Explainable AI-Based Clinical Decision Support for Knee Osteoarthritis Surgery Prediction

Dr. Manimegalai P[1], Dr Kalpana Nagpal[2]
[1] Lincoln University; [2] Pharmacy Amity University, Noida
pdf.manimegalaip@lincoln.edu.my , kalpananagpal@gmail.com

Abstract: Knee osteoarthritis (OA), one of the most common degenerative joint disorders affecting the elderly population, is a source of pain, disability, and decreased quality of life. A portion of individuals need surgical intervention, such as total knee arthroplasty, as the illness worsens. For better clinical results and individualized treatment planning, it is essential to determine which patients are most likely to require surgery. In order to forecast whether patients with osteoarthritis in their knees would require surgery, this study suggests a clinical decision support framework based on explainable artificial intelligence (XAI) and organized clinical data. Several machine learning models were created and assessed using accuracy, F1-score, and area under the receiver operating characteristic curve. The Random Forest classifier outperformed the models in prediction, demonstrating its ability to capture non-linear correlations within clinical variables. SHapley Additive exPlanations (SHAP) were incorporated into the framework to address the lack of transparency connected to machine learning models. SHAP analysis offered the significance of global features and patient-specific justifications. The suggested approach improves clinical trust by combining interpretability with prediction accuracy. This study demonstrates the potential of explainable AI-driven decision support systems to facilitate surgical decision-making in clinical practice.

Keywords: Knee Osteoarthritis, Surgery Prediction, Explainable Artificial Intelligence, Clinical Decision Support System, Machine Learning

---

Introduction

One of the most prevalent degenerative joint conditions, knee osteoarthritis (OA) is a leading cause of persistent pain, stiffness, and functional impairment, especially in the elderly. The illness develops gradually, and at more advanced stages, surgery, including a total knee replacement, is frequently required. Early detection of patients who may need surgery is essential for efficient clinical decision-making, individualized treatment planning, and preventing needless delays or interventions[1]. Using structured clinical data, this study proposes an explainable AI-based clinical decision support framework for predicting whether knee OA patients would require surgery. After developing and comparing several ML models, Random Forest shows the best results. In order to improve interpretability and clinical trust, SHAP-based explanations are included, which supports the proposed system's possible application in the management of knee OA in the actual world.

Related work

The use of machine learning (ML) approaches for knee osteoarthritis (OA) analysis has garnered a lot of attention in recent years, especially for diagnosis, severity grading, and disease progression prediction. Using longitudinal clinical data, Jamshidi et al. used machine learning models to forecast the course of knee OA. They found that ensemble-based models performed better than conventional regression techniques[6]. Explainable artificial intelligence (XAI) techniques have been progressively incorporated into healthcare machine learning models in order to tackle this problem. Despite these developments, little research has been done expressly on explainable machine learning models for leveraging structured clinical data to predict surgical intervention in knee OA. The current study addresses this gap by combining machine learning prediction with SHAP-based explainability

Key Contribution

In order to plan surgical intervention in patients with osteoarthritis of the knee using structured clinical data, this research suggests a clinical decision support system based on explainable artificial intelligence (XAI). Following the development and comparison of several machine learning models, Random Forest was shown to have the best prediction performance.

Method

Dataset Description

The knee osteoarthritis patient records gathered from regular clinical evaluations make up the structured clinical dataset used by the suggested framework. Age, body mass index (BMI), pain scores, range of motion, morning stiffness length, presence of crepitus and edema, difficulty walking, inflammatory markers (CRP level), and Kellgren-Lawrence (KL) grade are important characteristics. Surgery_Required, a binary outcome that indicates whether surgical intervention is necessary (1) or not (0), is the goal variable.

Data Preprocessing

The dataset was checked for discrepancies and missing values before the model was developed. To ensure class balance between surgery and non-surgery patients, the dataset was then divided into training and testing subsets using an 80:20 stratified split.

Machine Learning Models

In order to predict whether patients with knee osteoarthritis will need surgery, many machine learning models were created and assessed. All models were trained using the training dataset, and their evaluation was done using the test dataset, which was hidden. Performance was assessed using common classification measures, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC-AUC).

Explainable AI Framework

SHapley Additive exPlanations (SHAP) were used to incorporate an explainable AI framework into the suggested system in order to improve transparency and interpretability. To measure each feature's contribution to the model's predictions, SHAP values were calculated for the Random Forest model that had been trained.

Results

Performance Comparison of ML Models

The performance comparison across multiple machine learning models for the prediction of knee osteoarthritis surgery is compiled in Table 1. Random Forest and CatBoost outperformed other advanced models in terms of accuracy (97.5%) and F1-score (0.9748), demonstrating a robust balance between recall and precision.

Table 1. Performance Comparison across multiple ML Models for Knee OA Surgery Prediction

S.No.	Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
1	Logistic Regression	0.846	0.837	0.858	0.848	0.918
2	Linear SVM	0.850	0.850	0.850	0.850	0.915
3	Random Forest	0.975	0.983	0.967	0.975	0.976
4	Extra Trees	0.967	0.983	0.950	0.966	0.968
5	XGBoost	0.971	0.983	0.958	0.970	0.979
6	LightGBM	0.971	0.975	0.967	0.971	0.978
7	CatBoost	0.975	0.983	0.967	0.975	0.976
8	Stacking Ensemble	0.971	0.983	0.958	0.970	0.977

Random Forest was chosen for more examination because of its strong overall performance, resilience to noise, and compatibility with explainable AI approaches, even though boosting-based models displayed somewhat higher ROC-AUC values.

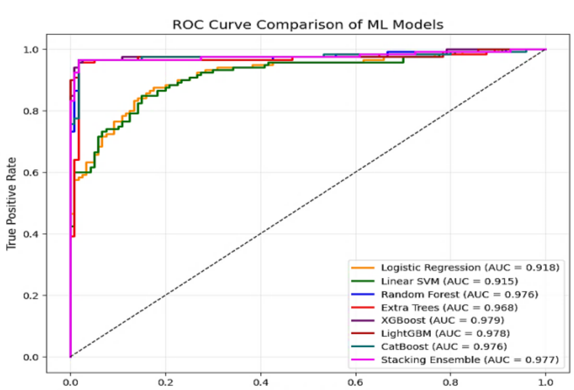

Figure 1. ROC Comparison of ML Models

Figure 1 displays the receiver operating characteristic (ROC) curves for each of the evaluated machine learning models used to forecast knee osteoarthritis surgery. In clinical settings, a high ROC-AUC is necessary to reduce false-negative predictions since incorrectly classifying patients who need surgery could worsen outcomes and delay necessary intervention. The observed ROC performance supports the use of Random Forest as a reliable model for additional explainability analysis and validates that ensemble-based learning techniques are well-suited for OA surgery prediction.

Explainability Analysis

The Random Forest model's global feature importance analysis is shown in Figure.2. Kellgren-Lawrence (KL) grade is the most significant feature, generating the greatest relevance score among all variables, according to the Random Forest feature importance plot. Since KL grade is a key factor in surgical decision-making and represents the severity of radiographic illness, this observation is clinically intuitive. The next most important indicators are walking difficulties and pain score, underscoring the significance of functional restrictions and patient-reported symptoms in assessing the need for surgery.

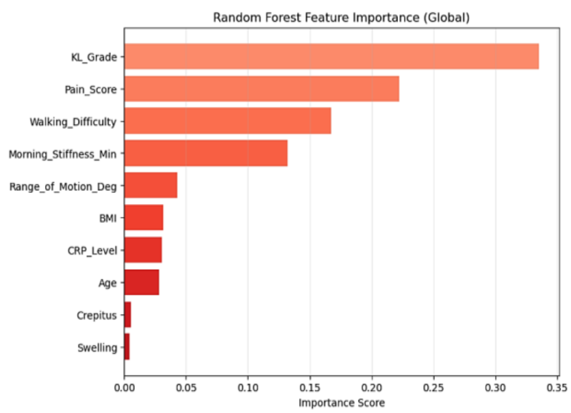

Figure 2. Feature Importance Chart

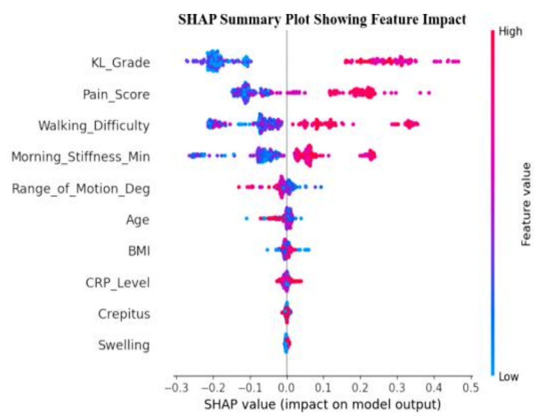

Figure 3. SHAP Summary Plot

The SHAP summary (beeswarm) graphic in Figure.3 illustrates the intricate interaction between feature values and their directional influence on the model output. Positive SHAP values are linked to higher KL Grade, Pain Score, and Walking Difficulty (highlighted in red), suggesting a strong push in the direction of the surgery prediction class.

Discussions

This study introduced a clinical decision support system based on explainable artificial intelligence that uses structured clinical data to forecast whether patients with osteoarthritis in their knees will need surgery. High accuracy, F1-score, and ROC–AUC were specifically attained by the Random Forest classifier. By incorporating SHapley Additive exPlanations (SHAP), the suggested approach prioritizes clinical interpretability and transparency in addition to predictive performance. The most significant determinants of surgical decision outcomes, according to global explainability analysis, were radiographic severity (KL grade), pain score, walking difficulty, and morning stiffness duration.

Conclusions

Using structured clinical data, this study created an explainable AI-based clinical decision support system for predicting surgical intervention in patients with osteoarthritis of the knee. Random Forest produced the greatest prediction performance with high accuracy and ROC-AUC among the assessed machine learning models. The most significant clinical characteristics were determined by SHAP-based explainability to be KL grade, pain score, walking difficulties, and morning stiffness duration. The suggested paradigm enhances openness and facilitates early risk assessment, individualized treatment planning, and precise surgical decision-making. Future research will concentrate on real-world clinical deployment, multi-center validation, and the integration of imaging and longitudinal data.

References

1. C. Chen *et al.*, "Heliyon Global years lived with disability for musculoskeletal disorders in adults 70 Years and older from 1990 to 2019 , and projections to 2040," *Heliyon*, vol. 10, no. 15, p. e35026, 2024, doi: 10.1016/j.heliyon.2024.e35026.
2. B.-Z. S. Hunter DJ, "Osteoarthritis," *Lancet*, 2019, doi: 10.1016/S0140-6736(19)30417-9.
3. L. Arbeeva, M. C. Minnig, K. A. Yates, and A. E. Nelson, "Machine learning approaches to the prediction of osteoarthritis phenotypes and outcomes," *Curr. Rheumatol. Rep.*, vol. 25, no. 11, pp. 213–225, 2024, doi: 10.1007/s11926-023-01114-9.Machine.
4. Z. Sadeghi et al., "A review of Explainable Artificial Intelligence in healthcare," Comput. Electr. Eng., vol. 118, no. PA, p. 109370, 2024, doi: 10.1016/j.compeleceng.2024.109370.
5. H. et al. Lundberg, S.M., Erion, G., Chen, "From local explanations to global understanding with explainable AI for trees," Nat. Mach. Intell., vol. 2, pp. 56–67, 2020, [Online]. Available: https://doi.org/10.1038/s42256-019-0138-9
6. R. Jose et al., "Machine - learning - based diagnosis and progression analysis of knee osteoarthritis," Discov. Data, 2025, doi: 10.1007/s44248-025-00026-6.
7. C. Kokkotis, C. Ntakolia, S. Moustakidis, G. Giakas, and D. Tsaopoulos, "Explainable machine learning for knee osteoarthritis diagnosis based on a novel fuzzy feature selection methodology," Phys. Eng. Sci. Med., vol. 45, no. 1, pp. 219–229, 2022, doi: 10.1007/s13246-022-01106-6.
8. R. Ahmed, A. L. I. Shariq, and I. Member, "Knee Osteoarthritis Analysis Using Deep Learning and XAI on X-rays," IEEE Access, vol. PP, p. 1, 2024, doi: 10.1109/ACCESS.2024.3400987.
9. Y. Dong, Y. Yan, J. Zhou, Q. Zhou, and H. Wei, "Evidence on risk factors for knee osteoarthritis in middle older aged : a systematic review and meta analysis," J. Orthop. Surg. Res., vol. 6, pp. 1–16, 2023, doi: 10.1186/s13018-023-04089-6.
10. M. D. Kohn, A. A. Sassoon, and N. D. Fernando, "Classifications in Brief," Clin. Orthop. Relat. Res., vol. 474, no. 8, pp. 1886–1893, 2016, doi: 10.1007/s11999-016-4732-4.