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

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

***Abstract*— Biomechanical movement data are highly correlated multivariate time-series for which a variety of machine learning and deep neural network classification techniques are possible. For image classification, convolutional neural networks have reshaped the field, but have been challenging to apply to 3D movement data with its intrinsic multidimensional nonlinear correlations. Deep neural networks afford the opportunity to reduce feature engineering effort, remove model-based approximations that can introduce systematic errors, and reduce the manual data processing burden which is often a bottleneck in biomechanical data acquisition. What classification techniques are most appropriate for biomechanical movement data? Baseline performance for 3D joint centre trajectory classification using a number of traditional machine learning techniques are presented. Our framework and dataset support a robust comparison between classifier architectures over 416 athletes (professional, college, and amateur) from five primary 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*— Non-sport-specific movement screens are a generalized tool that can identify training deficiencies and performance limitations. Athlete movement patterns can be measured by 3D motion capture and evaluated 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$^{\mathrm{TM}}$, Functional Movement Systems, USA). FMS$^{\mathrm{TM}}$ has been used at the NFL Scouting Combine since 2011 to screen college football players for preexisting performance limiting dysfunction(s). Identification of specific movement limitations is typically based on exceeding a threshold score for each movement as assigned by a trained observer. There are known issues with inter- and intra-rater reliability in commonly used movement screening techniques [4]–[8]. We hypothesize that movement data are much richer than the relatively crude assessments typically used today. Potential applications of this work are: to more systematically classify movement quality, to identify specific joints and time points (using "explainable AI" techniques), and to improve non-automated movement screening by optimizing observer location and focus.

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 participant's dynamic movement at resolutions under 2 mm [9], [10]. Preprocessing of 3D marker time-series allows labelling of markers based on relative positions. With a cleaned set of marker trajectories, the labelled markers can 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 that 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

This work was supported by a MITACS Postdoctoral Fellowship.

Alistair Boyle, Gwyneth Ross, and Ryan Graham are with 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."

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

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 are 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 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 discriminant analysis. A multi-layer perceptron and naïve (select the largest class) classifier round out the alternatives.

Which type of machine learning or deep neural network classifier performs best for classifying human movement quality? For this work, we define the best performing classifier as the architecture that 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 and 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 (127), baseball (77), soccer (59), golf (60), football (49) and other sports (44: track and field, tennis, lacrosse, cricket, volleyball, squash) were recorded. Participants were approximately balanced between professional sports (148: MBA, MLB, NFL, PGA, FIFA), college (119), and amateur (149). 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 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.

TABLE I
MOVEMENTS

Acronym	Activity	# Athletes Left	# Athletes 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	

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 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. I). 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. 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 Principal Component Analysis

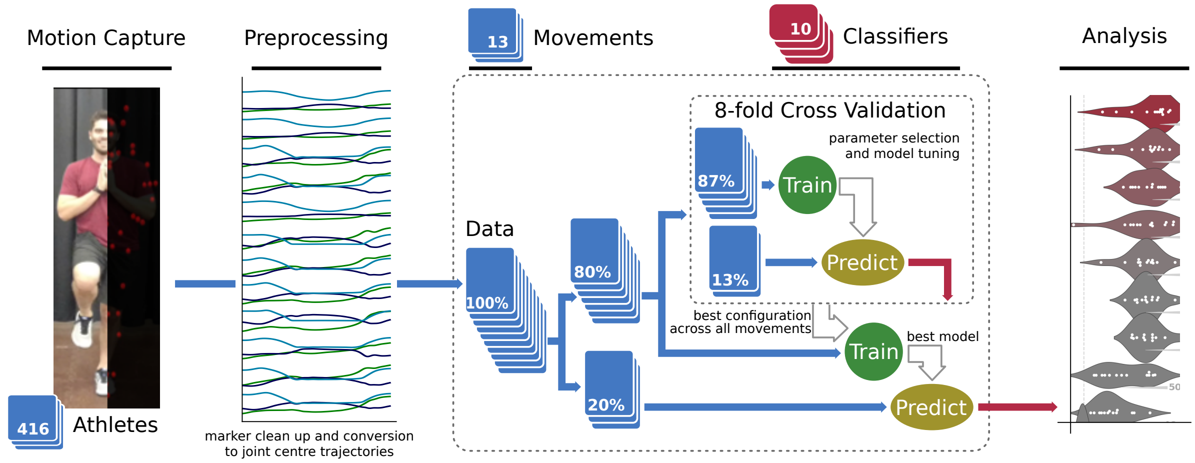

Fig. 1. Data flow *(left-to-right)* from motion capture, through preprocessing, classifier training and testing (with 8-fold cross-validation for classifier tuning), and analysis of the results grouped by movement.

(PCA) for preprocessing (**PCA+SVM** and **PCA+LDA**). 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. I for a list of the 13 movements and Fig. 1 for a summary of the classification workflow and dataset splitting. 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 was selected for each classifier across all movements and predictors. The complete training data were used to train a finalized classifier for each movement and evaluated on the testing data. Data was reused for final training because there is insufficient data for some movements to further split the dataset. Results were recorded and analysis of performance across the classifiers, movements, and predictors was carried out. We have reported median accuracy in these results.

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. The classifier performance in general was quite good, considering the multiclass and non-sport-specific nature of the data. Classification of this type of data by human observers is challenging. The performance of the PCA+LDA classification was comparable, or better than, reported in [13]. Various feature engineering methods were attempted to reduce the number of features: automated feature selection methods were defeated by the highly correlated data, typically achieving one third reduction in features, while manually constructing a number of relative movement features such as model-based joint angles, movement relative to proximal joints, or left-right imbalance did not change, or reduced, performance. The regularizing/sparsifying controls in our selected classification techniques achieved better performance than our attempts at extensive feature engineering. Untuned classifiers generally performed poorly. Tuning of the classic machine learning techniques has been completed. The tuning of the deep neural network classifiers is currently underway. We anticipate competitive performance from these classifiers.

Predicting the level of an athlete (pro, college, amateur) shows SVM as the best classifier with a median accuracy of 64.15%. 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 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, varying by approximately 5–10%, with the exception of the untuned Resnet classifier, which we expect to improve after further refinement. By plotting confusion matrices we observed 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, some movements such as BDL and T-balance right (TBR) better identified college versus professional athletes.

For the prediction of sport, our results show that

PCA+SVM was the best classifier with 58.75% median accuracy. 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. By plotting confusion matrices, we observed that classifying basketball players was quite successful across all movements. The accuracy for a classifier, for example PCA+SVM, varied by approximately 10% depending on which movement was classified.

IV. Discussion

In predicting athlete level and sport, classifier performance varied between movements. An ensemble classifier might take advantage of these variations to improve accuracy by comparing an athlete's movements amongst themselves. Combining information across movements could be decisive, though we are somewhat restricted by the number of "complete" movement sets in this data which would reduce our dataset from 416 to 200 participants. This is the subject of a future investigation.

Predicting the level of an athlete (pro, college, amateur) appeared to be a fairly linear task, where the linear LDA, SVM, and PCA+SVM classified at roughly the same accuracy. All tuned classifiers significantly outperformed a naïve classifier. PCA+SVM gave very similar results to SVM since the tuned PCA parameters gave best performance at 99.99% of explained variance, the greatest explained variance ratio we tested.

The prediction of sport showed promising preliminary results with tuned classifiers significantly outperforming a naïve classifier. Basketball players' movement patterns appeared to be quite distinct. Surprisingly, golf, for which the majority of the class were amateur athletes, was relatively identifiable. College and amateur athletes were left in the datasets, though these athletes have almost certainly not specialized as much as a professional would have. We speculate that using only the professional athletes for classifier training, or providing additional groupings for cross-trained and untrained athletes, may lead to more consistent results.

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 presented. Tuning of the more complex deep neural network architectures is on-going. Upon completion, we expect to be able to provide advice on the most promising classification architectures for 3D movement data.

ACKNOWLEDGEMENT

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.

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] G. Cook, L. Burton, B. Hoogenboom, and M. Voight, "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)™: 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.