Joint kinematics alone can distinguish hip or knee osteoarthritis patients from asymptomatic controls with high accuracy

Emmerzaal, Jill; Van Rossom, Sam; van der Straaten, Rob; De Brabandere, Arne; Corten, Kristoff; De Baets, Liesbet; Davis, Jesse; Jonkers, Ilse; Timmermans, Annick; Vanwanseele, Benedicte

Published in:
Journal of orthopaedic research : official publication of the Orthopaedic Research Society

DOI:
10.1002/jor.25269

Publication date:
2022

Document Version:
Accepted author manuscript

Joint kinematics alone can distinguish hip or knee osteoarthritis patients from asymptomatic controls with high accuracy

Authors

Jill S. Emmerzaal[1,2], Sam van Rossom[1], Rob van der Straaten[2], Arne De Brabandere[3], Kristoff Corten[4], Liesbet De Baets[5], Jesse Davis[3], Ilse Jonkers[1], Annick Timmermans[2], Benedicte Vanwanseele[1]

Affiliations

1. Human Movement Biomechanics Research Group, Department of movement sciences, KU Leuven, Belgium
2. REVAL Rehabilitation Research Centre, Faculty of Rehabilitation Sciences, Hasselt University, Belgium
3. Declarative Languages and Artificial Intelligence Group, Department of Computer Science, KU Leuven, Belgium
4. Department of Orthopaedics, Ziekenhuis Oost Limburg, Belgium
5. Pain in Motion Research Group (PAIN), Department of Physiotherapy, Human Physiology and Anatomy, Vrije Universiteit Brussel, Belgium

Corresponding author:

Jill S. Emmerzaal:

Tervuursevest 101 – box 1501, 3001 Leuven, Belgium

jill.emmerzaal@kuleuven.be

+3216373738

Running title

Classifying osteoarthritis patients and controls

Author contribution:

Study concept and design: Jill Emmerzaal, Rob van der Straaten, Liesbet De Baets, Kristoff Corten, Jesse Davis, Ilse Jonkers, Annick Timmermans, Benedicte Vanwanseele

Data collection and analysis: Jill Emmerzaal, Rob van der Straaten, Sam Van Rossom, Arne De Brabandere

Results interpretation: Jill Emmerzaal, Sam Van Rossom, Arne De Brabandere, Jesse Davis, Annick Timmermans, Ilse Jonkers, Benedicte Vanwanseele

Manuscript draft: Jill Emmerzaal

Critical review, edit and approval of manuscript: All authors

Abstract:

Osteoarthritis is one of the leading musculoskeletal disabilities worldwide, and several interventions intend to change the gait pattern in osteoarthritis patients to more healthy patterns. However, an accessible way to follow up the biomechanical changes in a clinical setting is still missing. Therefore, this study aims to evaluate whether we can use biomechanical data collected from a specific activity of daily living to help distinguish hip OA patients from controls and knee OA patients from controls using features that potentially could be measured in a clinical setting. To achieve this goal, we considered three different classes of statistical models with different levels of data complexity. Class 1 is kinematics based only (clinically applicable), class 2 includes joint kinetics (semi applicable under the condition of access to a force plate or prediction models), and class 3 uses data from advanced musculoskeletal modelling (not clinically applicable). We used a machine learning pipeline to determine which classification model was best. We found 100% classification accuracy for KneeOA-vs-Asymptomatic and 93.9% for HipOA-vs-Asymptomatic using seven features derived from the lumbar spine and hip kinematics collected during ascending stairs. These results indicate that kinematical data alone can distinguish hip or knee OA patients from asymptomatic controls. However, to enable clinical use, we need to validate if the classifier also works with IMU-based kinematical data and whether the probabilistic outcome of the logistic regression model can be used in the follow-up of patients with OA.

osteoarthritis — biomechanics – daily activities- classification model – machine learning

Introduction

Osteoarthritis (OA) is one of the most common musculoskeletal disorders worldwide. Approximately 18% of women and 9.6% of men over 60 years suffer from symptomatic OA[1], which is characterised by pain, physical disability, and difficulties performing activities of daily life[1]. Both non-mechanical (e.g., age, inflammation, genetics) and mechanical factors (abnormal joint anatomy, abnormal joint loading, body mass index (BMI)) might contribute to the onset and progression of OA[2].

People with osteoarthritis show a range of biomechanical adaptations in their gait patterns compared to asymptomatic individuals[3,4]. Those observed differences in kinetics and contact forces make OA patients possibly more susceptible to OA progression[5]. In knee OA patients, the baseline knee adduction moment has been associated with cartilage loss after five years[5]. Similarly, in hip OA patients, an increase in cumulative hip loading was related to increased cartilage loss[6]. Moreover, hip OA patients have adjusted their gait pattern to reduce the loading on the cartilage[7] regardless of muscle co-contraction[8]. Therefore biomechanical analysis has been suggested to provide insight into the person's progress and the effectiveness of an intervention (e.g. gait retraining[9,10]). However, conventional lab-based methods to capture joint kinematics, joint kinetics, and joint contact forces are time-consuming and require specialised equipment and specialised expertise, which makes them infeasible in a clinical context.

Alternatively, we can use advanced statistical models (i.e. machine learning or deep learning) to find distinctive patterns in the biomechanical data to help classify gait patterns or patients accordingly[11,12]. Accurate classification models based on motion data could be used for multiple purposes. First, they could possibly identify undiagnosed individuals when hip or knee OA is suspected as a fist screening tool by identifying osteoarthritis specific movement patterns.

Secondly, they could be used to evaluate a person's progress towards a normalised movement pattern after a surgical intervention or a rehabilitation program[12]. Suppose the classification of a person changes from OA to asymptomatic; one might conclude with some degree of certainty that the intervention was successful with regards to changing the biomechanical pattern to resemble that of an asymptomatic individual. However, first, an accurate model needs to be created that could be used in a clinical setting. Previous research used gait kinematics and kinetics to classify OA patients from asymptomatic controls and found classification accuracies ranging from 45% to 97.62% using various methodologies[13–19]. Until now, Jones et al. (2008) developed the most accurate classification method with an in- and out-sample classification accuracy of 97.62%[17] to classify knee OA patients. They used 12 principal components derived from kinematics and ground reaction force waveforms, as well as spatiotemporal and anthropometric data as input in their classification process. However, the need for force plate data complicates translating this approach to a clinical setting. To classify hip OA subjects, Laroche et al. (2014) found accuracies between 93% and 97% using Support Vector Machines (SVM)[14]. However, one of the drawbacks to SVM is that it does not show uncertainty, which might make it harder to monitor changes.

There exists an opportunity to develop alternative approaches for classifying patients that are suitable in clinical practice, e.g. using mobile sensors, or methods that report some degree of certainty. Therefore, instead of using a hard classifier like SVM, a soft, simple classifier like a logistic regression model (LR) might be more relevant. Compared to the SVM, an LR model is easier to interpret and also indicates the probability of a subject belonging to that class, which might be relevant in the evaluation of a person's progress. Contrarily, an LR model usually has lower predictive power, which might cause a risk when only including less challenging movement tasks such as gait to discriminate subjects with (hip or knee) OA from asymptomatic controls, therefore

we included more challenging tasks. Moreover, Komnik et al. (2015) highlighted the importance of investigating more challenging tasks of daily living to detect potential (mal)adaptive movement patterns in patients with OA[20]. Due to the differences in biomechanical adaptation for hip or knee OA patients, other activities that produce a higher load on the affected joint might be more useful for classification purposes. Therefore, it is of interest to investigate challenging exercises alongside gait, such as stair climbing, to determine which exercise is most effective in differentiating asymptomatic controls from subjects suffering from either hip or knee OA.

To classify hip and knee OA patients in a clinical setting, one needs features that could be derived from inexpensive mobile sensors (e.g. kinematical data using IMUs). However, to date, most classification models still require force plate data to classify OA patients from controls. Therefore, we want to explore the minimum number of movement features derived from activities of daily living that enable classification of (either hip or knee) OA patient groups from asymptomatic controls, indicating which parameters are of interest to measure and thereby proving the generalisability of the approach. We consider three different classes of statistical models with different levels of data collection and processing complexity. Class 1 is kinematics based only (clinical applicable), class 2 includes joint kinetics (which would still rely on access to a force plate or use of ML-based prediction models), and class 3 uses data from advanced musculoskeletal modelling (not clinically applicable). Moreover, as the gait pattern might not be sensitive enough, we will create these three statistical models using different exercises with varying difficulty levels. Accordingly, this study aims to evaluate whether we can use biomechanical data collected from a specific activity of daily living to help distinguish either hip or knee OA patients from asymptomatic controls using features that potentially could be measured in a clinical setting exploring the use of a general method for both patient groups.

Methods

Participants

This is a controlled laboratory study in which 51 people participated: 12 asymptomatic controls, 20 unilateral end-stage hip OA patients and 19 unilateral end-stage knee OA patients (see Table 1 for participant characteristics). This is an explorative study based on a secondary analysis of a more extensive prospective follow-up (S59857) that evaluated hip and knee joint contact forces in people with hip or knee osteoarthritis and following total knee arthroplasty. The sample size for that study was based on joint contact forces measured in people with an instrumented knee prosthesis (1.61 +/- 0.305 bodyweight during gait)[21]. Assuming that a difference of one standard deviation is significant and to achieve a power of 0.8, a sample size of 14 subjects per group is needed. Taking a possible loss to follow-up of 15-20% into account, we recruited 18-20 participants per group. Because for the larger study both legs of the asymptomatic controls were considered as independent, only 12 asymptomatic control subjects were recruited.

Patients awaiting a joint replacement surgery (Kellgren-Lawrence grade III (N=1)-IV (N=38)) were recruited from two local hospitals (Ziekenhuis Oost Limburg, Genk and Jessa Hospital, Hasselt, Belgium). Inclusion criteria for the patients were: age between 50 and 75 years; unilateral hip or knee OA; BMI<30kg/m^2; able to walk 10 meters; able to navigate stairs; no corticosteroid injection at least three months prior to inclusion; no joint replacement in other lower limb joints; no neurological or musculoskeletal disorders that could affect the movement pattern and no history of pathological osteoporotic fractures. Asymptomatic controls were included aged between 50 and 75 years old; they had a BMI < 30kg/m^2, no neurological or musculoskeletal disorders that could

affect their movement pattern, and were recruited from a local seniors network in Leuven, Belgium. Table 1 summarises the participant demographics and the patient-reported outcome of the Hip/Knee Disability and Osteoarthritis Outcome Score (HOOS or KOOS). The local ethics committee of the academic hospital Leuven approved the study protocol (s-59857), and the participants provided written informed consent before the start of the study.

Study protocol

All participants performed five repetitions of nine exercises in the Movement and posture Analysis Laboratory Leuven: level walking; forward lunge; sideward lunge; single-leg stance; single-leg squat; standing up from a chair; sitting down; ascending stairs; descending stairs. These tasks were selected based on their relevance for clinical practice. They resemble physiotherapy exercises (lunges) and parts of movements that are relevant for daily life functioning (single-leg-squat and single-leg-stance) or are repeatedly performed during everyday life (walking, sit/stand transitions and stair climbing). Further details about task specifications and standardisation can be found in supplementary materials A.

Reflective markers were attached to each participant conforming to the full-body Plug-in Gait (Oxford Metrics), including 38 markers[22]. In addition, three marker clusters replaced the single markers on the segments (shank, thigh, upper and lower arms), and one extra three-marker cluster was placed on the superior aspect of the left iliac crest of the pelvis. We placed additional anatomical markers on the sacrum, medial femur epicondyle and medial malleoli[22]. 3D marker trajectories were collected using 13 optoelectronic cameras (Vicon, Oxford Metrics, UK, 100Hz). Ground reaction forces (GRF) were measured synchronously using three ground-embedded force plates (AMTI, Watertown, MA, USA, sampling at 1000Hz).

Musculoskeletal modelling

The motion capture data were processed using a standard, musculoskeletal modelling workflow implemented in OpenSim3.3[14]. We used the generic model gait2392[14] (—model specifications and comprehensive information on the musculoskeletal workflow, see Supplementary material B*)*. The generic OpenSim model was scaled to match the body dimensions and bodyweight of the subject by using a measurement-based scaling approach within OpenSim. A personalised knee joint axis orientation and position were implemented within each scaled model to allow for a more complex knee joint description. The functional axis of rotation was calculated using the SARA algorithm based on a standing flexion/extension range of motion task[23,24]. All other coordinate frames were used directly within the OpenSim model. After that, joint angles were derived from the measured marker trajectories using the Kalman smoothing algorithm described by De Groote et al. (2008) and available from SimTK. The Kalman smoother algorithm is an alternative to the standard Inverse Kinematics Tool available in OpenSim, which improves the estimation of the joint kinematics and kinetics by using prior knowledge with the measured marker trajectories while minimising the estimation error statistically[26]. Joint moments were calculated using a standard inverse dynamics approach available in OpenSim. Afterwards, muscle forces and activations were calculated using a static optimisation routine that minimised the total muscle activation squared. Finally, joint contact forces were calculated using the vector sum of the estimated muscle forces and reaction forces in the joint using the standard OpenSim pipeline[27].

Feature construction

For each subject and each exercise, features were derived from the kinematics, kinetics and contact force time series using a multivariate feature construction tool, TSFuse[28]. Table 2 shows the variables of which the time series were used as input. The joint moments and joint contact forces

were normalised body weight. For the asymptomatic controls, both legs were analysed; for both OA groups, only the affected sides were analysed since altered kinematics and kinetics might have been related to contralateral, secondary OA involvement.

Given these time series, TSFuse generates a new time series by fusing multiple input time series, for example, by computing the ratio of two-time series. The system then extracts features from both the input time series and the generated time series. These features include statistical features (mean, variance, minimum, maximum, etc.), Fourier transform coefficients, number of peaks, zero crossings, etc. Feature construction was performed in an unsupervised manner, i.e. independent of the groups.

After extracting the features from each trial, the features were averaged over all trials per activity per participant. For the healthy participants, the features were also averaged over both legs. Note that we only use the affected leg for the hip and knee OA patients, and hence averaging over the legs is not necessary for the patients.

Statistical analysis

Separate analyses were used to compare people with hip OA versus asymptomatic controls as well as people with knee OA versus asymptomatic controls for all nine activities. Figure 1 also shows an overview of the statistical analysis. We used Python (version 3.7.8) with the TSFuse package (version 1.0dev) to construct the features and the scikit-learn package (version 0.21.2) to train the models. In total, we trained, tested, and evaluated 54 different models: hipOA-vs-Asymptomatic and kneeOA-vs-Asymptomatic models, for nine exercises with three levels of processing complexity of the input data.

To estimate the classification accuracy of a model on future (i.e., unseen) subjects a stratified five-fold cross-validation procedure was used. This procedure partitions the data into five disjoint folds (i.e., subsets of the data), where each fold has an identical ratio of examples between each group. Then four of the folds are used to train the model, and the model is used to make predictions on the held aside fold. This procedure is repeated five times, with each fold serving as the held aside test set one time.

To predict whether a given person belongs to the OA group, an L1-regularisation logistic regression (LR)[21] was trained on the training set and tested on the unseen test set[29]. LR returns a score between 0 and 1, representing the likelihood that an individual belongs to the OA group[30] (figure 2). If the probability score is below 0.5, the subject is predicted as belonging to the asymptomatic control. Otherwise, it is predicted as belonging to the OA patient group. The input of the model consists of the features constructed by TSFuse. Since we use regularised logistic regression, all features are normalised. All participants with missing values for the exercise that was analysed were removed.

The L1-regularization strength of the LR model was tuned based on the training data of each fold. To select a reduced number of features, we compared different values for the regularisation strength hyperparameter and selected the value that resulted in the lowest number of non-zero coefficients but still had a good area under the receiver operator characteristic curve (AUC) (i.e. the smallest C to the right of the largest change in AUC). Specifically, we trained LR models for ten different values of the inverse regularisation strength C, spaced logarithmically between 0.01 and 1. For each C, we evaluated the AUC using an inner 5-fold cross-validation procedure. As illustrated in Figure 3, we searched for the largest drop in AUC (computed as the difference in AUC between each consecutive pair of C values) and selected the smallest C to the right of this drop.

The usefulness of the LR model is evaluated in two different ways. First, the performance of the LR model is analysed by computing the accuracy, AUC, recall, miss, and fallout. Accuracy measures the model's ability to differentiate between asymptomatic controls and OA patients. The AUC is another standard measure of a model's classification ability[31]. Fallout represents the number of false alarms the model gives, i.e. when an asymptomatic individual is misclassified as symptomatic. The recall denotes how many of the patients are correctly classified by the model. The miss indicates the proportion of patients that are misclassified as asymptomatic controls, i.e. the patients that are missed by the model. Preferably, fallout and miss are low, recall and accuracy high. However, it is most often a trade-off between the metrics. Those metrics can be calculated using the true positive (TP), false positive (FP), true negative (TN), and false-negative (FN) predictions. Second, the features selected by the LR model are considered.

Results

Classification accuracies

The average classification results are shown in Table 3 and Table 4 for HipOA-vs-Asymptomatic and KneeOA-vs-Asymptomatic, respectively. The best performing classifier for HipOA-vs-Asymptomatic was found using class 2 of the statistical models during ascending stairs (Table 3). Overall accuracy was 0.970, fallout = 0, which means no asymptomatic controls were misclassified as symptomatic, 92.3% of the HipOA subjects were correctly classified (recall = 0.923) and 7.7% were misclassified as asymptomatic (miss = 0.076). Using kinematics only (class 1) a slightly lower overall accuracy was obtained during ascending stairs and gait. During ascending stairs, the accuracy was 0.939, with a slightly higher number of HipOA patients that were misclassified (miss = 0.143). At the same time, all asymptomatic controls were still correctly classified (fallout = 0). During gait, the accuracy is slightly lower (accuracy = 0.879). However, the recall was higher

(0.900), and miss was lower (0.100). That means that it was better at classifying HipOA patients; however, it performed poorer on the asymptomatic controls (fallout = 0.130). The most important time series to distinguish between HipOA-vs-Asymptomatic using class 1 during ascending stairs are from the lumbar spine and hip kinematics (Table S2).

To distinguish KneeOA-vs-Asymptomatic controls, we found perfect classification accuracies using kinematical data (class 1) during ascending stairs (Table 4). Overall accuracy = 1, fallout = 0, recall = 1, and miss = 0. Indicating that the model misclassified no asymptomatic controls and KneeOA subjects. No other activity performed as well as ascending stairs. The most important time series to distinguish between KneeOA-vs-Asymptomatic during ascending stairs were features derived from the lumbar spine and hip kinematics; particularly, hip flexion, hip adduction and lumbar extension. (Table S3). In particular, it was the variance between the different repetitions that was important.

Discussion

This study aims to evaluate whether we can use biomechanical data collected from a specific activity of daily living to help distinguish hip OA patients from controls and knee OA patients from controls using features that potentially could be measured in a clinical setting. The three different classes of statistical models applied in this study contained different levels of complexities: (1) kinematics only (clinically applicable); (2) includes joint kinetics (semi clinically applicable); (3) using advanced musculoskeletal modelling (not clinically applicable). To independently classify both OA groups from controls (hipOA-vs-Asymptomatic and kneeOA-vs-Asymptomatic),

ascending stairs using class 1 of the statistical models was most accurate also showing the generalisability of the approach across two different populations.

The kinematics only statistical model distinguished hip OA patients from asymptomatic controls with an overall accuracy of 93.9% using data collected during ascending stairs. Gait showed a slightly lower accuracy; however, the recall increased. Indicating that gait is marginally better at classifying HipOA patients but performs poorer on asymptomatic individuals. Knee OA patients could be perfectly distinguished from asymptomatic controls using kinematical data during ascending stairs.

To indicate whether a classification model is accurate enough depends on the number of subjects per group. A good classification model should always perform better than simply predicting the class that is represented most. For our HipOA-vs-Asymptomatic model, it should outperform the overall accuracy of 0.64. Our KneeOA-vs-Asymptomatic model should reach at least 0.59, and in both cases, our model far outperformed that threshold. Furthermore, our HipOA-vs-Asymptomatic model performed similarly to the model of Laroche et al. (2014) that reported accuracy levels between 93% and 97%[14]. However, we used a far simpler LR model, which has two advantages over an SVM model. First, by using an L1-regularization instead of an L2-regularization, we use less input data, which improves the interpretability of our model. Secondly, the LR model gives a probabilistic outcome, indicating how confident the model is that a subject belongs to a specific group. This probability score could, in theory, be used in the follow-up of patients. Changes in the probability of a person belonging to the OA class after an intervention might mean that the intervention was successful with regards to changing the biomechanical pattern. However, how well the classification model is able to evaluate progression after an intervention should be investigated in future work.

Our KneeOA-vs-Asymptomatic model surpassed the classification accuracies of models already existing in the literature [13,15–19]. Jones et al. (2008) showed classification accuracies of 97.62%, only minimally lower than our classification accuracy[17]. However, one disadvantage of Jones' model is the need for force plate data, limiting the use of that model in the clinical setting. Considering that we only need kinematical data makes the translation to the clinic easier. However, in order to obtain high accuracies using our LR-model, the variance between the repetitions is of importance. Therefore, measuring the activity multiple times is necessary.

There are still some limitations to the applicability of this study in the clinic. We only included a limited number of subjects in this study. By only using such a small number of subjects, we obtained a "wide data set", i.e., more features than subjects, risking the overfitting of our model. We reduced the risk of overfitting by only including the most relevant features using regularisation. In future research, we will add more subjects to our machine learning model to investigate whether there is an added benefit of more complex input data. Moreover, the inclusion of additional subjects will improve the generalisability of the model. Furthermore, we included unilateral end-stage hip and knee OA patients. Therefore, at this point, we cannot generalise the results to populations with lower KL grades or people suffering from bi-lateral OA. Considering that previous research found significant differences in gait patterns between people with mild to moderate OA and asymptomatic individuals and between different stages of OA severity (e.g. Astephen et al. 2007; Foucher et al. 2012), models need to be trained and tested using appropriate input data (i.e. lower KL-grades or multi-joint OA). Even though our results show that kinematic data alone are sufficient to distinguish OA patients from asymptomatic controls, we still used laboratory-based data as input. Recognising that differences in absolute angles were previously reported between IMU data and lab-based[32,33], we are uncertain how well our model performs with kinematic data derived from IMU sensors. Good results have been found by Teufl et al. (2019) that used kinematics derived

from IMU sensors to distinguish healthy gait from postoperative hip arthroplasty gait—indicating that it is possible to use IMU derived kinematics in a classification problem with good accuracy and validity[34]. Alternatively, future research could also focus on determining which features derived from raw acceleration data are useful in classifying OA patients from asymptomatic controls using a machine learning pipeline. In running and gait research, many variables have been identified to be able to distinguish between patient populations and asymptomatic controls using data derived from the IMU sensor[29]. However, if they can be used to classify new unseen patients is still largely undetermined. Lastly, we excluded participants with a BMI above 30 kg/m^2. As BMI is an important risk factor for OA, this might influence the generalisability of our model.

In conclusion, features derived from the lumbar spine and hip kinematics during ascending stairs are sufficient in classifying HipOA-vs-Asymptomatics and KneeOA-vs-Asymptomatic controls. However, to enable clinical use, we need to validate if IMU-based kinematical data works as well and whether the probabilistic outcome of the logistic regression model can be used in the follow-up of patients with OA.

Acknowledgements

This research was funded by Research Foundation Flanders (FWO) grant number T004716N. The authors declared that they have no competing interests.

References

1. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. *Bull World Health Organ*. 2003;81(9):646-656. doi:S0042-96862003000900007 [pii]

2. Felson DT. Osteoarthritis as a disease of mechanics. *Osteoarthr Cartil*. 2013;21(1):10-15. doi:10.1016/j.joca.2012.09.012

3. Mills K, Hunt MA, Ferber R. Biomechanical deviations during level walking associated with knee osteoarthritis: A systematic review and meta-analysis. *Arthritis Care Res*. 2013;65(10):1643-1665. doi:10.1002/acr.22015

4. Meyer CAG, Corten K, Fieuws S, et al. Biomechanical Gait Features Associated With Hip Osteoarthritis : Towards a Better Definition of Clinical Hallmarks. 2015;(October):1498-1507. doi:10.1002/jor.22924

5. 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):617-622. doi:10.1136/ard.61.7.617

6. Tateuchi H, Koyama Y, Akiyama H, et al. Daily cumulative hip moment is associated with radiographic progression of secondary hip osteoarthritis. *Osteoarthr Cartil*. 2017;25(8):1291-1298. doi:10.1016/j.joca.2017.02.796

7. Meyer CAG, Wesseling M, Corten K, et al. Hip movement pathomechanics of patients with hip osteoarthritis aim at reducing hip joint loading on the osteoarthritic side. *Gait Posture*. 2018;59:11-17. doi:10.1016/j.gaitpost.2017.09.020

8. Diamond LE, Hoang HX, Barrett RS, et al. Individuals with mild-to-moderate hip osteoarthritis walk with lower hip joint contact forces despite higher levels of muscle co-contraction compared to healthy individuals. *Osteoarthr Cartil*. 2020;28(7):924-931. doi:10.1016/j.joca.2020.04.008

9. Shull PB, Shultz R, Silder A, et al. Toe-in gait reduces the first peak knee adduction moment in patients with medial compartment knee osteoarthritis. *J Biomech*. 2013;46(1):122-128. doi:10.1016/J.JBIOMECH.2012.10.019

10. Ardestani MM, Chen Z, Wang L, et al. A neural network approach for determining gait modifications to reduce the contact force in knee joint implant. *Med Eng Phys*. 2014;36(10):1253-1265. doi:10.1016/j.medengphy.2014.06.016

11. 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. doi:10.1016/J.JBIOMECH.2018.09.009

12. Figueiredo J, Santos CP, Moreno JC. Automatic recognition of gait patterns in human motor disorders using machine learning: A review. *Med Eng Phys*. 2018;53:1-12. doi:10.1016/j.medengphy.2017.12.006

13. Mezghani N, Husse S, Boivin K, et al. Automatic classification of asymptomatic and osteoarthritis knee gait patterns using kinematic data features and the nearest neighbor classifier. *IEEE Trans Biomed Eng*. 2008;55(3):1230-1232. doi:10.1109/TBME.2007.905388

14. Laroche D, Tolambiya A, Morisset C, et al. A classification study of kinematic gait trajectories in hip osteoarthritis. *Comput Biol Med*. 2014;55:42-48. doi:10.1016/j.compbiomed.2014.09.012

15. Beynon MJ, Jones L, Holt CA. Classification of osteoarthritic and normal knee function using three-dimensional motion analysis and the Dempster-Shafer theory of evidence. *IEEE Trans Syst Man, Cybern Part ASystems Humans*. 2006. doi:10.1109/TSMCA.2006.859098

16. Jones L, Beynon MJ, Holt CA, Roy S. An application of the Dempster–Shafer theory of evidence to the classification of knee function and detection of improvement due to total

knee replacement surgery. *J Biomech*. 2006;39(13):2512-2520. doi:10.1016/J.JBIOMECH.2005.07.024

17. Jones L, Holt CA, Beynon MJ. Reduction, classification and ranking of motion analysis data: an application to osteoarthritic and normal knee function data. *Comput Methods Biomech Biomed Engin*. 2008;11(1):31-40. doi:10.1080/10255840701550956

18. Deluzio KJ, Astephen JL. Biomechanical features of gait waveform data associated with knee osteoarthritis. An application of principal component analysis. *Gait Posture*. 2007;25(1):86-93. doi:10.1016/j.gaitpost.2006.01.007

19. Kotti M, Duffell LD, Faisal AA, McGregor AH. Detecting knee osteoarthritis and its discriminating parameters using random forests. *Med Eng Phys*. 2017;43:19-29. doi:10.1016/j.medengphy.2017.02.004

20. Komnik I, Weiss S, Fantini Pagani CH, Potthast W. Motion analysis of patients after knee arthroplasty during activities of daily living - A systematic review. *Gait Posture*. 2015;41(2):370-377. doi:10.1016/j.gaitpost.2015.01.019

21. Fregly BJ, Besier TF, Lloyd DG, et al. Grand challenge competition to predict in vivo knee loads. *J Orthop Res*. 2012;30(4):503-513. doi:10.1002/jor.22023

22. Davis RB, Õunpuu S, Tyburski D, Gage JR. A gait analysis data collection and reduction technique. *Hum Mov Sci*. 1991;10(5):575-587. doi:10.1016/0167-9457(91)90046-Z

23. Ehrig RM, Taylor WR, Duda GN, Heller MO. A survey of formal methods for determining functional joint axes. *J Biomech*. 2007;40(10):2150-2157. doi:10.1016/j.jbiomech.2006.10.026

24. Meireles S, De Groote F, Van Rossom S, Verschueren S, Jonkers I. Differences in knee adduction moment between healthy subjects and patients with osteoarthritis depend on the knee axis definition. *Gait Posture*. 2017;53:104-109. doi:10.1016/J.GAITPOST.2017.01.013

25. Van Campen A, De Groote F, Bosmans L, Scheys L, Jonkers I, De Schutter J. Functional knee axis based on isokinetic dynamometry data: Comparison of two methods, MRI validation, and effect on knee joint kinematics. *J Biomech*. 2011;44(15):2595-2600. doi:10.1016/J.JBIOMECH.2011.08.022

26. De Groote F, De Laet T, Jonkers I, De Schutter J. Kalman smoothing improves the estimation of joint kinematics and kinetics in marker-based human gait analysis. *J Biomech*. 2008;41(16):3390-3398. doi:10.1016/J.JBIOMECH.2008.09.035

27. Steele KM, DeMers MS, Schwartz MH, Delp SL. Compressive tibiofemoral force during crouch gait. *Gait Posture*. 2012;35(4):556-560. doi:10.1016/J.GAITPOST.2011.11.023

28. De Brabandere A, Robberechts P, Op T, Beéck D, Davis J. *Automating Feature Construction for Multi-View Time Series Data*. https://lirias.kuleuven.be/retrieve/550593. Accessed December 18, 2020.

29. Pedregosa FABIANPEDREGOSA F, Michel V, Grisel OLIVIERGRISEL O, et al. *Scikit-Learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot*. Vol 12.; 2011. http://scikit-learn.sourceforge.net. Accessed December 4, 2019.

30. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform*. 2002;35(5-6):352-359.

doi:10.1016/S1532-0464(03)00034-0

31. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29-36. doi:10.1148/radiology.143.1.7063747

32. Poitras I, Dupuis F, Bielmann M, et al. Validity and Reliability of Wearable Sensors for Joint Angle Estimation: A Systematic Review. *Sensors (Basel)*. 2019;19(7). doi:10.3390/s19071555

33. Cuesta-Vargas AI, Galán-Mercant A, Williams JM. The use of inertial sensors system for human motion analysis. *Phys Ther Rev*. 2010;15(6):462-473. doi:10.1179/1743288X11Y.0000000006

34. Teufl, Taetz, Miezal, et al. Towards an Inertial Sensor-Based Wearable Feedback System for Patients after Total Hip Arthroplasty: Validity and Applicability for Gait Classification with Gait Kinematics-Based Features. *Sensors*. 2019;19(22):5006. doi:10.3390/s19225006

List of figures

Figure 1 Schematic summary of the methodological analysis; The first step is the data acquisition consisting of lab-based data collection and calculating the parameters of interest using a musculoskeletal workflow. The second step is building the machine learning pipeline. Automatic feature construction from the time series is performed with TSFresh. After that, the complete data set is run through a 5-fold cross-validation method. The first step in the kth-fold is to split the data set in a training (80%), and a test (20%) set containing the same percentage of healthy controls and OA patients. On the training set, a classification model is trained using the following

steps: (1) normalising features (2) feature selection using hyperparameter tuning of the L1-regularization strength, (3) train the logistic regression model. the model is evaluated on the unseen test set. The classification model is evaluated on the unseen test set. This procedure was repeated for every fold.

Figure 2: Schematic representation of the sigmoid function of the logistic regression analysis and class allocation.

Figure 3: Hyperparameter tuning for the regularisation strength. The selected regularisation strength C is the smallest C to the right of the largest change in AUC.