

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

Peer-reviewed author version

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 (2022) Joint kinematics alone can distinguish hip or knee osteoarthritis patients from asymptomatic controls with high accuracy. In: JOURNAL OF ORTHOPAEDIC RESEARCH, 40 (10) , p. 2229-2239.

DOI: 10.1002/jor.25269

Handle: <http://hdl.handle.net/1942/36622>

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

3 **Authors**

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

7 **Affiliations**

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

17 **Corresponding author:**

18 Jill S. Emmerzaal:

19 Tervuursevest 101 – box 1501, 3001 Leuven, Belgium

20 jill.emmerzaal@kuleuven.be

21 +3216373738

22 **Running title**

23 Classifying osteoarthritis patients and controls

24 **Author contribution:**

25 Study concept and design: Jill Emmerzaal, Rob van der Straaten, Liesbet De Baets, Kristoff Corten,

26 Jesse Davis, Ilse Jonkers, Annick Timmermans, Benedicte Vanwanseele

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

28 Brabandere

29 Results interpretation: Jill Emmerzaal, Sam Van Rossom, Arne De Brabandere, Jesse Davis,

30 Annick Timmermans, Ilse Jonkers, Benedicte Vanwanseele

31 Manuscript draft: Jill Emmerzaal

32 Critical review, edit and approval of manuscript: All authors

33

34 **Abstract:**

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

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

## 54 **Introduction**

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

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

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

77 Secondly, they could be used to evaluate a person's progress towards a normalised movement  
78 pattern after a surgical intervention or a rehabilitation program<sup>12</sup>. Suppose the classification of a  
79 person changes from OA to asymptomatic; one might conclude with some degree of certainty that  
80 the intervention was successful with regards to changing the biomechanical pattern to resemble  
81 that of an asymptomatic individual. However, first, an accurate model needs to be created that  
82 could be used in a clinical setting. Previous research used gait kinematics and kinetics to classify  
83 OA patients from asymptomatic controls and found classification accuracies ranging from 45% to  
84 97.62% using various methodologies<sup>13-19</sup>. Until now, Jones et al. (2008) developed the most  
85 accurate classification method with an in- and out-sample classification accuracy of 97.62%<sup>17</sup> to  
86 classify knee OA patients. They used 12 principal components derived from kinematics and ground  
87 reaction force waveforms, as well as spatiotemporal and anthropometric data as input in their  
88 classification process. However, the need for force plate data complicates translating this approach  
89 to a clinical setting. To classify hip OA subjects, Laroche et al. (2014) found accuracies between  
90 93% and 97% using Support Vector Machines (SVM)<sup>14</sup>. However, one of the drawbacks to SVM  
91 is that it does not show uncertainty, which might make it harder to monitor changes.

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

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

107 To classify hip and knee OA patients in a clinical setting, one needs features that could be derived  
108 from inexpensive mobile sensors (e.g. kinematical data using inertial measurement units (IMUs)).  
109 However, to date, most classification models still require force plate data to classify OA patients  
110 from controls. Therefore, we want to explore the minimum number of movement features derived  
111 from activities of daily living that enable classification of (either hip or knee) OA patient groups  
112 from asymptomatic controls, indicating which parameters are of interest to measure and thereby  
113 proving the generalisability of the approach. We consider three different classes of statistical  
114 models with different levels of data collection and processing complexity. Class 1 is kinematics  
115 based only (clinical applicable), class 2 includes joint kinetics (which would still rely on access to  
116 a force plate or use of machine learning-based prediction models<sup>21</sup>), and class 3 uses data from  
117 advanced musculoskeletal modelling (not clinically applicable). Moreover, as the gait pattern  
118 might not be sensitive enough, we will create these three statistical models using different exercises  
119 with varying difficulty levels. Accordingly, this study aims to evaluate whether we can use  
120 biomechanical data collected from a specific activity of daily living to help distinguish either hip  
121 or knee OA patients from asymptomatic controls. Thus, exploring the use of a general workflow  
122 for both patient groups that could potentially be used in a clinical setting.

## 123 **Methods**

### 124 *Participants*

125 This is a controlled laboratory study in which 51 people participated: 12 asymptomatic controls,  
126 20 unilateral end-stage hip OA patients and 19 unilateral end-stage knee OA patients (see Table 1  
127 for participant characteristics). This is an explorative study based on a secondary analysis of a more  
128 extensive prospective follow-up (S59857) that evaluated hip and knee joint contact forces in people  
129 with hip or knee osteoarthritis and following total knee arthroplasty. The sample size for that study  
130 was based on joint contact forces measured in people with an instrumented knee prosthesis (1.61  
131 +/- 0.305 bodyweight during gait)<sup>22</sup>. Assuming that a difference of one standard deviation is  
132 significant and to achieve a power of 0.8, a sample size of 14 subjects per group is needed. Taking  
133 a possible loss to follow-up of 15-20% into account, we recruited 18-20 participants per group.  
134 Because for the larger study both legs of the asymptomatic controls were considered as  
135 independent, only 12 asymptomatic control subjects were recruited.

136 Patients awaiting a joint replacement surgery (Kellgren-Lawrence grade III (N=1)-IV (N=38))  
137 were recruited from two local hospitals (Ziekenhuis Oost Limburg, Genk and Jessa Hospital,  
138 Hasselt, Belgium). Inclusion criteria for the patients were: age between 50 and 75 years; unilateral  
139 hip or knee OA; BMI<30kg/m<sup>2</sup>; able to walk 10 meters; able to navigate stairs; no corticosteroid  
140 injection at least three months prior to inclusion; no joint replacement in other lower limb joints;  
141 no neurological or musculoskeletal disorders that could affect the movement pattern and no history  
142 of pathological osteoporotic fractures. Asymptomatic controls were included aged between 50 and  
143 75 years old; they had a BMI < 30kg/m<sup>2</sup>, no neurological or musculoskeletal disorders that could  
144 affect their movement pattern and were recruited from a local senior's network in Leuven, Belgium.  
145 Table 1 summarises the participant demographics and the patient-reported outcome of the



146 Hip/Knee Disability and Osteoarthritis Outcome Score (HOOS or KOOS). Separate analysis for  
147 the group differences between the patient reported outcome scores were calculated using an  
148 independent samples t-test, significance level was set at 0.05. The local ethics committee of the  
149 academic hospital Leuven approved the study protocol (s-59857), and the participants provided  
150 written informed consent before the start of the study.

### 151 *Study protocol*

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

160 Reflective markers were attached to each participant conforming to the full-body Plug-in Gait  
161 (Oxford Metrics), including 38 markers<sup>23</sup>. In addition, three marker clusters replaced the single  
162 markers on the segments (shank, thigh, upper and lower arms), and one extra three-marker cluster  
163 was placed on the superior aspect of the left iliac crest of the pelvis. We placed additional  
164 anatomical markers on the sacrum, medial femur epicondyle and medial malleoli<sup>23</sup>. 3D marker  
165 trajectories were collected using 13 optoelectronic cameras (Vicon, Oxford Metrics, UK, 100Hz).  
166 Ground reaction forces (GRF) were measured synchronously using three ground-embedded force  
167 plates (AMTI, Watertown, MA, USA, sampling at 1000Hz).

168 *Musculoskeletal modelling*

169 The motion capture data were processed using a standard, musculoskeletal modelling workflow  
170 implemented in OpenSim3.3<sup>24</sup>. We used the generic model gait2392 (—model specifications and  
171 comprehensive information on the musculoskeletal workflow, see Supplementary material B). The  
172 generic OpenSim model was scaled to match the body dimensions and bodyweight of the subject  
173 by using a measurement-based scaling approach within OpenSim. A personalised knee joint axis  
174 orientation and position were implemented within each scaled model to allow for a more complex  
175 knee joint description. The functional axis of rotation was calculated using the SARA algorithm  
176 based on a standing flexion/extension range of motion task<sup>25,26</sup>. All other coordinate frames were  
177 used directly within the OpenSim model. After that, joint angles were derived from the measured  
178 marker trajectories using the Kalman smoothing algorithm described by De Groote et al. (2008)  
179 and available from SimTK<sup>27</sup>. The Kalman smoother algorithm is an alternative to the standard  
180 Inverse Kinematics Tool available in OpenSim, which improves the estimation of the joint  
181 kinematics and kinetics by using prior knowledge with the measured marker trajectories while  
182 minimising the estimation error statistically<sup>27</sup>. Joint moments were calculated using a standard  
183 inverse dynamic approach available in OpenSim. Afterwards, muscle forces and activations were  
184 calculated using a static optimisation routine that minimised the total muscle activation squared.  
185 Finally, joint contact forces were calculated using the vector sum of the estimated muscle forces  
186 and reaction forces in the joint using the standard OpenSim pipeline<sup>28</sup>.

187 *Feature construction*

188 For each subject and each exercise, features were derived from the kinematics, kinetics and contact  
189 force time series using a multivariate feature construction tool, TSFuse<sup>29</sup>. Table 2 shows the  
190 variables of which the time series were used as input. The joint moments and joint contact forces

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

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

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

#### 204 *Statistical analysis*

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

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

218 To predict whether a given person belongs to the OA group, an L1-regularisation logistic regression  
219 (LR)<sup>30</sup> was trained on the training set and tested on the unseen test set<sup>31</sup>. LR returns a score between  
220 0 and 1, representing the likelihood that an individual belongs to the OA group<sup>32</sup> (figure 2). If the  
221 probability score is below 0.5, the subject is predicted as belonging to the asymptomatic control.  
222 Otherwise, it is predicted as belonging to the OA patient group. The input of the model consists of  
223 the features constructed by TSFuse. Since we use regularised logistic regression, all features are  
224 normalised. All participants with missing values for the exercise that was analysed were removed.

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

234 The usefulness of the LR model is evaluated in two different ways. First, the performance of the  
235 LR model is analysed by computing the accuracy, AUC, recall, miss, and fallout. Accuracy  
236 measures the model's ability to differentiate between asymptomatic controls and OA patients. The  
237 AUC is another standard measure of a model's classification ability<sup>33</sup>. Fallout represents the  
238 number of false alarms the model gives, i.e. when an asymptomatic individual is misclassified as  
239 symptomatic. The recall denotes how many of the patients are correctly classified by the model.  
240 The miss indicates the proportion of patients that are misclassified as asymptomatic controls, i.e.  
241 the patients that are missed by the model. Preferably, fallout and miss are low, recall and accuracy  
242 high. However, it is most often a trade-off between the metrics. Those metrics can be calculated  
243 using the true positive (TP), false positive (FP), true negative (TN), and false-negative (FN)  
244 predictions. Second, the features selected by the LR model are considered.

## 245 **Results**

### 246 *Classification accuracies*

247 The average classification results are shown in Table 3 and Table 4 for HipOA-vs-Asymptomatic  
248 and KneeOA-vs-Asymptomatic, respectively. The best performing classifier for HipOA-vs-  
249 Asymptomatic was found using class 2 of the statistical models during ascending stairs (Table 3).  
250 Overall accuracy was 0.970, fallout = 0, which means no asymptomatic controls were misclassified  
251 as symptomatic, 92.3% of the HipOA subjects were correctly classified (recall = 0.923) and 7.7%  
252 were misclassified as asymptomatic (miss = 0.076). Using kinematics only (class 1) a slightly lower  
253 overall accuracy was obtained during ascending stairs and gait. During ascending stairs, the  
254 accuracy was 0.939, with a slightly higher number of HipOA patients that were misclassified (miss  
255 = 0.143). At the same time, all asymptomatic controls were still correctly classified (fallout = 0).  
256 During gait, the accuracy is slightly lower (accuracy = 0.879). However, the recall was higher

257 (0.900), and miss was lower (0.100). That means that it was better at classifying HipOA patients;  
258 however, it performed poorer on the asymptomatic controls (fallout = 0.130). The most important  
259 time series to distinguish between HipOA-vs-Asymptomatic using class 1 during ascending stairs  
260 are from the lumbar spine and hip kinematics (Table S2).

261 To distinguish KneeOA-vs-Asymptomatic controls, we found perfect classification accuracies  
262 using kinematical data (class 1) during ascending stairs (Table 4). Overall accuracy = 1, fallout =  
263 0, recall = 1, and miss = 0. Indicating that the model misclassified no asymptomatic controls and  
264 KneeOA subjects. No other activity performed as well as ascending stairs. The most important time  
265 series to distinguish between KneeOA-vs-Asymptomatic during ascending stairs were features  
266 derived from the lumbar spine and hip kinematics; particularly, hip flexion, hip adduction and  
267 lumbar extension. (Table S3). In particular, it was the variance between the different repetitions  
268 that was important.

## 269 **Discussion**

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

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

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

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

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

307 There are still some limitations to the applicability of this study in the clinic. We only included a  
308 limited number of subjects in this study. By only using such a small number of subjects, we  
309 obtained a “wide data set”, i.e., more features than subjects, risking the overfitting of our model.  
310 We reduced the risk of overfitting by only including the most relevant features using regularisation.  
311 However, we did see a decrease in the test accuracies compared to the training accuracies for some  
312 of the activities however this was not the case for our best classifier. Therefore, in future research,  
313 when other activities are included, more subjects need to be added to our machine learning model  
314 to investigate whether there is an added benefit of more complex input data. Moreover, the  
315 inclusion of additional subjects will improve the generalisability of the model. Furthermore, we  
316 included unilateral end-stage hip and knee OA patients. Therefore, at this point, we cannot  
317 generalise the results to populations with lower KL grades or people suffering from bi-lateral OA.  
318 Considering that previous research found significant differences in gait patterns between people  
319 with mild to moderate OA and asymptomatic individuals and between different stages of OA  
320 severity (e.g. Astephen et al. (2008)<sup>34</sup>; Foucher et al. (2012)<sup>35</sup>), models need to be trained and tested  
321 using appropriate input data (i.e. lower KL-grades or multi-joint OA). Even though our results  
322 show that kinematic data alone are sufficient to distinguish OA patients from asymptomatic  
323 controls, we still used laboratory-based data as input. Recognising that differences in absolute



324 angles were previously reported between IMU data and lab-based<sup>36,37</sup>, we are uncertain how well  
325 our model performs with kinematic data derived from IMU sensors. Good results have been found  
326 by Teufl et al. (2019) that used kinematics derived from IMU sensors to distinguish healthy gait  
327 from postoperative hip arthroplasty gait—indicating that it is possible to use IMU derived  
328 kinematics in a classification problem with good accuracy and validity<sup>38</sup>. Alternatively, future  
329 research could also focus on determining which features derived from raw acceleration data are  
330 useful in classifying OA patients from asymptomatic controls using a machine learning pipeline.  
331 In running and gait research, many variables have been identified to be able to distinguish between  
332 patient populations and asymptomatic controls using data derived from the IMU sensor<sup>39</sup>.  
333 However, if they can be used to classify new unseen patients is still largely undetermined. Lastly,  
334 we excluded participants with a BMI above 30 kg/m<sup>2</sup>. As BMI is an important risk factor for OA,  
335 this might influence the generalisability of our model.  
336 In conclusion, features derived from the lumbar spine and hip kinematics during ascending stairs  
337 are sufficient in classifying HipOA-vs-Asymptomatics and KneeOA-vs-Asymptomatic controls.  
338 However, to enable clinical use, we need to validate if IMU-based kinematical data works as well  
339 and whether the probabilistic outcome of the logistic regression model can be used in the follow-  
340 up of patients with OA.

#### 341 **Acknowledgements**

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

#### 344 **References**

- 345 1. Woolf AD, Pfleger B. 2003. Burden of major musculoskeletal conditions. Bull. World  
346 Health Organ. 81(9):646–656.

- 347 2. Felson DT. 2013. Osteoarthritis as a disease of mechanics. *Osteoarthr. Cartil.* 21(1):10–5.  
348 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23041436>.
- 349 3. Mills K, Hunt MA, Ferber R. 2013. Biomechanical deviations during level walking  
350 associated with knee osteoarthritis: A systematic review and meta-analysis. *Arthritis Care*  
351 *Res.* 65(10):1643–1665. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23554153>.
- 352 4. Meyer CAG, Corten K, Fieuws S, et al. 2015. Biomechanical gait features associated with  
353 hip osteoarthritis: Towards a better definition of clinical hallmarks. *J. Orthop. Res.*  
354 33(10):1498–1507.
- 355 5. Miyazaki T, Wada M, Kawahara H, et al. 2002. Dynamic load at baseline can predict  
356 radiographic disease progression in medial compartment knee osteoarthritis. *Ann. Rheum.*  
357 *Dis.* 61(7):617–622. Available from: <https://pubmed.ncbi.nlm.nih.gov/12079903/>.
- 358 6. Tateuchi H, Koyama Y, Akiyama H, et al. 2017. Daily cumulative hip moment is  
359 associated with radiographic progression of secondary hip osteoarthritis. *Osteoarthr. Cartil.*  
360 25(8):1291–1298. Available from:  
361 <https://linkinghub.elsevier.com/retrieve/pii/S1063458417308634>.
- 362 7. Meyer CAG, Wesseling M, Corten K, et al. 2018. Hip movement pathomechanics of  
363 patients with hip osteoarthritis aim at reducing hip joint loading on the osteoarthritic side.  
364 *Gait Posture* 59:11–17. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28968547>.
- 365 8. Diamond LE, Hoang HX, Barrett RS, et al. 2020. Individuals with mild-to-moderate hip  
366 osteoarthritis walk with lower hip joint contact forces despite higher levels of muscle co-  
367 contraction compared to healthy individuals. *Osteoarthr. Cartil.* 28(7):924–931.
- 368 9. Shull PB, Shultz R, Silder A, et al. 2013. Toe-in gait reduces the first peak knee adduction

- 369 moment in patients with medial compartment knee osteoarthritis. *J. Biomech.* 46(1):122–  
370 128. Available from:  
371 <https://www.sciencedirect.com/science/article/pii/S0021929012006173>.
- 372 10. Ardestani MM, Chen Z, Wang L, et al. 2014. A neural network approach for determining  
373 gait modifications to reduce the contact force in knee joint implant. *Med. Eng. Phys.*  
374 36(10):1253–1265.
- 375 11. Halilaj E, Rajagopal A, Fiterau M, et al. 2018. Machine learning in human movement  
376 biomechanics: Best practices, common pitfalls, and new opportunities. *J. Biomech.* 81:1–  
377 11. Available from:  
378 <https://www.sciencedirect.com/science/article/pii/S0021929018307309>.
- 379 12. Figueiredo J, Santos CP, Moreno JC. 2018. Automatic recognition of gait patterns in  
380 human motor disorders using machine learning: A review. *Med. Eng. Phys.* 53:1–12.
- 381 13. Mezghani N, Husse S, Boivin K, et al. 2008. Automatic classification of asymptomatic and  
382 osteoarthritis knee gait patterns using kinematic data features and the nearest neighbor  
383 classifier. *IEEE Trans. Biomed. Eng.* 55(3):1230–1232.
- 384 14. Laroche D, Tolambiya A, Morisset C, et al. 2014. A classification study of kinematic gait  
385 trajectories in hip osteoarthritis. *Comput. Biol. Med.* 55:42–48.
- 386 15. Beynon MJ, Jones L, Holt CA. 2006. Classification of osteoarthritic and normal knee  
387 function using three-dimensional motion analysis and the Dempster-Shafer theory of  
388 evidence. *IEEE Trans. Syst. Man, Cybern. Part A Systems Humans* .
- 389 16. Jones L, Beynon MJ, Holt CA, Roy S. 2006. An application of the Dempster–Shafer  
390 theory of evidence to the classification of knee function and detection of improvement due

- 391 to total knee replacement surgery. *J. Biomech.* 39(13):2512–2520 [cited 2019 Jun 18]  
392 Available from:  
393 <https://www.sciencedirect.com/science/article/pii/S0021929005003635?via%3Dihub>.
- 394 17. Jones L, Holt CA, Beynon MJ. 2008. Reduction, classification and ranking of motion  
395 analysis data: an application to osteoarthritic and normal knee function data. *Comput.*  
396 *Methods Biomech. Biomed. Engin.* 11(1):31–40 [cited 2019 Jul 25] Available from:  
397 <http://www.tandfonline.com/doi/abs/10.1080/10255840701550956>.
- 398 18. Deluzio KJ, Astephen JL. 2007. Biomechanical features of gait waveform data associated  
399 with knee osteoarthritis: An application of principal component analysis. *Gait Posture*  
400 25(1):86–93 [cited 2019 Jun 18] Available from:  
401 <https://www.sciencedirect.com/science/article/pii/S0966636206000300?via=ihub>.
- 402 19. Kotti M, Duffell LD, Faisal AA, McGregor AH. 2017. Detecting knee osteoarthritis and its  
403 discriminating parameters using random forests. *Med. Eng. Phys.* 43:19–29.
- 404 20. Komnik I, Weiss S, Fantini Pagani CH, Potthast W. 2015. Motion analysis of patients after  
405 knee arthroplasty during activities of daily living - A systematic review. *Gait Posture*  
406 41(2):370–377.
- 407 21. Boswell MA, Uhlrich SD, Kidziński, et al. 2021. A neural network to predict the knee  
408 adduction moment in patients with osteoarthritis using anatomical landmarks obtainable  
409 from 2D video analysis. *Osteoarthr. Cartil.* 29(3):346–356.
- 410 22. Fregly BJ, Besier TF, Lloyd DG, et al. 2012. Grand challenge competition to predict in  
411 vivo knee loads. *J. Orthop. Res.* 30(4):503–513. Available from:  
412 <http://www.ncbi.nlm.nih.gov/pubmed/22161745>.

- 413 23. Davis RB, Öunpuu S, Tyburski D, Gage JR. 1991. A gait analysis data collection and  
414 reduction technique. *Hum. Mov. Sci.* 10(5):575–587.
- 415 24. Delp SL, Anderson FC, Arnold AS, et al. 2007. OpenSim : Open-Source Software to  
416 Create and Analyze Dynamic Simulations of Movement.54(11):1940–1950.
- 417 25. Ehrig RM, Taylor WR, Duda GN, Heller MO. 2007. A survey of formal methods for  
418 determining functional joint axes. *J. Biomech.* 40(10):2150–2157.
- 419 26. Meireles S, De Groote F, Van Rossom S, et al. 2017. Differences in knee adduction  
420 moment between healthy subjects and patients with osteoarthritis depend on the knee axis  
421 definition. *Gait Posture* 53:104–109. Available from:  
422 <https://www.sciencedirect.com/science/article/pii/S096663621730022X#bib0105>.
- 423 27. De Groote F, De Laet T, Jonkers I, De Schutter J. 2008. Kalman smoothing improves the  
424 estimation of joint kinematics and kinetics in marker-based human gait analysis. *J.*  
425 *Biomech.* 41(16):3390–3398. Available from:  
426 <https://www.sciencedirect.com/science/article/pii/S0021929008004685>.
- 427 28. Steele KM, DeMers MS, Schwartz MH, Delp SL. 2012. Compressive tibiofemoral force  
428 during crouch gait. *Gait Posture* 35(4):556–560. Available from:  
429 <https://www.sciencedirect.com/science/article/pii/S0966636211007910?via%3Dihub>.
- 430 29. De Brabandere A, Robberechts P, Op T, et al. 2019. Automating Feature Construction for  
431 Multi-View Time Series Data. Available from: <https://lirias.kuleuven.be/retrieve/550593>.
- 432 30. Lee S-I, Lee H, Abbeel P, Ng AY. 2006. Efficient L 1 Regularized Logistic Regression.  
433 Available from: [www.aaai.org](http://www.aaai.org).

- 434 31. Pedregosa FABIANPEDREGOSA F, Michel V, Grisel OLIVIERGRISEL O, et al. 2011.  
435 Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent  
436 Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL.  
437 Matthieu Perrot. 2825–2830. Available from: <http://scikit-learn.sourceforge.net>.
- 438 32. Dreiseitl S, Ohno-Machado L. 2002. Logistic regression and artificial neural network  
439 classification models: a methodology review. *J. Biomed. Inform.* 35(5–6):352–359.  
440 Available from: <https://www.sciencedirect.com/science/article/pii/S1532046403000340>.
- 441 33. Hanley JA, McNeil BJ. 1982. The meaning and use of the area under a receiver operating  
442 characteristic (ROC) curve. *Radiology* 143(1):29–36.
- 443 34. Astephen JL, Deluzio KJ, Caldwell GE, et al. 2008. Gait and neuromuscular pattern  
444 changes are associated with differences in knee osteoarthritis severity levels. *J. Biomech.*  
445 41(4):868–876.
- 446 35. Foucher KC, Schlink BR, Shakoor N, Wimmer MA. 2012. Sagittal plane hip motion  
447 reversals during walking are associated with disease severity and poorer function in  
448 subjects with hip osteoarthritis. *J. Biomech.* 45(8):1360–1365.
- 449 36. Poitras I, Dupuis F, Biemann M, et al. 2019. Validity and Reliability of Wearable Sensors  
450 for Joint Angle Estimation: A Systematic Review. *Sensors (Basel)*. 19(7). Available from:  
451 <http://www.ncbi.nlm.nih.gov/pubmed/30935116>.
- 452 37. Cuesta-Vargas AI, Galán-Mercant A, Williams JM. 2010. The use of inertial sensors  
453 system for human motion analysis. *Phys. Ther. Rev.* 15(6):462–473. Available from:  
454 <http://www.ncbi.nlm.nih.gov/pubmed/23565045>.
- 455 38. Teufl, Taetz, Miezal, et al. 2019. Towards an Inertial Sensor-Based Wearable Feedback

456 System for Patients after Total Hip Arthroplasty: Validity and Applicability for Gait  
457 Classification with Gait Kinematics-Based Features. *Sensors* 19(22):5006. Available from:  
458 <https://www.mdpi.com/1424-8220/19/22/5006>.

459 39. Benson LC, Clermont CA, Bošnjak E, Ferber R. 2018. The use of wearable devices for  
460 walking and running gait analysis outside of the lab: A systematic review. *Gait Posture*  
461 63:124–138.

462

463

#### 464 **List of figures**

465 Figure 1 Schematic summary of the methodological analysis; The first step is the data acquisition  
466 consisting of lab-based data collection and calculating the parameters of interest using a  
467 musculoskeletal workflow. The second step is building the machine learning pipeline. Automatic  
468 feature construction from the time series is performed with TSFresh. After that, the complete data  
469 set is run through a 5-fold cross-validation method. The first step in the  $k^{\text{th}}$ -fold is to split the data  
470 set in a training (80%), and a test (20%) set containing the same percentage of healthy controls  
471 and OA patients. On the training set, a classification model is trained using the following steps:  
472 (1) normalising features (2) feature selection using hyperparameter tuning of the L1-  
473 regularization strength, (3) train the logistic regression model. the model is evaluated on the  
474 unseen test set. The classification model is evaluated on the unseen test set. This procedure was  
475 repeated for every fold.

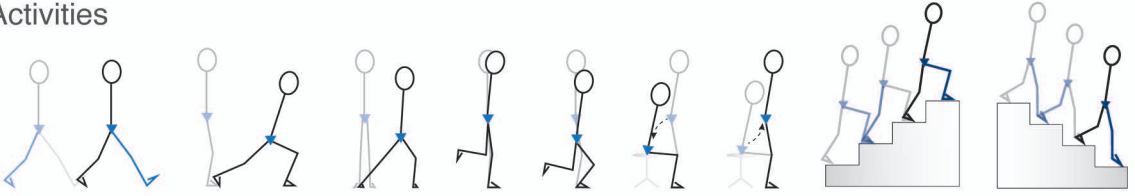
476 Figure 2: Schematic representation of the sigmoid function of the logistic regression analysis and  
477 class allocation.

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



# Data acquisition

## Activities



## Musculoskeletal modeling workflow

Lab data | Scaling | Inverse kinematics | Inverse dynamics | Static optimization | Joint reaction load



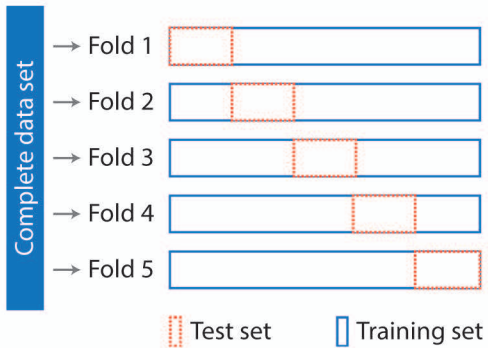
# Machine learning pipeline

## Feature construction TSFuse

- Class 1  
Kinematical features
- Class 2  
Kinematical features  
Kinetics features
- Class 3  
Kinematical features  
Kinetics features  
Reaction force features

## k-fold cross validation procedure

20% test | 80% train split



## Analysis per k-fold per exercise

### Train the classification model

Training set

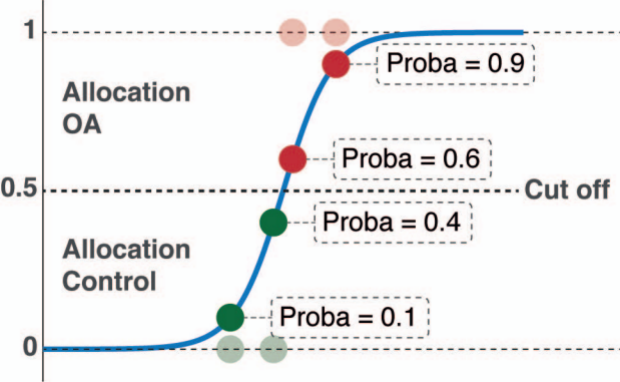
- 1: Normalizing features
- 2: Feature selection by hyperparameter tuning
- 3: Train L1-regularization Logistic Regression model

### Evaluate the classification model

Test set

Classification model

- 4: Test L1-regularization Logistic Regression model
- 5: Calculate Receiver Operator Characteristic



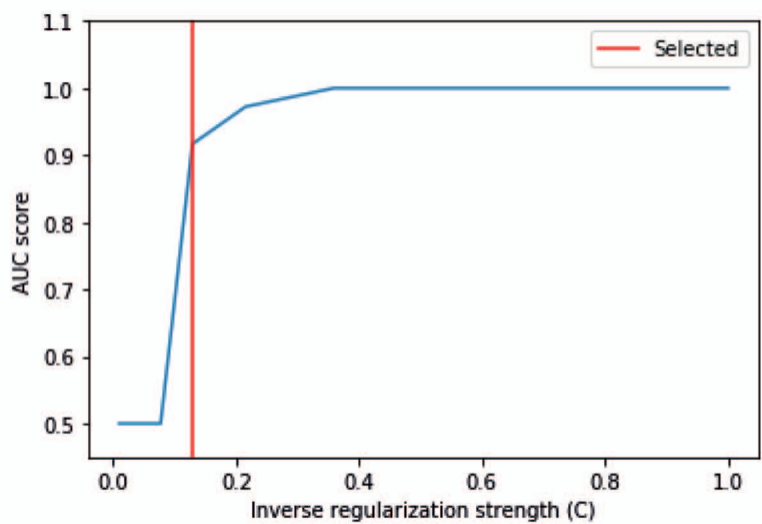


Table 1 Patient demographics and HOOS/KOOS scores mean [+/- 95% confidence interval]. \* indicates a significant difference ( $p < 0.05$ ) between the indicated OA group and the asymptomatic control group

	<b>Asymptomatic</b> N=12	<b>Hip OA</b> N=20	<b>Knee OA</b> N=19	<b>95%CI difference between groups</b>	
Age (yrs)	59.7 [55.3 – 64.2]	63.1 [60.0 – 66.2]	65.1 [62.6 – 67.6]		
Weight (kg)	74.2 [64.8 – 83.7]	75.4 [70.1 – 80.7]	79.8 [75.7 – 83.9]		
Height (m)	1.71 [1.65 – 1.77]	1.75 [1.71 – 1.79]	1.75 [1.71 – 1.79]		
BMI (kg/m)	25.1 [23.0 – 27.3]	24.5 [23.3 – 25.8]	26.0 [24.9 – 27.0]		
Gender/sex	6 female	9 female	7 female		
<b>PROM</b>	<b>KOOS</b>	<b>HOOS</b>	<b>KOOS</b>	<b>Hip OA - asym</b>	<b>knee OA - asym</b>
<i>Pain</i>	95.4 [91.8 - 98.9]	56.8 [48.8 – 64.7]*	50.9 [45.0 – 56.8]*	[29.0-48.8]	[37.8- 51.1]
<i>Symptoms</i>	96.9 [93.7 – 100]	51.8 [42.6 – 60.9]*	52.3 [43.4 – 61.1]*	[34.5- 57.0]	[35.4- 53.9]
<i>ADL</i>	98.7 [97.2 - 100]	57.9 [48.5 – 67.4]*	56.4 [48.8 – 64.1]*	[32.3- 53.6]	[34.5- 50.0]
<i>Sport</i>	93.3 [88.0 – 98.6]	35.8 [23.7 – 47.9]*	24.1 [12.2 – 36.0]*	[44.9- 71.9]	[56.5- 82.0]
<i>QOL</i>	92.1 [97.7 – 86.5]	41.7 [33.8 - 49.5]*	28.0 [20.1 – 35.8]*	[38.4- 59.7]	[54.8- 73.4]

*Table Variables of which time series were used as input for TSFuse*

<b>Kinematics</b>	<b>Kinetics</b>	<b>Contact forces</b>
ankle plantar-dorsiflexion	ankle plantar-dorsiflexion moment	ankle x
knee flexion-extension	knee flexion-extension moment	ankle y
knee ab-adduction	knee ab-adduction moment	ankle z
hip rotation	hip rotation moment	knee x
hip ab-adduction	hip ab-adduction moment	knee y
hip flexion-extension	hip flexion-extension moment	knee z
lumbar rotation		hip x
lumbar bending		hip y
lumbar flexion-extension		hip z

*Table3* Model accuracy to differentiate HipOA-vs-Asymptomatic controls for the different exercises and different combinations of input variables were used. The three different classes are different statistical models with different levels of data complexity. Class 1 is kinematics based only (clinical applicable), class 2 includes force plate data to the kinematics (semi applicable under the condition of, e.g. access to a force plate or prediction models), and class 3 uses data from advanced musculoskeletal modelling (kinematics, kinetics and contact forces; not clinically applicable). Acc is the ratio of how many subjects were predicted correctly to the total number of subjects. Constr. shows the number of features obtained by the feature constriction tool TSFuse and selected shows the number of features that were selected by the Logistic Regression model. TP, FP, FN, and TN are true positive, false positive, false positive and true negative respectively.

<b>HipOA-vs-Asymptomatic</b>												
		acc.	constr.	selected	TP	FP	FN	TN	Fallout	Recall	Miss	AUC
<b>Gait</b>												
	Class 1	0.879	27221	24	9	3	1	20	0.130	0.900	0.100	0.925
	Class 2	0.818	57490	39	9	3	3	18	0.143	0.750	0.250	0.810
	Class 3	0.758	107688	20	8	4	4	17	0.190	0.667	0.333	0.833
<b>Forward lunge</b>												
	Class 1	0.844	29147	22	11	1	4	16	0.059	0.733	0.267	0.925
	Class 2	0.750	64028	17	8	4	4	16	0.200	0.667	0.333	0.808
	Class 3	0.719	115754	25	10	2	7	13	0.133	0.588	0.412	0.788
<b>Sideward Lunge</b>												
	Class 1	0.848	15915	15	11	1	4	17	0.056	0.733	0.267	0.877
	Class 2	0.727	32581	11	10	2	7	14	0.125	0.588	0.412	0.837
	Class 3	0.758	61140	16	8	4	4	17	0.190	0.667	0.333	0.865
<b>Single-leg-stance</b>												
	Class 1	0.818	31076	65	11	1	5	16	0.059	0.688	0.313	0.948
	Class 2	0.818	56779	52	10	2	4	17	0.105	0.714	0.286	0.913
	Class 3	0.788	89404	25	11	1	6	15	0.063	0.647	0.353	0.845
<b>Single-leg-squat</b>												
	Class 1	0.813	27992	21	11	1	5	15	0.063	0.688	0.313	0.875
	Class 2	0.844	54448	21	11	1	4	16	0.059	0.733	0.267	0.912
	Class 3	0.844	95865	15	11	1	4	16	0.059	0.733	0.267	0.883
<b>Sit down</b>												
	Class 1	0.667	13042	31	7	5	6	15	0.250	0.538	0.462	0.758
	Class 2	0.697	27797	22	7	5	5	16	0.238	0.583	0.417	0.698
	Class 3	0.606	49254	32	5	7	6	15	0.318	0.455	0.545	0.679
<b>Stand up</b>												
	Class 1	0.879	13228	13	10	2	2	19	0.095	0.833	0.167	0.804
	Class 2	0.788	26829	12	8	4	3	18	0.182	0.727	0.273	0.780
	Class 3	0.758	48100	36	9	3	5	16	0.158	0.643	0.357	0.752
<b>Ascending stairs</b>												
	Class 1	0.939	28750	6	12	0	2	19	0.000	0.857	0.143	1.000
	Class 2	0.970	63997	6	12	0	1	20	0.000	0.923	0.077	0.996
	Class 3	0.939	119961	7	12	0	2	19	0.000	0.857	0.143	1.000
<b>Descending stairs</b>												
	Class 1	0.939	31042	16	12	0	2	19	0.000	0.857	0.143	0.968
	Class 2	0.909	55563	17	12	0	3	18	0.000	0.800	0.200	0.976
	Class 3	0.909	117300	18	12	0	3	18	0.000	0.800	0.200	0.988

*Table 2* Model accuracy to differentiate KneeOA-vs-Asymptomatic controls for the different exercises and different combinations of input variables were used. The three different classes are different statistical models with different levels of data complexity. Class 1 is kinematics based only (clinical applicable), class 2 includes force plate data to the kinematics (semi applicable under the condition of, e.g. access to a force plate or prediction models), and class 3 uses data from advanced musculoskeletal modelling (kinematics, kinetics and contact forces; not clinically applicable). Accuracy is the ratio of how many subjects were predicted correctly to the total number of subjects. Constructed shows the number of features obtained by the feature constriction tool TSFuse and selected shows the number of features that were selected by the Logistic Regression model. TP, FP, FN, and TN are true positive, false positive, false positive and true negative respectively.

<b>KneeOA-vs-Asymptomatic</b>												
		acc.	constr.	selected	TP	FP	FN	TN	Fallout	Recall	Miss	AUC
<b>Gait</b>												
	Class 1	0.833	27221	31	10	2	3	15	0.118	0.769	0.231	0.889
	Class 2	0.833	57490	31	11	1	4	14	0.067	0.733	0.267	0.894
	Class 3	0.800	107688	15	10	2	4	14	0.125	0.714	0.286	0.843
<b>Forward lunge</b>												
	Class 1	0.586	29147	38	8	4	8	9	0.308	0.500	0.500	0.696
	Class 2	0.517	64028	54	8	4	10	7	0.364	0.444	0.556	0.532
	Class 3	0.690	115754	15	9	3	6	11	0.214	0.600	0.400	0.711
<b>Sideward Lunge</b>												
	Class 1	0.444	15915	82	6	6	9	6	0.500	0.400	0.600	0.561
	Class 2	0.444	32581	29	5	7	8	7	0.500	0.385	0.615	0.489
	Class 3	0.704	61140	32	9	3	5	10	0.231	0.643	0.357	0.683
<b>Single-leg-stance</b>												
	Class 1	0.800	31076	59	10	2	4	14	0.125	0.714	0.286	0.843
	Class 2	0.733	56779	39	9	3	5	13	0.188	0.643	0.357	0.778
	Class 3	0.733	89404	98	9	3	5	13	0.188	0.643	0.357	0.764
<b>Single-leg-squat</b>												
	Class 1	0.724	27992	23	8	4	4	13	0.235	0.667	0.333	0.760
	Class 2	0.621	54448	28	7	5	6	11	0.313	0.538	0.462	0.662
	Class 3	0.655	95865	80	9	3	7	10	0.231	0.563	0.438	0.750
<b>Sit down</b>												
	Class 1	0.483	13042	67	7	5	10	7	0.417	0.412	0.588	0.613
	Class 2	0.448	27797	32	7	5	11	6	0.455	0.389	0.611	0.495
	Class 3	0.448	49254	25	4	8	8	9	0.471	0.333	0.667	0.426
<b>Stand up</b>												
	Class 1	0.690	13228	33	7	5	4	13	0.278	0.636	0.364	0.760
	Class 2	0.724	26829	76	10	2	6	11	0.154	0.625	0.375	0.755
	Class 3	0.690	48100	37	9	3	6	11	0.214	0.600	0.400	0.828
<b>Ascending stairs</b>												
	Class 1	1.000	28750	7	12	0	0	16	0.000	1.000	0.000	1.000
	Class 2	1.000	63997	7	12	0	0	16	0.000	1.000	0.000	1.000
	Class 3	1.000	119961	7	12	0	0	16	0.000	1.000	0.000	1.000
<b>Descending stairs</b>												
	Class 1	0.852	31042	17	11	1	3	12	0.077	0.786	0.214	0.867
	Class 2	0.852	55563	23	11	1	3	12	0.077	0.786	0.214	0.922
	Class 3	0.815	117300	37	11	1	4	11	0.083	0.733	0.267	0.867

## Supplementary materials

A.

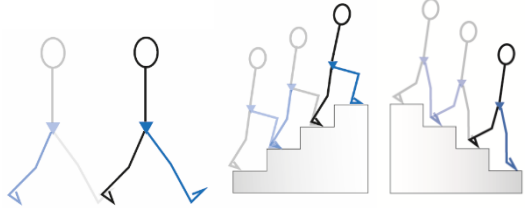
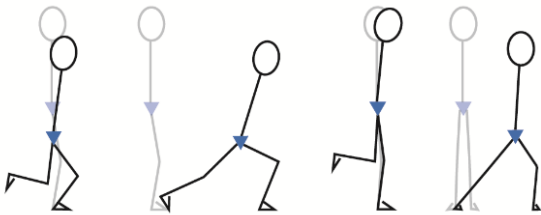


All participants performed five repetitions of nine functional exercises in the Movement and posture Analysis Laboratory of the KU Leuven:

- Level walking; self-selected pace. A trial was successful when the entire foot contacted the force plate.
- Forward lunge and sideward lunge; Step length and step width was standardised at 70% of the leg length. Leg length was defined as the distance from the trochanter major to the floor. The participant was instructed to go as far as possible while remaining stable
- Single-leg stance and single-leg squat; Hands were fixed at the side. For the single-leg squat, the participant was instructed to go as far as possible while remaining stable
- Stand up and sit-down transition; chair height was standardised as knee height. Knee height was defined as the distance from the tibiofemoral joint line to the floor. The participant was instructed not to look back before sitting down.
- Ascending stairs and descending stairs; at self-selected speed without handheld support on a standardised 4-step staircase (step height = 0.16m and tread length = 0.31m). The first, second, and third step was placed on the floor embedded force plates. These steps were loose blocks within the staircase. During the placement of the stairs, special care was taken to put the steps on the force plates properly. For the analysis, the second and third step was analysed.

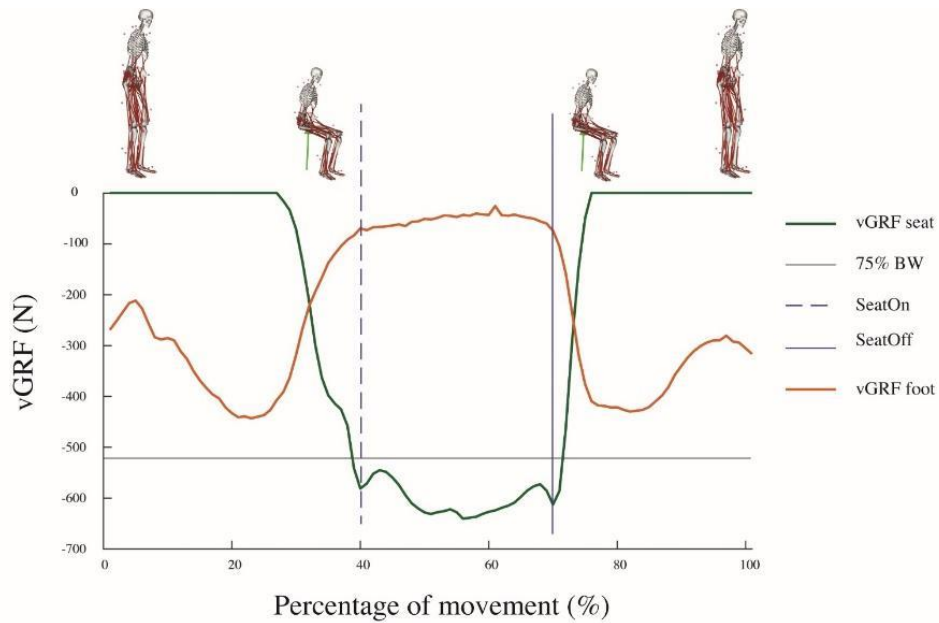


Supplementary materials

Table S1 shows the definitions of the start and end of the trials that were analysed

Exercises	Start	End
 <p data-bbox="204 593 790 660">Level walking; Ascending stairs; Descending stairs</p>	<p data-bbox="842 342 1085 414">Force on force plate exceeds 40N</p>	<p data-bbox="1133 342 1385 448">Consecutive contact of the same foot</p>
 <p data-bbox="204 963 766 1041">Single leg squat; Forward lunge; Single leg stance; Sideward lunge</p>	<p data-bbox="842 683 1101 784">The velocity of the toe marker exceeds 0.15 m/s</p>	<p data-bbox="1133 683 1388 784">The velocity of the toe marker is below 0.15 m/s</p>
 <p data-bbox="204 1366 454 1400">Sit down transition</p>	<p data-bbox="842 1097 1093 1198">The velocity of the clavicular marker exceeds 0.15 m/s</p>	<p data-bbox="1133 1097 1380 1276">The first peak after the force under the seat exceeds 75% bodyweight (see figure 1)</p>
 <p data-bbox="204 1713 454 1747">Stand up transition</p>	<p data-bbox="842 1456 1101 1601">Peak before the force under the seat is below 75% (see figure 1)</p>	<p data-bbox="1133 1456 1388 1556">The velocity of the clavicular marker is below 0.15 m/s</p>

## Supplementary materials



### B.

Gait2392 is a generic lower limb model (no arms)<sup>24</sup>. The original model has 92 musculotendon accentuators that represent 76 muscles of the lower limbs and torso (Figure S2). The lower extremities are represented by seven rigid-body segments (pelvis, femur, tibia, patella, talus, calcaneus, and toe); the torso and head are considered as one. The relative segmental motions are described by the lumbar spine, hip, knee, ankle, subtalar, and metatarsophalangeal joints. We added a DOF for the ab-adduction of the knee joint with a constraint of 20 degrees. We removed the DOFs of the subtalar and metatarsophalangeal joints. The modified model's DOFs were as follows: 6-DOF pelvis (pelvis wrt ground), 3-DOF hip joints (pelvis wrt femur), 2-DOF knee joints (femur wrt tibia), 1-DOF ankle joints (tibia wrt foot), and 3-DOF lumbar spine joint (pelvis wrt torso).

The extended generic model was scaled to match the body dimensions and bodyweight of the subject by using a measurement-based scaling approach within OpenSim. Scaling factors per segment were calculated using the relative distance between the markers positioned on the anatomical landmarks recorded in a static trial and the corresponding virtual markers in the

## Supplementary materials

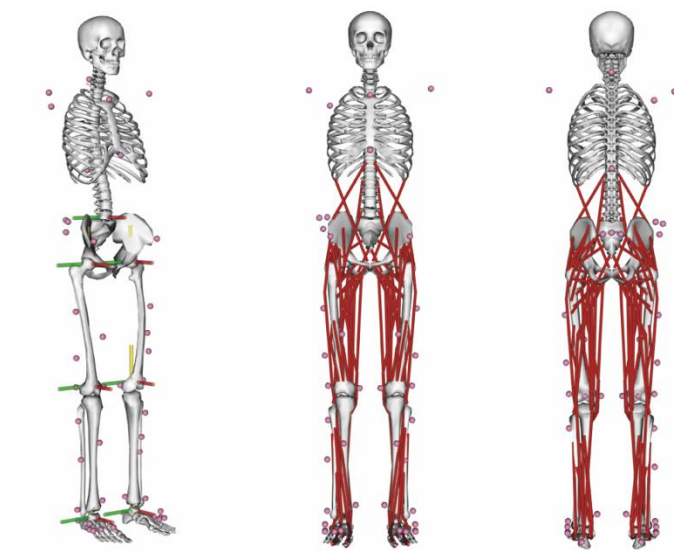


Figure S2 Generic model gait2392 with displayed joint axis and virtual marker positions (far left). Middle and far right are front and back view respectively of the muscles used in this model.

musculoskeletal model. A personalised knee joint axis orientation and position was 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<sup>25,26</sup>. For more details on the SARA algorithm, see Ehrig et al. (2007)<sup>25</sup> and evaluated for use in knee OA patients in Meireles (2017)<sup>26</sup>. In short, the algorithm uses marker trajectory data from the marker on the lateral epicondyle combined with the cluster marker set of the shank and thigh. Based on the marker trajectories throughout the motion, the averaged orientation and position of the knee flexion/extension axis (i.e. functional axis) are calculated in both the femoral reference frame and the tibial reference frame. The knee joint centre is then defined as the intersection of the functional axis and the XY-planes of the femoral reference frame and tibial reference frame. The orientation of the ab/adduction axis was defined as the cross product of the unit vector of the functional axis and the axis pointing from the hip joint centre to the knee joint center<sup>25</sup>. The functional axis of rotation was implemented in each scaled OpenSim model by changing the joint axis definition relative to the femoral and tibial reference frame such that it corresponds to the calculated location and

## Supplementary materials

orientation of the knee joint functional axis<sup>26</sup>. 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<sup>27</sup>. 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 on the measured marker trajectories while minimising the estimation error statistically<sup>27</sup>. The following angles were calculated for each activity: hip flexion/extension, hip ab/adduction, hip internal/external rotation; knee flexion/extension, knee ab/adduction; ankle flexion/extension; lumbar spine flexion/extension, lumbar spine rotations, lumbar spine bending. Considering that the head and thorax are considered as a single segment, all motion from the upper body are assumed to take place in the lumbar spine. We followed the ISB recommendation for kinematic data convention, meaning the left and right hip rotation angles, knee flexion angles, left knee ab/adduction angle calculated by OpenSim were inverted.

Joint moments were calculated using an inverse dynamics approach using the standard implementation in OpenSim. Afterwards, muscle forces and activations were calculated using a static optimisation routine within the OpenSim tool 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 tool<sup>28</sup>.

## Supplementary materials

C.

Table S2: List of features needed to classify HipOA-vs-Asymptomatic control subjects and the number of fold in which they are used. Variance represents the variance of the feature over the five repetitions of that exercise. ArgMax indicates the difference in the location of the maximum in the time curve. Difference represents the newly constructed time curve of the difference between, e.g. lumbar bending and hip flexion. The input signals of each feature are highlighted in bold

Feature	Number of folds
Variance[HighStandardDeviation(Difference(lumbar bending, hip flexion), r=0.3)]	5
Variance[ArgMax(Difference(hip flexion, hip adduction), first=true, rel=true)]	1
Variance[ArgMax(hip flexion, first=true, rel=true)]	1
Variance[ArgMax(hip flexion, first=false, rel=true)]	1
Variance[HighStandardDeviation(Difference(hip flexion, hip adduction), r=0.3)]	1
Variance[ArgMax(Difference(hip flexion, hip adduction), first=false, rel=true)]	1

## Supplementary materials

Table S3: List of features needed to classify KneeOA-vs-Asymptomatic control subjects and the number of fold in which they are used. Variance represents the variance of the feature over the five repetitions of that exercise. ArgMax indicates the difference in the location of the maximum. Difference means the newly constructed time curve of the difference between, e.g. lumbar bending and hip flexion. The input signals of each feature are highlighted in bold.

Feature	Number of folds
Variance[ArgMax(Difference(hip flexion, hip adduction), first=false, rel=true)]	5
Variance[ArgMax(hip flexion, first=true, rel=true)]	5
Variance[ArgMax(Difference(hip flexion, hip adduction), first=true, rel=true)]	5
Variance[ArgMax(hip flexion, first=false, rel=true)]	4
Variance[ArgMax(Difference(lumbar extension, hip flexion), first=true, rel=true)]	1
Variance[ArgMax(Difference(lumbar extension, hip flexion), first=false, rel=true)]	1
Variance[HighStandardDeviation(Difference(hip flexion, hip adduction), r=0.3)]	1