Osteoarthritis and Cartilage 29 (2021) 346–356

Osteoarthritis and Cartilage

A neural network to predict the knee adduction moment in patients with osteoarthritis using anatomical landmarks obtainable from 2D video analysis

M.A. Boswell † * [a], S.D. Uhlrich ‡ # ** [a], Ł. Kidziński †, K. Thomas §, J.A. Kolesar † #, G.E. Gold ||, G.S. Beaupre † #, S.L. Delp † ‡ ¶

† *Department of Bioengineering, Stanford University, Stanford, CA, USA*
‡ *Department of Mechanical Engineering, Stanford University, Stanford, CA, USA*
§ *Department of Biomedical Data Science, Stanford University, Stanford, CA, USA*
|| *Department of Radiology, Stanford University, Stanford, CA, USA*
¶ *Department of Orthopaedic Surgery, Stanford University, Stanford, CA, USA*
# *Musculoskeletal Research Lab, VA Palo Alto Healthcare System, Palo Alto, CA, USA*

ARTICLE INFO

Article history:
Received 20 April 2020
Accepted 28 December 2020

Keywords:
Knee adduction moment
Osteoarthritis
Gait
Machine learning
Neural network
Video motion analysis

SUMMARY

Objective: The knee adduction moment (KAM) can inform treatment of medial knee osteoarthritis; however, measuring the KAM requires an expensive gait analysis laboratory. We evaluated the feasibility of predicting the peak KAM during natural and modified walking patterns using the positions of anatomical landmarks that could be identified from video analysis.

Method: Using inverse dynamics, we calculated the KAM for 86 individuals (64 with knee osteoarthritis, 22 without) walking naturally and with foot progression angle modifications. We trained a neural network to predict the peak KAM using the 3-dimensional positions of 13 anatomical landmarks measured with motion capture (*3D neural network*). We also trained models to predict the peak KAM using 2-dimensional subsets of the dataset to simulate 2-dimensional video analysis (*frontal* and *sagittal plane neural networks*). Model performance was evaluated on a held-out, 8-person test set that included steps from all trials.

Results: The *3D neural network* predicted the peak KAM for all test steps with r^2(Murray et al., 2012) 2 = 0.78. This model predicted individuals' average peak KAM during natural walking with r^2(Murray et al., 2012) 2 = 0.86 and classified which 15° foot progression angle modifications reduced the peak KAM with accuracy = 0.85. The *frontal plane neural network* predicted peak KAM with similar accuracy (r^2(Murray et al., 2012) 2 = 0.85) to the *3D neural network*, but the *sagittal plane neural network* did not (r^2(Murray et al., 2012) 2 = 0.14).

Conclusion: Using the positions of anatomical landmarks from motion capture, a neural network accurately predicted the peak KAM during natural and modified walking. This study demonstrates the feasibility of measuring the peak KAM using positions obtainable from 2D video analysis.

Introduction

Knee osteoarthritis (OA) affects approximately 14 million people in the United States[1] and is a leading cause of disability worldwide[2]. The medial compartment of the knee is most commonly affected[3], likely because it bears a greater proportion of total knee contact force compared to the lateral compartment[4]. Joint loading cannot be directly measured *in vivo* in an osteoarthritic knee, so the peak knee adduction moment (KAM) is a common surrogate measure of medial knee loading due to its relationship to the medio-lateral

* Address correspondence and reprint requests to: M.A. Boswell, Department of Bioengineering, Stanford Human Performance Lab, 341 Galvez St, Lower Level, Stanford, CA, 94305, USA. Tel.: 1-650-721-2547.
** Address correspondence and reprint requests to: S.D. Uhlrich, Department of Mechanical Engineering, Stanford Human Performance Lab, 341 Galvez St, Lower Level, Stanford, CA, 94305, USA.
E-mail addresses: boswellm@stanford.edu (M.A. Boswell), suhlrich@stanford.edu (S.D. Uhlrich), lukasz.kidzinski@stanford.edu (Ł. Kidziński), kathoma@stanford.edu (K. Thomas), julie14@stanford.edu (J.A. Kolesar), gold@stanford.edu (G.E. Gold), garybeaupre@gmail.com (G.S. Beaupre), delp@stanford.edu (S.L. Delp).
[a] These authors contributed equally to this work.

https://doi.org/10.1016/j.joca.2020.12.017
1063-4584/

distribution of force in the knee[5]. The peak KAM, which typically occurs during the first half of the stance phase, is correlated with the presence[6], severity[7,8], and progression[9] of medial compartment knee osteoarthritis.

Measurements of the peak KAM enhance clinical decision making. Measuring the peak KAM during natural walking aids in diagnosing OA[10–12], evaluating a patient's risk of OA progression[9], and predicting surgical outcomes[13]. Gait modifications such as increasing trunk lean[14,15], avoiding contralateral pelvic drop[16], medializing the knee (medial thrust)[17,18], and changing the foot progression angle[19,20] can reduce the peak KAM and pain. However, personalizing interventions based on how they affect each individual's peak KAM is critical for maximizing the achievable reduction in loading and avoiding interventions that cause a harmful increase in loading[18,21–23]. Personalized foot progression angle modifications (toe-in or toe-out), in particular, are an effective and subtle way to reduce the peak KAM[23].

The KAM is calculated using inverse dynamics from ground reaction forces and kinematics measured in a gait analysis laboratory equipped with force plates and a motion capture system [Fig. 1(A)]. The expensive equipment and technical expertise necessary to operate a gait laboratory are inaccessible to most clinicians, which excludes potentially valuable KAM measurements from routine clinical practice. The peak KAM has been calculated using mobile sensors (inertial measurement units and force-instrumented shoes) and an ID approach[24]; however, this method still requires expensive equipment, limiting its scalability. A simpler, cheaper way to assess the peak KAM is needed.

The ability to predict the peak KAM with 2-dimensional (2D) video analysis would make these measurements available to clinicians, researchers, and patients using only a video camera. Recently-developed automated keypoint detection algorithms such as OpenPose[25] extract the positions of anatomical landmarks (e.g., joint locations) from 2D video [Fig. 1(C)]. But the small number of keypoints on each body segment and the imperfect anatomical accuracy of keypoints (ranging from 14 to 30 mm)[26,27] limit their utility as a replacement for motion capture in an ID analysis. Machine learning models are a promising solution for predicting complex biomechanical outputs using low-fidelity inputs like video-based keypoints[28,29].

Machine learning models have been trained to predict the KAM, but many of these models rely on input features from equipment that is not readily available in clinical settings, such as force plates, inertial measurement systems, pressure-sensing insoles, and electromyography systems[30–33]. Other models have used kinematic features such as synthesized images of motion capture trajectories[34] and Euler angles[35] to predict the KAM curve in healthy individuals and individuals with alkaptonuria. These models demonstrate the promise of using kinematics alone to predict the KAM curve; however, the ability to predict the peak KAM in individuals with knee OA using kinematic features identifiable from video analysis remains unknown.

The objective of our study is to evaluate the feasibility of predicting the peak KAM in individuals with and without knee OA using keypoints that could be extracted from video. In the absence of a dataset with synchronized 2D video and joint kinetics, we

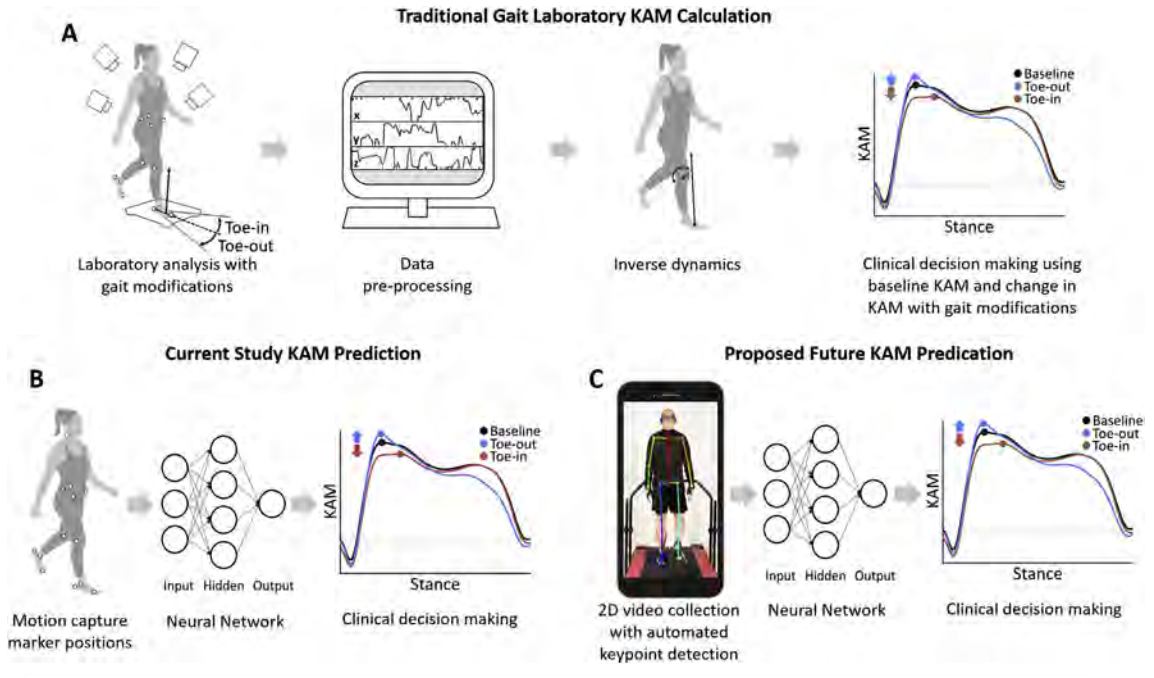

Fig. 1

A) The gold standard, laboratory-based workflow for measuring the knee adduction moment (KAM). After motion capture data is collected and semi-manually pre-processed, it is combined with force plate data to compute the KAM using inverse dynamics. **B**) The workflow for the current study. We use the coordinates of 13 anatomical landmarks from motion capture (to simulate video keypoints) as inputs into a neural network trained to predict the peak KAM. **C**) Our proposed future workflow for the automated measurement of the KAM. After collecting 2D video of gait, keypoints (e.g., joint positions) could be detected automatically using OpenPose25. A neural network would predict the peak KAM using these keypoints as input.

Osteoarthritis and Cartilage

simulate video-based keypoints using the positions of anatomical landmarks measured with motion capture from a previously-collected dataset. Furthermore, we simulate keypoints from 2D video by projecting 3-dimensional (3D) anatomical landmark positions onto a 2D plane. Anatomical keypoint detection is a rapidly maturing field, so evaluating a machine learning model's ability to predict the peak KAM using motion-capture-based keypoints will demonstrate the potential for a video-based solution in the future.

We aimed to evaluate the feasibility of predicting the first KAM peak using machine learning models that use only the positions of anatomical landmarks as inputs. We evaluated model performance on two clinical decision-making tasks. The first task was predicting an individual's first peak KAM during natural walking with a mean absolute error (MAE) less than 0.5% bodyweight*height (BW*H); this threshold was chosen because it is the minimum of the clinically-relevant range of 0.5–2.2%BW*H related to diagnosing OA and evaluating the risk of progression[9–12]. The second task was classifying whether a foot progression angle modification increases or decreases the first peak KAM. To evaluate the feasibility of a video-based solution that uses a single camera, we compared the performance of models that use only sagittal or frontal plane anatomical landmark positions to the model that uses 3D positions.

Methods

Data collection

Eighty-six individuals (64 with medial knee OA and 22 without OA[23]) participated in this study after providing written consent in compliance with the Stanford University Institutional Review Board (Table I). We included individuals with and without knee OA to establish a larger dataset and to train a model that is generalizable to a broad spectrum of OA severities. While assessing the KAM in individuals without OA is not currently a part of clinical practice, an inexpensive screening tool, like the pipeline proposed here [Fig. 1(C)], could help identify asymptomatic individuals with a large KAM who may be at risk of developing medial knee OA[12]. Inclusion criteria for the medial knee OA group were (1) Kellgren–Lawrence grade 1–3 and smaller medial compared to lateral joint space width assessed from an anterior-posterior weight-bearing radiograph by a radiologist (GEG); (2) medial pain of three or higher out of 10 on the numeric rating scale and medial pain greater than patellofemoral or lateral pain; (3) the ability to walk safely for 25 min on a treadmill without ambulatory aids; and (4) body mass index (BMI) less than 35. Individuals without osteoarthritis were included if they had no lower-extremity pain or history of lower-extremity injury in the past year.

Characteristic	Without OA	With OA
Number of subjects	22	64
Gender	9F, 13M	43F, 21M
Age (years)	24.7 ± 3.2	64.6 ± 8.81
Height (m)	1.75 ± 0.13	1.68 ± 0.10
Weight (kg)	69.4 ± 15.3	77.2 ± 15.0
BMI (kg/m^2)	22.3 ± 2.2	27.2 ± 3.7
Preferred Walking Speed (m/s)	1.15 ± 0.10	1.16 ± 0.13
Kellgren Lawrence grade	N/A	I: 12, II: 32, III: 20

Table I Population information for participants with and without osteoarthritis (OA). Data are presented as mean ± standard deviation

Osteoarthritis and Cartilage

Participants walked on an instrumented treadmill (Bertec Corporation, Columbus, OH, USA) at their self-selected speed in an 11-camera motion capture volume (Motion Analysis Corporation, Santa Rosa, CA, USA) with a retro-reflective marker set sufficient to compute the 3D kinematics of the lower limbs and trunk[23]. Participants first completed a static trial to identify ankle and knee joint axes and joint centers and then performed bilateral hip circumduction trials to determine the hip joint centers[36]. After 5 min of treadmill familiarization, participants completed a 2-min baseline trial of natural walking. Participants then received vibrotactile biofeedback (described in Uhlrich *et al.*[23]) that instructed them to practice toeing-in and toeing-out by 5° and 10° relative to their natural FPA for a minimum of 1 min per angle. They subsequently performed four 2-min modification trials where they targeted each of the FPA modifications. The baseline and modification trials were used for analysis.

Ground reaction forces and marker positions were low-pass filtered at 8 Hz using a fourth-order, zero lag, Butterworth filter. The KAM was calculated using inverse dynamics in MATLAB (Mathworks Corporation, Natick, MA, USA) as the frontal plane component of the 3D knee moment expressed in the proximal tibial reference frame[23] and normalized by bodyweight and height. The peak KAM was defined as the maximum value of the KAM curve during the first half of the stance phase; the stance phase was defined as the time during which the vertical ground reaction force was greater than 30N.

Predictive statistical models

The inputs to our machine learning model were the 3D positions of 13 anatomical landmarks from motion capture and a binary value identifying the stance leg [Fig. 1(B)]. From the motion capture marker set, we selected a subset of anatomical landmarks that are similar to the keypoints tracked using 2D video analysis algorithms like OpenPose[25]. The anatomical landmarks chosen included four from each lower extremity (2nd metatarsal head, posterior calcaneus, lateral malleolus, lateral femoral epicondyle), four pelvic landmarks (right and left anterior superior iliac spine and posterior superior iliac spine), and the C7 vertebrae. We defined the midpoint of the posterior superior iliac spine markers as the origin of the anatomical landmark positions to make these input positions independent of the participant's location in the laboratory frame. Positions were then normalized by participant height. Additionally, the medio-lateral positions of left steps were reflected across the body midline, so that all data appeared to be from the right leg, making the input more consistent[37]. Since our goal was to predict the first peak KAM, which occurs during the first half of the stance phase, we sampled the input positions 8 times during the first half of the stance phase. This sample rate was sufficient to capture the frequency content of the 8Hz-lowpass-filtered inputs during the stance phase, which lasts less than 1 s. The final input matrix for each step was of size 40×8 ((13 markers · three dimensions + one leg binary) · eight timesteps).

Participants were randomly divided 80%-10–10% into training, development, and test sets, such that all of a participant's data resided in only one set. The training set consisted of 91,245 steps from 70 individuals (18 without OA, 52 with OA), the development set consisted of 9,810 steps from eight individuals (2 without OA, 6 with OA), and the test set consisted of 11,675 steps from eight individuals (2 without OA, 6 with OA). The performance of the model on the development set was used to select the model architecture

and hyperparameters. The test set was used only to evaluate the accuracy of the final model.

Models were trained in Python (version 3.6.8) using Keras (version 2.2.4) with one GPU (NVIDIA GeForce GTX 960). We evaluated the following architectures: linear regression, a convolutional neural network (three 1-dimensional convolutional layers, three fully connected layers)[38], a long short-term memory network (2 long short-term memory layers, one fully connected layer)[39], and a fully connected neural network (10 hidden layers with 100 neurons each). Neural network (NN) weights were initialized with Xavier initialization[40], and Adam gradient descent[41] was used to minimize the root-mean-squared-error. The final layer of each model had a single neuron with linear activation to output the single peak KAM value. We selected a fully connected NN architecture due to its superior performance on the development set (Table S1, *Supplemental Information*). We refer to this model as the *3D NN*, as its inputs are 3D anatomical landmark positions.

We used lasso regression[42] to reduce redundant information in the flattened input vector before training the *3D NN* (glmnet package[43] in R[44]). To select the lambda value for lasso regression, we performed 10 runs of 10-fold cross-validation using the training set. We used the one-standard-error rule for selecting lambda: for each run, we selected the lambda value that yielded a prediction error that was one standard error greater than the minimum prediction error, thereby removing more features than the lambda that minimized prediction error. We then removed all but one binary leg feature from the flattened input vector. The input vector size was reduced from 320 to 299 features.

After a random search, we selected the following *3D NN* hyperparameters based on model performance on the development set: one hidden layer (tested 1–20), 800 neurons in the hidden layer (tested 0–1,000), and 0.01 probability of dropout (tested 0–0.25). Neurons in the hidden layer used a rectified linear unit activation function[45], while the output neuron had a linear activation function. To reduce overfitting, we used early stopping with five epochs of patience and used model weights from the epoch with the best development-set performance.

To identify which anatomical landmark positions were most influential on the peak KAM predictions, we computed the saliency of the input features. In general, saliency analysis identifies how changes in each input feature relate to changes in the output by computing the partial derivative of the output with respect to each input feature[46]. For this analysis, we trained a model with the same architecture and input features as the *3D NN* but without lasso regression in order to account for every feature at each timestep. We averaged saliency (Python Keras-vis package[47]) over 2000 examples randomly selected from the test set, then averaged over all timesteps for each of the 39 anatomical landmark positions (Eq. (1)):

$$Saliency_{x_i} = \text{mean}_k\left(\text{mean}_j\left(\left|\frac{\partial KAM^{(j)}}{\partial x_i^{(j,k)}}\right|\right)\right) \quad (1)$$

where $x_i = 1{:}39$ input positions, $j = 1{:}2000$ examples, and $k = 1{:}8$ timesteps. Finally, we normalized the saliency value for each feature by the saliency value of the most salient feature (Eq. (2)).

$$Saliency_{x_i,norm} = \frac{Saliency_{x_i}}{\max_i(Saliency_{x_i})} \quad (2)$$

We then evaluated the performance of our models on the clinical decision-making tasks. To test a model's ability to predict an individual's baseline peak KAM during natural walking, we predicted the peak KAM for all steps during the baseline trial, computed the average value for an individual leg, then compared this value to the average of the peak KAM values from ID. We analyzed both legs for the eight individuals in the test set, yielding 16 conditions. To test the model's ability to identify changes in the peak KAM from varying degrees of foot progression angle modifications, we identified steps for which individuals toed-in or toed-out by 5 ± 2.5°, 10 ± 2.5°, or 15 ± 2.5°. For each modification, there were 32 possible conditions (8 subjects, two legs, two angles), but not all legs achieved the target angles, leaving between 20 and 31 conditions for analysis. Changes in the peak KAM were computed by subtracting a leg's average baseline peak KAM from the peak KAM from a modified step, using KAM values from either the NN or ID.

For our final aim, we simulated 2D video input by using only sagittal or frontal plane anatomical landmark positions as model inputs. We removed the anterior-posterior landmark positions (in the laboratory frame) for the *frontal plane NN* and the medio-lateral components for the *sagittal plane NN*. There were 209 inputs for these planar models (13 markers · two dimensions · eight timesteps + one leg binary). We used the same model architecture

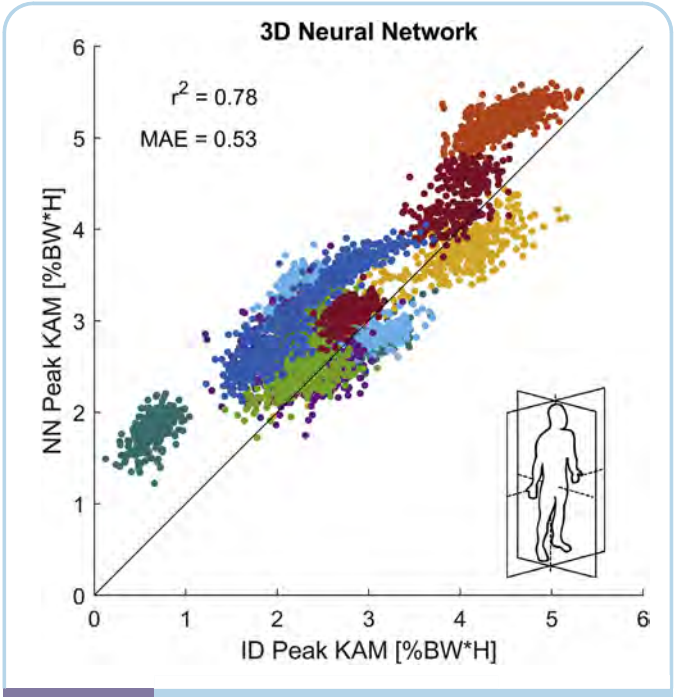

Fig. 2 The predicted peak knee adduction moment (KAM) from the neural network (NN) using 3D anatomical landmark positions as input (3D neural network) vs. the peak KAM calculated from inverse dynamics (ID) plotted against the y = x line. Presented data are for test subjects from the baseline and foot progression angle modification trials. Each point represents a single step, and a single color represents the steps from both legs of a subject.

Osteoarthritis and Cartilage

		Feature_coordinate (Leg)	$Saliency_{norm}$
3D Neural Network	1	PSIS_z (Stance)	1.00
	2	PSIS_z (Swing)	1.00
	3	Knee_y (Stance)	0.88
	4	C7 Vertebrae_y	0.76
	5	Knee_x (Swing)	0.43
Frontal Plane Neural Network	1	Toe_y (Swing)	1.00
	2	Knee_y (Stance)	1.00
	3	Knee_y (Swing)	0.74
	4	C7 Vertebrae_z	0.38
	5	Toe_z (Swing)	0.21
Sagittal Plane Neural Network	1	Toe_x (Swing)	1.00
	2	Knee_x (Stance)	0.98
	3	PSIS_z (Swing)	0.59
	4	ASIS_z (Swing)	0.50
	5	PSIS_x (Stance)	0.18

Fig. 3 The top five most salient features (features that, when changed, have the greatest effect on the predicted peak KAM) normalized by the most salient feature for the 3D, frontal plane, and sagittal plane neural networks (left). The positions and Cartesian coordinate directions of the most salient features (right) where x corresponds to the anterior-posterior direction, y to medio-lateral, and z to superior-inferior.

and training procedures for the planar models as the *3D NN*, excluding lasso regression.

Analytical statistics

All statistical analyses were performed in MATLAB. We evaluated model performance using r^2 (Pearson's correlation coefficient) and the mean absolute error (MAE; Eq. (3)):

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|y_i - \hat{y}_i\right| \quad (3)$$

where m = number of examples; y_i = ID peak KAM; and $\hat{y}_i$ = NN-predicted peak KAM. As a supplemental analysis, we added three types of virtual noise to the dataset to simulate various sources of error in video-based keypoint identification, with error magnitudes (14–30 mm) based on the previously-reported accuracy of pose recognition algorithms[26,27] (details in Table S2).

To estimate the uncertainty in our performance estimates based on test-set subject selection, we used percentile bootstrapping[48] to compute 95% confidence intervals (CI) for r^2 and MAE. For all bootstrapped distributions, we trained one model and resampled (10,000 times) the subjects in the test set with replacement from the eight subjects allocated to the test set, then calculated the test statistics for each resampled set. When evaluating model performance on all steps in the test set (Figs. 2 and 6), we down-sampled the test set to have an equal number of steps (400, randomly selected) for each leg, which gave each individual an equal effect on the test statistic, regardless of walking speed or step frequency. We computed the sample r^2 and MAE from this down-sampled test set. To evaluate if our model predicted the baseline peak KAM with MAE less than 0.5%BW*H (Fig. 4), we computed the average baseline peak KAM for each leg in the test set and computed the 95% CI from a bootstrapped distribution with per-subject resampling.

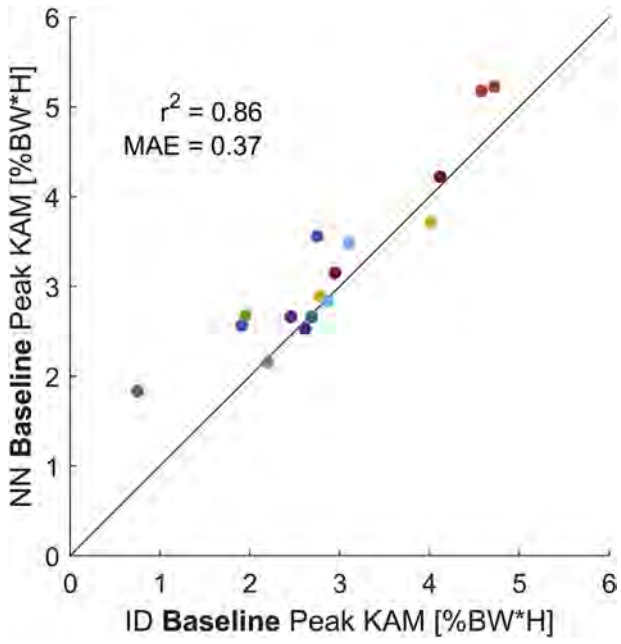

Fig. 4 The peak knee adduction moment (KAM) estimated by the 3D neural network (NN) and inverse dynamics (ID) from the baseline (natural walking) trial. Data are averaged over all baseline steps for each leg of each subject (represented by a color) in the test set plotted against the y = x line.

Results

The *3D NN* predicted the first peak KAM for all steps in the test set (Fig. 2) with $r^2 = 0.78$ (95% CI = [0.44, 0.89]) and MAE = 0.53% BW*H (95% CI = [0.39, 0.67]). The training set performed with $r^2 = 0.88$ and the development set with $r^2 = 0.72$. The saliency analysis (Fig. 3) showed that for the *3D NN*, the peak KAM prediction was most sensitive to changes in positions of the swing and stance-leg anterior superior iliac spines (anterior-posterior), the stance-leg knee (medio-lateral), the C7 vertebrae (medio-lateral), and the swing-leg anterior superior iliac spine (anterior-posterior). From the supplemental virtual noise simulation (Table S2), the *3D NN* performance did not degrade substantially with constant error added to a coordinate across all subjects or with 14 mm of error that changed on a subject-by-subject basis. Model performance did degrade with 30 mm of subject-by-subject noise and any magnitude of noise that changed randomly at each timestep.

The *3D NN* predicted the average peak KAM during baseline (natural) walking with $r^2 = 0.86$ (95% CI = [0.62, 0.94]) and MAE = 0.37%BW*H (95% CI = [0.23, 0.51]) (Fig. 4). The MAE confidence interval minimally overlapped with our most stringent clinically-meaningful accuracy threshold of 0.5%BW*H. Predictions of the baseline peak KAM were, on average, 0.31%BW*H (95% CI = [0.10, 0.51]) greater than the ID values (Bland Altman analysis in Fig. S1) with an absolute error range of 0.02–1.09%BW*H. The *3D NN* classified if 5°, 10°, and 15° toe-in or toe-out gait modifications increased or reduced the peak KAM with accuracies of 0.65, 0.71, and 0.85, respectively (Fig. 5, Table II).

The *frontal plane NN* predicted the peak KAM for all steps in the test set [Fig. 6(A)] with $r^2 = 0.85$ (95% CI = [0.56, 0.91]) and MAE = 0.49%BW*H (95% CI = [0.39, 0.59]), which was not statistically different from the performance of the *3D NN*. The *frontal plane NN* predicted the average peak KAM during baseline walking with $r^2 = 0.86$ (95% CI = [0.49, 0.96]) and MAE = 0.40%BW*H (95% CI = [0.23, 0.59]), and predicted the change in KAM resulting from 5°, 10°, and 15° FPA modifications with accuracy of 0.58, 0.79, and 0.80, respectively (Table II).

The *sagittal plane NN* predicted the peak KAM for all steps in the test set [Fig. 6(B)] with $r^2 = 0.14$ (95% CI = [0.02, 0.45]) and MAE = 0.85%BW*H (95% CI = [0.63, 1.06]), which was less accurate than the *frontal plane NN* and had minimal confidence-interval overlap with the *3D NN*.

Discussion

The purpose of this study was to evaluate the feasibility of using video to predict the peak KAM in individuals with and without knee OA during walking. To do this, we trained machine learning models to predict the peak KAM using the positions of anatomical landmarks from motion capture that are similar to those that could be measured with video analysis in the future. We found that a NN that uses 3D positions accurately predicted the peak KAM for people with and without knee OA ($r^2 = 0.78$). This model accurately predicted individuals' peak KAM during natural walking as well as

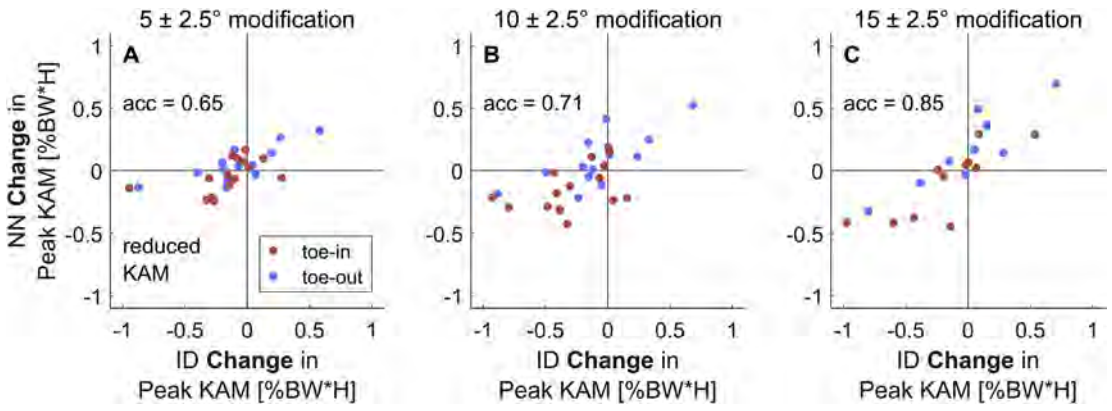

Fig. 5 The average change in the peak knee adduction moment (KAM) estimated by the 3D neural network (NN) vs inverse dynamics (ID) for 5°, 10°, and 15° foot progression angle modifications for each leg of each subject in the test set. The accuracy (acc.) of classification is increases with increasing degrees of foot progression angle modification.

changes in peak KAM that resulted from large (15°) FPA modifications, suggesting that it performs well enough to inform clinical decision making in many cases. A NN that only uses frontal plane positions also predicted the peak KAM accurately, suggesting that it might be feasible to use a front-facing camera and video analysis tools to estimate the peak KAM in the future.

The *3D* and *frontal plane NNs* predicted the peak KAM with $r^2 = 0.78$ and $r^2 = 0.85$, which is similar to previous techniques that predicted the peak KAM with a reduced set of inputs compared to a full gait analysis laboratory. The 95% CI around the performance of our models overlap with the performance of machine learning approaches that used inertial measurement units ($r^2 = 0.71$)[32] or anthropometric measures and force plate data ($r^2 = 0.59$)[30] as well as an ID approach that used instrumented force shoes and inertial measurement units ($r^2 = 0.80$)[24]. Other studies have predicted the KAM curve during walking using a reduced set of inputs[34,35,49], but it is worth noting that predicted curves that are highly correlated with the reference curve do not always yield accurate peak predictions. For example, Favre *et al.* predicted the KAM curve with $r^2 = 0.94$, but the peak KAM extracted from the curve was less accurate ($r^2 = 0.59$)[30], demonstrating the importance of evaluating model performance on clinically-meaningful outcome metrics, like the peak KAM. Although our models do not directly enable mobile measurements, the *frontal plane NN* demonstrates the feasibility of

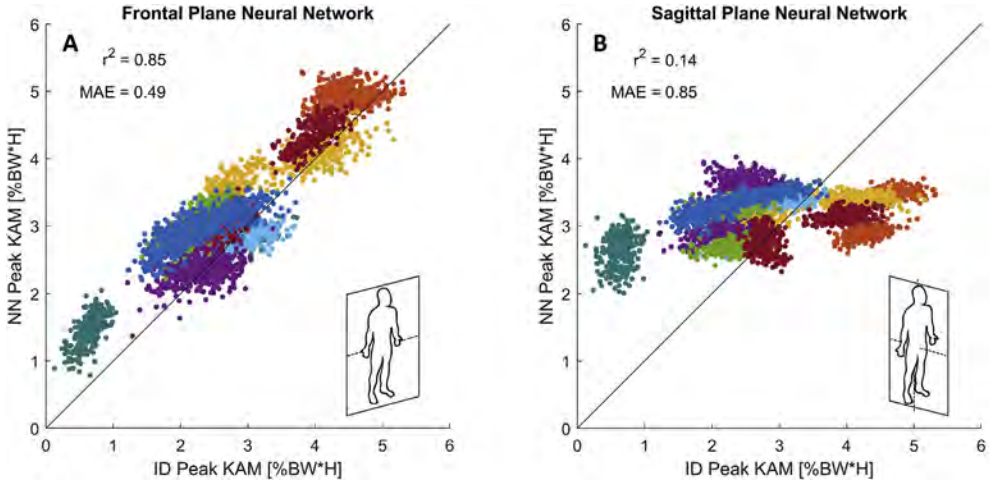

Fig. 6 The performance of neural networks that use planar projections of anatomical landmark positions as inputs. **A**) The frontal plane neural network predicts the peak KAM with similar accuracy to the 3D neural network ($r2 = 0.85$). **B**) The sagittal plane neural network is less accurate ($r2 = 0.14$) than the 3D or frontal plane neural networks. Presented data are for test subjects from the baseline and foot progression angle modification trials. A point represents a single step, and a single color represents the steps from both legs of a subject.

predicting the peak KAM from a limited set of anatomical landmark positions that are identifiable from 2D video. Secondarily, the model could serve as the basis for a transfer learning approach that uses keypoints from front-facing video as inputs.

The 3D input positions that the saliency analysis identified as most influential for KAM predictions correspond with kinematic changes that have been shown to affect the KAM. First, the superior-inferior positions of the posterior superior iliac spines relates to contralateral pelvic drop, which increases the KAM[16]. Second, the medial position of the stance-limb knee relates to the frontal-plane knee angle. Medializing the knee is a suggested mechanism for KAM-reducing interventions such as medial-thrust gait[17], toe-in gait[50], and variable stiffness shoes[51]. Third, the medio-lateral position of the C7 vertebrae relates to the trunk-sway angle. Trunk-sway influences the KAM by altering the medio-lateral position of the center of mass[15]. Finally, the anterior-posterior position of the swing-leg knee could relate to stride length. A slower walking speed or decreased stride length can reduce the peak KAM[10,17]. Some of the salient features in the *frontal* and *sagittal plane* models were different from the *3D* model's salient features in these planes, which may be explained by the redundancy of input features, complex nonlinear relationships between features, and the stochastic nature of model training. Even small perturbations to input data can lead to dramatically different saliency results, and thus, different interpretations of the model[52], so the relationships between gait modifications and the salient features remains speculative.

Both the *3D* and *frontal plane NNs* performed with sufficient accuracy to inform clinical decision making in most cases. The models predicted the peak KAM during natural walking with MAE = 0.37–0.49%BW*H, which is accurate enough to evaluate the risk of OA progression (differences between the peak KAM of progressors and non-progressors is 2.1%BW*H)[9]. Our models also predicted the baseline peak KAM with sufficient accuracy to classify between patients who are likely or unlikely to benefit from a high tibial osteotomy (individuals with a 2.2%BW*H lower pre-surgical peak KAM had better post-surgical outcomes)[13,53]. Additionally, our models classified whether 15° FPA modifications increased or reduced the peak KAM with an accuracy of 0.80–0.85. However, they were less accurate in predicting smaller 5° and 10° modifications, potentially due to subtler changes to both input kinematics and the output peak KAM. Notably, toeing-in typically reduces the first peak KAM[50], but it is not effective for all individuals[23]; the *3D NN* correctly identified the five of the six legs that reduced the first peak KAM by toeing-in by 15° and the four of the five that did not. If a future video-based model performs with similar accuracy, any clinician with a smartphone would be able to make clinically-actionable biomechanical measurements without purchasing expensive equipment.

It is important to identify the limitations of this study. First, machine learning models may not generalize well to conditions not represented in the training data. Our models were trained on individuals with and without knee OA, with a BMI below 35, who were walking with varying FPAs. Thus, our models will likely lose accuracy when predicting the peak KAM for new populations performing different activities. Fortunately, we trained accurate models with a reasonably-sized dataset ($n = 86$), suggesting that new models predicting different parameters can likely be trained using pre-existing datasets. Second, the inputs to our models were motion capture marker positions that were very similar, but not identical to the positions commonly used in video pose-recognition algorithms (e.g., lateral femoral epicondyle vs knee joint center). However, model performance did not change dramatically when a constant offset was added to the motion capture positions (Table S2), indicating that the models may not be sensitive to the small differences in the definitions of anatomical landmarks between motion capture and video-based keypoints. Finally, our models do not directly facilitate mobile measurement of the peak KAM since they rely on anatomical landmark positions from motion capture rather than keypoints from video. The robustness of our pipeline to several sources of error that may arise from video keypoints (Table S2) adds confidence that a video-based approach may be feasible. Nonetheless, the performance of our models trained on motion capture positions will likely exceed the accuracy of a similar model trained with a similarly-sized dataset of video keypoints. Even with a slight decrease in performance, an inexpensive, video-based solution would provide clinical value compared to other mobile solutions that require equipment that costs thousands of dollars[24,30,54].

In summary, we developed a model to predict the first peak KAM in individuals with and without knee OA using the positions of anatomical landmarks. Our model accurately predicts the peak KAM during natural walking as well as changes in the KAM that result from gait modifications. Since anatomical landmark positions projected onto a plane are similar to the output from 2D video keypoint detection algorithms, our results support the feasibility of predicting the peak KAM from a 2D frontal plane video as a promising next step. These results support the utility of computer vision and machine learning as tools that can bring biomechanical

Table II

The performance of the 3D and frontal plane neural networks in classifying whether changes in foot progression angle (FPA) increase or reduce the peak knee adduction moment (Fig. 5). A reduction in peak KAM was considered a positive result. For example, sensitivity represents the number of cases that the neural network predicted as a reduction in the peak KAM divided by the number of cases that ID identified as a reduction in the peak KAM. Both models predict the effects of larger FPA modifications more accurately than small modifications

	Accuracy	Sensitivity	Specificity	Precision
3D Neural Network				
5° FPA modification	0.65	0.59	0.78	0.87
10° FPA modification	0.71	0.70	0.75	0.87
15° FPA modification	0.85	0.73	1.0	1.0
Frontal Plane Neural Network				
5° FPA modification	0.58	0.68	0.33	0.71
10° FPA modification	0.79	0.90	0.50	0.80
15° FPA modification	0.80	1.0	0.73	0.73

measurements into the clinic or home. We envision that models like the ones presented here will soon enable scientists to easily monitor joint loading in large cohorts and clinicians to prescribe personalized treatments for musculoskeletal pathologies.

Contributions

JAK and SDU collected the experimental data. MAB and SDU trained and analyzed the statistical models. LK and KT advised the statistical analyses. All authors contributed to the study design and interpretation of the data. GSB and SLD secured funding for the project. MAB and SDU drafted the manuscript and all authors revised and approved the final manuscript.

Conflict of interest

The authors have no conflicts of interest to declare.

Role of the funding source

The funding sources did not play a role in study design, collection, analysis and interpretation of data; writing of the manuscript; nor the decision to submit the manuscript for publication.

Data and code availability

The models, code, and a portion of the dataset are available on GitHub (https://github.com/stanfordnmbl/predictKAM).

Acknowledgements

The authors would like to thank Patrick Cho and Andrew Ng for their advice on the machine learning models. We would also like to thank Amy Silder and Brittany Presten for their assistance in data collection. This work was supported by the Graduate Research Fellowship Program from the United States (U.S.) National Science Foundation under Grant No. DGE-114747 and Grant No. DGE-1656518. This work was also supported by Merit Review Award Number I01 RX001811 from the U.S. Department of Veterans Affairs Rehabilitation R&D (Rehab RD) Service, and Grants U54EB020405 and P41EB027060 from the U.S. National Institutes of Health.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joca.2020.12.017.

References

1. Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, Messier SP, *et al*. The number of persons with symptomatic knee osteoarthritis in the United States: impact of race/ethnicity, age, sex, and obesity. Arthritis Care Res 2016;68(12):1743–50, https://doi.org/10.1002/acr.22897.
2. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, *et al*. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380(9859):2197–223, https://doi.org/10.1016/S0140-6736(12)61689-4.
3. Wise BL, Niu J, Yang M, Lane NE, Harvey W, Felson DT, *et al*. Patterns of compartment involvement in tibiofemoral osteoarthritis in men and women and in whites and African Americans. Arthritis Care Res 2012;64(6):847–52, https://doi.org/10.1002/acr.21606.
4. Mündermann A, Dyrby CO, D'Lima DD, Colwell CW, Andriacchi TP. In vivo knee loading characteristics during activities of daily living as measured by an instrumented total knee replacement. J Orthop Res 2008;26(9):1167–72, https://doi.org/10.1002/jor.20655.
5. Kutzner I, Trepczynski A, Heller MO, Bergmann G. Knee adduction moment and medial contact force-facts about their correlation during gait. PloS One 2013;8(12):8–15, https://doi.org/10.1371/journal.pone.0081036.
6. Hurwitz DE, Ryals AB, Case JP, Block JA, Andriacchi TP. 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 2002;20(1):101–7, https://doi.org/10.1016/S0736-0266(01)00081-X.
7. Sharma L, Hurwitz DE, Ej-Ma Thonar, Sum JA, Lenz ME, Dunlop DD, *et al*. Knee adduction moment, serum hyaluronan level, and disease severity in medial tibiofemoral osteoarthritis. Arthritis Rheum 1998;41(7):1233–40, https://doi.org/10.1002/1529-0131(199807)41:7<1233::AID-ART14>3.0.CO;2-L.
8. Foroughi N, Smith R, Vanwanseele B. The association of external knee adduction moment with biomechanical variables in osteoarthritis: a systematic review. Knee 2009;16(5):303–9, https://doi.org/10.1016/j.knee.2008.12.007.
9. Miyazaki T, Wada M, Kawahara H, Sato M, Baba H, Shimada S. Dynamic load at baseline can predict radiographic disease progression in medial compartment knee osteoarthritis. Ann Rheum Dis 2002;61(7):617–22, https://doi.org/10.1136/ard.61.7.617.
10. Mündermann A, Dyrby CO, Hurwitz DE, Sharma L, Andriacchi TP. Potential strategies to reduce medial compartment loading in patients with knee osteoarthritis of varying severity: reduced walking speed. Arthritis Rheum 2004;50(4):1172–8, https://doi.org/10.1002/art.20132.
11. Mündermann A, Dyrby CO, Andriacchi TP. Secondary gait changes in patients with medial compartment knee osteoarthritis: increased load at the ankle, knee, and hip during walking. Arthritis Rheum 2005;52(9):2835–44, https://doi.org/10.1002/art.21262.
12. Amin S, Luepongsak N, McGibbon CA, LaValley MP, Krebs DE, Felson DT. Knee adduction moment and development of chronic knee pain in elders. Arthritis Care Res 2004;51(3):371–6, https://doi.org/10.1002/art.20396.
13. Prodromos CC, Andriacchi TP, Galante JO. A relationship between gait and clinical changes following high tibial osteotomy. J Bone Jt Surg 1985;67(8):1188–94, https://doi.org/10.2106/00004623-198567080-00007.
14. Simic M, Hunt MA, Bennell KL, Hinman RS, Wrigley TV. Trunk lean gait modification and knee joint load in people with medial knee osteoarthritis: the effect of varying trunk lean angles. Arthritis Care Res 2012;64(10):1545–53, https://doi.org/10.1002/acr.21724.
15. Mündermann A, Asay JL, Mündermann L, Andriacchi TP. Implications of increased medio-lateral trunk sway for ambulatory mechanics. J Biomech 2008;41(1):165–70, https://doi.org/10.1016/j.jbiomech.2007.07.001.
16. Dunphy C, Casey S, Lomond A, Rutherford D. Contralateral pelvic drop during gait increases knee adduction moments of asymptomatic individuals. Hum Mov Sci 2016;49:27–35, https://doi.org/10.1016/j.humov.2016.05.008.
17. Fregly BJ, Reinbolt JA, Rooney KL, Mitchell KH, Chmielewski TL. Design of patient-specific gait modifications for knee osteoarthritis rhabilitation. IEEE Trans Biomed Eng 2007;54(9):1687–95, https://doi.org/10.1109/TBME.2007.891934.
18. Gerbrands TA, Pisters MF, Theeven PJR, Verschueren S, Vanwanseele B. Lateral trunk lean and medializing the knee as

gait strategies for knee osteoarthritis. Gait Posture 2017;51:247–53, https://doi.org/10.1016/j.gaitpost.2016.11.014.
19. Shull PB, Silder A, Shultz R, Dragoo JL, Besier TF, Delp SL, *et al.* Six-week gait retraining program reduces knee adduction moment, reduces pain, and improves function for individuals with medial compartment knee osteoarthritis. J Orthop Res 2013;31(7):1020–5, https://doi.org/10.1002/jor.22340.
20. Hunt MA, Takacs J. Effects of a 10-week toe-out gait modification intervention in people with medial knee osteoarthritis: a pilot, feasibility study. Osteoarthritis Cartilage 2014;22(7):904–11, https://doi.org/10.1016/j.joca.2014.04.007.
21. Gerbrands TA, Pisters MF, Vanwanseele B. Individual selection of gait retraining strategies is essential to optimally reduce medial knee load during gait. Clin Biomech 2014;29(7):828–34, https://doi.org/10.1016/j.clinbiomech.2014.05.005.
22. Shull PB, Huang Y, Schlotman T, Reinbolt JA. Muscle force modification strategies are not consistent for gait retraining to reduce the knee adduction moment in individuals with knee osteoarthritis. J Biomech 2015;48(12):3163–9, https://doi.org/10.1016/j.jbiomech.2015.07.006.
23. Uhlrich SD, Silder A, Beaupre GS, Shull PB, Delp SL. Subject-specific toe-in or toe-out gait modifications reduce the larger knee adduction moment peak more than a non-personalized approach. J Biomech 2018;66:103–10, https://doi.org/10.1016/j.jbiomech.2017.11.003.
24. van den Noort J, van der Esch M, Steultjens MPM, Dekker J, Schepers M, Veltink PH, *et al.* Ambulatory measurement of the knee adduction moment in patients with osteoarthritis of the knee. J Biomech 2013;46(1):43–9, https://doi.org/10.1016/j.jbiomech.2012.09.030.
25. Cao Z, Hidalgo G, Simon T, Wei S, Sheikh Y. OpenPose : realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans Pattern Anal Mach Intell 2019:1–14, https://doi.org/10.1109/tpami.2019.2929257. Published online.
26. Iskakov K, Burkov E, Lempitsky V, Malkov Y. Learnable triangulation of human pose. Proc IEEE Int Conf Comput Vis 2019:7717–26, https://doi.org/10.1109/ICCV.2019.00781. 2019-Octob.
27. Nakano N, Sakura T, Ueda K, Omura L, Kimura A, Iino Y, *et al.* Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras. Front Sport Act Living 2020;2(May):1–9, https://doi.org/10.3389/fspor.2020.00050.
28. Halilaj E, Rajagopal A, Fiterau M, Hicks JL, Hastie TJ, Delp SL. Machine learning in human movement biomechanics: best practices, common pitfalls, and new opportunities. J Biomech 2018;81:1–11, https://doi.org/10.1016/j.jbiomech.2018.09.009.
29. Ł Kidziński, Yang B, Hicks JL, Rajagopal A, Delp SL, Schwartz MH. Deep neural networks enable quantitative movement analysis using single-camera videos. Nat Commun 2020;11(4054), https://doi.org/10.1038/s41467-020-17807-z.
30. Favre J, Hayoz M, Erhart-Hledik JC, Andriacchi TP. A neural network model to predict knee adduction moment during walking based on ground reaction force and anthropometric measurements. J Biomech 2012;45(4):692–8, https://doi.org/10.1007/s11103-011-9767-z.Plastid.
31. Ardestani MM, Zhang X, Wang L, Lian Q, Liu Y, He J, *et al.* Human lower extremity joint moment prediction: a wavelet neural network approach. Expert Syst Appl 2014;41(9):4422–33, https://doi.org/10.1016/j.eswa.2013.11.003.
32. Stetter BJ, Krafft FC, Ringhof S, Stein T, Sell S. A machine learning and wearable sensor based approach to estimate external knee flexion and adduction moments during various locomotion tasks. Front Bioeng Biotechnol 2020;8(January), https://doi.org/10.3389/fbioe.2020.00009.
33. Wang C, Chan PPK, Lam BMF, Wang S, Zhang JH, Chan ZYS, *et al.* Real-time estimation of knee adduction moment for gait retraining in patients with knee osteoarthritis. IEEE Trans Neural Syst Rehabil Eng 2020;28(4):888–94, https://doi.org/10.1109/TNSRE.2020.2978537.
34. Johnson WR, Mian A, Lloyd DG, Alderson JA. On-field player workload exposure and knee injury risk monitoring via deep learning. J Biomech 2019;93:185–93, https://doi.org/10.1016/j.jbiomech.2019.07.002.
35. Aljaaf AJ, Hussain AJ, Fergus P, Przybyla A, Barton GJ. Evaluation of machine learning methods to predict knee loading from the movement of body segments. Proc Int Jt Conf Neural Networks 2016;2016-Octob:5168–73, https://doi.org/10.1109/IJCNN.2016.7727882.
36. Piazza SJ, Erdemir A, Okita N, Cavanagh PR. Assessment of the functional method of hip joint center location subject to reduced range of hip motion. J Biomech 2004;37(3):349–56, https://doi.org/10.1016/S0021-9290(03)00288-4.
37. Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. Sci Rep 2018;8:1–15, https://doi.org/10.1038/s41598-018-20132-7.
38. Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: a strong baseline. arXiv 2016;1611:6455, https://doi.org/10.1109/IJCNN.2017.7966039.
39. Kidziński L. Event-Detector-Train 2018. Published, https://github.com/kidzik/event-detector-train.
40. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks Xavier. In: Proc 13th Int Conf Artif Intell Stat 2010:249–56, https://doi.org/10.1109/ijcnn.1993.716981. Published online.
41. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Int Conf Learn Represent 2015:1–15. Published online, http://arxiv.org/abs/1412.6980.
42. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B 1996;58(1):267–88, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
43. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Software 2010;33(1):1–22, https://doi.org/10.18637/jss.v033.i01.
44. R Core Team. A Language and Environment for Statistical Computing 2019. Published online, https://www.r-project.org/.
45. Nair V, Hinton GE. Rectified linear units improve Restricted Boltzmann machines. In: ICML 2010 - Proceedings, 27th International Conference on Machine Learning 2010.
46. Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. Vis Res 2000;40(10–12):1489–506, https://doi.org/10.1016/S0042-6989(99)00163-7.
47. Kotikalapudi R, and Contributors. Keras-Vis. Published 2017. https://github.com/raghakot/keras-vis.
48. Efron B, Tibshirani R. An Introduction to the Bootstrap. Chapman & Hall; 1993, https://doi.org/10.1007/978-1-4899-4541-9.
49. Karatsidis A, Jung M, Schepers HM, Bellusci G, de Zee M, Veltink PH, *et al.* Predicting kinetics using musculoskeletal modeling and inertial motion capture. arXiv, 2018. Published online, http://arxiv.org/abs/1801.01668.
50. Shull PB, Shultz R, Silder A, Dragoo JL, Besier TF, Cutkosky MR, *et al.* Toe-in gait reduces the first peak knee adduction moment in patients with medial compartment knee osteoarthritis. J Biomech 2013;46(1):122–8, https://doi.org/10.1016/j.jbiomech.2012.10.019.

51. Jenkyn TR, Erhart JC, Andriacchi TP. An analysis of the mechanisms for reducing the knee adduction moment during walking using a variable stiffness shoe in subjects with knee osteoarthritis. J Biomech 2011;44(7):1271–6, https://doi.org/10.1016/j.jbiomech.2011.02.013.
52. Ghorbani A, Abid A, Zou J. Interpretation of neural networks is fragile. Proc Thirty-Third AAAI Conf Artif Intell. 2019;33(1):3681–8, https://doi.org/10.1609/aaai.v33i01.33013681.
53. Wang JW, Kuo KN, Andriacchi TP, Galante JO. The influence of walking mechanics and time on the results of proximal tibial osteotomy. J Bone Joint Surg Am 1990;72(6):905–9, https://doi.org/10.2106/00004623-199072060-00017.
54. Konrath JM, Karatsidis A, Martin Schepers H, Bellusci G, de Zee M, Andersen MS. Estimation of the knee adduction moment and joint contact force during daily living activities using inertial motion capture. Sensors 2019;19(7), https://doi.org/10.3390/s19071681.