Machine Learning Approach For Prediction Of Knee Osteoarthritis Progression Using Demographic And Clinical Variables

[1]Wisely Koay Zhi-Tang; [1]Shishkina, Margarita; [1]Tan Sik-Loo; [2]Wong Pooi-Fong; [3,4]Chang Siow-Wee; [1]Teo Seow-Hui; [1]Abbas Azlina Amir; [5]Roebuck Margaret M.; [1, 6]Kamarul Tunku

[1]NOCERAL, Department of Orthopaedic Surgery, Faculty of Medicine, Universiti Malaya (UM), Lembah Pantai, Kuala Lumpur, Malaysia, [2]Department of Pharmacology, Faculty of Medicine, UM, Lembah Pantai, Kuala Lumpur, Malaysia, [3]Bioinformatics Programme, Institute of Biological Sciences, Faculty of Science, UM, Kuala Lumpur, Malaysia, [4]Centre of Research in System Biology, Structural, Bioinformatics and Human Digital Imaging (CRYSTAL), UM, Kuala Lumpur, Malaysia, [5]Department of Musculoskeletal & Ageing Science, Institute of Life Course & Medical Sciences, University of Liverpool, [6]Advanced Medical and Dental Institute (AMDI), Universiti Sains Malaysia, Bertam, Kepala Batas, Pulau Pinang.

INTRODUCTION:

Knee osteoarthritis (KOA), characterized by progressive cartilage deterioration, poses a significant burden on global healthcare systems due to its high prevalence. Despite advancements in treatment modalities, accurately predicting the trajectory of OA progression remains a challenge and early identification of individuals at risk of rapid OA progression is crucial for implementing timely interventions to mitigate disease severity[1]. This study presents a machine learning (ML) approach to predict Kellgren-Lawrence (KL) graded KOA progression at 48 months based on baseline demographic and clinical data.

METHODS:

The dataset utilized in this study comprises demographic factors along with clinical variables obtained from the Osteoarthritis Initiative (OAI) database. The data were pre-processed and encoded before using them as input for the learning of machine learning algorithms. A variety of advanced classifiers such as Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Naïve Bayes (NB) were trained and evaluated for their performance.

RESULTS:

The classification with the RF algorithm emerged as the top-performing model. Our results demonstrate the efficacy of the proposed approach, with the RF classifier achieving an average area under the receiver operating characteristic curve (AUC) of 0.78 and an average accuracy of 0.71.

Table 1 shows the performance metrics of different ML classifiers.

ML model	Average AUC	Average accuracy
RF	**0.78** (0.77 - 0.80)	**0.71** (0.70 - 0.72)
LR	0.64 (0.62 - 0.65)	0.59 (0.58 - 0.60)
KNN	0.62 (0.61 - 0.64)	0.59 (0.58 - 0.61)
SVM	0.66 (0.64 - 0.67)	0.61 (0.59 - 0.62)
NB	0.61 (0.60 - 0.63)	0.58 (0.56 - 0.59)

DISCUSSIONS:

The superior RF classifier's performance emphasizes ensemble learning techniques' importance in handling complex, high-dimensional datasets, i.e. nonlinear relationships among variables in KOA progression modeling. In this study, this RF modelling is better than the GBM and LR models reported by Tiulpin et al[2].

CONCLUSION:

The RF classifier is the best model for KL grade progression prediction based on demographic and clinical data at 48 months. The findings of this study contribute to the advancement of prognostic modeling in knee OA. Continuous refinement and validation of the predictive model are required to further enhance its accuracy and applicability in clinical settings.

REFERENCES:

1. Machine learning in knee osteoarthritis: A review. Osteoarthritis and Cartilage Open, 2(3), 100069.

2. Tiulpin et al. (2019). Multimodal Machine Learning-based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data. Scientific Reports, 9(1), 20038.