PREDICTING 3D GROUND REACTION FORCE FROM 2D VIDEO VIA NEURAL NETWORKS IN SIDESTEPPING TASKS

Corey Morris[1], Marion Mundt[1], Molly Goldacre[1], Jason Weber[1], Ajmal Mian[2], Jacqueline Alderson[1,3]

Minderoo Tech & Policy Lab, The University of Western Australia, Perth, Australia[1]
School of Computer Science, The University of Western Australia, Perth, Australia[2]
Auckland University of Technology, Sports Performance Research Institute New Zealand (SPRINZ), Auckland, New Zealand[3]

Sports science practitioners often measure ground reaction forces (GRFs) to assess performance, rehabilitation and injury risk. However, recording of GRFs during dynamic tasks has historically been limited to lab settings. This work aims to use neural networks (NN) to predict three-dimensional (3D) GRF via pose estimation keypoints as inputs, determined from 2D video data. Two different NN were trained on a dataset containing 1474 samples from 14 participants and their prediction accuracy compared with ground truth force data. Results for both NN showed correlation coefficients ranging from 0.936 to 0.954 and normalised root mean square errors from 11.05% to 13.11% for anterior-posterior and vertical GRFs, with poorer results found in the medio-lateral direction. This study demonstrates the feasibility and utility of predicting GRFs from 2D video footage.

KEYWORDS: workloads, machine learning, biomechanics, athlete.

INTRODUCTION: An understanding of the external mechanical loads (GRFs) athletes are exposed to on the field has the potential to provide sport scientists with an improved understanding of contributors to injury and performance (Bahr & Krosshaug, 2005; Verheul, Gregson, Lisboa, Vanrenterghem & Robinson, 2019). Direct measurement of GRFs has historically been restricted to lab-based environments due to the requirement of ground-embedded force plates (Johnson, Mian, Donnelly, Lloyd & Alderson, 2018), limiting ecological validity, and neglecting external factors such as the influence of opponents or tactics. Recent efforts have witnessed the development of machine learning models to predict GRFs using wearable sensors such as inertial measurement units and 3D optical motion capture data (Johnson, Mian, Robinson, Verheul, Lloyd & Alderson, 2021; Mundt, David, Koeppe, Bamer, Markert & Potthast, 2019; Verheul et al., 2019). However, the collection of 3D motion capture data is also constrained to the laboratory, while wearable sensors are limited in application due to safety, logistical and regulatory restrictions – especially when considering their on-field use.
With the increasing availability of open-source human pose estimation models, automated video-based analysis of human motion outside the constraints of the lab is becoming a reality. Pose estimation models allow for the estimation of joint centres and other key landmarks in images. When used in conjunction with additional machine learning techniques, GRFs may be predicted solely from 2D video poses. The purpose of this research was to predict 3D GRF components of unanticipated sidestepping from 2D video. OpenPose (Cao, Hidalgo Martinez, Simon, Wei & Sheikh, 2021) was used to estimate anatomical keypoints which served as input into two different neural networks for GRF prediction. The performance of a dense neural network (DNN) and a recurrent long short-term memory neural network (LSTM) was compared.

METHODS: Fourteen semi-professional and professional female Australian Rules Football players (23 ± 3.9 years, 63.4 ± 5.4 kg) participated in this study (ethics approval RA/4/1/2593).

Each participant completed 21 unanticipated 45-degree sidestepping trials, consisting of both left and right foot strikes with an approach speed of 4.5 - 5.5 ms^{-1}, as measured via wireless timing gates (SmartSpeed Pro; Fusion Sports, QLD, Australia) placed three metres in front of a force plate (Advanced Mechanical Technology Inc., Watertown, MA, USA, 2000 Hz). A large projector screen indicated to participants which direction to cut (left or right) when participants were ~ 0.5 m from the force plate.
Three sagittal video cameras were positioned on the right-hand side of the participant (Sony HDR-CX700, 25 Hz, 1920x1080 pixels): true sagittal, one offset 3 m anterior to true sagittal, and one 3 m posterior to provide different vantage points for NN training. Trials where participants were deemed to target, have double foot contact, sidestep and/or stutter prior to force plate contact were excluded. Stance phase of each trial was defined from when vertical force exceeded a threshold of 20N. Force data was normalised to participant's body mass (N/Kg). Trials exceeding $\pm$ 3 standard deviations of the mean, applied to all GRF components, to exclude trials showing measurement errors. To enlarge the dataset, video data was augmented by cropping the frame in three different ways (centre 90% of frame dimension, 100 pixels from the left of frame, and 60 pixels from the top of frame). Videos were cropped to stance phase based on manually identified heel strike and toe-off. OpenPose (Cao et al., 2021) was used to determine 25 pose estimation keypoints (Figure 1). Trials which showed errors in identification of these keypoints were excluded.

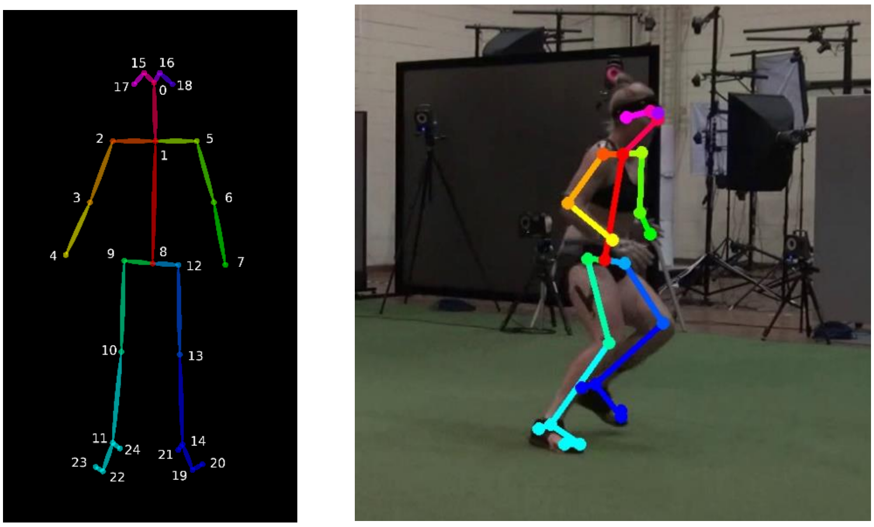

Figure 1: OpenPose 25 keypoint pose estimation outputs.

The final video dataset, based on the three camera views and augmented video data, comprised of 1474 samples. The total number of unique force trials was 125. Stance phases of OpenPose keypoint data and force data were time normalised (30 and 300 samples, respectively). Keypoint data were used as input features for both LSTM and DNN architectures, with output features being 3D force data. The data was randomly split, subject-wise, with 80% for training, 10% validation and the remaining 10% held over for testing. An automated hyperband search was performed (Koeppe, Hernandez Padilla, Voshage, Schleifenbaum & Markert, 2018) to determine optimal network architecture and hyperparameters in a three-fold cross-validation (Table 1). After determining optimal network parameters, a leave-one-subject-out validation was conducted to further assess model performance. This approach trains the NN on all subjects but one and tests the performance of the trained algorithm on the unseen subject. Metrics calculated to assess LSTM and DNN prediction accuracy were nRMSE and correlation coefficients (*r*) for all three force components (vertical, anterior-posterior, medio-lateral).

RESULTS: The mean *r* of predicted and ground-truth GRFs over the leave-one-subject-out validation for the DNN were 0.954, 0.946, and 0.576 for vertical, anterior-posterior and medio-lateral, forces, respectively. The LSTM predictions achieved correlations of 0.939, 0.936 and 0.128. The mean nRMSE for the DNN was 13.07%, 13.11% and 44.21% and for the LSTM 11.05%, 11.15% and 59.03% in vertical, anterior-posterior and medio-lateral direction, respectively. The comparison of nRMSE between the measured and predicted values for all GRF components of individual participants is displayed in Figure 2.

Table 1: Neural network hyperparameter search conditions and optimal results.

NN Architecture	Hyperband Search	DNN	LSTM
Dropout	0.1 - 0.8	0.7	0.7
Layer 1 size	8 - 650	128	550
Layer 2 size	8 - 650	550	300
Learning rate	3E-02 & 1E-02 - 3E-05 & 1E-05	1E-04	3E-04
No. epochs	11, 21, 31, 41, 51, 61, 71	61	21

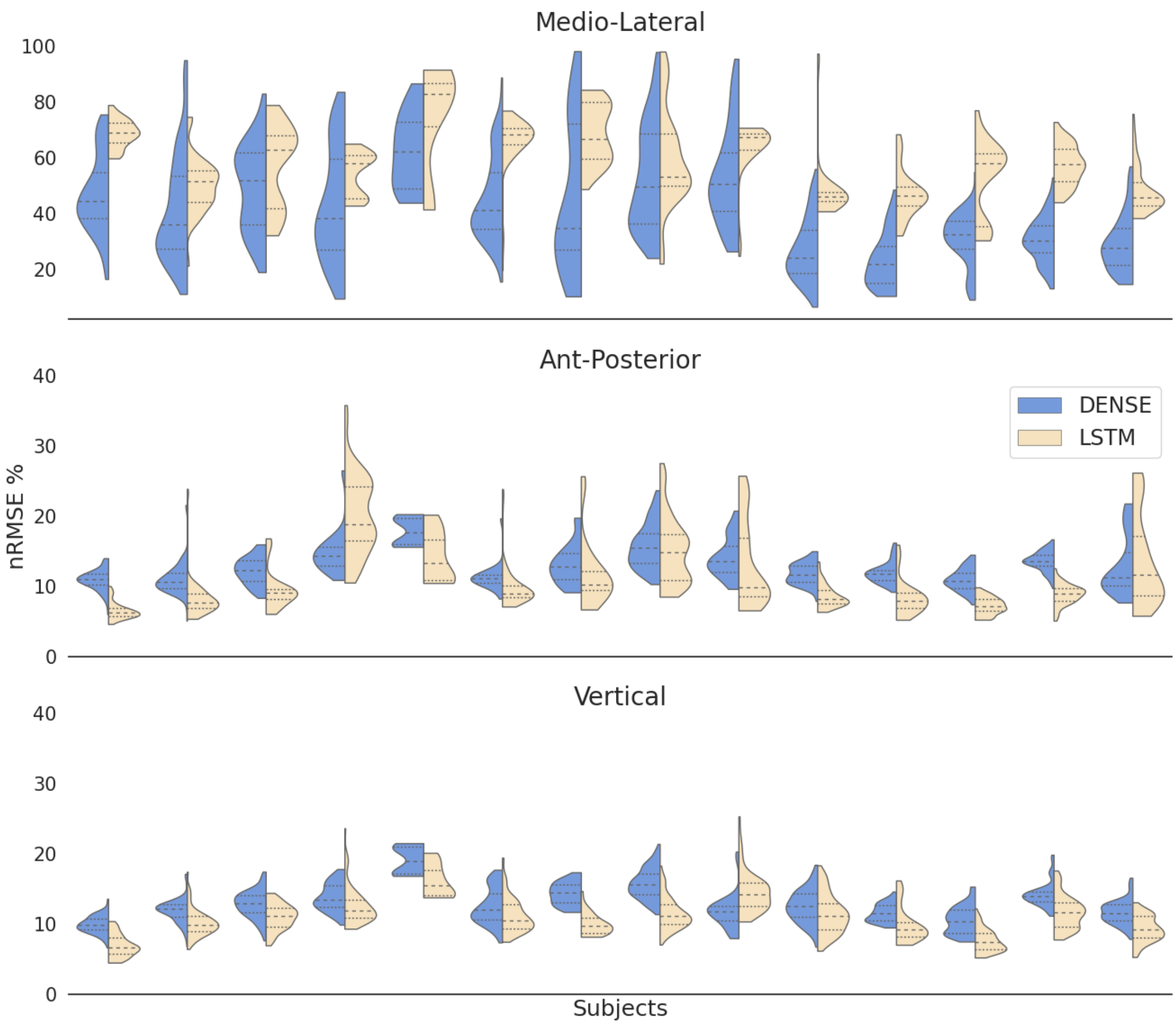

Figure 2: Violin plot of the nRMSE of predicted GRF against ground-truth GRF values for leave-one-subject-out validation comparing DNN and LSTM neural networks.

DISCUSSION: The aim of this study was to predict 3D GRF from 2D video during 45-degree sidestepping manoeuvres. Two different machine learning algorithms were compared based on accuracy of predicted and ground truth measured values. In contrast to DNN, LSTM networks do not require time-normalised inputs which facilitates real-time feedback. LSTM results may be further improved using non time-normalised trials (Mundt et al., 2020). Overall, both NNs achieved high mean agreement (r > 0.920) and low nRMSE of < 13.11% between predicted and ground-truth anterior-posterior and vertical force components. Although the DNN achieved higher agreement than LSTM predictions, correlation coefficients do not account for offsets between ground-truth and predicted values. A more sensitive indication of differences between these offsets is nRMSE, with LSTM predictions showing a lower nRMSE than DNN predictions when compared with ground-truth anterior-posterior and vertical force data. Unsurprisingly, the worst prediction for both NN was observed for the medio-lateral force.

In comparison with recent studies, Johnson et al. (2021) used a NN to predict GRF from wearables in running and sidestepping, achieving nRMSE values of 17.06% and 13.92% for anterior-posterior and vertical forces, respectively, and 21.56% for medio-lateral forces. The large difference (DNN 22.56%, LSTM 37.47%) in medio-lateral force prediction in that study compared with ours could be attributed to the loss of spatial information associated with going from 3D to 2D video – that is, a loss of depth information. This might be overcome by increasing the number of camera views, especially in the frontal plane. The results show a higher anterior-posterior and vertical force prediction accuracy compared with research using wearable sensors.

Although the results show high accuracy, differences between single subjects can be noted, which indicates that a larger dataset with higher variability will further improve the generalisability, especially in the medio-lateral force component prediction. The sample population in the present study was relatively homogenous, all being female team-sport athletes of similar age, with the testing environment constrained (e.g., running speed, volume). To improve the prediction accuracy and robustness of future NN models a wider range of athletes and testing conditions e.g., sex, age and ethnic diversity, varying approach speeds and cutting angles, different camera setups, should be included.

CONCLUSION: This research trained a NN to successfully predict vertical and anterior-posterior ground reaction forces during sidestepping, via pose estimation from 2D video data. These results build on current literature, which used wearable sensors or motion capture data, to remotely predict GRFs and has the potential to provide practitioners with a more relevant on-field tool for monitoring athlete GRFs, considered a surrogate measure of external load.

REFERENCES

Bahr, R., & Krosshaug, T. (2005). Understanding injury mechanisms: a key component of preventing injuries in sport. *British Journal of Sports Medicine, 39*(6), 324-329. doi:10.1136/bjsm.2005.018341.

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S.E., & Sheikh, Y.A. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1), 172-186, doi:10.1109/TPAMI.2019.2929257.

Johnson, W. R., Mian, A., Donnelly, C. J., Lloyd, D., & Alderson, J. A. (2018). Predicting athlete ground reaction forces and moments from motion capture. *Medical & Biological Engineering & Computing*, 56(10), 1781–1792, doi.org/10.1007/s11517-018-1802-7.

Johnson, W.R., Mian, A., Robinson, M.A., Verheul, J., Lloyd, D.G., & Alderson, J. A. (2021). Multidimensional ground reaction forces and moments from wearable sensor accelerations via deep learning. *IEEE Transactions on Biomedical Engineering*, 68(1), 289-297, doi:10.1109/TBME.2020.3006158.

Koeppe, A., Hernandez Padilla, C. A., Voshage, M., Schleifenbaum J.H. & Markert, B. (2018). Efficient numerical modeling of 3D-printed lattice-cell structures using neural networks. *Manufacturing Letters*, 15, 147-150, doi.org/10.1016/j.mfglet.2018.01.002.

Mundt, M., David, S., Koeppe, A., Bamer, F., Markert, B. & Potthast, W. (2019). Intelligent prediction of kinetic parameters during cutting manoeuvres. *Medical & Biological Engineering & Computing*, 57(8), 1833–1841, doi: 10.1007/s11517-019-02000-2.

Mundt, M., Koeppe, A., Bamer, F., David, S., & Markert, B. (2020). Artificial Neural Networks in Motion Analysis—Applications of Unsupervised and Heuristic Feature Selection Techniques. *Sensors*, *20*(16), 4581, doi:10.3390/s20164581.

Verheul, J., Gregson, W., Lisboa, P., Vanrenterghem, J., & Robinson, M.A. (2019). Whole-body biomechanical load in running-based sports: The validity of estimating ground reaction forces from segmental accelerations. *Journal of Science and Medicine in Sport*, 22(6), 716-22, doi:10.1016/j.jsams.2018.12.007.