Machine Learning and Deep Neural Network Architectures for 3D Motion Capture Datasets

Alistair Boyle, Gwyneth B. Ross and Ryan B. Graham

***Abstract*— Baseline performance for 3D joint centre trajectory classification using a number of traditional machine learning techniques are presented. This framework supports a robust comparison between classifier architectures over a 416 subject dataset of athletes (professional, college, and amateur) from five primary sports and six non-primary sports performing thirteen non-sport specific movements. A variety of deep neural networks specifically intended for time-series data are currently being evaluated.**

***Clinical/sports relevance*— Patient and athlete movement patterns can be measured by 3D motion capture and evaluated by systematically using machine learning. By providing a distributable "expert", issues with inter- and intra-rater variability may be reduced. This work explores a variety of machine learning techniques to evaluate which methods are most appropriate for motion capture data.**

I. INTRODUCTION

Movement screens are a set of non-sport specific movements that are typically evaluated by human observers to classify athletes and identify training deficiencies [1]–[3]. One example is the Functional Movement Screen (FMS$^{\text{TM}}$, Functional Movement Systems, USA). There are known issues with inter- and intra-rater reliability in commonly used movement screening techniques [4]–[8].

Three-dimensional (3D) motion capture technology uses markers applied to an athlete and a number of cameras to capture a 3D time-series of a subject's movement at resolutions under 2 mm for dynamic movements [9], [10]. Preprocessing of the 3D marker time-series allows labelling of markers based on relative (approximate) positions. With a cleaned set of marker trajectories, the labelled markers can then be used to construct frames of reference for each limb or body segment. A calibrated anatomical system technique (CAST) was used to find dynamic joint centres by placing additional markers on anatomical landmarks and then using regression to estimate joint centres based on the original marker placement (without the extra markers on anatomical landmarks) throughout dynamic movements [11], [12]. This processing produces joint centre locations in 3D over time. Marker trajectories or joint centres can be stored in the C3D data format alongside sampled analog data such as force plate measurements.

Recent work by Ross *et al.* [13] used 3D joint centre trajectory data to classify athletes into elite and novice levels[1]. We use the same dataset in this work. Ross *et al.* used principal component analysis (PCA) followed by linear discriminant analysis (LDA) trained for each of the thirteen movements. However, the most appropriate deep neural network architecture for 3D marker or joint centre trajectories is unclear.

Many types of image classification problems have benefited from the development of convolutional neural networks (CNN) [14], a technique which scans a picture for two-dimensional patterns in each of the three red-green-blue (RGB) colour channels and merges the results to perform classification or prediction tasks. The efficiency of the CNN comes from training small filters which are scanned across the whole image. This approach implies that a feature should be consistent across different regions of the image. The results are mixed and pooled to achieve a blending across colour channels and at different scales. Research into the classification of activities of daily living (ADL) have made use of "colouring" data by assigning accelerometer xyz data streams to the red-green-blue (RGB) channels and formatting data as small RGB images of rectangular shape [15]–[17]. This structuring of the data enables the use of image-based CNN architectures followed by recurrent neural networks such as the long short term memory (LSTM) network [18] to classify successive chunks of a time-series. The placement of adjacent time-series rows in the image implies a relationship between adjacent time-series that may not be appropriate. Some classification results suggest that three measurements are "optimal," but we note that this is the limit beyond which ordering of time-series rows becomes critical to maintaining structural relationships, for example the relationship between the hip, knee, and ankle on each leg. One could order the rows by left side, then right side with the hips in the middle rows to maintain ordered structure but adding a torso introduces an ordering problem that requires more than one-dimension. We speculate that clever ordering or repetition may overcome these structural limitations. On the other hand, new architectures may be required to make the most of our data, rather than reusing general purpose image classifiers.

Recently, Fawaz *et al.* [19] have evaluated a number of deep neural network architectures on univariate and multivariate time-series data. Our time-series data is more structured than general multivariate time-series data. For joint centre or marker data, the specific relationships between the three spatial axes and the highly correlated and structured

This work was supported by a MITACS Postdoctoral Fellowship.

Alistair Boyle, Gwyneth Ross, and Ryan Graham are with the School of the School of Human Kinetics, University of Ottawa, Ottawa, Ontario, Canada `alistair.boyle@uottawa.ca`

[1] In this work, Ross *et al.* 's "elite" corresponds to our "professional" and "college" combined, and their "novice" corresponds to our "amateur."

TABLE I

CLASSIFIERS

Acronym	Classifier
MLP	Multi-Layer Perceptron
FCN	Fully Convolutional Neural Network
TCNN	Time Convolutional Neural Network
Resnet	Residual Network
Encoder	Auto-Encoder
SVM	Support Vector Machine
LDA	Linear Discriminant Analysis
PCA+X	Principle Component Analysis followed by $\ldots X$
Naive	Naïve Baseline (majority class)

TABLE II

DATASET SUMMARY — MOVEMENT SCREENED ATHLETES

n	Type	Categories
416	Subjects	
13	Movements	See Tab. III
3	Levels	Professional/**Pro** (148), **College** (119), **Amateur** (149)
7+	Sports	**Basketball** (59+54+14=127), **Baseball** (61+3+13=77), **Golf** (1+3+56=60), **Soccer** (13+18+28=59), **Football** (2+31+16=49), **Other** (12+8+24=44) [Track & Field (9+2+3=14), Tennis (3+3+10=16), Lacrosse (0+3+4=7), Cricket (0+0+1=1), Volleyball (0+0+1=1), Squash (0+0+1=1), un-reported (0+0+4=4)]

20.4 ± 4.3	Mean Age	Professional: 23.4 ± 3.7;
		College: 21.1 ± 1.9;
		Amateur: 16.8 ± 3.7

6.2:1.0	Gender M:F	Male (143+110+105=358), Female (5+7+46=58), other/un-reported (0)

subject counts by sport and gender are listed as (pro + college + amateur = total)

TABLE III

MOVEMENTS

		# Athletes	
Acronym	Activity	Left	Right
DJ	Drop Jump	—	275
BDL, **BDR**	Bird Dog Left, Right	381	387
HDL, **HDR**	Hop Down Left, Right	401	400
LHL, **LHR**	L-Hop Left, Right	268	267
LL, **LR**	Lunge Left, Right	400	401
SDL, **SDR**	Step Down Left, Right	399	403
TBL, **TBR**	T-Balance Left, Right	392	395
Athletes with at least one “good” movement		416	
Athletes with all thirteen movements		200	

nature of the relationships between joints suggest that it should be possible to provide this prior information to a neural network. There is no existing architecture that bridges this gap. We have implemented variations of the best performing architectures from [19] (fully convolutional network, time convolutional neural network, auto-encoder). A convolutional neural network structure commonly used in image recognition (residual network) was also included. These techniques were compared to general machine learning techniques using principal component analysis, support vector machines, and linear discriminate analysis. A multi-layer perceptron and naïve (select the largest class) classifier round out the alternatives. (Classifiers and their acronyms are listed in Tab. I.)

For this work, we define the best performing classifier as the architecture which achieves the greatest median accuracy across movement tasks. We also consider classifier training time and the quantity of trainable parameters, ultimately looking for a promising classifier that avoids over/under fitting.

II. METHODS

Athletes from a range of sports had 45 markers applied which were tracked at 120 Hz (Raptor-E, Motion Analysis, Santa Rosa, USA) during the performance of thirteen dynamic movements that challenge balance and stability. Data were collected from 416 athletes by Motus Global (Rockville Centre, New York). Athletes from basketball, baseball, soccer, golf, football and other sports (track and field, tennis, lacrosse, cricket, volleyball, squash) were recorded. Subjects were approximately balanced between professional sports (MBA, MLB, NFL, PGA, FIFA), college, and amateur. A summary of the athletes is available in Tab. II. Prior to data collection, participants read and signed consent forms permitting future use of the data for research. The University of Ottawa Research Ethics Board (Ottawa, Canada) approved[2] the secondary use of the data. Data were preprocessed by gap-filling (Cortex, Motion Analysis, Rohnert Park, USA) and a whole-body kinematic model (Visual3D, C-Motion Inc., Germantown, USA) was used to produce joint centre trajectories from the gap-filled marker data. The results of preprocessing were 32 trajectories in xyz for each athlete. Trajectories included the head (4), spine at T_2 and T_8 (2), pelvis anterior and posterior mid-points (2), sternum (1), trunk centre of gravity (1), proximal and distal ends for upper and lower arm (8), proximal and distal ends for upper and lower leg segments (8), and feet (6). Collectively, we refer to these 32 trajectories as the “joint centre trajectories“ in this work. Selected movements were cropped from the time-series, and low-pass 4^{th}-order Butterworth zero-phase filtered (f_c=15 Hz). Each movement was normalized by linear interpolation to the median number of frames across athletes. Athletes were allowed to repeat movements until they were satisfied with their performance. We used only the “best” self-reported attempt for each movement.

Ten types of classifiers were trained to predict either the level of the athlete or their sport. The same data were used for predicting level or sport. PCA was the only feature

[2] Ethics file number H-08-18-1085.

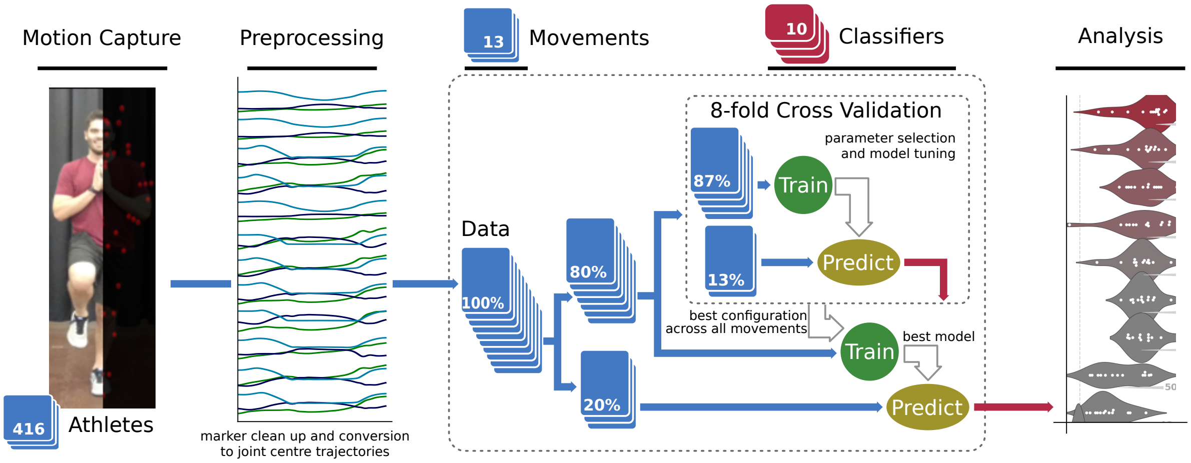


Fig. 1. Data flow *(left-to-right)* from motion capture, through preprocessing, classifier training and testing (with 8-fold cross-validation for classifier tuning), and analysis of the results grouped by movement.

selection used, after preliminary attempts with some wrapper and filter methods were inconclusive. Not all participants had high quality data captured across all movements: only 200 athletes had a complete set of the thirteen movements. The total number of athletes available for each movement varied from 267 (L-hop right) to 403 (step down right) (Tab. III). To avoid discarding much of our data, each movement was used separately, leading to thirteen trained classifiers for each predictor (level or sport), over ten classifiers: in total, 260 classifiers were evaluated. Classifier performance was grouped by movement so that each classifier was ranked based on its overall median accuracy across all thirteen movements. The analysis of classifier performance for predicting sport and level were performed separately using the same criteria.

Ten classifiers were compared (Tab. I). Bold text has been used to identify labels used in tables and figures. A baseline naïve classifier (selecting the majority class) set a minimum performance threshold (**Naive**). Classifiers using Support Vector Machines (**SVM**), and Linear Discriminant Analysis (**LDA**) were compared to classifiers using Principle Component Analysis (PCA) for preprocessing (**PCA+SVM** and **PCA+LDA**). Deep neural networks (DNNs) are currently being tuned, and preliminary results are reported here. A Multi-Layer Perceptron (**MLP**) was compared to the leading techniques from Fawaz *et al.* [19] on multivariate classification. These classifiers included the Fully Convolutional Network (**FCN**), Time Convolutional Neural Network (**TCNN**), residual network (**Resnet**), and auto-encoder (**Encoder**). Details of these DNN architectures for time-series data can be found in [19].

Each classifier was trained independently on a single type of movement to predict either the athlete's sport or level. (See Tab. III for a list of the 13 movements. Fig. 1 summarizes the classification workflow.) For each movement, the data were separated into an 80:20 train:test split. The training data were used in an 8-fold cross-validation grid search to tune classifier configuration parameters such as LDA shrinkage and MLP hidden layers. A single set of configuration parameters were selected for each classifier across all movements and predictors. This set of configuration parameters was used with the same training data to recalculate an 8-fold cross-validated estimate of expected accuracy by bootstrap confidence bounds. The training data were already used for tuning so that the estimate was biased and likely to over estimate performance. Finally, all the training data were used together to train a finalized classifier for each movement and evaluated on the testing data. Results were recorded and analysis of performance across the classifiers, movements, and predictors was carried out.

We have reported median accuracy in these results. Accuracies for each movement were plotted in the violin plots (Fig. 2, Fig. 3) as white dots. Confusion matrices for the top performing classifier have been provided for each movement. Variability in the accuracy across movements for all classifiers within 5% of the top performing classifier were shown in bar charts.

The Resnet, MLP, FCN, TCNN, and Encoder classifiers have not yet been tuned using the cross-validation grid search.

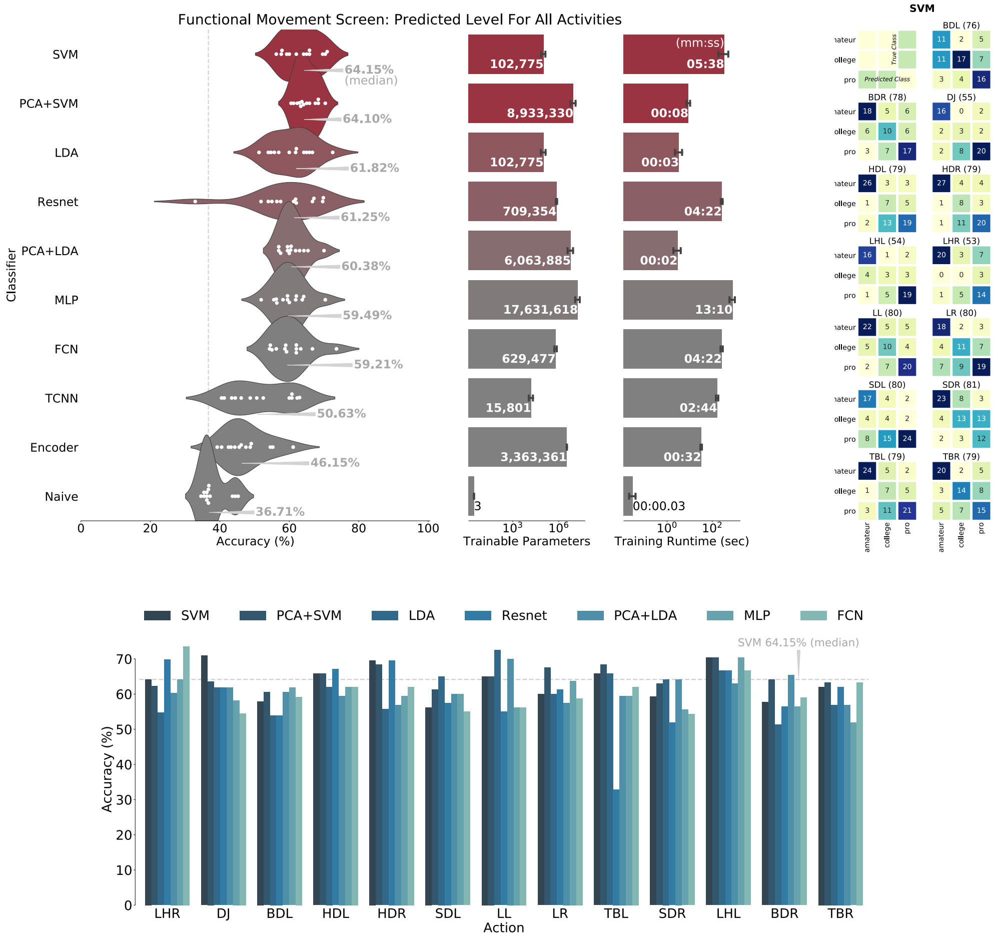
III. Results

The results for classification by level and by sport are summarized next.

As expected, the un-tuned deep neural networks generally performed poorly. The tuning of these networks continues.

A. *Classification by Level*

Predicting the level of an athlete (pro, college, amateur) shows SVM as the best classifier with a median accuracy of 64.15% (Fig. 2, violin plots). The PCA+SVM, LDA, Resnet, PCA+LDA, and MLP classifiers are within 5% of the leading classifier. The naïve classifier which selects the largest class had a median accuracy of 37%, as expected for a roughly balanced three class problem. The best classification rate was nearly double this naïve classification rate. The PCA+SVM



	amateur	college	pro
BDL (76) amateur	11	2	5
BDL (76) college	11	17	7
BDL (76) pro	3	4	16
BDR (78) amateur	18	5	6
BDR (78) college	6	10	6
BDR (78) pro	3	7	17
DJ (55) amateur	16	0	2
DJ (55) college	2	3	2
DJ (55) pro	2	8	20
HDL (79) amateur	26	3	3
HDL (79) college	1	7	5
HDL (79) pro	2	13	19
HDR (79) amateur	27	4	4
HDR (79) college	1	8	3
HDR (79) pro	1	11	20
LHL (54) amateur	16	1	2
LHL (54) college	4	3	3
LHL (54) pro	1	5	19
LHR (53) amateur	20	3	7
LHR (53) college	0	0	3
LHR (53) pro	1	5	14
LL (80) amateur	22	5	5
LL (80) college	5	10	4
LL (80) pro	2	7	20
LR (80) amateur	18	2	3
LR (80) college	4	11	7
LR (80) pro	7	9	19
SDL (80) amateur	17	4	2
SDL (80) college	4	4	2
SDL (80) pro	8	15	24
SDR (81) amateur	23	8	3
SDR (81) college	4	13	13
SDR (81) pro	2	3	12
TBL (79) amateur	24	5	2
TBL (79) college	1	7	5
TBL (79) pro	3	11	21
TBR (79) amateur	20	2	5
TBR (79) college	3	14	8
TBR (79) pro	5	7	15

Fig. 2. Classification by level; *(top)* violin plots show per movement classification accuracy (white dots) ordered by median accuracy for each classifier; the number of parameters and training time for each classifier are shownin log-scaled bar charts with 95% confidence intervals; *(middle)* confusion matrices for the top classifier (best median accuracy) are presented; *(bottom)* bar charts for the classification accuracy of each movement over all classifiers within 5% of the top classifier show that some movements are classified more successfully than others.

and SVM perform to nearly the same accuracy because after a grid search, the PCA keeps 99.99% of the explained variance (approximately 150 principal components from approximately 500 data frames) so that almost all the information was retained in a slightly compressed form. SVM performance increased markedly when using shrinkage, at the cost of considerably extended runtimes. Interestingly, PCA+SVM has significantly more trainable parameters, but because of the efficient singular value decomposition used in PCA, training was very quick. Accuracy for the classifiers was fairly uniform across movements (Fig. 2, bar chart), varying by approximately 5–10%, with the exception of the un-tuned residual network classifier, which we expect to improve after further refinement. The confusion matrices (Fig. 2, confusion matrices) illustrate that identifying amateur level athletes was generally not difficult using any single movement except for bird dog left (BDL) and step down left (SDL). In addition, we can observe that some movements such as bird dog left and T-balance right (BDL, TBR) better identify college versus professional athletes.

B. Classification by Sport

For the prediction of sport, our results show that PCA+SVM was the best classifier with 58.75% median accuracy (Fig. 3, violin plots). The best classifier performed much better than the naïve classifier which had a median accuracy of 30.00%. There was more class imbalance in the sport classification than in the level classification explaining the greater median naïve accuracy on this six class problem (nominally 16.67% for a balanced 6-class problem). The PCA+SVM, PCA+LDA, MLP, and SVM all have approximately the same performance with PCA+LDA being the fastest to train and SVM having the smallest number of trainable parameters. LDA performance was low without shrinkage: we are in the process of testing LDA with shrinkage which is likely to boost the LDA classifier accuracy. The confusion matrices (Fig. 3, confusion matrices) show that classifying basketball players was quite successful across all movements. Accuracy across movements (Fig. 3, bar chart) varies by approximately 10% between movements.

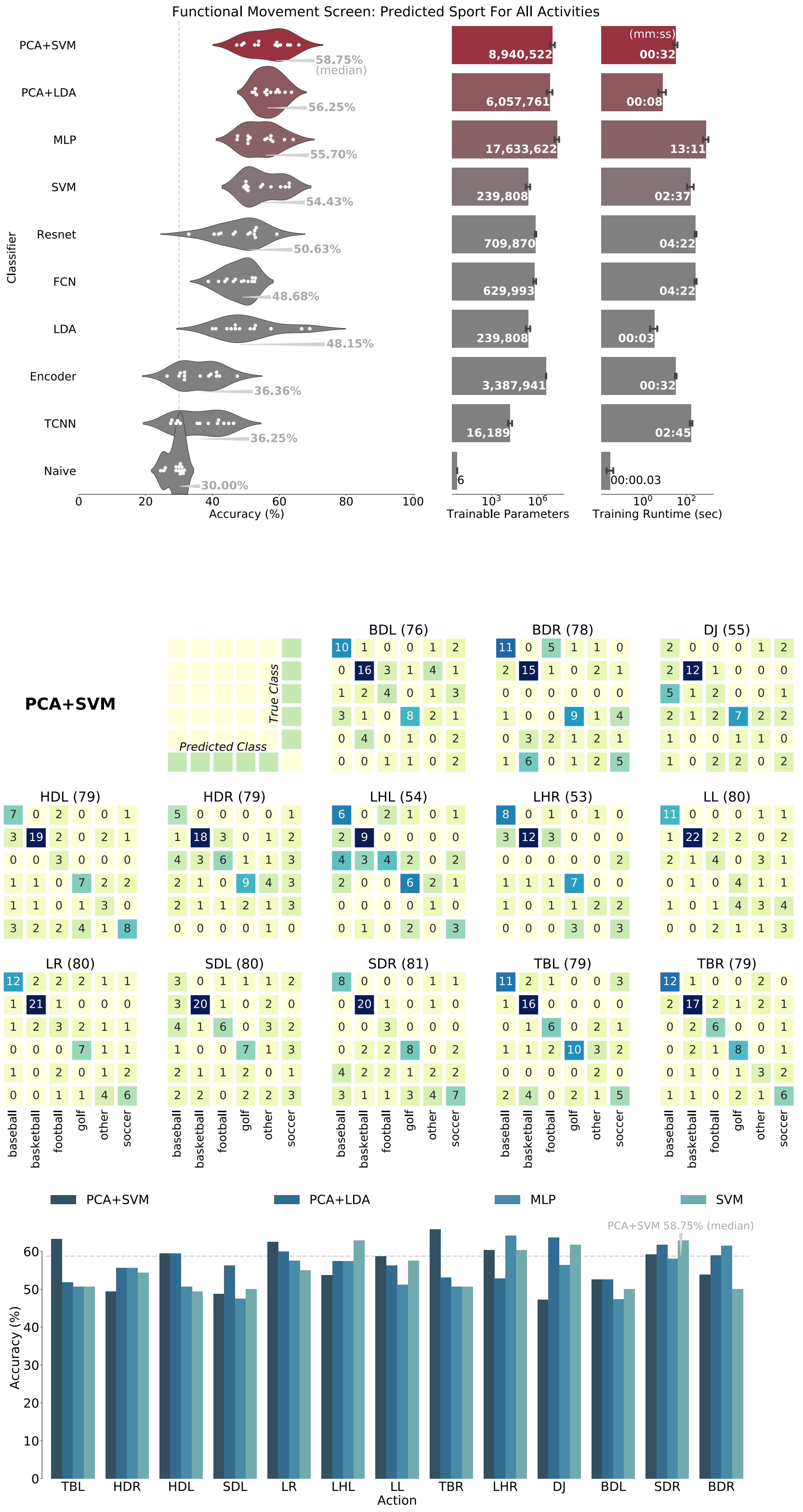


Fig. 3. Classification by sport; *(top)* violin plots show per movement classification accuracy (white dots) ordered by median accuracy for each classifier; the number of parameters and training time for each classifier are shownin log-scaled bar charts with 95% confidence intervals; *(middle)* confusion matrices for the top classifier (best median accuracy) are presented; *(bottom)* bar charts for the classification accuracy of each movement over all classifiers within 5% of the top classifier show that some movements are classified more successfully than others.

IV. Discussion

In predicting athlete level and sport, classifier performance varies between movements. The confusion matrices illustrate that this performance is not uniform across classes. This presents the opportunity to leverage this information to construct an ensemble classifier which may be more successful at separating the groups (pro, college, and amateur; baseball, basketball, football, golf, soccer, and other).

Predicting the level of an athlete (pro, college, amateur) appears to be a fairly linear task (Fig. 2), where the linear LDA, SVM, and PCA+SVM classify at roughly the same accuracy. PCA+SVM gives very similar results to SVM since the tuned PCA parameters give best performance at 99.99% of explained variance, the greatest explained variance ratio tested to date.

The prediction of sport shows promising preliminary results (Fig. 3). Basketball players' movement patterns appear to be quite distinct. Surprisingly, golf, for which the majority of the class are amateur athletes, is relatively identifiable. College and amateur athletes were left in the data sets, though these athletes have almost certainly not specialized as much as a professional would have. We speculate that using the professional athletes for classifier training or providing additional groupings for cross-trained and untrained athletes may lead to more consistent results.

Feature selection may improve these results, though DNNs are likely to benefit less than the SVM/LDA-based classifiers. The automated feature selectors we used required the data to be "flattened" into a two-dimensional array which hides xyz joint centre groupings in the trajectories. This blindness to higher level relationships between joint centre locations at each time interval resulted in comprehensive feature selection methods that experienced a combinatorial explosion in the number of possible features that were to be tested, making them impractical. Random search methods plateaued after removing approximately one third of the features and did not measurably improve overall classifier performance. Our preliminary attempts at automated grouped feature selection by dropping joint centres were inconclusive. With classifiers that are now better tuned, it is possible that feature selection by joint centre will be more successful in the future.

V. Conclusions

A framework for comparison of classifier performance on joint centre trajectory datasets was presented. Preliminary results showing traditional machine learning classification performance on two types of classification problems were illustrated. Tuning of the more complex deep neural network architectures is on-going.

Acknowledgment

We would like to acknowledge the excellent 3D marker dataset collected by Motus Global Inc. and their thoroughness in measuring athletes across such a wide range of high-level sports. A. B. deeply appreciates the countless hours spent by G. R. cleaning the dataset and preprocessing the marker data into joint centre trajectories.

References

[1] G. Cook, L. Burton, B. Hoogenboom, and M. Voight, "Functional movement screening: the use of fundamental movements as an assessment of function - part 1," *International Journal of Sports Physical Therapy*, vol. 9, no. 3, pp. 396–409, 2014.

[2] ——, "Functional movement screening: the use of fundamental movements as an assessment of function - part 2," *International Journal of Sports Physical Therapy*, vol. 9, no. 4, pp. 549–563, 2014.

[3] R. McCunn, K. Fünten, H. Fullagar, I. McKeown, and T. Meyer, "Reliability and association with injury of movement screens: A critical review," *Sports Medicine*, vol. 46, no. 1, pp. 763–781, 2016.

[4] K. Minick, K. Kiesel, L. Burton, A. Taylor, P. Plisky, and R. Butler, "Interrater reliability of the functional movement screen," *Journal of Strength and Conditioning Research*, vol. 24, no. 2, pp. 479–486, 2010.

[5] J. Onate, T. Dewey, R. Kollock, K. Thomas, B. V. Lunen, M. DeMaio, and S. Ringleb, "Real-time intersession and interrater reliability of the functional movement screen," *Journal of Strength and Conditioning Research*, vol. 26, no. 2, pp. 408–415, 2012.

[6] P. Gribble, J. Brigle, B. Pietrosimone, K. Pfile, and K. Webster, "Intrarater reliability of the functional movement screen," *Journal of Strength and Conditioning Research*, vol. 27, no. 4, pp. 978–981, 2013.

[7] C. Smith, N. Chimera, N. Wright, and M. Warren, "Interrater and intrarater reliability of the functional movement screen," *Journal of Strength and Conditioning Research*, vol. 27, no. 4, pp. 982–987, 2013.

[8] H. Gulgin and B. Hoogenboom, "The functional movement screening (FMS)$^{\text{TM}}$: an inter-rater reliability study between raters of varied experience," *International Journal of Sports Physical Therapy*, vol. 9, no. 1, pp. 14–20, 2014.

[9] P. Eichelberger, M. Ferraro, U. Minder, T. Denton, A. Blasimann, and F. K. andH. Baur, "Analysis of accuracy in optical motion capture — a protocol for laboratory setup evaluation," *J. Biomech.*, vol. 49, no. 10, pp. 2085–2088, 2016.

[10] P. Merriaux, Y. Dupuis, R. Boutteau, P. Vasseur, and X. Savatier, "A study of vicon system positioning performance," *Sensors*, vol. 17, no. 7, p. 18, 2017.

[11] A. Cappozzo, F. Catani, U. Croce, and A. Leardini, "Position and orientation in space of bones during movement: anatomical frame definition and determination," *Clinical Biomechanics*, vol. 10, no. 4, pp. 171–178, 1995.

[12] F. Buczek, M. Rainbow, K. Cooney, M. Walker, and J. Sanders, "Implications of using hierarchical and six degree-of-freedom models for normal gait analyses," *Gait & Posture*, vol. 31, no. 1, pp. 57–63, 2010.

[13] G. Ross, B. Dowling, N. Troje, S. Fischer, and R. Graham, "Objectively differentiating movement patterns between elite and novice athletes," *Medicine & Science in Sports & Exercise*, vol. 50, no. 7, pp. 1457–1464, 2018.

[14] Y. LeCunn, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 1, pp. 436–444, 2015.

[15] J. Yang, M. Nguyen, P. San, X. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time seriesfor human activity recognition," in *Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI2015)*. Palo Alto, California, USA: AAAI Press, 2015, pp. 3995–4001.

[16] F. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 25, 2016.

[17] A. Clouthier, G. Ross, and R. Graham, "Sensor data required for automatic recognition of athletic tasks using deep neural networks," *Front. Bioeng. Biotechnol.*, vol. 7, no. 473, p. 10, 2019.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[19] H. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, 2019.