Machine learning-based prediction of knee pain severity over 8 years – Data from the Osteoarthritis Initiative

Teemu A. T. Nurmirinta[1,3], Mikael J. Turunen[1], Jussi Tohka[2], Mimmi K. Liukkonen[3], Mika E. Mononen[1]
[1]Department of Technical Physics, University of Eastern Finland, Kuopio, Finland; [2]AI Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland; [3]Diagnostic Imaging Centre, Kuopio University Hospital, Kuopio, Finland
teemu.nurmirinta@uef.fi

Disclosures: T.A.T. Nurmirinta (N), M.J. Turunen (4-Aikoa Technologies Oy), J. Tohka (N), M.K. Liukkonen (N), M.E. Mononen (4-Aikoa Technologies Oy)

INTRODUCTION: Patients experiencing knee pain often feel compelled to seek medical advice due to the negative impact it has on their functional abilities and overall quality of life [1]. For patients with unbearable knee pain, total knee joint replacement surgery is usually considered the only clinical intervention to relieve knee pain. To avoid expensive surgery, the focus should be on preventive actions. This is only possible if pain progression can be predicted, and patients can be targeted and motivated to change their lifestyle before irreversible symptoms appear. Although some prediction models have been published [2,3], there are no clinically available tools or applications for this purpose. The challenge in developing such models is the implementation of features that are practical to measure and give high enough prediction performance. Knee pain is commonly predicted as a binary outcome [3], which makes prediction easier but does not capture the varying degrees of impact on patients' lives. Such information is critical for developing personalized treatment plans and interventions. Therefore, the aim of this study was to predict the severity of knee pain over an 8-year period among radiographically healthy individuals. Based on a previous study [4], we hypothesized that knee pain is related to knee joint shape and cartilage thickness obtained from radiographs. In addition to the knee geometry, the basic subject characteristics and baseline knee pain severity were utilized as predictor variables.

METHODS: 8-year follow-up data of 663 subjects (1188 knees) from a total of 4796 subjects from the Osteoarthritis Initiative Database (OAI, https://nda.nih.gov/oai) was selected. Exclusion criteria are shown in Fig. 1a. Each knee was categorized into a group based on how intense (0 no pain to 10 intense pain) and frequently (0 never, 1 monthly, 2 weekly, 3 daily, 4 always) the subjects felt pain. Groups were defined as: **Class 0** (no pain or low pain occasionally): intensity < 5 and frequency < 3, **Class 1** (high pain occasionally): intensity > 5 and frequency < 3, and **Class 2** (frequent pain): frequency ≥ 3. The number of knees in Classes 0, 1, and 2 was 946, 83, and 158, respectively. The model training process involved the utilization of a two-stage Balanced Random Forest (BRF) [5]. In the two-stage model, Predictor 1 underwent training by merging Classes 1 and 2 to differentiate them from Class 0. Predictor 2 underwent training without Class 0 to differentiate between Class 1 and Class 2. Fig. 1b displays a simplified flowchart illustrating the two-stage model. Two distinct models were trained. In **Model 1**, predictor variables included subject characteristics such as age (45-67 years), height, weight, gender, and baseline pain severity, whereas in **Model 2**, information from knee joint dimensions and angles was added to the predictor variables (Fig. 1c). In both models, the target variable was pain severity after the 8-year follow-up period. Joint dimensions and angles were analyzed from baseline knee radiographs using Matlab (v. R2019b, MathWorks Inc.). Machine learning was implemented in Python 3.9.7. The BRF predictor was trained using 100 trees. To achieve consistent outcomes, we performed twenty-five repetitions of 10-fold stratified cross-validation and computed the average confusion matrix for evaluation. In the model performance analysis, we used Scikit-learn's [6] Balanced Accuracy (BA) and Weighted F1 (WF1) metrics. Area under Receiver Operating Characteristic (AUC) [7] was calculated for the models first predictors (Class 0 vs. Class 1 and 2).

RESULTS: In Model 1, the scores were 70.9% (WF1) and 58.9% (BA) (Fig. 1d). In Model 2, these scores were observed to be 67.6% and 51.2%, respectively (Fig. 1e). AUC for the first predictor (Class 0 vs. Classes 1 and 2) in Model 1 and Model 2 was found to be 76% and 73%, respectively.

DISCUSSION: The presented two-state model demonstrates its potential as a valuable tool for predicting knee pain, thereby facilitating the identification of suitable candidates for drug development trials or motivating individuals at risk for unbearable knee pain to seek preventive treatments. The model's performance can be considered acceptable since the reference work [3] with a binary predictor of knee pain yielded an AUC of 71%, and our Model 1 achieved an AUC of 76%. The lack of multi-class knee pain prediction studies poses challenges for comparing our findings more extensively. However, based on the results between Model 1 and Model 2, it can be argued that contrary to the prediction of the risk of severe osteoarthritis [4], it appears that neither knee geometry nor cartilage thickness is indicative of future knee pain. A possible explanation could be that the radiographic features introduce noise that hinders the algorithm's ability to identify significant patterns across a diverse range of subjects. Much more training data is needed for the algorithm to effectively generalize with an increased number of features [8]. An alternative interpretation could posit that the influence of knee dimensions and angles on pain prediction in healthy knees may be relatively insignificant. We also analyzed how knee pain classes and predictions correspond to Kellgren-Lawrence (KL) grades of KOA (Fig. 1f). It was evident that a substantial number of participants classified as healthy (KL0 and 1) experienced continuous knee pain, whereas many of the participants diagnosed with severe knee osteoarthritis (KL3 and 4) reported no pain or occasional low pain. This is most likely due to the dissimilarities in the epidemiological characteristics of KOA and general knee pain [1]. For this reason, knee pain should be considered equal to predicted/diagnosed KOA when developing preventive actions for KOA management.

SIGNIFICANCE/CLINICAL RELEVANCE: The developed model enables a fast and cost-effective tool for knee pain prediction and could be used when developing personalized preventive actions for KOA or selecting suitable subjects for drug development in the future.

REFERENCES: [1] Hadler, Ann Intern Med, 116(7): 598-599, 1992; [2] Guan, Skelet Radiol, 1: 1-11, 2022; [3] Landsmeer, JABFM, 32(4): 575-584, 2019; [4] Nurmirinta, ORS, 1690, 2023; [5] Chen, Using Random Forest to Learn Imbalanced Data, UC Berkeley, 2005; [6] Scikit-learn: Machine learning in python, 2011; [7] Linden, J Eval Clin Pract, 12(2): 132-139, 2006; [8] Altman, Nat Methods, 15(6): 399-400, 2018.

ACKNOWLEDGEMENTS: Funding from the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding (5654242), the Research Council of Finland (324994, 328920, 352666) and Sigrid Juselius Foundation.

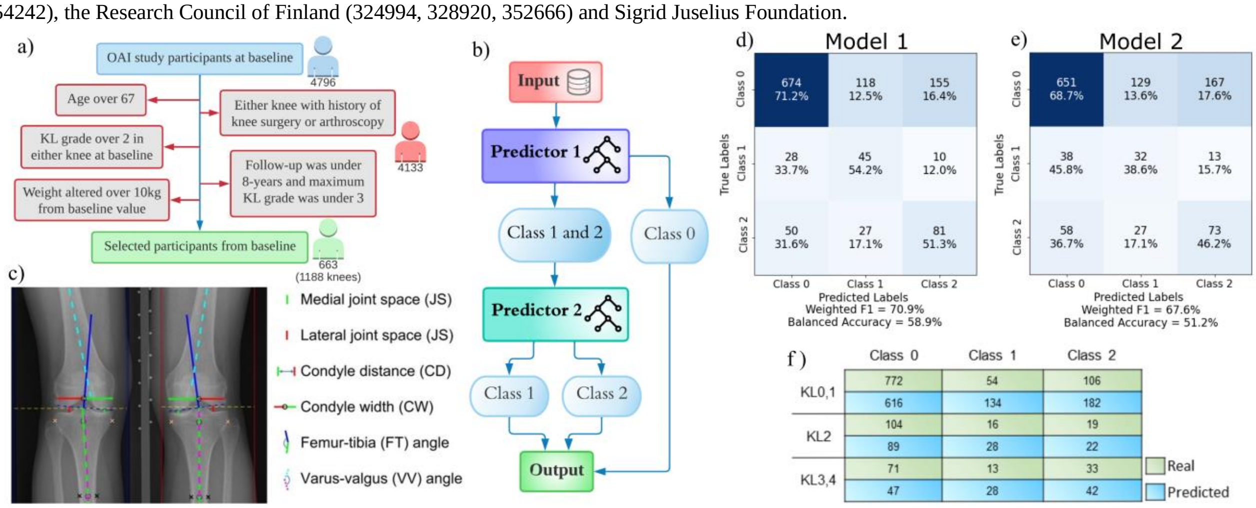

Fig 1. a) Exclusion criteria for OAI subjects, b) flowchart of a two-state model, c) knee dimensions and angles in radiograph, d) confusion matrix of Model 1 test set predictions, e) confusion matrix of Model 2 test set predictions, and f) table showing how our knee pain classes and predictions correspond to KL grades.