FEATURE SELECTION FOR THE APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN MOTION ANALYSIS

Marion Mundt[1], Arnd Koeppe[1], Franz Bamer[1], Wolfgang Potthast[2], Bernd Markert[1]

Institute of General Mechanics, RWTH Aachen University, Aachen, Germany [1]
Institute of Biomechanics and Orthopaedics, German Sport University Cologne, Cologne, Germany [2]

The application of IMUs and artificial neural networks have shown their potential in estimating joint moments in various motion tasks. In this study, IMU data collected with five sensors during gait was used as input data to estimate hip, knee and ankle joint moments using artificial neural networks. Additionally, the original 30 features of the sensors' data were reduced to their ten most relevant principal components and also used as input to the neural networks to evaluate the influence of feature selection. The prediction accuracy of the networks was lower for the reduced dataset. Research with a larger dataset needs to be undertaken to further understand the influence of a reduced number of features on the prediction accuracy.

KEYWORDS: joint moments, IMUs, artificial intelligence, principle component analysis

INTRODUCTION: The biomechanical analysis of motion during daily life gains more and more relevance in our society. For this purpose, wearable inertial sensor based motion analysis excels due to its low cost, small size and easy applicability. However, there is no commercial inertial sensor based system for the analysis of joint moments available yet. In previous work, we have shown that it is possible to train artificial neural networks for this task (Mundt et al., 2019, 2020). Additionally, optical motion capture data can be used to estimate resulting joint moments (e.g. Ardestani, Zhang, et al., 2014; Johnson, Alderson, Lloyd, & Mian, 2019; Johnson, Mian, Lloyd, & Alderson, 2018). Different inputs and different types of neural networks such as fully-connected feedforward, long short-term memory or convolutional networks achieved good results.
Computational intelligence models tend to become large with an increasing number of input data. If there are irrelevant features present in the inputs, this might cause the model to become larger than if it could focus on the relevant features only. This might be particularly relevant for fully-connected networks where the time sequence needs to be unrolled. To decrease the size of the models, it might be favourable to decrease the number of inputs to the most relevant features. For this purpose, an exhaustive greedy algorithm has been used to find the most relevant features (Choi, Jung, Lee, Lee, & Mun, 2019; Joo, Oh, & Mun, 2016; Zhang, 2011), the correlation between different inputs has been determined to exclude highly correlated inputs (Aljaaf, Hussain, Fergus, Przybyla, & Barton, 2016), and the vector norm has been used to concentrate data (Ardestani, Chen, et al., 2014).
In this study, the number of inputs to the neural network is reduced by applying principle component analysis on the input data to investigate the influence of a reduced number of features on the prediction accuracy of fully-connected and long short-term memory (LSTM) neural networks. A reduced number of inputs might increase the generalisability of a model to new data.

METHODS: Thirty healthy subjects (12 female, 28.1±6.0 years, 72.3±12.7 kg, 1.77±0.07 m) participated in this study that was approved by the Ethical Committee of the German Sport University Cologne. All participants provided their informed written consent. Each subject performed ten level walking trials at five different speeds: 0.8 ms^{-1}, 1.1 ms^{-1}, 1.4 ms^{-1}, 1.7 ms^{-1} and 2.0 ms^{-1} ±10 % on a 5 m walkway. Each participant was equipped with 28 retro-reflective markers to capture the motion by twelve infrared cameras (125 Hz, VICONTM, MX F40, 233 Oxford, UK). Simultaneously, the participants were equipped with five sensors of a custom low cost IMU system (100 Hz, TinyCircuits, Akron, OH, USA). The marker set and sensor placement are displayed in Figure 1. The data of seven subjects was excluded from this study due to connectivity issues (Mundt et al., 2020a). The stance phases of each trial were extracted from the data using a threshold of 10 N measured by the force plates. After extraction, the stance phases were time-normalised to 101 time steps indicating 100% of the stance phase. The dataset contained 1751 samples of 23 participants.

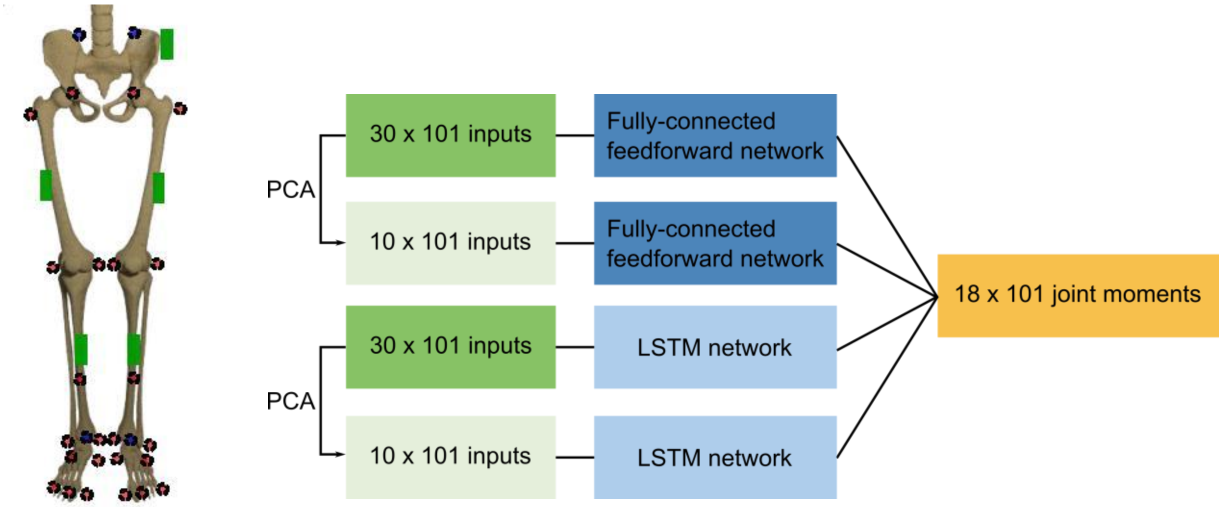

Figure 1 Placement of sensors (green boxes) and markers (circles) on the left and overview of the different inputs and networks trained on the right. Either the first ten principal components of the sensor data or all 30 features of the sensor data were used as inputs to train a fully-connected feedforward or LSTM neural network. The output of all networks are the 18 joint moments of the lower limbs.

To be able to compare the influence of the reduced input features on the estimation, two different sets of input parameters have been used. A principal component analysis was performed on the linear accelerations and angular rates of all five sensors. The first ten components explaining more that 95% of the variance were chosen as inputs to the neural network. Thereby, the input was decreased to one third of the original data. Additionally, the original dataset containing 30 inputs (5 sensors x (3 linear accelerations + 3 angular rates)) was used as input data. To evaluate whether the influence of the input features is dependent on the neural network type, a fully-connected feedforward neural network and a long short-term memory neural network were trained. The architecture and hyperparameters were optimised individually for the fully-connected and LSTM model. The final models' architectures and hyperparameters are displayed in Table 1.

Table 1 Overview of the models' architectures and hyperparameters

	Layers	Dropout	Epochs	Batch size	Learning rate
Fully-connected	600-200-400-600	30 %	70	4-8-16-32-64	10^{-5}
LSTM	300-300	40 %	100	4-8-16-32-64	10^{-5}

The neural network architectures were implemented using a Python framework (Koeppe, Bamer & Markert, 2019) supported by the Tensorflow library. All architectures were tested on their performance in estimating the joint moments of the stance leg based on different input data. For regularisation, dropout and early stopping were used. A five-fold cross validation was undertaken to find the optimum architecture and hyperparameters. For this purpose, a fixed test set (15%) was split off the data. The remaining data was split into training (70%) and validation (15%) for each split. The stance phases of the supporting leg were extracted from the data and the estimation of the resulting nine joint moments was analysed based on the RMSE normalised to the range of each joint moment (nRMSE) and the Pearson correlation coefficient.

RESULTS: The estimation revealed high accuracies for all nine joint moments independent of the inputs and neural networks used. The fully-connected network performed better than the LSTM network showing mean correlation coefficients of 0.909 (PCA) and 0.919 (no PCA) compared to 0.869 (PCA) and 0.886 (no PCA). In general, the error is distributed similarly for all networks and inputs. The joint moments of the ankle joint show the largest mean error and variance. The effect of the reduced input data is very small. For the fully-connected network the mean correlation coefficient is decreased by 0.010 and the nRMSE increased by 1.29 % when using the reduced input data. For the LSTM network the correlation coefficient is decreased by 0.022 and the nRMSE increased by 1.14 %. The mean nRMSE of the fully-connected network is 14.22 % for the full dataset and 15.52 % for the reduced dataset. For the LSTM network the full dataset resulted in a mean nRMSE of 15.12 % and 16.26 % for the reduced dataset. One example of the prediction by the different networks on one random sample of the test set is displayed in Figure 2.

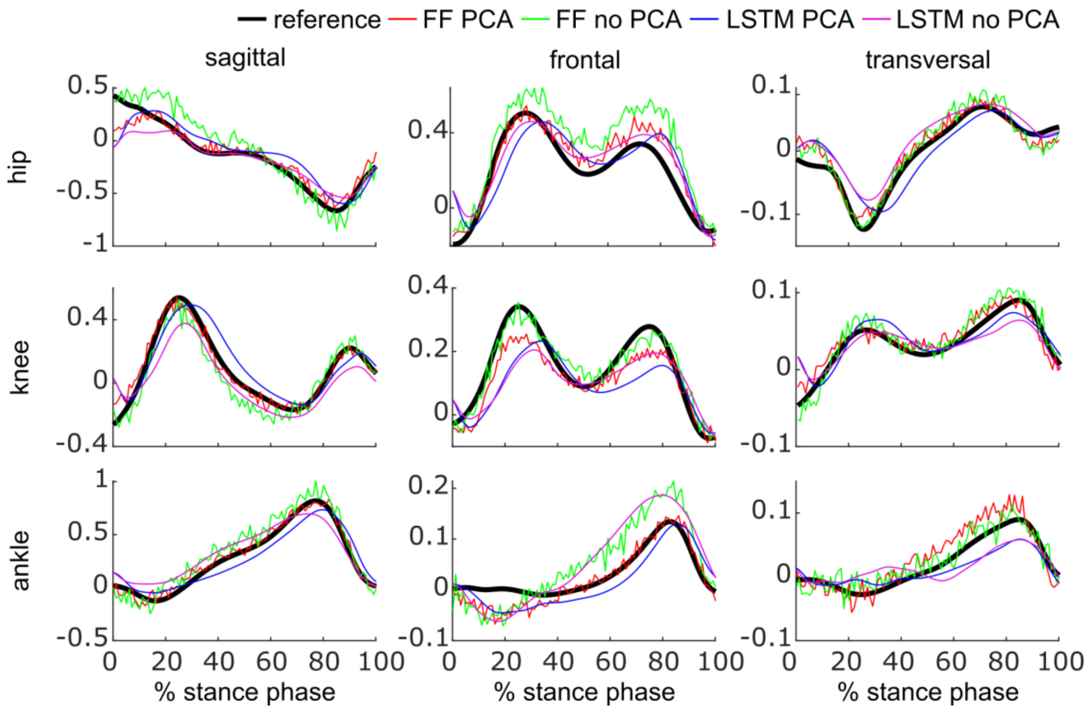

Figure 2 One example of the predicted and the reference joint moment. The fully-connected (FF) network's predictions are noisy because the model does not know about the time dependency of the data but just predicts each time step. The mean correlation coefficient of this example is higher than 0.93 and the nRMSE is smaller than 10 %.

DISCUSSION: The aim of this study was to analyse the influence of the application of PCA for feature selection on the input data to different neural networks. We hypothesised that the reduced number of inputs may help the model to generalise to new data and prevent it from overfitting. An analysis of the loss curves (not shown) of the validation set did not reveal differences with regard to the number of training phases before overfitting occurred for both datasets. However, the training dataset used in this study consisted of only about 1100 samples of 15 participants. The effect of the PCA might become more relevant using a larger dataset.

The correlation coefficient and the nRMSE indicate that neural networks are able to estimate the joint moments of new data for both the full and reduced input data. The LSTM network performed slightly worse than the fully-connected network, which might also be attributed to the small dataset (Um et al., 2017). The impact of the reduced dataset to both networks was very similar, although it is necessary to unroll the input data to apply a fully-connected network which results in a much larger network than when using an LSTM network. In future work, a model with less parameters should be trained on the reduced dataset for further evaluation. The fully-connected network is not able to learn from information covered in the time domain, which might lead to a higher redundancy in the data. The effect of redundancy needs further investigation. The use of PCA might support neural networks to deal with different sensor orientations during the experiment. This might be advantageous since especially the misalignment in orientation has shown to highly influence the accuracy of neural network predictions (Tan, Chiasson, Hu, & Shull, 2019).

CONCLUSION: This study showed that fully-connected and LSTM neural networks can be used to predict joint moments based on a reduced number of input features measured by inertial sensors. The use of PCA for feature selection had a slightly negative influence on the prediction accuracy. However, further exploration of the application of feature selection on the prediction accuracy is necessary to gain insight into the black box neural networks. For this purpose, research based on larger datasets should be undertaken.

REFERENCES:

Aljaaf, A. J., Hussain, A. J., Fergus, P., Przybyla, A., & Barton, G. J. (2016). Evaluation of machine learning methods to predict knee loading from the movement of body segments. In *Proceedings of the International Joint Conference on Neural Networks* (Vol. 2016-Octob, pp. 5168–5173). https://doi.org/10.1109/IJCNN.2016.7727882

Ardestani, M. M., Chen, Z., Wang, L., Lian, Q., Liu, Y., He, J., … Jin, Z. (2014). Feed forward artificial neural network to predict contact force at medial knee joint: Application to gait modification. *Neurocomputing*, *139*, 114–129. https://doi.org/10.1016/j.neucom.2014.02.054

Ardestani, M. M., Zhang, X., Wang, L., Lian, Q., Liu, Y., He, J., … Jin, Z. (2014). Human lower extremity joint moment prediction: A wavelet neural network approach. *Expert Systems with Applications*, *41*(9), 4422–4433. https://doi.org/10.1016/j.eswa.2013.11.003

Choi, A., Jung, H., Lee, K. Y., Lee, S., & Mun, J. H. (2019). Machine learning approach to predict center of pressure trajectories in a complete gait cycle: a feedforward neural network vs. LSTM network. *Med. Biol. Eng. Comput.* https://doi.org/10.1007/s11517-019-02056-0

Johnson, W. R., Alderson, J., Lloyd, D., & Mian, A. (2019). Predicting Athlete Ground Reaction Forces and Moments From Spatio-Temporal Driven CNN Models. *IEEE Transactions on Biomedical Engineering*, *66*(3), 689–694. https://doi.org/10.1109/TBME.2018.2854632

Johnson, W. R., Mian, A., Lloyd, D., & Alderson, J. (2018). On-field player workload exposure and knee injury risk monitoring via deep learning, 1–16.

Joo, S. Bin, Oh, S. E., & Mun, J. H. (2016). Improving the ground reaction force prediction accuracy using one-axis plantar pressure: Expansion of input variable for neural network. *Journal of Biomechanics*, *49*(14), 3153–3161. https://doi.org/10.1016/j.jbiomech.2016.07.029

Koeppe, A., Bamer, F., & Markert, B. (2019). An efficient Monte Carlo strategy for elasto-plastic structures based on recurrent neural networks. *Acta Mechanica*, *230*(9), 3279–3293. https://doi.org/10.1007/s00707-019-02436-5

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

Mundt, M., Thomsen, W., Witter, T., Koeppe, A., David, S., Bamer, F., … Markert, B. (2020). Prediction of lower limb joint angles and moments during gait using artificial neural networks. *Med. Biol. Eng. Comput*, *58*, 211–225. https://doi.org/10.1007/s11517-019-02061-3

Mundt, M., Koeppe, A., David, S., Witter, T., Bamer, F., Potthast, W. & Markert, B. (2020a). IMU-based estimation of joint kinematics and kinetics during gait. *Front. Bioeng. Biotechnol. - Biomechanics,* https://doi.org/10.3389/fbioe.2020.00041

Tan, T., Chiasson, D. P., Hu, H., & Shull, P. B. (2019). Influence of IMU position and orientation placement errors on ground reaction force estimation. *Journal of Biomechanics*, *97*(December), 109416. https://doi.org/10.1016/j.jbiomech.2019.109416

Um, T. T., Pfister, F. M. J., Pichler, D., Endo, S., Lang, M., Hirche, S., … Kulic, D. (2017). Data augmentation of wearable sensor data for Parkinson’s disease monitoring using convolutional neural networks. *ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 216–220. https://doi.org/10.1145/3136755.3136817

Zhang, T. (2011). Adaptive Forward-Backward Greedy Algorithm for Learning Sparse Representations. *IEEE Transactions on Information Theory*, *57*(7), 4689–4708.