Merging Expert Knowledge and Machine-Learning to Provide Personalised Just-In-Time Advice in an Osteo-Arthritis Self-Management Application

Abdallah Al-Janabi[0009−0002−0651−3264]

AI & Behaviour Group, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Abstract. Osteoarthritis (OA) is a prevalent joint disorder and a significant cause of disability in older adults [1, 2]. Although guidelines for non-surgical treatment options for hip and knee OA exist, their management remains underutilised [3, 4].

This research project has the aim to develop an application that delivers personalised and just-in-time advice for patients with OA by merging expert systems (ES) and machine learning (ML). ES provide the transparency and explainability that is often necessary to gain the user-trust that is of particular relevance in clinical settings. However, recent advances in ML offer superior predictive power and the ability to model complex, non-linear relationships. This is particularly useful when utilising clinical longitudinal trial data that is not independent and identically distributed (i.i.d.). We will explore hybrid approaches that retain transparency without compromising predictive power.

To this end we will collect data from 600 OA patients trying out a simplified web-based application for OA self-management over a 12-month period. The collected data includes self-report measures on 47 prognostic factors that we identified to be associated with changes in three OA outcomes variables in a previous systematic review [6]. We will assess scores on the three outcome variables (pain, physical functioning and participation) on a biweekly basis to determine eligibility for three OA self-management modules (addressing movement, sleep or weight management) upon which a randomised choice for one of the eligible interventions is triggered (see Figure 1 for an illustration of the movement module). The intervention modules follow a decision tree structure and resulted from a structured process in which a expert-led focus-group designed clinical cut-offs and recommendations based on patient characteristics and complaints. The data will be used to train and compare several models that take the temporal dependency of the data into account, such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs).

Furthermore, we are interested in reducing the user's measurement burden by reducing the amount of self-report instruments. In order to accurately assess feature importance we need to account for potential confounding introduced through treatment effects of the administered interventions. One approach is modelling such treatment effects as features in

our dataset, which will also enable us to detect which subgroups respond best to which intervention. Additionally, causal machine learning techniques, such as structural causal models (SCMs), can be used to estimate the causal effect of predictors on outcomes by adjusting for confounders introduced by interventions. These models isolate the true effect of predictors by modeling interventions as treatment variables and estimating counterfactual outcomes using methods like inverse probability weighting or do-calculus.

Whether such models effectively deal with confounding treatment effects will be the subject of simulation experiments in which we employ a method introduced by [5]. This method uses restricted permutations to test if an algorithm has learned the relationship of interest in the presence of confounding factors and allows us to quantify and correct for the influence of observed confounders.

Lastly, we will integrate our ML model with our ES. To this end we will compare several options such as using the ML model to predict changes in outcome scores that require preventative action through the ES, or adjusting the parameters of the decision tree. Additionally, we will explore model-agnostic explanation methods for the employed ML models such as local interpretable model-agnostic explanations (LIME) [7] and SHapley Additive exPlanations (SHAP) [8] as well as model specific explanation methods such as visual explainable LSTMs (VixLSTM) [9]. The efficiency, explainability and feasibility of the hybrid system will be assessed and compared to an ES-only control condition in a randomised controlled trial.

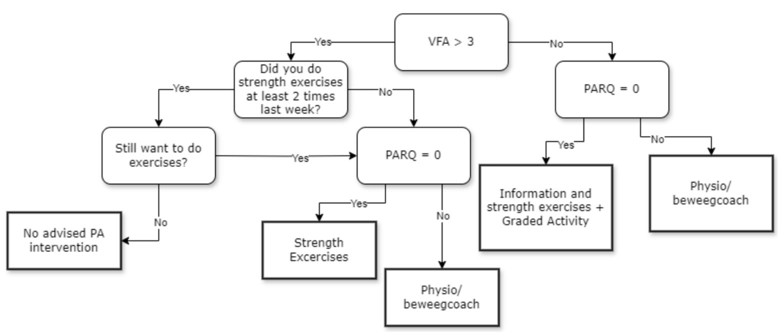

Fig. 1. Decision tree of the movement module. *Vragenlijst Fysieke Activiteit* (VFA, to measure physical activity); *Physical Activity Readiness Questionnaire* (PARQ, to test whether the patient is fit take part in an excercise programm); physical activity (PA).

Keywords: Personalised Healthcare · Explainable AI · Symbolic AI · Neuro-symbolic AI · Osteoarthritis.

References

1. Hunter, D. J., & Bierma-Zeinstra, S. (2019). Osteoarthritis. *The Lancet*, 393(10182), 1745-1759.
2. Hunter, D. J., March, L., & Chew, M. (2020). Osteoarthritis in 2020 and beyond: A Lancet Commission. *The Lancet*, 396(10264), 1711-1712.
3. Smink, A. J., van den Ende, C. H. M., Vliet Vlieland, T. P. M., Swierstra, B. A., Kortland, J. H., Bijlsma, J. W. J., Voorn, T. B., Schers, H. J., Bierma-Zeinstra, S. M. A., & Dekker, J. (2011). "Beating osteoARThritis": Development of a stepped care strategy to optimize utilization and timing of non-surgical treatment modalities for patients with hip or knee osteoarthritis. *Clinical Rheumatology*, 30(12), 1623–1629. https://doi.org/10.1007/s10067-011-1835-x
4. Nelson, A. E., Allen, K. D., Golightly, Y. M., Goode, A. P., & Jordan, J. M. (2014). A systematic review of recommendations and guidelines for the management of osteoarthritis: The Chronic Osteoarthritis Management Initiative of the U.S. Bone and Joint Initiative. *Seminars in Arthritis and Rheumatism*, 43(6), 701–712. https://doi.org/10.1016/j.semarthrit.2013.11.012
5. Neto, E. C. (2018). Using permutations to detect, quantify and correct for confounding in machine learning predictions. arXiv:1805.07465. https://doi.org/10.48550/arXiv.1805.07465
6. Cijs, B., Stekelenburg, R., Veenhof, C., Knoop, J., Boymans, T., de Rooij, M., & Kloek, C. (2024). Prognostic factors on changes in pain, physical functioning and participation in patients with hip- and/or knee OA: A systematic review. https://doi.org/10.1002/acr.25428
7. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). 'Why Should I Trust You?': Explaining the Predictions of Any Classifier. https://doi.org/10.48550/ARXIV.1602.04938
8. Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. arXiv:1705.07874. https://doi.org/10.48550/arXiv.1705.07874
9. Dang, T., Nguyen, H. N., & Nguyen, N. V. T. (2021). VixLSTM: Visual Explainable LSTM for Multivariate Time Series. *Proceedings of the 12th International Conference on Advances in Information Technology*, 1–5. https://doi.org/10.1145/3468784.3471603