Bertomeu, J.M., Durá-Gil, J.V., Piérola-Orcero, A., Parrilla Bernabé, E., Ballester, A., Alemany-Munt, S., 2016. A PCA-based bio-motion generator to synthesize new patterns of human running. PeerJ Comput. Sci. 2, e102. doi:10.7717/peerj-cs.102

Begg, R., Kamruzzaman, J., 2005. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. J. Biomech. 38, 401–408. doi:10.1016/j.jbiomech.2004.05.002

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Networks 5, 157–166.

Bergmann, J.H.M., Langdon, P.M., Mayagoitia, R.E., Howard, N., 2014. Exploring the use of sensors to measure behavioral interactions: An experimental evaluation of using hand trajectories. PLoS One 9, 1–10. doi:10.1371/journal.pone.0088080

Berrar, D., 2019. Bayes' theorem and naive bayes classifier, in: Encyclopedia of Bioinformatics and Computational Biology.

Bianchi, F.M., Scardapane, S., Lokse, S., Jenssen, R., 2021. Reservoir computing approaches for representation and classification of multivariate time series. IEEE Trans. Neural Networks Learn. Syst. 32, 2169–2179.

Bishop, C.M., 2006. Pattern recognition and machine learning. Springer Science + Business Media, LLC.

Blanchard, N., Skinner, K., Kemp, A., Scheirer, W., Flynn, P., 2019. "Keep me in, Coach!": A computer vision perspective on assessing ACL injury risk in female athletes. 2019 IEEE Winter Conf. Appl. Comput. Vis. 1366–1374. doi:10.1109/wacv.2019.00150

Blockeel, H., 2011. Hypothesis language. Encycl. Mach. Learn. Data Min. doi:10.1007/978-1-4899-7687-1_372

Bolink, S.A.A.N., Naisas, H., Senden, R., Essers, H., Heyligers, I.C., Meijer, K., Grimm, B., 2018. Validity of an inertial measurement unit to assess pelvic orientation angles during gait, sit-stand transfers and step-up transfers: Comparison with an optoelectronic motion capture system. Med. Eng. Phys. 13, 1–7. doi:10.1016/j.medengphy.2015.11.009

Butler, R.J., Contreras, M., Burton, L.C., Plisky, P.J., Goode, A., Kiesel, K., 2013. Modifiable risk factors predict injuries in firefighters during training academies. Work 46, 11–17. doi:10.3233/WOR-121545

Caggiari, S., Worsley, P.R., Payan, Y., Bucki, M., Bader, D.L., 2020. Biomechanical monitoring and machine learning for the detection of lying postures. Clin. Biomech. 80, 105181. doi:10.1016/j.clinbiomech.2020.105181

Carse, B., Meadows, B., Bowers, R., Rowe, P., 2013. Affordable clinical gait analysis: An assessment of the marker tracking accuracy of a new low-cost optical 3D motion analysis system. Physiotherapy 99, 347–351. doi:10.1016/j.physio.2013.03.001

Cavalheiro, G.L., Almeida, M.F.S., Pereira, A.A., Andrade, A.O., 2009. Study of age-related changes in postural control during quiet standing through linear discriminant analysis. Biomed. Eng. Online 8, 1–13. doi:10.1186/1475-925X-8-35

Chen, J., Ahn, C.R., Han, S., 2014. Detecting the hazards of lifting and carrying in construction through a coupled 3D sensing and IMUs sensing system. Comput. Civ. Build. Eng. 1110–1117.

Chimera, N.J., Smith, C.A., Warren, M., 2015. Injury history, sex, and performance on the Functional Movement Screen and Y Balance Test. J. Athl. Train. 50, 475–485. doi:10.4085/1062-6050-49.6.02

Chorba, R.S., Chorba, D.J., Bouillon, L.E., Overmyer, C.A., Landis, J.A., 2010. Use of a functional movement screening tool to determine injury risk in female collegiate athletes. North Am. J. Sport. Phys. Ther. 5, 47–54.

Clermont, C.A., Phinyomark, A., Osis, S.T., Ferber, R., 2019. Classification of higher- and lower-

mileage runners based on running kinematics. J. Sport Heal. Sci. 8, 249–257. doi:10.1016/j.jshs.2017.08.003

Clouthier, A.L., Ross, G.B., Graham, R.B., Graham, R.B., 2020. Sensor data required for automatic recognition of athletic tasks using deep neural networks. Front. Bioeng. Biotechnol. 7, 473. doi:10.3389/fbioe.2019.00473

Clouthier, A.L., Ross, G.B., Mavor, M.P., Coll, I., Boyle, A., Graham, R.B., 2021. Development and validation of a deep learning algorithm and open-source platform for the automatic labelling of motion capture markers. IEEE Access 9, 36444–36454.

Cook, G., Burton, L., Hoogenboom, B., 2006. Pre-participation screening: the use of fundamental movements as an assessment of function–Part 1. North Am. J. Sport. Phys. Ther. NAJSPT 1, 62. doi:10.1055/s-0034-1382055

Cook, G., Burton, L., Hoogenboom, B.J., 2014. Functional movement screening : The use of fundamental movements as an assssment of function- Part 2. Int. J. Sports Phys. Ther. 9, 549–563. doi:10.1111/j.1600-0838.2010.01267.x

Cristianini, N., Shawe-Taylor, J., 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press.

De Wit, B., De Clercq, D., Aerts, P., 2000. Biomechanical analysis of the stance phase during barefoot and shod running. J. Biomech. 33, 269–278. doi:10.1016/S0021-9290(99)00192-X

Deluzio, K.J., Astephen, J.L., 2007. Biomechanical features of gait waveform data associated with knee osteoarthritis. An application of principal component analysis. Gait Posture 25, 86–93. doi:10.1016/j.gaitpost.2006.01.007

Dodson, B., Hammett, P.C., Klerx, R., 2014. Binary logistic regression, in: Probabilistic Design for Optimzation and Robustness for Engineers. pp. 202–224. doi:10.1007/978-1-4614-3417-7_8

Domingos, P., 2012. A few useful things to know about machine learning. Commun. ACM 55, 78. doi:10.1145/2347736.2347755

Domire, Z.J., Challis, J.H., 2017. The influence of squat depth on maximal vertical jump performance 414. doi:10.1080/02640410600630647

Donà, G., Preatoni, E., Cobelli, C., Rodano, R., Harrison, A.J., 2009. Application of functional principal component analysis in race walking: An emerging methodology. Sport. Biomech. 8, 284–301. doi:10.1080/14763140903414425

Eichelberger, P., Ferraro, M., Minder, U., Denton, T., Blasimann, A., Krause, F., Baur, H., 2016. Analysis of accuracy in optical motion capture – A protocol for laboratory setup evaluation. J. Biomech. 49, 2085–2088. doi:10.1016/j.jbiomech.2016.05.007

Emken, J.L., Benitez, R., Sideris, A., Bobrow, J.E., Reinkensmeyer, D.J., 2007. Motor adaptation as a greedy optimization of error and effort. J. Neurophysiol. 97, 3997–4006. doi:10.1152/jn.01095.2006

Ermes, M., Pärkkä, J., Mäntyjärvi, J., Korhonen, I., 2008. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. IEEE Trans. Inf. Technol. Biomed. 12, 20–26. doi:10.1109/TITB.2007.899496

Eskofier, B.M., Federolf, P., Kugler, P.F., Nigg, B.M., 2013. Marker-based classification of young-elderly gait pattern differences via direct PCA feature extraction and SVMs. Comput. Methods Biomech. Biomed. Engin. 16, 435–442. doi:10.1080/10255842.2011.624515

Esling, P., Agon, C., 2012. Time-series data mining. ACM Comput. Surv. 45, Article 12. doi:10.1145/2379776.2379788

Faber, I.R., Elferink-Gemser, M.T., Faber, N.R., Oosterveld, F.G.J., Nijhuis-Van Der Sanden, M.W.G., 2016. Can perceptuo-motor skills assessment outcomes in young table tennis players (7-11 years) predict future competition participation and performance? An observational prospective study. PLoS One 11, 1–9. doi:10.1371/journal.pone.0149037

Federolf, P., Reid, R., Gilgien, M., Haugen, P., Smith, G., 2014. The application of principal component analysis to quantify technique in sports. Scand. J. Med. Sci. Sport. 24, 491–499. doi:10.1111/j.1600-0838.2012.01455.x

Field, A., 2009. Discovering Statistics Using SPSS, 3rd ed. SAGE Publications Ltd., London, UK.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Hum. Genet. 7, 179–188.

Fortier, S., Basset, F.A., Mbourou, G.A., Favérial, J., Teasdale, N., 2005. Starting block

performance in sprinters: A statistical method for identifying discriminative parameters of the performance and an analysis of the effect of providing feedback over a 6-week period. J. Sport. Sci. Med. 4, 134–143.

Fox, D., Malley, E.O., Blake, C., 2014. Normative data for the Functional Movement Screen in male Gaelic field sports. Phys. Ther. Sport 15, 194–199. doi:10.1016/j.ptsp.2013.11.004

Frost, D., Andersen, J., Lam, T., Finlay, T., Darby, K., McGill, S., 2012. The relationship between general measures of fitness, passive range of motion and whole-body movement quality. Ergonomics 139, 1–16. doi:10.1080/00140139.2011.620177

Frost, D.M., Beach, T.A., Callaghan, J.P., McGill, S.M., 2015a. FMS^{TM} scores change with performers' knowledge of the grading criteria - Are general whole-body movement screens capturing "dysfunction"? J. Strength Cond. Res. 29, 3037–3044. doi:10.1519/JSC.0b013e3182a95343

Frost, D.M., Beach, T.A.C., Campbell, T.L., Callaghan, J.P., Mcgill, S.M., 2015b. An appraisal of the Functional Movement Screen^{TM} grading criteria - Is the composite score sensitive to risky movement behavior? Phys. Ther. Sport 16, 324–330. doi:10.1016/j.ptsp.2015.02.001

Garrison, M., Westrick, R., Johnson, M.R., Benenson, J., 2015. Association between the Functional Movement Screen and injury development in college athletes. Int. J. Sports Phys. Ther. 10, 21–8.

Girard, C.I., Warren, C.E., Romanchuk, N.J., Bel, M.J. Del, Carsen, S., Chan, A.D.C., Benoit, D.L., 2020. Decision tree learning algorithm for classifying knee injury status using return-to-activity criteria. IEEE Eng. Med. Biol. Soc. 5494–5497.

Glaws, K.R., Juneau, C.M., Becker, L.C., Di Stasi, S.L., Hewett, T.E., 2014. Intra- and inter-rater reliability of the Selective Functional Movement Assessment (SFMA). Int. J. Sports Phys. Ther. 9, 195–207.

Gløersen, Ø., Myklebust, H., Hallén, J., Federolf, P., 2017. Technique analysis in elite athletes using principal component analysis. J. Sports Sci. 0, 1–9. doi:10.1080/02640414.2017.1298826

Gribble, P.A., Brigle, J., Piertrosimone, B.G., Pfile, K.R., Webster, K.A., 2013. Intrarater

reliability of the function movement screen. J. Strength Cond. Res. 27, 978–981.

Gu, X., Deligianni, F., Lo, B., Chen, W., Yang, G.Z., 2018. Markerless gait analysis based on a single RGB camera. 2018 IEEE 15th Int. Conf. Wearable Implant. Body Sens. Networks, BSN 2018 2018–January, 42–45. doi:10.1109/BSN.2018.8329654

Gulgin, H., Hoogenboom, B., 2014. The Functional Movement Screening (FMS)^{TM}: An inter-rater reliability study Between raters of varied experience. Int. J. Sports Phys. Ther. 9, 14–20.

Gunn, S.R., 1998. Support vector machines for classification and regression. ISIS technical report.

Guskiewicz, K.M., Mihalik, J.P., Shankar, V., Marshall, S.W., Crowell, D.H., Oliaro, S.M., Ciocca, M.F., Hooker, D.N., 2007. Measurement of head impactcs in collegiate football players: Relationship between head impact biomechanics and acute clinical outcome after concussion. Neurosurgery 61, 1244–1253. doi:10.1227/01.NEU.0000280146.37163.79

Hadjidj, A., Souil, M., Bouabdallah, A., Challal, Y., Owen, H., 2013. Wireless sensor networks for rehabilitation applications: Challenges and opportunities. J. Netw. Comput. Appl. 36, 1–15. doi:10.1016/j.jnca.2012.10.002

Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J.L., Hastie, T.J., Delp, S.L., 2018. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. J. Biomech. 81, 1–11. doi:10.1016/j.jbiomech.2018.09.009

Harris-Hayes, M., Van Dillen, L.R., 2009. The inter-tester reliability of physical therapists classifying low back pain problems based on the Movement System Impairment Classification System. PM R 1, 117–126. doi:10.1016/j.pmrj.2008.08.001

Hastie, T., Tibshirani, R., Friedman, J., 2008. The elements of statistical learning: Data mining, inference, and prediction, 2nd ed. Springer Series in Statistics.

Helsen, W.F., Starkes, J.L., Hodges, N.J., 1998. Team sports and the theory of deliberate practice. J. Sport Exerc. Psychol. 20, 12–34.

Hewett, T.E., Myer, G.D., Ford, K.R., Heidt, R.S., Colosimo, A.J., McLean, S.G., van den Bogert, A.J., Paterno, M. V, Succop, P., 2005. Biomechanical measures of neuromuscular control and valgus loading of the knee predict Anterior cruciate Ligament injury risk in female athletes: A prospective study. Am. J. Sports Med. 33, 492–501. doi:10.1177/0363546504269591

Hochreiter, S., 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. Int. J. Uncertainty, Fuzziness, Knowledge-Based Syst. 6, 107–116.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.

Hong, Y., Kim, I., Ahn, S.C., Kim, H., 2008. Activity recognition using wearable sensors for elder care, in: Second International Conference on Future Generation Communication and Networking. IEEE, pp. 302–305. doi:10.1109/FGCN.2008.165

Hua, J., Xiong, Z., Lowey, J., Suh, E., Dougherty, E.R., 2005. Optimal number of features as a function of sample size for various classification rules. Bioinformatics 21, 1509–1515. doi:10.1093/bioinformatics/bti171

Izenman, A.J., 2013. Modern multivariate statistical techniques: Regression, classification, and manifold learning, 2nd ed. Springer Science + Business Media, LLC, New York, New York.

Jain, A.K., Mao, J., Mohiuddin, K.M., 1996. Artificial neural networks: A tutorial. Computer (Long. Beach. Calif). 29, 31–44.

Jirak, D., Tietz, S., Ali, H., Wermter, S., 2020. Echo state networks and long short-term memory for continuous gesture recognition: A comparative study. Cognit. Comput. 1–13.

Johnston, W., O'Reilly, M., Dolan, K., Reid, N., Coughlan, G., Caulfield, B., 2016. Objective classification of dynamic balance using a single wearable sensor, in: 4th International Congress on Sports Sciences Research and Technology Support 2016. Porto, Portugaul, pp. 15–24.

Johnston, W., Reilly, M.O., Duignan, C., Liston, M., Mcloughlin, R., Coughlan, G.F., Caulfield, B., 2019. Association of dynamic balance with sports-related concussion: A prospective cohort study. Am. J. Sports Med. 47, 197–205. doi:10.1177/0363546518812820

Karuc, J., Mišigoj-Duraković, M., Šarlija, M., Marković, G., Hadžić, V., Trošt-Bobić, T., Sorić, M., 2021. Can injuries be predicted by Functional Movement Screen in adolescents? The application of machine learning. J. Strength Cond. Res. 35, 910–919.

Kerr, M.S., Frank, J.W., Shannon, H.S., Norman, R.W.K., Wells, R.P., Neumann, W.P., Bombardier, C., Group, O.U.B.P.S., 2001. Biomechanical and psychosocial risk factors for low back pain at work. Am. J. Public Health 91, 1069–1075.

Kiesel, K., Plisky, P.J., Voight, M.L., 2007. Can Serious Injury in Professional Football be Predicted by a Preseason Function Movement Screen? N. Am. J. Sports Phys. Ther. 2, 147–158. doi:10.1186/2052-1847-5-11

Kiesel, K.B., Butler, R.J., Plisky, P.J., 2014. Prediction of injury by limited and asymmetrical fundamental movement patterns in American football players. J. Sport Rehabil. 23, 88–94. doi:10.1123/JSR.2012-0130

Knapik, J.J., Cosio-Lima, L.M., Reynolds, K.L., Shumway, R.S., 2015. Efficacy of function movement screening for predicting injuries in coast guard cadets. J. Stength Cond. Res. 29, 1157–1162.

Kobsar, D., Osis, S.T., Hettinga, B.A., Ferber, R., 2014. Classification accuracy of a single tri-axial accelerometer for training background and experiece level in runners. J Biomech. 47, 2508-2511.

Kritz, M., 2012. Development, reliability and effectiveness of the Movement Competency Screen (MCS). AUT University.

Kritz, M., Cronin, J., Hume, P., 2009. The bodyweight squat: A movement screen for the squat pattern. Strength Cond. J. 31, 76–85. doi:10.1519/SSC.0b013e318195eb2f

Lee, M., Roan, M., Smith, B., Lockhart, T.E., 2009. Gait analysis to classify external load conditions using linear discriminant analysis. Hum. Mov. Sci. 28, 226–235. doi:10.1016/j.humov.2008.10.008

Leeder, J.E., Horsley, I.G., Herrington, L.C., 2013. The Inter-rater reliability of the Functional Movement Screen within an athletic population using untrated raters. J. Strength Cond. Res. 30, 2591–2599.

Lees, A., Vanrenterghem, J., De Clercq, D., 2004. Understanding how an arm swing enhances performance in the vertical jump 37, 1929–1940. doi:10.1016/j.jbiomech.2004.02.021

Lelas, J.L., Merriman, G.J., Riley, P.O., Kerrigan, D.C., 2003. Predicting peak kinematic and kinetic parameters from gait speed. Gait Posture 17, 106–112. doi:10.1016/S0966-6362(02)00060-7

Lim, T., Loh, W., 2000. A comparison of prediction accuracy , complexity, and training time of

thirty-three old and new classification algorithms. Mach. Learn. 40, 203–228.

Liu, Y., Shih, S., Tian, S., Zhong, Y., Li, L., 2009. Lower extremity joint torque predicted by using artificial neural network during vertical jump. J. Biomech. 42, 906–911. doi:10.1016/j.jbiomech.2009.01.033

Loper, M., Mahmood, N., Black, M.J., 2014. MoSh : Motion and shape capture from sparse markers. ACM Trans. Graph. 33, 1–13.

Loper, M., Mahmood, N., Romero, D., Pons-Moll, G., Black, M.J., 2015. SMPL: A skinned multi-person linear model. ACM Trans. Graph. 34, 1–16.

Lukoševicius, M., Jaeger, H., 2009. Reservoir computing approaches to recurrent neural network training. Comput. Sci. Rev. 3, 127–149. doi:10.1016/j.cosrev.2009.03.005

Lumley, T., McNamara, T.F., 1993. Rater characteristics and rater bias: Implications for training. Lang. Test. 12, 54–71.

Madgwick, S.O.H., Harrison, A.J.L., Vaidyanathan, R., 2011. Estimation of IMU and MARG orientation using a gradient descent algorithm, in: IEEE International Conference on Rehabilitation Robotics (ICORR). Zurich, Switzerland, pp. 1–7.

Malus, J., Skypala, J., Silvernail, J.F., Uchytil, J., Hamill, J., Barot, T., Jandacka, D., 2021. Marker placement reliability and objectivity for biomechanical cohort study: Healthy aging in industrial environment (HAIE - Program 4). Sensors 21, 1830.

Mammone, A., Turchi, M., Cristianini, N., 2009. Support vector machines 1, 283–289. doi:10.1002/wics.49

Marey, E., 1874. Animal mechanism: A treatise on terrestrial and aerial locomtion. Henry S. King & Co., London.

Maroof, D.A., 2012. Statistical methods in neuropsychology: Common procedures made comprehensible. Springer Science + Business Media, LLC.

Marras, W.S., Lavender, S.A., Leurgans, S.E., Fathallah, F.A., Ferguson, S.A., Allread, W.G., Rajulu, S.L., 1995. Biomechanical risk factors for occupationally related low back disorders. Ergonomics 38, 377–410. doi:10.1080/00140139508925111

Marras, W.S., Lavender, S.A., Leurgans, S.E., Rajulu, S.L., Allread, W.G., Fathallah, F.A., Ferguson, S.A., 1993. The role of dynamic three-dimensional trunk motion in occupationally-related low back disorders: The effects of workplace factors, trunk position, and trunk motion characteristics on risk of injury. Spine (Phila. Pa. 1976). doi:10.1097/00007632-199304000-00015

Masoud, H.I., Zerehsaz, Y., Jin, J., 2017. Analysis of human motion variation patterns using UMPCA. Appl. Ergon. 59, 401–409. doi:10.1016/j.apergo.2016.09.016

Mavor, M.P., Ross, G.B., Clouthier, A.L., Karakolis, T., Graham, R.B., 2020. Validation of an IMU suit for military-based tasks. Sensors 20, 4280.

McCall, A., Carling, C., Nedelec, M., Davison, M., Le Gall, F., Berthoin, S., Dupont, G., 2014. Risk factors, testing and preventative strategies for non-contact injuries in professional football: Current perceptions and practices of 44 teams from various premier leagues. Br. J. Sports Med. 48, 1352–7. doi:10.1136/bjsports-2014-093439

McCulloch, W.S., Pitts, W.H., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5, 115–133.

McCunn, R., aus der Fünten, K., Fullagar, H.H.K., McKeown, I., Meyer, T., 2016. Reliability and association with injury of movement screens: A critical review. Sport. Med. 46, 763–781. doi:10.1007/s40279-015-0453-1

McCunn, R., Fünten, K. Der, Govus, A., Julian, R., Schimpchen, J., Meyer, T., 2017. The intra- and inter-rater reliability of the Soccer Injury Movement Screen (SIMS). Int. J. Sport Phys. Ther. 12, 53–66.

Mcgill, S., Frost, D., Lam, T., Finlay, T., Darby, K., Mcgill, S., Frost, D., Lam, T., Finlay, T., Darby, K., 2015. Can fitness and movement quality prevent back injury in elite task force police officers? A 5-year longitudinal study. Ergonomics 58, 1682–1689. doi:10.1080/00140139.2015.1035760

McGill, S.M., Andersen, J.T., Horne, A.D., 2012. Predicting Performance and Injury Resilience From Movement Quality and Fitness Scores in a Basketball Team Over 2 Years. J. strength Cond. Res. 26, 1731–1739.

Mcginnis, R.S., Cain, S.M., Davidson, S.P., Vitali, R. V, Mclean, S.G., Perkins, N.C., 2014. Validation of complementary filter based IMU data fusion for tracking torso angle and rifle orientation, in: ASME 2014 International Mechanical Engineering Congress and Exposition IMECE. Montreal, Quebec. doi:10.1115/IMECE201436909

McHugh, M.L., 2012. Interrater reliability: The kappa statistic. Biochem. Medica 22, 276–282.

McPherson, A.L., Dowling, B., Tubbs, T.G., Paci, J.M., 2016. Sagittal plane kinematic differences between dominant and non-dominant legs in unilateral and bilateral jump landings. Phys. Ther. Sport 22, 54–60.

Menolotto, M., Komaris, D., Tedesco, S., Flynn, B.O., Walsh, M., 2020. Motion capture technology in industrial applications: A systematic review. Sensors 20, 5687.

Merriaux, P., Dupuis, Y., Boutteau, R., Vasseur, P., Savatier, X., 2017. A study of Vicon system positioning performance. Sensors 17, 1591–1608. doi:10.3390/s17071591

Meskers, C.G.M., Van Der Helm, F.C.T., Rozendaal, L.A., Rozing, P.M., 1997. In vivo estimation of the glenohumeral joint rotation center from scapular bony landmarks by linear regression. J. Biomech. 31, 93–96. doi:10.1016/S0021-9290(97)00101-2

Milner, C.E., Ferber, R., Pollard, C.D., Hamill, J., Davis, I.S., 2006. Biomechanical factors associated with tibial stress fracture in female runners. Med. Sci. Sports Exerc. 38, 323–328. doi:10.1249/01.mss.0000183477.75808.92

Minick, K.I., Kiesel, K.B., Burton, L., Taylor, A., Plisky, P., Butler, R.J., 2010. Interrater Reliability of the Functional Movement Screen. Strength Cond. 24, 479–486.

Mischiati, C.R., Comerford, M., Gosford, E., Swart, J., Ewings, S., Botha, N., Stokes, M., Mottram, S.L., 2015. Intra and inter-rater reliability of screening for movement impairments : movement control tests from The Foundation Matrix. J. Sport. Sci. Med. 14, 427–440.

Mostaert, M., Deconinck, F., Pion, J., Lenoir, M., 2016. Anthropometry, physical fitness and coordination of young figure skaters of different levels. Int. J. Sports Med. 37, 531–538. doi:10.1055/s-0042-100280

Mündermann, L., Corazza, S., Andriacchi, T.P., 2006. The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. J.

Neuroeng. Rehabil. 3, 1–11. doi:10.1186/1743-0003-3-6

Muybridge, E., 1887. Animal locomotion. J.B. Lippincott Company, Philadelphia.

Myles, A.J., Feudale, R.N., Liu, Y., Woody, N.A., Brown, S.D., 2004. An introduction to decision tree modeling. J. Chemom. 18, 275–285. doi:10.1002/cem.873

Noehren, B., Manal, K., Davis, I., 2010. Improving between-day kinematic reliability using a marker placement device. J. Orthop. Res. 28, 1405–1410. doi:10.1002/jor.21172

Norouzi, M., Collins, M.D., Johnson, M., Fleet, D.J., Kohli, P., 2015. Efficient non-greedy optimization of decision trees. Adv. Neural Inf. Process. Syst. 1729–1737.

Onate, J.A., Dewey, T., Kollock, R.O., Thomas, K.S., Van Lunen, B.L., DeMaio, M., Ringleb, S.I., 2012. Real-time intersession and interrater reliability of the Functional Movement Screen. J. Strength Cond. Res. 26, 408–415. doi:10.1519/JSC.0b013e318220e6fa

Ow, P.S., Morton, T.E., 1988. Filtered beam search in scheduling. Int. J. Prod. Res. 26, 35–62. doi:10.1080/00207548808947840

Owings, T.M., Grabiner, M.D., 2004. Step width variability, but not step length variability or step time variability, discriminates gait of healthy young and older adults during treadmill locomotion. J. Biomech. 37, 935–938. doi:10.1016/j.jbiomech.2003.11.012

Padua, D.A., Marshall, S.W., Boling, M.C., Thigpen, C.A., Garrett, W.E., Beutler, A.I., 2009. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. Am. J. Sports Med. 37, 1996–2002. doi:10.1177/0363546509343200

Panza, G.A., 2018. Weight bias among exercise and nutrition professionals: A systematic review. Obeisity Rev. 19, 1492–1503. doi:10.1111/obr.12743

Patel, S., Park, H., Bonato, P., Chan, L., Rodgers, M., 2012. A review of wearable sensors and systems with application in rehabilitation. J. Neuroeng. Rehabil. 9, 1–17.

Pedregosa, F., Weiss, R., Brucher, M., 2011. Scikit-learn : Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pion, J.A., Fransen, J., Deprez, D.N., Segers, V.I., Vaeyens, R., Philippaerts, R.M., Lenoir, M.,

2015. Stature and jumping height are required in female volleyball, but motor coordination is a key factor for future elite success. J. Strength Cond. Res. 29, 1480–1485.

Pion, J., Fransen, J., Lenoir, M., Segers, V., 2014. The value of non-sport-specific characteristics for talent orientation in young male judo, karate and taekwondo athletes. Arch. Budo 10, 147–154.

Rekhi, N.S., Arora, A.S., Singh, S., Singh, D., 2009. Multi-Class SVM Classification of Surface EMG Signal for Upper Limb Function, in: Bioinformatics and Biomedical Engineering, 2009.ICBBE 2009. 3rd Inernational Conference. IEEE, pp. 1–4. doi:10.1109/ICBBE.2009.5163093

Richards, J.G., 1999. The measurement of human motion: A comparison of commercially available systems. Hum. Mov. Sci. 18, 589–602.

Roetenberg, D., Luinge, H., Slycke, P., 2013. Xsens MVN: Full 6DOF human motion tracking using miniature inertial sensors. Enschede, The Netherlands.

Ross, G.B., Dowling, B., Troje, N.F., Fischer, S.L., Graham, R.B., 2018. Objectively differentiating movement patterns between elite and novice athletes. Med. Sci. Sport. Exerc. 50, 1457–1464. doi:10.1249/MSS.0000000000001571

Ross, G.B., Dowling, B., Troje, N.F., Fischer, S.L., Graham, R.B., Graham, R.B., 2020. Classifying elite From novice athletes using simulated wearable sensor data. Front. Bioeng. Biotechnol. 8, 1–10. doi:10.3389/fbioe.2020.00814

Saeys, Y., Abeel, T., Van de Peer, Y. Van, 2008. Robust feature selection using ensemble feature selection techniques, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidleberg, pp. 313–325.

Schilaty, N.D., Bates, N.A., Kruisselbrink, S., Krych, A.J., Hewett, T.E., 2020. Linear discriminant analysis successfully predicts knee injury outcome from biomechanical variables. Am. J. Sports Med. 48, 2447–2455. doi:10.1177/0363546520939946

Schöllhorn, W.I., Jäger, J.M., Janssen, D., 2008. Artificial neural network models of sports motions, in: Handbook of Biomechanics and Human Movement Science. pp. 50–64.

Sgro, F., Mango, P., Pignato, S., Schembri, R., Licari, D., Lipoma, M., 2017. Assessing standing

long jump developmental levels using an inertial measurement unit. Percept. Mot. Skills 124, 15–24.

Sgro, F., Mango, P., Pignato, S., Schembri, R., Licari, D., Lipoma, M., 2017. Assessing standing long jump developmental levels using an inertial measurement unit. Percept. Mot. Skills 124, 21–38. doi:10.1177/0031512516682649

Shultz, R., Anderson, S.C., Matheson, G.O., Marcello, B., Besier, T., 2013. Test-retest and interrater reliability of the Functional Movement Screen. J. Athl. Train. 48, 331–336. doi:10.4085/1062-6050-48.2.11

Smith, C.A., Chimera, N.J., Wright, N.J., Warren, M., 2013. Interrater and intrarater reliability of the Functional Movement Screen. J. Strength Cond. Res. 27, 982–987.

Smith, R.M., Spinks, W.L., 1995. Discriminant analysis of biomechanical differences between novice , good and elite rowers. J. Sport Sci. 13, 377–385. doi:10.1080/02640419508732253

Song, Y., Lu, Y., 2015. Decision tree methods: Applications for classification and prediction. Biostat. Psychiatry 27, 130–135.

Stagni, R., Fantozzi, S., Cappello, A., Leardini, A., 2005. Quantification of soft tissue artefact in motion analysis by combining 3D fluoroscopy and stereophotogrammetry: A study on two subjects. Clin. Biomech. 20, 320–329. doi:10.1016/j.clinbiomech.2004.11.012

Subasi, A., 2013. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. Comput. Biol. Med. 43, 576–586. doi:10.1016/j.compbiomed.2013.01.020

Tarara, D.T., Hegedus, E.J., 2014. Reliability of select physical performance measures in physically active college-aged students. Int. J. Sports Phys. Ther. 9, 874–887.

Thaler, A., Bieg, A., Black, M.J., Troje, N.F., 2020. Attractiveness and confidence in walking style of male and female virtual characters, in: IEEE Conference on Virtual Reality and 3D User Interfaces. pp. 678–679. doi:10.1109/VRW50115.2020.00190

Troje, N.F., 2008. Retrieving information from human movement patterns, in: Thomas, F., Zacks, J. (Eds.), Understanding Events: From Perception to Action. Oxford University Press, Oxford, pp. 308–335.

Troje, N.F., 2002a. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. J. Vis. 2, 371–387. doi:10.1167/2.5.2

Troje, N.F., 2002b. The little difference: Fourier based synthesis of gender-specific biological motion, in: Würtz, R., Lappe, M. (Eds.), Dynamic Perception. Aka Press, Berlin, pp. 115–120.

Tsushima, H., Morris, M.E., McGinley, J., 2003. Test-retest reliability and inter-tester reliability of kinematic data from a three-dimensional gait analysis system. J. Japanese Phys. Ther. Assoc. 6, 9–17.

Vandorpe, B., Vandendriessche, J.B., Vaeyens, R., Pion, J., Lefevre, J., Philippaerts, R.M., Lenoir, M., 2012. The value of a non-sport-specific motor test battery in predicting performance in young female gymnasts. J. Sport Sci. 30, 497–505. doi:10.1080/02640414.2012.654399

Vapnik, V.N., 1995. The Sature of Statistical Learning Theory, 1st ed. Springer Science + Business Media, LLC, New York, New York.

Warren, M., Smith, C.A., Chimera, N.J., 2015. Association of the Functional Movement Screen with injuries in Division I athletes. J. Sport Rehabil. 24, 163–170. doi:10.1123/jsr.2013-0141

Weng, X., Shen, J., 2008. Knowledge-based systems classification of multivariate time series using two-dimensional singular value decomposition. Knowledge-Based Syst. 21, 535–539. doi:10.1016/j.knosys.2008.03.014

Whiteside, D., Deneweth, J.M., Pohorence, M.A., Sandoval, B., Russell, J.R., McLean, S.G., Zernicke, R.F., Goulet, G.C., 2014. Grading the Function Movement Screen: A comparison of manual (real-time) and objective methods. J. Strength Cond. Res. 30, 924–933.

Wittmann, F., Lambercy, O., Gassert, R., 2019. Magnetometer-based drift correction during rest in IMU arm motion tracking. Sensors 19, 1312–1330. doi:10.3390/s19061312

Wrigley, A.T., Albert, W.J., Deluzio, K.J., Stevenson, J.M., 2006. Principal component analysis of lifting waveforms. Clin. Biomech. 21, 567–578. doi:10.1016/j.clinbiomech.2006.01.004

Wu, J., Wang, J., 2008. PCA-based SVM for automatic recognition of gait patterns. J. Appl. Biomech. 24, 83–87. doi:10.1123/jab.24.1.83

Xanthopoulos, P., Pardalos, P.M., Trafalis, T., 2012. Robust Data Mining. Springer Science + Business Media.

Xanthopoulos, P., Pardalos, P.M., Trafalis, T.B., 2013. Linear Discriminant Analysis, in: Robust Data Mining. Springer New York, New York, NY, pp. 27–33. doi:10.1007/978-1-4419-9878-1_4

Yang, Q., Wu, X., 2006. 10 challenging problems in data mining research. Int. J. Inf. Technol. Decis. Mak. 5, 597–604.

Young, A.D., 2010. From posture to motion: The challenge for real time wireless inertial motion capture, in: Proceedings of the Fifth International Conference on Body Area Networks. Corfu, Greece, pp. 131–137.

Young, C., Reinkensmeyer, D.J., 2014. Judging complex movement performances for excellence: A principal components analysis-based technique applied to competitive diving. Hum. Mov. Sci. 36, 107–122. doi:10.1016/j.humov.2014.05.009

Zaggelidis, G., Lazaridis, S., 2013. Muscle activation profiles of lower extremities in different throwing techniques and in jumping performance in elite and novice Greek judo athletes. J. Hum. Kinet. 37, 63–70. doi:10.2478/hukin-2013-0026

Zago, M., Sforza, C., Dolci, C., Tarabini, M., Galli, M., 2019. Use of machine learning and wearable sensors to predict energetics and kinematics of cutting maneuvers. Sensors 19, 3094–3105.

Zhou, H., Hu, H., 2008. Human motion tracking for rehabilitation — A survey. Biomed. Signal Process. Control 3, 1–18. doi:10.1016/j.bspc.2007.09.001

Appendix A: Ethics Certificates

QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD (HSREB)

HSREB Initial Ethics Clearance

December 10, 2015

Miss Gwyneth Ross
School of Kinesiology and Health Studies
Queen's University

ROMEO/TRAQ: #6017208
Department Code: PHE-157-15
Study Title: Quantifying whole-body movement dynamics using principal component analysis
Co-Investigators: Dr. S. Fischer, Dr. R. Graham
Review Type: Delegated
Date Ethics Clearance Issued: December 10, 2015
Ethics Clearance Expiry Date: December 10, 2016

Dear Miss Ross,

The Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board (HSREB) has reviewed the application and granted ethics clearance for the documents listed below. Ethics clearance is granted until the expiration date noted above.

- Protocol
- Nondisclosure Agreement
- Consent Waiver Form

Documents Acknowledged:

- CORE Certificate – G. Ross
- CORE Certificate – S. Fischer
- CORE Certificate – R. Graham

Amendments: No deviations from, or changes to the protocol should be initiated without prior written clearance of an appropriate amendment from the HSREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

Renewals: Prior to the expiration of your ethics clearance you will be reminded to submit your renewal report through ROMEO. Any lapses in ethical clearance will be documented on the renewal form.

Completion/Termination: The HSREB must be notified of the completion or termination of this study through the completion of a renewal report in ROMEO.

Reporting of Serious Adverse Events: Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 15 days after becoming aware of the information.

Reporting of Complaints: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the contact information for the Research Ethics Board.

Investigators please note that if your trial is registered by the sponsor, you must take responsibility to ensure that the registration information is accurate and complete.

Yours sincerely,

Chair, Health Sciences Research Ethics Board

The HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations, Canadian General Standards Board, and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is qualified through the CTO REB Qualification Program and is registered with the U.S. Department of Health and Human Services (DHHS) Office for Human Research Protection (OHRP). Federalwide Assurance Number: FWA#:00004184, IRB#:00001173

HSREB members involved in the research project do not participate in the review, discussion or decision.

16/12/2020

Université d'Ottawa
Bureau d'éthique et d'intégrité de la recherche

University of Ottawa
Office of Research Ethics and Integrity

CERTIFICAT D'APPROBATION ÉTHIQUE | CERTIFICATE OF ETHICS APPROVAL

Numéro du dossier / Ethics File Number	H-10-19-4983
Titre du projet / Project Title	The comparison of movement competency scores between an objective framework and expert assessors
Type de projet / Project Type	Thèse de doctorat / Doctoral thesis
Statut du projet / Project Status	Renouvelé / Renewed
Date d'approbation (jj/mm/aaaa) / Approval Date (dd/mm/yyyy)	20/12/2019
Date d'expiration (jj/mm/aaaa) / Expiry Date (dd/mm/yyyy)	19/12/2021

Équipe de recherche / Research Team

Chercheur / Researcher	Affiliation	Role
Gwyneth ROSS	École des sciences de l'activité physique / School of Human Kinetics	Chercheur Principal / Principal Investigator
Ryan GRAHAM	École des sciences de l'activité physique / School of Human Kinetics	Superviseur / Supervisor

Conditions spéciales ou commentaires / Special conditions or comments

550, rue Cumberland, pièce 154 550 Cumberland Street, Room 154
Ottawa (Ontario) K1N 6N5 Canada Ottawa, Ontario K1N 6N5 Canada

613-562-5387 • 613-562-5338 • ethique@uOttawa.ca / ethics@uOttawa.ca
www.recherche.uottawa.ca/deontologie | www.recherche.uottawa.ca/ethics

02/09/2020

Université d'Ottawa
Bureau d'éthique et d'intégrité de la recherche

University of Ottawa
Office of Research Ethics and Integrity

CERTIFICAT D'APPROBATION ÉTHIQUE | CERTIFICATE OF ETHICS APPROVAL

Numéro du dossier / Ethics File Number	H-08-18-1085
Titre du projet / Project Title	Developing a framework to quantify whole-body movement dynamics using principal component analysis and machine learning
Type de projet / Project Type	Recherche de professeur / Professor's research project
Statut du projet / Project Status	Renouvelé / Renewed
Date d'approbation (jj/mm/aaaa) / Approval Date (dd/mm/yyyy)	20/09/2018
Date d'expiration (jj/mm/aaaa) / Expiry Date (dd/mm/yyyy)	19/09/2021

Équipe de recherche / Research Team

Chercheur / Researcher	Affiliation	Role
Gwyneth ROSS	École des sciences de l'activité physique / School of Human Kinetics	Chercheur Principal / Principal Investigator
Ryan GRAHAM	École des sciences de l'activité physique / School of Human Kinetics	Co-chercheur principal / Co-principal investigator

Conditions spéciales ou commentaires / Special conditions or comments

550, rue Cumberland, pièce 154 Ottawa (Ontario) K1N 6N5 Canada
550 Cumberland Street, Room 154 Ottawa, Ontario K1N 6N5 Canada

613-562-5387 • 613-562-5338 • ethique@uOttawa.ca / ethics@uOttawa.ca
www.recherche.uottawa.ca/deontologie | www.recherche.uottawa.ca/ethics

Appendix B: Study 2 Supplementary Material

Table B1. The start and end point criteria used when trimming trial lengths for each movement.

	Start	End
Bird-Dog	The initiation of movement of either the ankle or the wrist	Once the velocity of both the wrist and the ankle reached zero and positional data were close to that of the starting position
T-Balance	The initiation of movement of either the ankle or the wrist	Once the velocity of both the wrist and the ankle reached zero and positional data were close to that of the starting position
Drop Jump	The initiation of movement of either the ankle or the wrist	Once the ankle and wrist reached a velocity of zero and reached peak jump height after the second jump
L-Hop	Once the ankle and wrist reached a velocity of zero and reached peak jump height after the initial horizontal hop	Once the ankle and wrist reached a velocity of zero and reached peak jump height after executing the lateral hop
Step-Down	The initiation of movement of either the ankle or the wrist	Once the velocity of both the wrist and the ankle reached zero and positional data were close to that of the starting position
Hop-Down	The initiation of movement of either the ankle or the wrist	Once the ankle and wrist reached a velocity of zero and reached peak jump height after the second hop
Lunge	The initiation of movement of either the knee or the ankle.	Once the velocity of both the knee and the ankle reached zero and positional data were close to that of the starting position

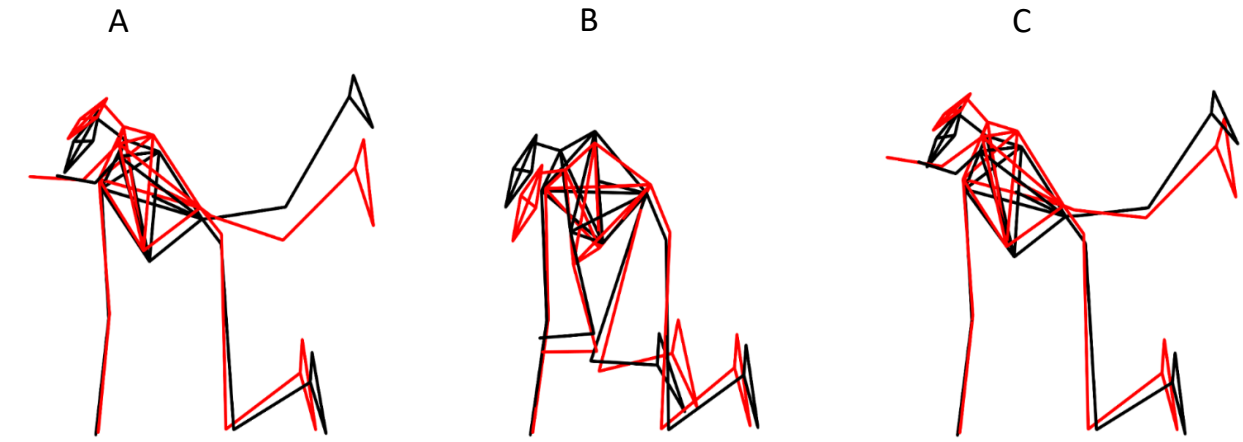


Figure B1. Bird Dog. During bird-dog, elite athletes (red) keep a more neutral spine compared to the novice athletes (black) during the flexion portion of the task (B). The novice athletes round their back in order to complete the task, whereas the elite athletes keep their backs flat during the entire movement. In addition, the novice athletes touch their elbow to their thigh closer to the knee joint, whereas the novice athletes touch their elbow more proximally on their thigh.

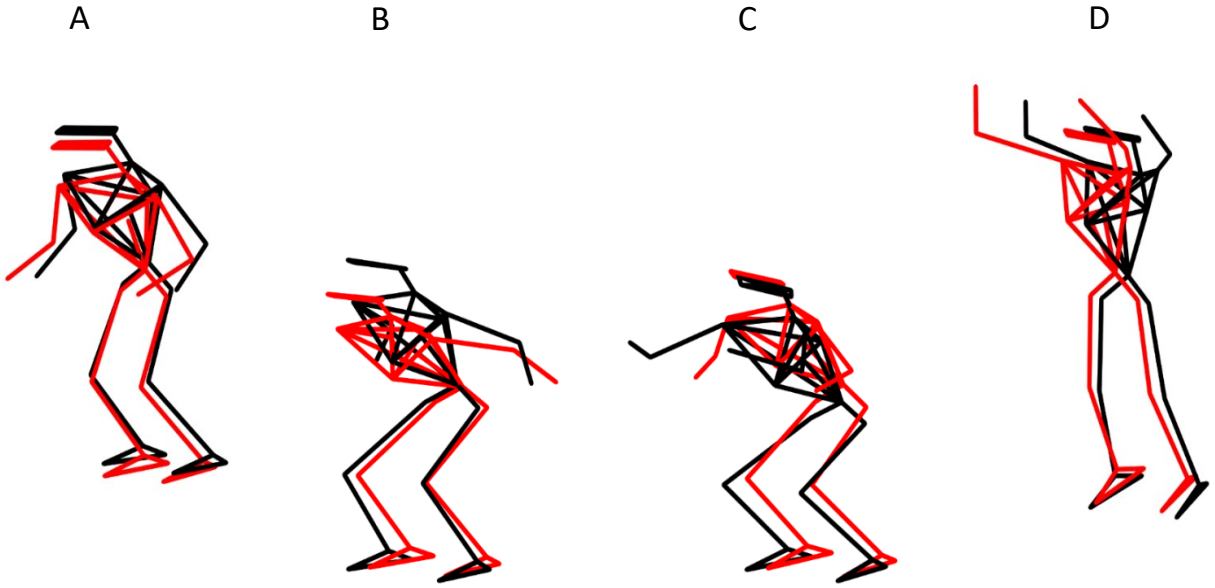


Figure B2. Drop-Jump. During the drop-jump, the timing of the arm swing during the landing phase (B) and transition phase (C) of the jump. Upon landing, the elite athletes (red) have greater shoulder extension, extending their arms further behind them than the novice (black) athletes. However, during the transition phase (changing from flexing the knee to extending the knee), the novice athletes have almost completed their arm swing, whereas the elite athletes are just transitioning from shoulder extension to flexion.

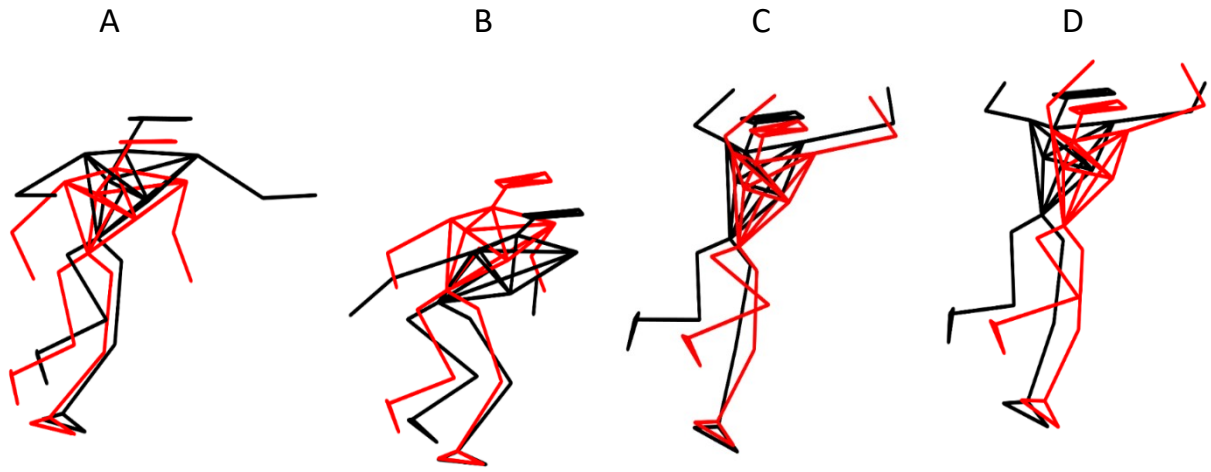


Figure B3. Hop-Down. During the hop-down, elite athletes (red) have less trunk and left knee flexion than novice athletes (black) during the landing phase of the initial hop down (B). During the take-off and flight phase (C & D, respectively), elite athletes have greater right hip flexion suggesting that the elite athletes use their non-stance leg (in this case the right leg) to increase upward momentum.

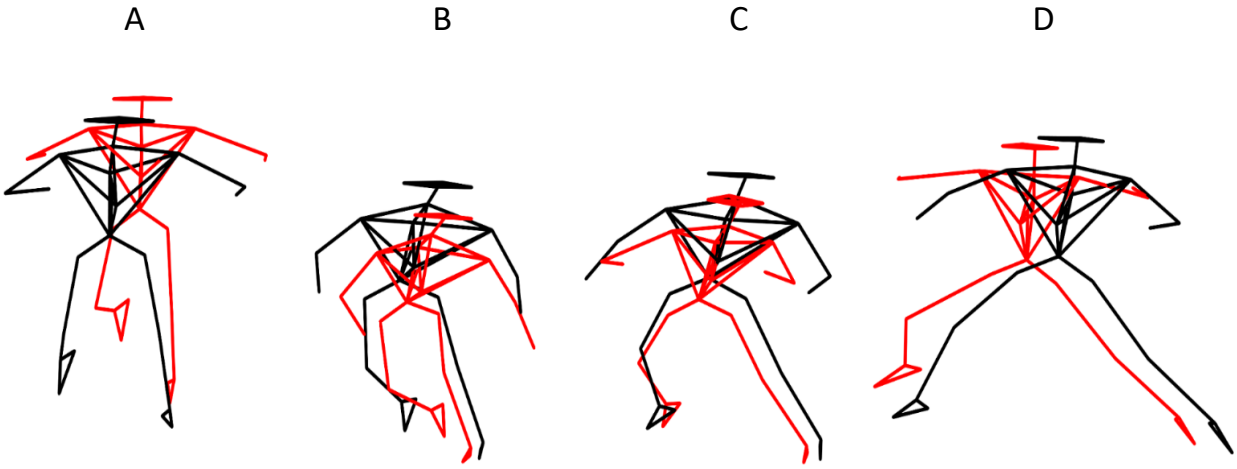


Figure B4. L-Hop. During the L-hop, elite athletes (red) have greater trunk flexion, greater left knee flexion and greater right hip adduction during the landing phase of the task (B) compared to novice athletes (black). In addition, elite athletes have greater knee flexion in the right knee when performing the lateral cut (C,D) than the novice athletes. This suggests that the elite athletes use their non-landing leg (in this case the right leg) to increase their momentum in the lateral direction in order to get a larger horizontal distance.

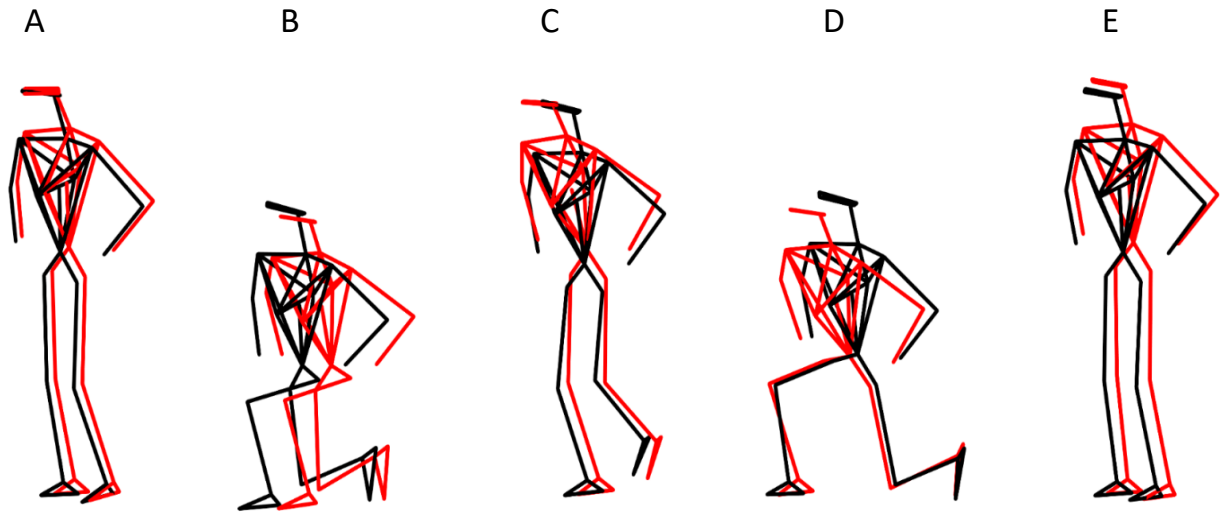


Figure B5. Lunges. During the lunge, elite athletes (red) have greater trunk flexion and, right knee flexion and right ankle dorsiflexion during the backwards lunge (D) compared to the novice athletes (black).

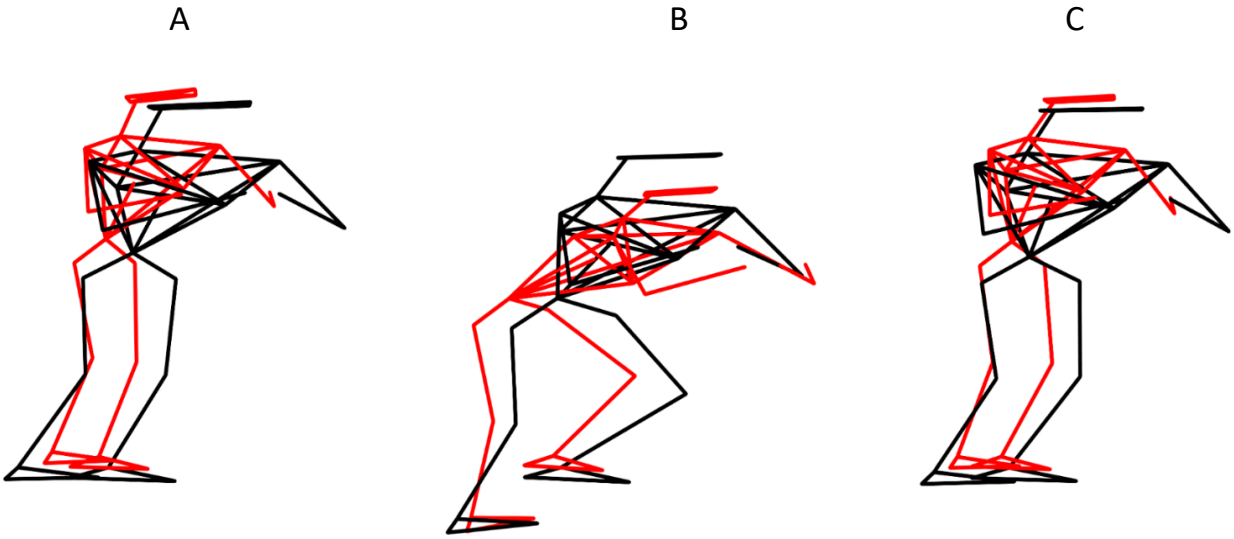


Figure B6. Step-Down. During the step-down, the elite athletes (red) have greater trunk flexion, greater left knee flexion, and less left ankle dorsiflexion while performing the step down (B) compared to the novice athletes (black). This suggests that the elite athletes shift their center of mass posteriorly in order to perform a deeper single-leg squat to complete the task compared to novice athletes.

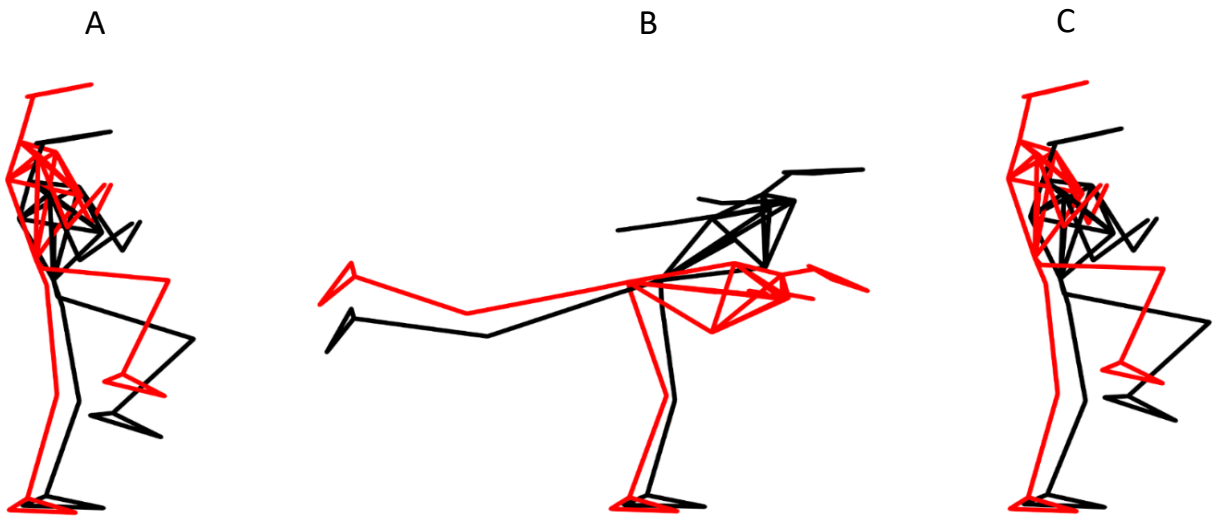


Figure B7. T-Balance. During the T-balance, the elite athletes (red) have greater right hip flexion and greater left knee extension during the initial and final pose (A & C, respectively) compared to the novice athletes (black). During the T-pose (B), elite athletes have greater left knee flexion and have greater forward rotation bringing their trunk more parallel to the ground than the novice athletes.

Appendix C: Study 4 Supplementary Material

Table C1. The number of athletes in the dataset and the average age, height and weight of the athletes broken down by competition level. Values are written as mean(std), where applicable.

		Elite	Novice	Elite & Novice
Male	n	306	174	480
	Age	22.5 (3.5)	16.8 (4.2)	20.05(4.7)
	Height	191.5 (19.8)	174.7 (13.0)	185.3 (19.4)
	Weight	95.7 (18.7)	69.7 (18.0)	86.1 (22.3)
Female	n	12	50	62
	Age	23.3 (4.0)	18.0 (5.6)	19.00 (5.7)
	Height	170.4 (7.2)	167.0 (8.7)	168.00 (8.9)
	Weight	64.3 (8.8)	60.0 (13.7)	61.1 (13.3)
Male & Female	n	318	224	542
	Age	22.6 (3.6)	17.1 (4.6)	20.2 (4.7)
	Height	190.7 (19.9)	172.9 (12.6)	183.3 (19.3)
	Weight	94.5 (19.4)	67.5 (17.6)	83.1 (22.9)

Table C2. The number of elite athletes in the dataset and the average age, height, and weight of the athletes broken down by sport played. Values are written as mean(std), where applicable.

		Basketball	Baseball	Football	Soccer	All
Elites	n	128	96	37	33	289
	Age	22.5 (3.6)	21.7 (2.9)	22.5 (1.5)	23.1 (4.2)	22.3 (3.3)
	Height	198.3 (14.4)	186.6 (6.4)	190.3 (7.8)	180.6 (7.2)	191.4 (12.5)
	Weight	97.9 (16.1)	93.4 (10.3)	116.3 (26.3)	77.7 (8.4)	96.5 (18.3)

Table C3. Number of athletes included in the analysis for each movement task broken down my competition level and sex.

		Male		Female	
	n	*Elite*	*Novice*	*Elite*	*Novice*
Bird-Dog Left	380	242	83	12	43
Bird-Dog Right	387	244	88	11	44
Drop-Jump	275	168	64	7	36
Hop-Down Left	396	242	99	10	45
Hop-Down Right	396	242	97	11	46
L-Hop Left	266	159	67	6	34
L-Hop Right	267	160	67	6	34
Lunge Left	399	246	97	12	44
Lunge Right	401	248	97	12	44
Step-Down Left	399	246	98	12	43
Step-Down Right	399	247	96	11	45
T-Balance Left	392	244	92	11	45
T-Balance Right	395	244	94	12	45

Table C4. Number of athletes included in the analysis for each movement task broken down my sport played.

	n	*Baseball*	*Basketball*	*Football*	*Soccer*
Bird-Dog Left	232	62	110	31	29
Bird-Dog Right	233	65	108	29	31
Drop-Jump	167	58	69	27	13
Hop-Down Left	230	64	102	33	31
Hop-Down Right	232	64	104	33	31
L-Hop Left	158	63	61	21	13
L-Hop Right	159	63	63	20	13
Lunge Left	235	66	109	29	31
Lunge Right	237	66	111	29	31
Step-Down Left	234	63	108	32	31
Step-Down Right	235	63	109	32	31
T-Balance Left	233	62	108	32	31
T-Balance Right	233	63	109	30	31

Table C5. Hypertuned LSTM parameters and the ranges of values tested.

Hyperparameter	Range
# of Epochs	[50, 100, 150, 300, 500, 1000, 2000]
Batch Size	[32, 64, 128, 256]
Momentum	[1, 2, 3, 4, 5, 6,7, 8, 9, 10]
# of Cells	[32, 64, 128, 256]
# of Layers	[1, 2, 3, 4, 5, 6,7, 8, 9, 10]
# of Nodes	[32, 64, 128, 256]
Learning Rate	[0.000001 - 0.06]
Dropout Rate	[0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]

Table C6. Hypertuned reservoir computing parameters and the ranges of values tested: ρ is the spectral radius, α is the leak rate, κ denotes connectivity, and λ is the regularization coefficient.

Hyperparameter	Range
Reservoir Size	[150, 225, 300, 375, 450, 525]
α	[0.1, 0.3, 0.5, 0.7, 0.9]
λ	[1,3,6,10]
ρ	[0.45, 0.59, 0.65]
κ	[0.125, 0.25, 0.375, 0.5]
Dimensionality Reduction Method	[None, PCA, Tensor PCA]
Readout for Classification	[MLR, MLP, SVM]

Table C7. The mean classification rates and standard deviations across the 10-cross fold validations for each machine learning algorithm for each task for classifying level. Values are written as mean(std), where applicable.

	Classification Rates (%)									
	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
Bird-Dog Left	80.7 (4.2)	64.3 (13.1)	77.2 (3.7)	77.0 (4.9)	77.6 (4.2)	71.1 (2.8)	64.1 (5.9)	78.2 (3.6)	75.3 (3.8)	67.1
Bird-Dog Right	78.8 (3.7)	69.5 (13.3)	79.4 (3.7)	80.3 (2.4)	80.5 (3.1)	72.6 (3.2)	66.0 (4.9)	76.4 (5.7)	78.5 (4.2)	65.4
Drop-Jump	77.8 (5.1)	80.0 (6.3)	79.6 (5.1)	79.1 (5.0)	79.5 (6.4)	64.9 (4.5)	64.4 (4.3)	78 (7.4)	72.2 (5.5)	63.6
Hop-Down Left	73.5 (4.1)	70.9 (12.3)	77.5 (3.8)	77.6 (3.5)	77.5 (3.8)	67.4 (3.2)	64.3 (4.0)	75.6 (4.3)	73 (5.5)	63.8
Hop-Down Right	76.5 (3.8)	69.1 (6.8)	75.9 (3.4)	75.4 (3.7)	75 (3.2)	65.5 (4.7)	63.7 (4.5)	71.5 (3.9)	70.6 (4.2)	63.8
L-Hop Left	80.2 (6.0)	76.9 (7.9)	80.2 (4.3)	79.8 (4.3)	80.0 (4.7)	72.4 (5.8)	63.7 (6.8)	74.8 (7.2)	75.4 (3.4)	61.1
L-Hop Right	78.7 (3.6)	75.0 (7.7)	77.4 (4.3)	77.6 (5.3)	77.4 (5.1)	66.9 (4.5)	63.0 (6.4)	73.3 (6.2)	76.7 (4.5)	63.0
Lunge Left	77.5 (4.0)	78.1 (4.0)	78.3 (4.3)	78.3 (4.0)	77.6 (3.9)	69.3 (4.0)	63.9 (4.9)	76.5 (2.1)	76.3 (2.8)	65.0
Lunge Right	77.9 (4.7)	79 (3.8)	76.9 (3.9)	77 (3.8)	76.3 (3.1)	66.0 (4.3)	62.6 (4.6)	72.6 (3.9)	73.6 (4.1)	65.4
Step-Down Left	79.1 (5.1)	76.1 (5.4)	75.5 (3.7)	78.4 (3.4)	77.9 (4.4)	71.9 (4.3)	63.5 (5.3)	71.6 (4.3)	76.9 (4.9)	65.0
Step-Down Right	78.7 (5.3)	75.5 (5.6)	77.5 (4.2)	77.5 (3.9)	78.8 (4.9)	72.7 (4.2)	63.0 (5.1)	75.4 (5.3)	76.1 (4.8)	65.0
T-Balance Left	72.5 (4.3)	66.2 (13.6)	74.1 (5.7)	74.1 (4.5)	72.8 (4.3)	66.8 (4.3)	61.4 (6.6)	72.7 (4.6)	71.1 (3.5)	64.6
T-Balance Right	73.2 (3.7)	71.9 (6.5)	74.2 (3.8)	76.1 (3.5)	75.3 (3.2)	69.5 (2.8)	63.3 (4.6)	74.6 (5.2)	71.3 (2.5)	64.6

Table C8. The mean classification rates and standard deviations across the 10-cross fold validations for each machine learning algorithm for each task for classifying sport. Values are written as mean(std), where applicable.

	Classification Rates (%)									
	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
Bird-Dog Left	63.1 (5.0)	60.2 (9.0)	63.8 (7.4)	63.6 (3.8)	63.6 (4.8)	53.0 (9.2)	49.1 (6.8)	62.6 (4.4)	62.6 (4.6)	46.8
Bird-Dog Right	64.9 (2.5)	76.2 (5.7)	64.9 (6.2)	65.1 (7.8)	66.6 (6.4)	57.7 (7.1)	49.8 (7.0)	63.6 (3.8)	62.1 (4.5)	46.8
Drop-Jump	62.9 (5.4)	67.6 (6.5)	61.5 (8.6)	61.8 (7.7)	59.7 (9.7)	51.2 (8.7)	41.2 (4.6)	55.0 (10.3)	57.1 (5.6)	41.2
Hop-Down Left	61.7 (4.7)	60.7 (7.8)	60.2 (6.5)	58.7 (5.8)	58.0 (5.2)	46.5 (4.3)	43.9 (5.1)	58.5 (6.7)	52.6 (5.8)	43.5
Hop-Down Right	63.4 (7.5)	57.4 (6.3)	62.6 (6.0)	63.4 (7.5)	61.3 (8.1)	47.7 (7.4)	41.3 (7.8)	57.7 (7.5)	56.4 (4.8)	44.7
L-Hop Left	61.3 (7.7)	71.9 (6.3)	59.4 (10.2)	59.4 (10.7)	57.2 (10.8)	45.6 (8.7)	48.1 (8.1)	55.9 (9.4)	51.2 (6.9	40.6
L-Hop Right	59.7 (4.3)	61.6 (9.9)	58.4 (4.2)	58.8 (6.2)	54.4 (5.9)	45.9 (7.7)	48.1 (9.6)	53.1 (10.0)	55.6 (7.9)	40.6
Lunge Left	69.6 (6.6)	68.5 (6.2)	61.5 (4.3)	61.3 (3.7)	60.4 (4.2)	50.4 (7.2)	50.9 (6.3)	58.7 (3.5)	61.7 (5.7)	46.8
Lunge Right	72.9 (6.1)	72.5 (6.5)	66.0 (6.9)	65.8 (7.2)	65.6 (5.6)	58.1 (5.2)	47.5 (9.0)	64.4 (7.0)	63.8 (2.0	47.9
Step-Down Left	66.2 (6.1)	70.2 (3.5)	65.1 (6.7)	65.3 (6.1)	61.9 (6.5)	54.9 (6.0)	47.4 (7.2)	61.3 (9.2)	58.1 (6.9)	46.8
Step-Down Right	71.5 (3.6)	60.2 (6.2)	66.0 (5.0)	65.7 (7.9)	63.8 (7.5)	52.3 (8.2)	41.7 (7.8)	55.7 (8.2)	58.3 (4.4)	46.8
T-Balance Left	71.5 (5.5)	71.1 (5.7)	63.8 (5.8)	63.0 (5.2)	62.3 (5.0)	56.6 (6.9)	45.7 (7.0)	61.7 (5.4)	64.7 (4.4)	46.8
T-Balance Right	66.8 (4.9)	70.0 (6.6)	64.9 (5.0)	63.0 (2.0)	62.3 (7.4)	53.6 (6.2)	44.7 (8.5)	60.0 (5.4)	58.3 (6.3)	46.8

Table C9. The results of the Tukey post-hoc tests for all movement tasks for level. The orange cells denote left, whereas the blue cells denote right, except for the drop-jump that is performed unilaterally. Red font denotes significance.

Bird-Dog

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.01	1	1	1	0.262	<0.001	0.992	1	<0.001
LSTM	<0.001		0.005	0.001	0.001	0.963	0.924	0.152	0.016	0.815
LDA	0.934	<0.001		1	1	0.171	<0.001	0.972	1	<0.001
MLR	0.899	<0.001	1		1	0.071	<0.001	0.865	0.999	<0.001
SVM	0.969	<0.001	1	1		0.054	<0.001	0.815	0.998	<0.001
KNN	0.008	0.196	0.297	0.358	0.219		0.213	0.865	0.347	0.12
DT	<0.001	1	<0.001	<0.001	<0.001	0.156		0.002	<0.001	1
NB	0.992	<0.001	1	1	1	0.139	<0.001		0.998	0.001
RBF	0.493	0.001	0.999	1	0.994	0.802	0.001	0.977		<0.001
Naïve	<0.001	0.983	0.004	0.006	0.002	0.855	0.969	0.001	0.048	

Drop-Jump

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.995	0.999	1	1	<0.001	<0.001	1	0.354	<0.001
LSTM			1	1	1	<0.001	<0.001	0.998	0.044	<0.001
LDA				1	1	<0.001	<0.001	1	0.066	<0.001
MLR					1	<0.001	<0.001	1	0.118	<0.001
SVM						<0.001	<0.001	1	0.081	<0.001
KNN							1	<0.001	0.081	1
DT								<0.001	0.044	1
NB									0.31	<0.001
RBF										0.018
Naïve										

Hop-Down

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.005	1	1	0.998	<0.001	<0.001	0.19	0.061	<0.001
LSTM	0.983		0.015	0.035	0.061	0.631	0.12	0.955	0.998	0.12
LDA	0.805	0.16		1	1	<0.001	<0.001	0.361	0.141	<0.001
MLR	0.776	0.142	1		1	<0.001	<0.001	0.538	0.25	<0.001
SVM	0.805	0.16	1	1		<0.001	<0.001	0.676	0.361	<0.001
KNN	0.248	0.901	0.002	0.002	0.002		0.994	0.051	0.164	0.994
DT	0.007	0.16	<0.001	<0.001	<0.001	0.949		0.003	0.013	1
NB	0.996	0.607	0.999	0.998	0.999	0.027	<0.001		1	0.003
RBF	1	0.996	0.678	0.643	0.678	0.363	0.014	0.983		0.013
Naïve	0.004	0.098	<0.001	<0.001	<0.001	0.881	1	<0.001	0.007	

L-Hop

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.838	1	1	1	<0.001	<0.001	0.374	0.969	<0.001
LSTM	0.936		0.988	0.98	0.988	0.021	<0.001	0.999	0.999	<0.001
LDA	1	0.936		1	1	0.001	<0.001	0.748	1	<0.001
MLR	1	0.969	1		1	<0.001	<0.001	0.698	1	<0.001
SVM	1	0.954	1	1		0.001	<0.001	0.748	1	<0.001
KNN	0.06	0.724	0.06	0.089	0.073		0.795	0.144	0.002	0.795
DT	<0.001	<0.001	<0.001	<0.001	<0.001	0.02		0.001	<0.001	1
NB	0.469	0.998	0.469	0.573	0.521	0.993	0.001		0.907	0.001
RBF	0.625	1	0.625	0.724	0.675	0.969	<0.001	1		<0.001
Naïve	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.987	<0.001	<0.001	

Lunge

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		1	1	1	0.995	<0.001	<0.001	0.076	0.276	<0.001
LSTM	1		0.967	0.978	0.855	<0.001	<0.001	0.012	0.063	<0.001
LDA	1	1		1	1	<0.001	<0.001	0.276	0.644	<0.001
MLR	1	1	1		1	<0.001	<0.001	0.241	0.595	<0.001
SVM	1	1	1	1		<0.001	<0.001	0.496	0.855	<0.001
KNN	<0.001	<0.001	<0.001	<0.001	<0.001		0.595	0.009	0.001	1
DT	<0.001	<0.001	<0.001	<0.001	<0.001	0.045		<0.001	<0.001	0.819
NB	1	0.992	0.986	0.986	1	0.001	<0.001		1	0.03
RBF	0.999	0.978	0.967	0.967	0.998	0.002	<0.001	1		<0.001
Naïve	<0.001	<0.001	<0.001	<0.001	<0.001	0.234	1	<0.001	<0.001	

Step-Down

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.85	1	1	1	0.112	<0.001	0.82	0.955	<0.001
LSTM	0.873		0.993	0.993	0.85	0.94	<0.001	1	1	<0.001
LDA	0.697	1		1	1	0.386	<0.001	0.989	1	<0.001
MLR	1	0.977	0.899		1	0.386	<0.001	0.989	1	<0.001
SVM	1	0.996	0.968	1		0.112	<0.001	0.82	0.955	<0.001
KNN	0.012	0.479	0.697	0.039	0.078		<0.001	0.955	0.82	0.01
DT	<0.001	<0.001	<0.001	<0.001	<0.001	0.002		<0.001	<0.001	0.993
NB	0.008	0.395	0.611	0.027	0.056	1	0.003		1	<0.001
RBF	0.977	1	0.999	0.999	1	0.251	<0.001	0.193		<0.001
Naïve	<0.001	<0.001	<0.001	<0.001	<0.001	0.022	<0.001	0.033	<0.001	

T-Balance										
	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.999	1	0.816	0.966	0.54	<0.001	0.999	0.995	<0.001
LSTM	0.385		0.952	0.352	0.639	0.933	<0.001	0.884	1	0.003
LDA	1	0.126		0.985	1	0.204	<0.001	1	0.816	<0.001
MLR	1	0.126	1		1	0.011	<0.001	0.997	0.175	<0.001
SVM	1	0.329	1	1		0.041	<0.001	1	0.396	<0.001
KNN	0.538	1	0.211	0.211	0.475		0.022	0.126	0.991	0.149
DT	0.004	0.754	<0.001	<0.001	0.003	0.602		<0.001	0.001	0.999
NB	1	0.356	1	1	1	0.507	0.003		0.687	<0.001
RBF	1	0.726	0.986	0.986	1	0.854	0.019	1		0.009
Naïve	0.113	1	0.025	0.025	0.09	0.998	0.976	0.101	0.329	

Table C10. The results of the Tukey post-hoc test for all movement tasks for sport. The orange cells denote left, whereas the blue cells denote right, except for the drop-jump that is performed unilaterally. Red font denotes significance.

Bird-Dog										
	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.001	1	1	1	0.125	<0.001	1	0.983	<0.001
LSTM	0.984		0.001	0.001	0.009	<0.001	<0.001	<0.001	<0.001	<0.001
LDA	1	0.943		1	1	0.125	<0.001	1	0.983	<0.001
MLR	1	0.961	1		1	0.102	<0.001	1	0.972	<0.001
SVM	1	0.961	1	1		0.02	<0.001	0.972	0.744	<0.001
KNN	0.01	0.203	0.005	0.06	0.006		0.066	0.353	0.744	0.002
DT	<0.001	0.004	<0.001	<0.001	<0.001	0.92		<0.001	<0.001	0.972
NB	1	0.997	1	1	1	0.022	<0.001		1	<0.001
RBF	1	0.997	1	1	1	0.022	<0.001	1		<0.001
Naïve	<0.001	<0.001	<0.001	<0.001	<0.001	0.415	0.997	<0.001	<0.001	

Drop-Jump										
	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.91	1	1	0.992	0.018	<0.001	0.32	0.732	<0.001
LSTM			0.674	0.732	0.32	<0.001	<0.001	0.007	0.05	<0.001
LDA				1	1	0.064	<0.001	0.613	0.938	<0.001
MLR					1	0.05	<0.001	0.551	0.91	<0.001
SVM						0.227	<0.001	0.91	0.998	<0.001
KNN							0.081	0.975	0.732	0.081
DT								0.002	<0.001	1
NB									1	0.002
RBF										<0.001
Naïve										

Hop-Down

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.609	1	1	0.999	<0.001	<0.001	0.657	0.371	<0.001
LSTM	1		0.79	0.609	0.956	0.047	<0.001	1	1	0.002
LDA	1	1		1	1	<0.001	<0.001	0.828	0.56	<0.001
MLR	0.966	0.999	1		0.999	<0.001	<0.001	0.657	0.371	<0.001
SVM	0.893	0.988	0.997	1		0.001	<0.001	0.969	0.828	<0.001
KNN	<0.001	<0.001	<0.001	<0.001	<0.001		0.511	0.038	0.117	0.992
DT	<0.001	<0.001	<0.001	<0.001	<0.001	0.988		<0.001	<0.001	0.979
NB	0.948	0.997	0.999	1	1	<0.001	<0.001		1	0.001
RBF	0.014	0.05	0.08	0.308	0.471	0.308	0.024	0.359		0.007
Naïve	<0.001	<0.001	<0.001	<0.001	<0.001	0.966	1	<0.001	0.014	

L-Hop

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		1	1	1	0.821	0.002	0.019	0.577	0.96	<0.001
LSTM	0.146		0.993	0.997	0.445	<0.001	0.003	0.225	0.707	<0.001
LDA	1	0.041		1	0.96	0.007	0.057	0.821	0.997	<0.001
MLR	1	0.041	1		0.937	0.005	0.043	0.767	0.993	<0.001
SVM	0.986	0.007	1	1		0.225	0.643	1	1	0.002
KNN	0.003	<0.001	0.015	0.015	0.08		1	0.445	0.094	0.821
DT	0.025	<0.001	0.098	0.098	0.336	1		0.868	0.383	0.383
NB	0.921	0.002	0.996	0.996	1	0.176	0.551		0.999	0.007
RBF	0.247	<0.001	0.551	0.551	0.892	0.856	0.996	0.976		<0.001
Naïve	<0.001	<0.001	<0.001	<0.001	0.001	0.945	0.608	0.004	0.12	

Lunge

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		1	0.273	0.235	0.201	<0.001	<0.001	0.068	0.036	<0.001
LSTM	1		0.358	0.314	0.273	<0.001	<0.001	0.1	0.056	<0.001
LDA	0.024	0.086		1	1	0.121	<0.001	1	0.998	<0.001
MLR	0.018	0.067	1		1	0.144	<0.001	1	0.999	<0.001
SVM	0.006	0.024	1	1		0.171	<0.001	1	1	<0.001
KNN	<0.001	<0.001	<0.001	<0.001	0.002		0.171	0.406	0.559	0.011
DT	<0.001	<0.001	0.001	0.001	0.003	1		<0.001	<0.001	1
NB	<0.001	0.002	0.971	0.983	0.999	0.018	0.031		1	<0.001
RBF	0.031	0.108	1	1	1	<0.001	<0.001	0.953		<0.001
Naïve	<0.001	<0.001	<0.001	<0.001	<0.001	0.86	0.763	<0.001	<0.001	

Step-Down

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.006	0.647	0.596	0.199	<0.001	<0.001	<0.001	0.001	<0.001
LSTM	0.91		0.596	0.647	0.959	0.17	<0.001	0.863	1	<0.001
LDA	1	0.719		1	0.999	<0.001	<0.001	0.02	0.199	<0.001
MLR	1	0.765	1		1	<0.001	<0.001	0.024	0.232	<0.001
SVM	0.881	0.104	0.979	0.968		0.005	<0.001	0.144	0.647	<0.001
KNN	0.004	<0.001	0.015	0.012	0.279		0.012	0.972	0.545	0.647
DT	<0.001	<0.001	<0.001	<0.001	<0.001	0.207		<0.001	<0.001	0.743
NB	0.765	0.058	0.934	0.91	1	0.413	<0.001		0.996	0.069
RBF	0.125	0.002	0.279	0.241	0.934	0.979	0.009	0.979		0.005
Naïve	<0.001	<0.001	<0.001	<0.001	<0.001	0.125	1	<0.001	0.004	

T-Balance

	RC	*LSTM*	*LDA*	*MLR*	*SVM*	*KNN*	*DT*	*NB*	*RBF*	*Naïve*
RC		0.978	0.999	0.897	0.779	<0.001	<0.001	0.186	0.045	<0.001
LSTM	1		0.677	0.221	0.129	<0.001	<0.001	0.008	0.001	<0.001
LDA	0.064	0.1		0.999	0.992	0.001	<0.001	0.622	0.26	<0.001
MLR	0.024	0.04	1		1	0.017	<0.001	0.965	0.73	<0.001
SVM	0.011	0.018	1	1		0.036	<0.001	0.992	0.864	<0.001
KNN	<0.001	<0.001	0.1	0.22	0.359		0.028	0.35	0.73	0.221
DT	<0.001	<0.001	<0.001	<0.001	<0.001	0.001		<0.001	<0.001	0.998
NB	0.004	0.008	0.997	1	1	0.531	<0.001		1	<0.001
RBF	0.151	0.22	1	1	0.994	0.04	<0.001	0.966		0.001
Naïve	<0.001	<0.001	<0.001	0.004	<0.001	0.004	1	<0.001	<0.001	

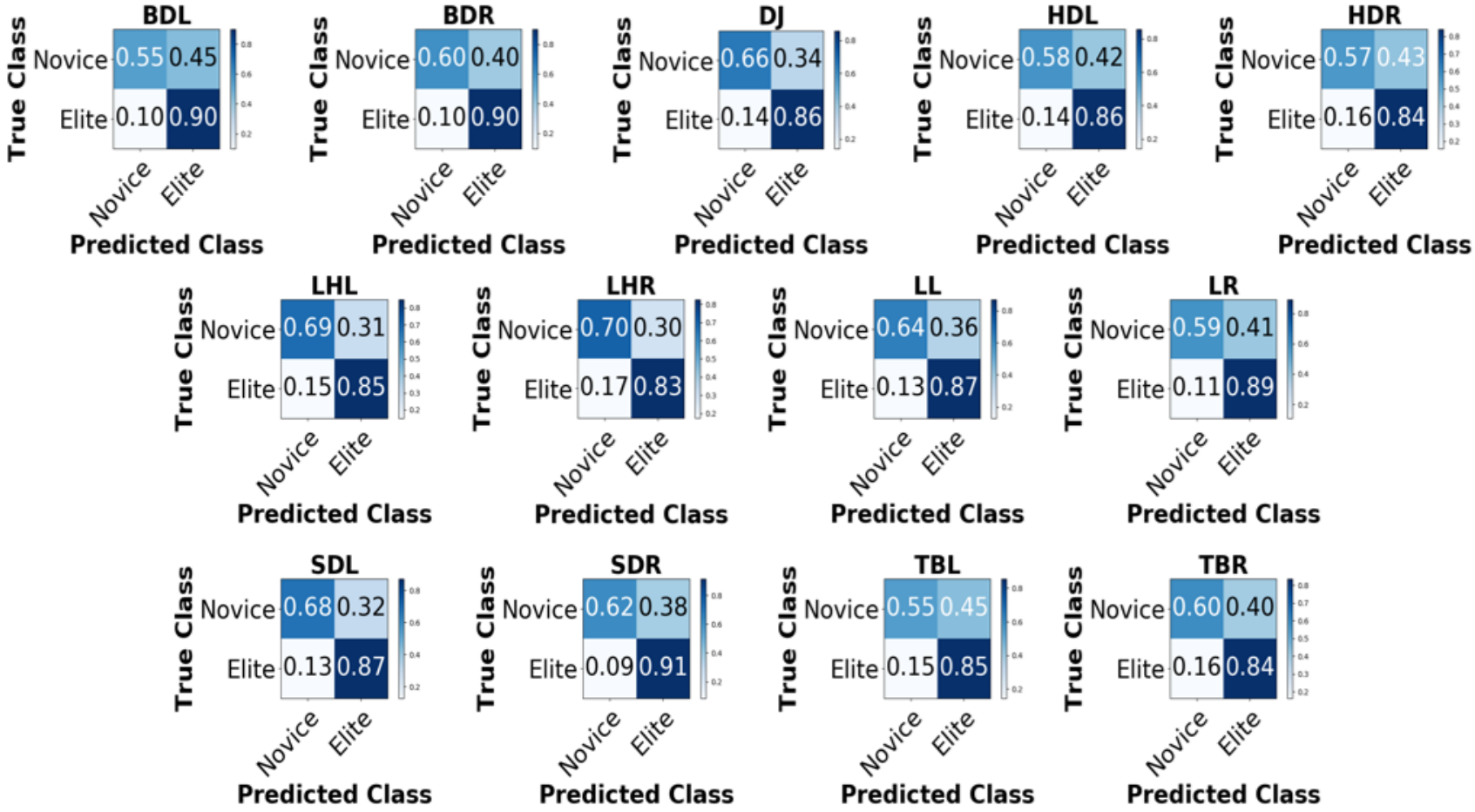


Figure C1. Confusion matrices for each task with level as the classifier. BDL = Bird-dog left, BDR = Bird-dog right, DJ = Drop jump, HDL = Hop-down left, HDR = Hop-down right, LHL = L-hop left, LHR = L-hop right, LL = Lunge left, LR = Lunge right, SDL = Step-down left, SDR = Step-down right, TBL = T-balance left, TBR = T-balance right.

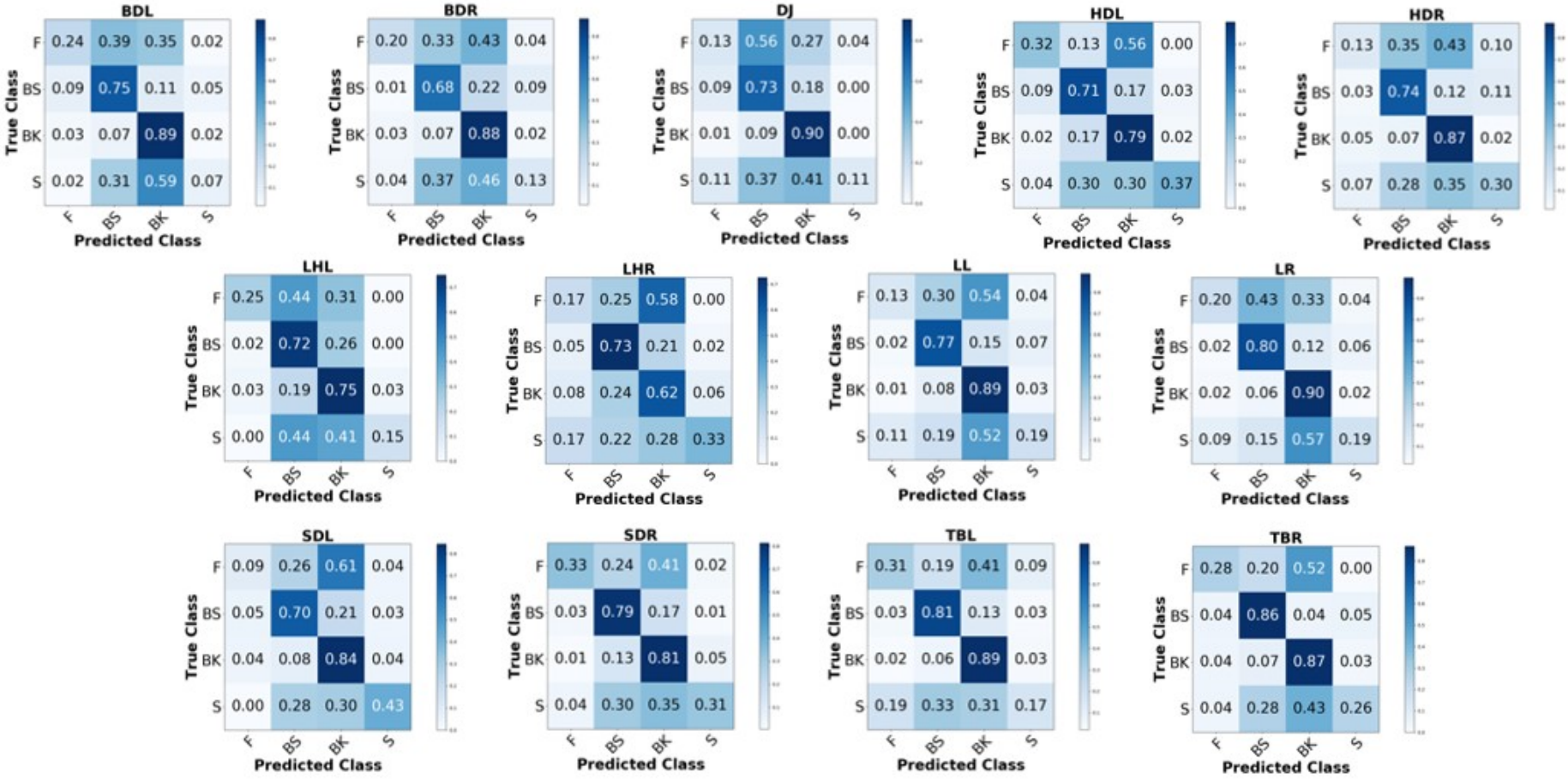
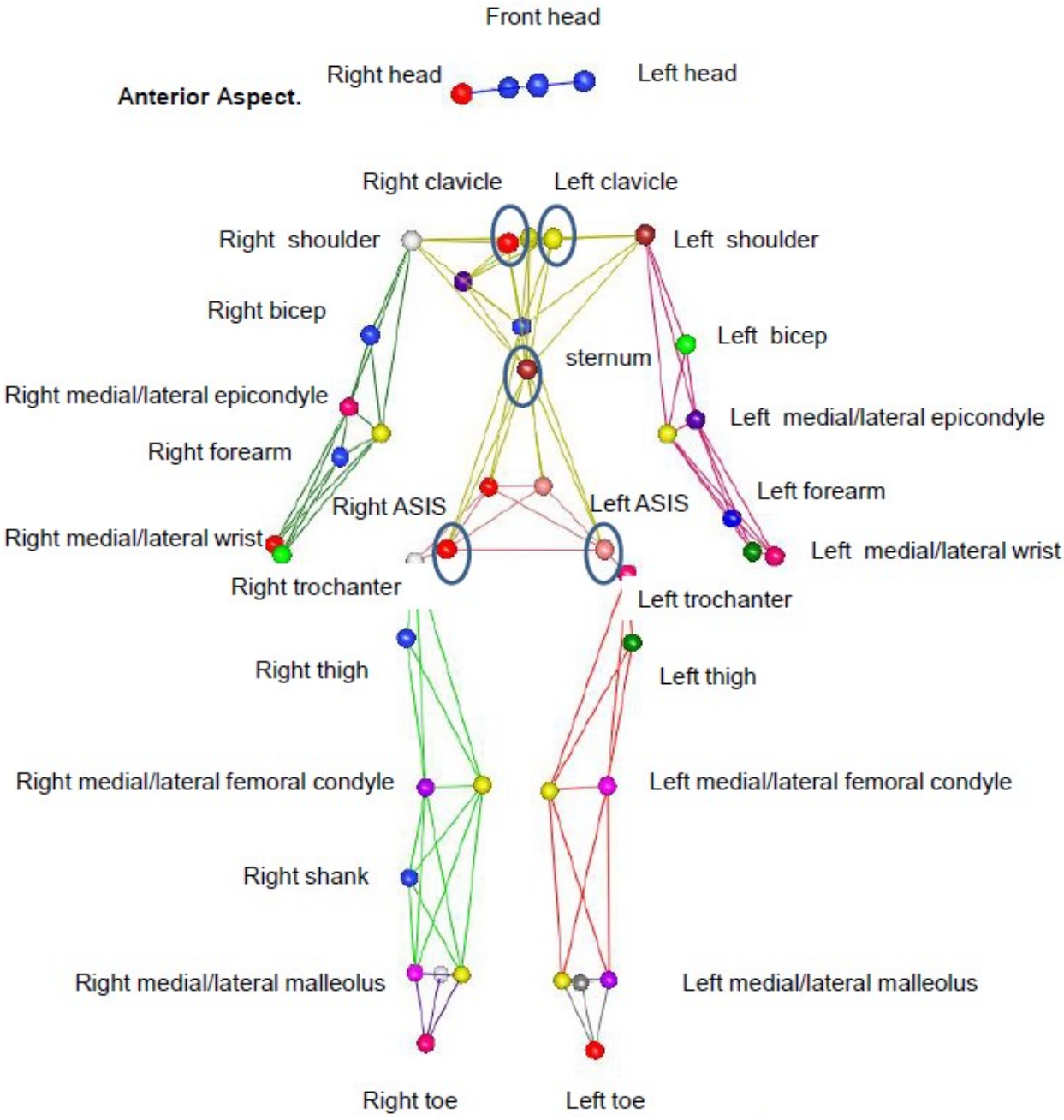
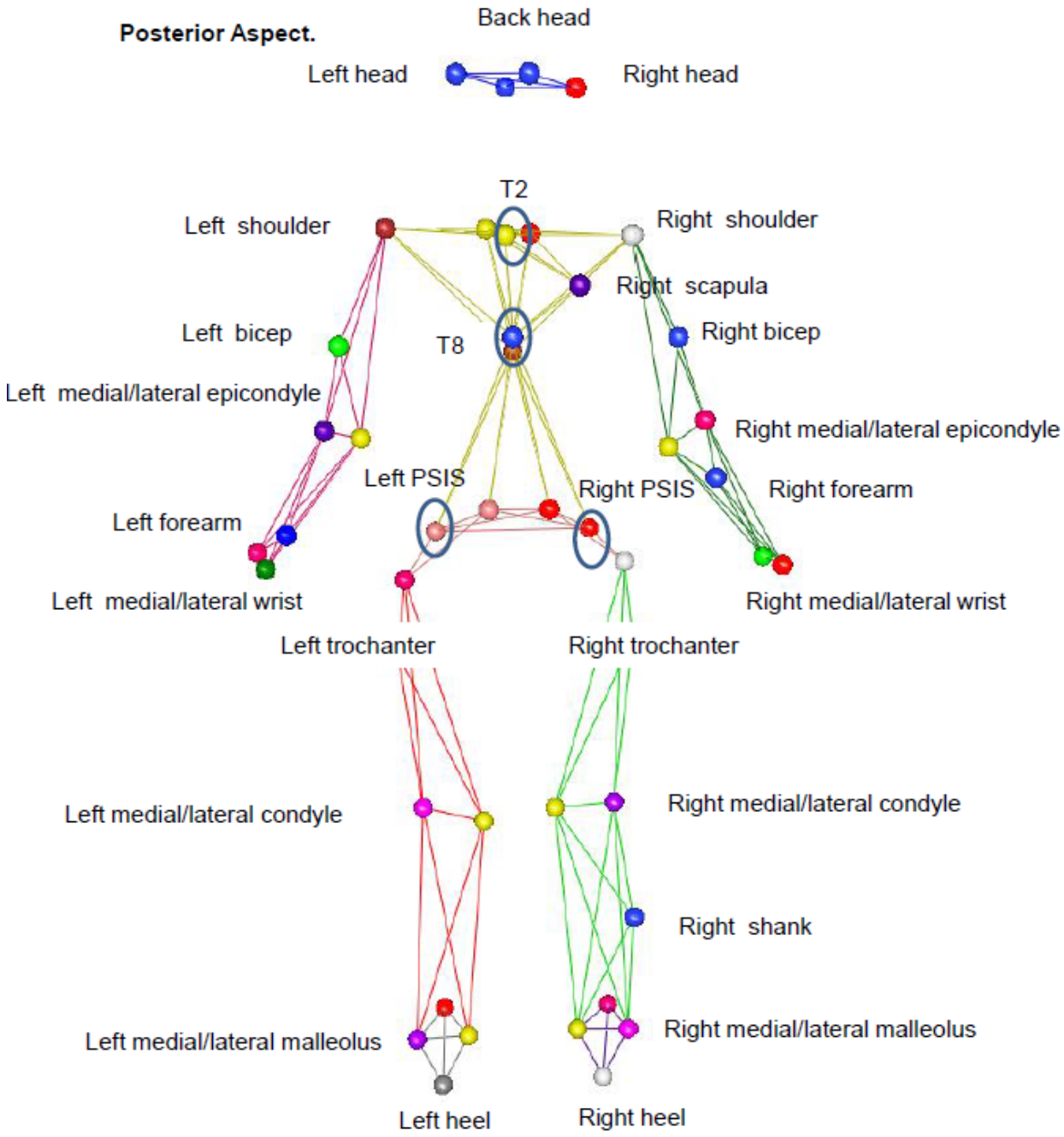


Figure C2. Confusion matrices for each task with sport as the classifier. F = Football, BS = Baseball, BK = Basketball, S = Soccer, BDL = Bird-dog left, BDR = Bird-dog right, DJ = Drop jump, HDL = Hop-down left, HDR = Hop-down right, LHL = L-hop left, LHR = L-hop right, LL = Lunge left, LR = Lunge right, SDL = Step-down left, SDR = Step-down right, TBL = T-balance left, TBR = T-balance right.

Appendix D: Marker Set




Posterior Aspect.
Back head
Left head
Right head
T2
Left shoulder
Right shoulder
Right scapula
Left bicep
T8
Right bicep
Left medial/lateral epicondyle
Right medial/lateral epicondyle
Left PSIS
Right PSIS
Right forearm
Left forearm
Left medial/lateral wrist
Right medial/lateral wrist
Left trochanter
Right trochanter
Left medial/lateral condyle
Right medial/lateral condyle
Right shank
Left medial/lateral malleolus
Right medial/lateral malleolus
Left heel
Right heel