A deep learning approach to predicting knee loading using 3D trajectories of anatomical landmarks

Melissa A. Boswell
Department of Bioengineering
Stanford University
boswellm@stanford.edu

Scott D. Uhlrich
Department of Mechanical Engineering
Stanford University
suhlrich@stanford.edu

Abstract - Knee osteoarthritis is a painful and debilitating disease that is accelerated by excessive loading in the joint. The knee adduction moment (KAM) is a surrogate measure of knee loading, however, measurement of KAM requires expensive motion capture equipment, forceplates, and trained engineers. Gait modifications that reduce the KAM can improve joint pain, but they require personalized prescription in the motion capture lab. Video-based extraction of 2D anatomical landmarks could be a scalable tool for estimating the effects of different gait modifications. In this study, 3D motion capture marker positions during walking with a variety of gait modifications were used to predict peak KAM per step. This data was collected for 98 people, giving a total of 125,415 steps. A fully connected neural network was able to predict the peak KAM on the test set of 8 subjects, on which the model did not train, with r^2 0.80. We have demonstrated the prediction of KAM from joint positions which can be extracted from 2D video analysis. This model will likely enable the personalization of gait modifications in any clinic with access to a 2D video camera.

I. INTRODUCTION

Knee osteoarthritis is a leading cause of years lost to disability worldwide (CDC, 2009). The medial compartment is affected 10 times more often than the lateral compartment (Ahlback, 1968), likely caused by the large medial-to-lateral loading ratio during gait (Schipplein and Andriacchi, 1991). The knee adduction moment (KAM) is commonly used as a surrogate measure of medial knee loading (Figure 1, Zhao et al., 2007), and is related to the presence (Hurwitz et al., 2002), severity (Sharma et al., 1998), and progression (Miyazaki et al., 2002) of medial compartment knee osteoarthritis.

Gait retraining that teaches people to change their walking kinematics such as increasing trunk sway, medial knee thrust, and changing the foot progression angle have been shown to reduce the KAM and reduce joint pain (Fregly et al., 2007; Guo et al., 2007; Mündermann et al., 2008; Richards et al., 2017; Shull et al., 2013a). In order to be effective, gait retraining must be personalized by assessing the effects of each modification on joint loading (Favre et al., 2016; Shull et al., 2015; Uhlrich, 2018). Joint loading is traditionally computed with Newtonian dynamics from force and acceleration data measured in an expensive motion capture lab. This requires force plates, motion capture cameras, and knowledge of how to use this technology. In order for this therapy to scale, a simpler personalization method is needed.

With recent improvements of pose estimation algorithms that use 2D video (Cao et al., 2016), the aim of our project is to predict joint loads using the positions of anatomical landmarks alone. More specifically, our goal is to predict the first peak KAM using time series 3D trajectories of anatomical landmarks measured from motion capture data without using force plate data or Newtonian mechanics.
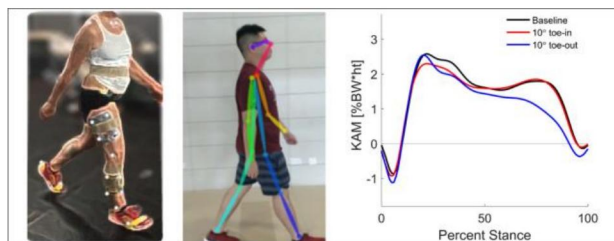
Figure 1: *Left*: Current means of calculating KAM requires motion capture and walking over force plates. *Center*: Joint position identification using OpenPose (Chiu, 2018). *Right*: Gait modifications such as changing the foot angle reduce the knee adduction moment peaks.

II. DATASET

We have motion capture data from 98 individuals (78 osteoarthritic, 20 healthy) walking at a self-selected speed with a variety of different gait modifications which alter the KAM (Figure 1). Using ground reaction forces from the force-instrumented treadmill and anatomic marker locations from optical motion capture, we compute the ground truth first peak of the KAM using Newtonian dynamics and normalize it to percent bodyweight x height (%BW*ht).

The input to our model is the 3D position, velocity, and acceleration timeseries (8 points) of 13 anatomical landmarks per step. Markers placed on the neck and bilaterally on the rear pelvis, front pelvis, lateral knee, lateral ankle, heel, were reported relative to the center of the rear pelvis and normalized by height. Since we are predicting the first KAM peak which, by definition, occurs during the first half of the stance phase, we only use input marker data from the first half of the stance phase. This clipped dataset originally comprised of 15 timesteps, but due to the frequency content of the timeseries input, we downsampled by a factor of 2 to end with 8 timesteps. We finally add a binary indication of which leg is taking the step to account for potential systematic bias of errors in the right and left forceplates. The input shape is (118 features x 8 timesteps x 125,415 steps. For the bilateral markers, the positions of the leg on the ground appear in the same position of the input matrix, regardless of the leg taking the step. The output is peak KAM 1x125,415. We split the data by subject with 80 subjects in the test set, 10 subjects in the dev set, and 8 subjects in the test set. The data was split by person so the model will not have seen the people in the test set during training. We also added held out 3,000 steps from the training set to the train-dev set to test how well the model performs predicting KAM for people it has trained on, but steps it has not.

III. METHODS

Model Architecture

We tested three different neural network (NN) architectures: a 1D Convolutional Neural Network (CNN), a bidirectional long short-term memory (LSTM) recurrent neural network, and a fully connected (FC) NN. All weights were initialized with Xavier initialization (Glorot et al., 2010) and ended with a single fully connected neuron with linear activation. Adam gradient descent (Kingma et al., 2014) was used to minimize our mean-squared-error loss function (Equation 1). Root mean square error and correlation coefficient (r^2) were used to assess model performance.

$$L = \sum_{i=1}^{m}(y - \hat{y})^2$$

Equation 1

Our CNN, motivated by Wang et al. (2017), included 4 convolutional layers with 20, 40, 60, and 80 channels, respectively. The full timeseries (30 timepoints) was used for this model. The 1D kernel was 5 timesteps wide, 118 features tall, and moved in the time dimension. Relu activation functions were used and batch normalization was performed between convolutional layers. Three fully connected layers followed with 15 neurons (Relu activation), 70 neurons (Relu activation), and 1 neuron (linear activation).

The LSTM model, motivated by Kidziński (2018), comprised of two bidirectional LSTM layers with 32 units with tanh activation. A single FC neuron followed with linear activation. The 8-timestep input was used for the LSTM model.

Our FC model after hyperparameter tuning (Figure 2) has 800 neurons in the first hidden layer with Relu activation followed by an output neuron with linear activation. Dropout with probability of 1% chance of zeroing is performed between layers.
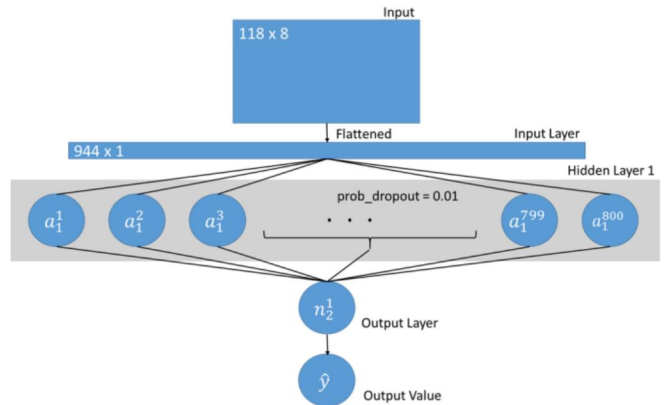

Figure 2: Diagram of best model (2-layer fully connected neural network)

Following a course hyperparameter search for all three models, we selected the model with the lowest MSE on the development set. The FC network performed best (Table 1), so a more systematic hyperparameter search was performed on this model.

Hyperparameter Search

Our initial FC model displayed low bias, but high variance, so the hyperparameter tuning was targeted at reducing variance. We tuned the model complexity (number of neurons per layer and number of hidden layers), probability of dropout, L_2 regularization. L_2 regularization did not improve model performance.

We randomly sampled parameters related to model complexity and dropout probability. After sampling over a broader range of parameters, the results of a more focussed search are displayed in Figure 3. This search led us to a network with only 1 hidden layer, a large number of neurons in the first layer, and a small probability of dropout (0.01).
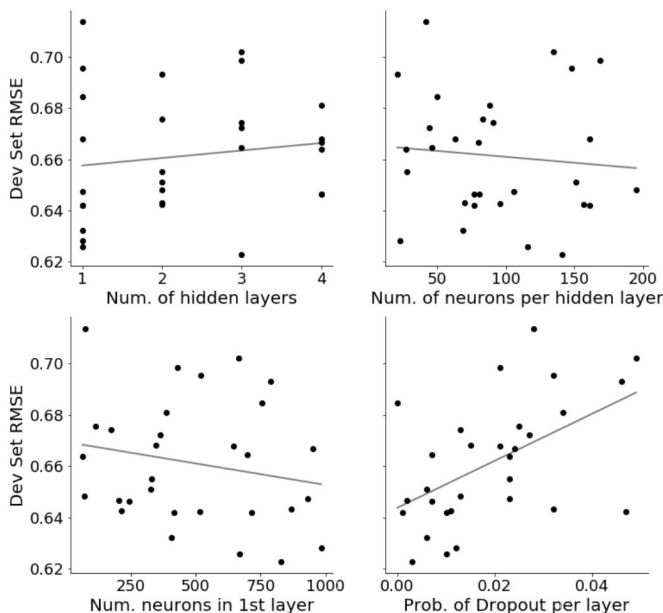
Figure 3: Results of the focussed hyperparameter search to determine model size and dropout hyperparameters. A small number of hidden layers with a large number of neurons in the first layer and a small dropout probability minimized RMSE on the development set.

Dataset Variation Experiments

To better understand the important input features, we varied the dataset inputs in two ways and computed a saliency map. We first removed marker accelerations, velocities, and both from the dataset to the performance of a limited dataset. Next, since the ultimate application of our network would input 2D joint position projections, we simulated side-view and front-view camera position by removing the third, depth dimension from our 3D motion capture data. For example, for a simulated front-view camera, we removed the front-back dimension and retained the side-side and up-down positions. Finally, we computed a saliency map (Kotikalapudi, 2017) for the position-only data normalized to have unit variance.

IV. RESULTS

The performance of the CNN, LSTM, and FC architectures are shown in Table 1. The FC model performed the best, with a test r^2 value and RMSE of 0.80 and 0.57 %BW*ht, respectively. This model performed equally well on the training set and train-dev sets (r^2=0.97, Figure 4). The model performs worse on the development and test sets (r^2=0.88 and $r^2$0.80). Despite the reduced overall performance on the test set, for many subjects (color coded), the slope of their true vs. predicted line is close to one. This indicates that despite a systematic offset from the true value, our model can accurately predict changes in peak KAM which is important for our application.

Table 1: The performance of different models.

Model	Details	Train r^2	Dev r^2	Test r^2	RMSE
CNN	3 Conv1D layers, 3 FC layers, 40 epochs	0.94	0.61	0.59	0.81
LSTM	2 LSTM layers, 1 FC layer, 40 epochs	0.84	0.54	0.56	0.81
FC	2 FC layers, 40 epochs (positions, velocities, accelerations)	0.97	0.88	0.80	0.57
	Front-view only (positions, velocities, accelerations)	0.95	0.87	0.78	0.60
	Side-view only (positions, velocities, accelerations)	0.92	0.56	0.44	0.97
	Positions only	0.95	0.83	0.76	0.63
	Positions, velocities only	0.96	0.88	0.81	0.56
	Positions, accelerations only	0.96	0.86	0.77	0.62

When only using the marker positions, performance only degraded slightly (test r^2=0.76) compared to the model that used positions, velocities, and accelerations (r^2=0.80, Table 1). Using only the positions and velocities or positions and accelerations yielded slight improvements compared to positions alone (r^2=0.81 and r^2=0.77, respectively). The front-view only model performed similarly (test r^2=0.78) to the full-feature model, but the side-view only model performed poorly (test r^2=0.44).
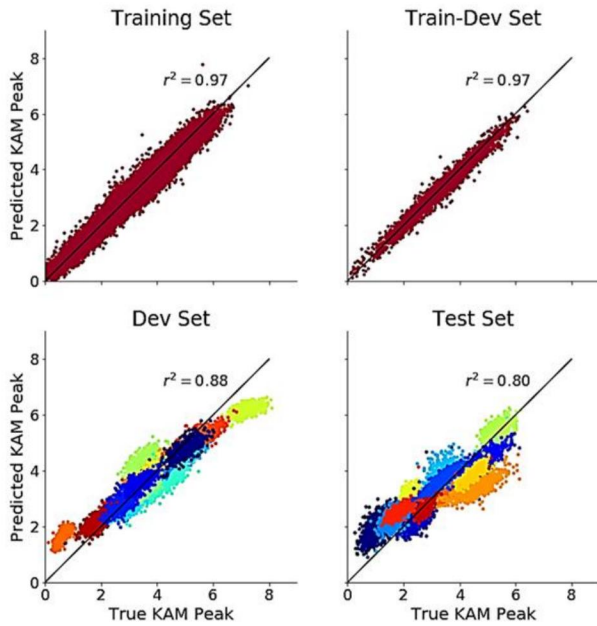
Figure 4: Predicted KAM vs. True KAM for all portions of the dataset and the y=x line.

The saliency map of the marker positions shows that changes in knee, heel, and neck side-side and pelvis up-down positions cause large changes in peak KAM (Figure 5). In general, Side-side and up-down positions are more important in the saliency map than front-back positions which corresponds with the findings of the front and side-view only models.

V. DISCUSSION AND FUTURE WORK

A 2-layer FC network with position, velocity, and acceleration inputs predicted the training set peak KAM with high accuracy (r^2=0.97) indicating that a relatively shallow NN is sufficient to solve this structured problem. The model performed with equivalent accuracy on the train-dev set indicating that if the model trained on steps from a person, it will accurately predict a new step. The development and test set performance was lower (r^2=0.88 and r^2=0.80, respectively) indicating that our model overfit to the training set. Given the relative performance of the train-dev set and test sets compared to the training set, our variance problem seems to arise from person-to-person variability instead of step-to-step variability. As an obvious example, there is a person in the development set with true peak KAM values between 6.5 and 8%BW*ht, but the training set did not train on any KAM values above 6.5%BW*ht (Figure 4). We attempted to solve this overfitting problem with L_2 regularization and dropout; both provided minimal benefits suggesting that more diverse training data is necessary for our model to generalize well to new people.

The FC network performed better than the 1D CNN or LSTM models. The data from the first half of the stance phase is most important for computing the first KAM peak, so we were able to cut off and downsample our original timeseries to only 8 timesteps for the FC model. This downsampling and cutting reduced the input data size (118 features * 8 timesteps) which proved to be small enough for a FC network with a reasonable number of parameters. Thus, the benefits of the CNN and LSTM of learning prior and future timesteps with a reduced number of parameters was less beneficial for our application. If we were predicting the entire timeseries output from a higher
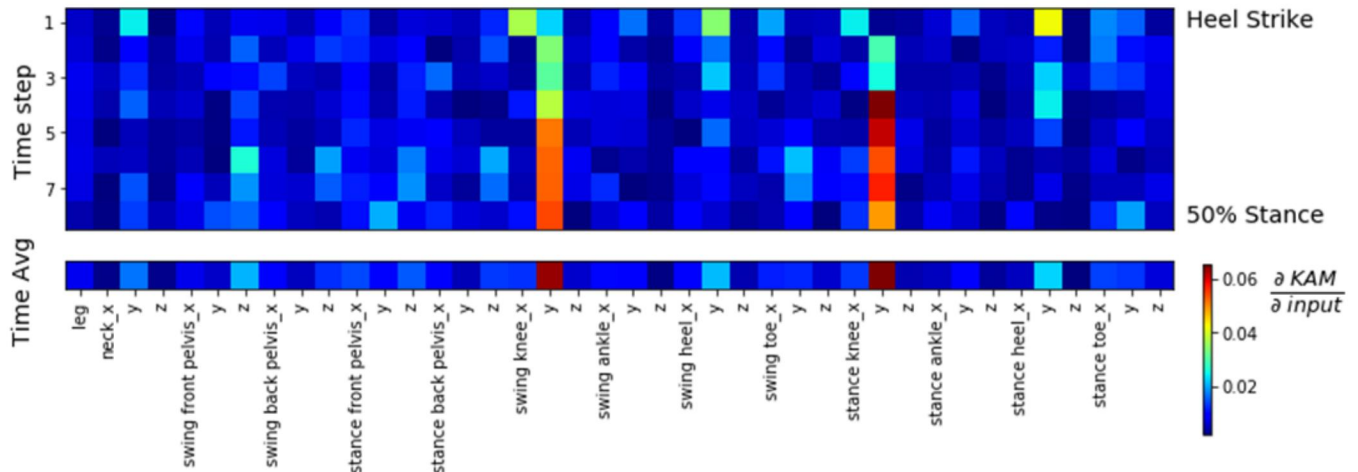
Figure 5: A saliency map shows which input changes cause the greatest change in output (peak KAM). Darker colors signify more important input features. The top map shows the saliency of all 118 input features, and the bottom map shows time-averaged saliency for the front-back (x), side-side (y), and up-down (z) marker positions. The side-side knee, heel, and neck positions as well as the up-down position of the swing leg front pelvis heavily influence output predictions. These positions align closely with important variables in the dynamic equations of motion for the KAM.

frequency, longer duration activity, the LSTM and CNN may have performed better than the FC network. The well-structured nature of our problem likely allowed for a shallow FC network to perform well. The input data were normalized by leg length and output data was normalized by body mass and leg length making the input and output data agnostic to body shape. Additionally, marker positions, velocities, and accelerations are related to KAM with few nonlinearities in the dynamic equations of motion, which likely allowed a network with only one hidden layer and nonlinear activation function to perform well.

Our model learned salient features that are similar to terms in the dynamic equations of motion. The KAM is the frontal plane moment about the knee joint (Figure 6), so the frontal plane (side-side and up-down) variables drive the equations of motion. The front-view only model performed with $r^2=0.78$ on the test set which was nearly equivalent 3D input model ($r^2=0.80$). The side-view model (test $r^2=0.44$) performed worse than the front-view model because the side-view input is devoid of the important side-side information. Additionally, the saliency map elucidated that changes in the side-side positions of the knees, heels, and neck led to large changes in the KAM peak which align with previous findings. "Toe-in gait" (Shull et al., 2013), which involves moving the heel outward, "medial knee thrust gait" (Fregly et al., 2007), which involves moving the knee towards the centerline of the body, and "trunk sway gait" (Mündermann et al., 2008), which involves side-to-side motion of the trunk are all previously identified KAM-reducing gait modifications. Finally, the up-down position of the swing leg frontal pelvis was an important predictor of changes in KAM in our model. This aligns with a common qualitative clinical assessment where clinicians watch for the swing-side pelvis to drop during gait as an indication of a large KAM. The saliency map corroborated previous observations on how to change the KAM, but it could also be used to identify new gait modifications.
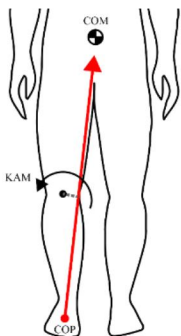
Figure 6: KAM can be estimated using Newtonian dynamics as the frontal plane moment of the ground reaction force (red) about the knee joint. The KAM is affected by the side-side position of the heel (center of pressure, COP) and knee as well as the side-side position of the center-of-mass (COM) which is in part driven by the side-side position of the neck and up-down position of the swing-side pelvis.

This study had several limitations. Our final model overfit to the training set, causing reduced test set accuracy. Data augmentation and adding more people to the training set may help this problem. Second, this study is motivated by the use of a 2D camera to identify joint positions and the KAM, but the inputs to our model were the high-fidelity marker positions from a motion capture system. Our model demonstrates that the KAM can be predicted without forceplates, but it remains unclear if video-based joint position recognition will provide high enough quality data to accurately predict the KAM. We have recently installed 2D cameras in our gait lab and plan to utilize the pre-trained weights from the present study to expedite training for a model that inputs 2D video data. Finally, our outcome metric, mean squared error, prioritized correctly predicting KAM magnitude for each step; however, intra-person change in KAM peak is the most clinically relevant outcome. We plan to train a new model that predicts the distance that each step lies from each person's mean KAM

In conclusion, a fully-connected neural network accurately predicted the peak KAM from the trajectories of anatomical landmarks. This study brings the field one step closer to performing personalized biomechanical assessments outside of a motion capture lab and in any clinic with access to a 2D camera. By making gait retraining available to the general population, we hope to reduce the debilitating effects on knee osteoarthritis on mobility worldwide.

CONTRIBUTIONS

Both authors pre-processed the data, built the initial models, and performed course hyperparameter tuning on the initial models. SDU fine-tuned the final model and performed model testing while MAB made the poster and github repository. Both authors wrote the paper.

GITHUB REPOSITORY

https://github.com/melboswell/predicting-KAM

REFERENCES

1. Ahlback S. 1968. Osteoarthrosis of the knee. A radiographic investigation. Acta Radiol. 277:70–72.
2. Centers for Disease Control and Prevention, USA. 2009. Prevalence and most common causes of disability among adults - United States. MMWR Morb. Mortal. Wkly. Rep. 58, 421–426.
3. Cao, Z., Simon, T., Wei, S., Sheikh, Y., 2016. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. CoRR. http://arxiv.org/abs/1611.08050.
4. Chiu, M., Shu, J., Hui, P., 2018. Emotion Recognition through Gait on Mobile Devices. IEEE EmotionAware'18 - 2nd International Workshop on emotion awareness for pervasive computing with mobile and wearable devices.
5. Favre, J., Erhart-Hledik, J.C., Chehab, E.F., Andriacchi, T.P., 2016. General scheme to reduce the knee adduction moment by modifying a combination of gait variables. J. Orthop. Res. 34, 1547–1556.
6. Fregly, B.J., Reinbolt, J.A., Chmielewski, T.L., 2008. Evaluation of a patient-specific cost function to predict the influence of foot path on the knee adduction torque during gait. Comput. Methods Biomech. Biomed. Engin. 11, 63–71.
7. Glorot, X., Bengio, Y. Understanding the difficulty of training deep feedforward neural networks, 2010. PMLR. 9: 249:256.
8. Guo, M., Axe, M.J., Manal, K., 2007. The influence of foot progression angle on the knee adduction moment during walking and stair climbing in pain free individuals with knee osteoarthritis. Gait Posture 26, 436–441.
9. Hurwitz, D.E., Ryals, A.B., Case, J.P., Block, J.A., Andriacchi, T.P., 2002. The knee adduction moment during gait in subjects with knee osteoarthritis is more closely correlated with static alignment than radiographic disease severity, toe out angle and pain. J. Orthop. Res. 20, 101–107.
10. Kidziński, L., 2018. GitHub repository. https://github.com/kidzik/event-detector-train.
11. Kingma, D.P., Ba, J.L. 2014. Adam: A Method for Stochastic Optimization. ICLR.
12. Kotikalapudi, R., et al., 2017. Keras-vis. GitHub repository. https://github.com/raghakot/keras-vis.
13. Miyazaki, T., Wada, M., Kawahara, H., Sato, M., Baba, H., Shimada, S., 2002. Dynamic load at baseline can predict radiographic disease progression in medial compartment knee osteoarthritis. Ann. Rheum. Dis. 61, 617–622.
14. Mündermann, A., Asay, J.L., Mündermann, L., Andriacchi, T.P., 2008. Implications of increased medio-lateral trunk sway for ambulatory mechanics. J. Biomech. 41, 165–170.
15. Richards, R., van den Noort, J.C., Dekker, J., Harlaar, J., 2017. Gait retraining with real-time biofeedback to reduce knee adduction moment: systematic review of effects and methods used. Arch. Phys. Med. Rehabil. 98, 137–150.
16. Schipplein OD, Andriacchi T. 1991. Interaction between active and passive knee stabilizers during level walking. J Orthop Res 9:113–119.
17. Shull, P.B., Shultz, R., Silder, A., Dragoo, J.L., Besier, T.F., Cutkosky, M.R., Delp, S.L., 2013a. Toe-in gait reduces the first peak knee adduction moment in patients with medial compartment knee osteoarthritis. J. Biomech. 46, 122–128.
18. Shull, P.B., Huang, Y., Schlotman, T., Reinbolt, J.A., 2015. Muscle force modification strategies are not consistent for gait retraining to reduce the knee adduction moment in individuals with knee osteoarthritis. J. Biomech. 48, 3163–3169.
19. Sharma, L., Hurwitz, D.E., Thonar, E.J., Sum, J.A., Lenz, M.E., Dunlop, D.D., Schnitzer, T.J., Kirwan-Mellis, G., Andriacchi, T.P., 1998. Knee adduction moment, serum hyaluronan level, and disease severity in medial tibiofemoral osteoarthritis. Arthritis Rheum. 41, 1233–1240.
20. Uhlrich S, Silder A, Beaupre G, Shull, P, Delp, S., 2018. Subject-specific toe-in or toe-out gait modifications reduce the larger knee adduction moment peak more than anon-personalized approach. Journal of Biomechanics 66:103–110.
21. Wang Z., Yan W., Oates T., 2017. Time series classification from scratch with deep neural networks: A strong baseline. *Proc of* IEEE.
22. Zhao D, Banks SA, Mitchell KH, et al. 2007. Correlation between the knee adduction torque and medial contact force for a variety of gait patterns. J Orthop Res 25:789–797.