Predicting Depression in Patients With Knee Osteoarthritis Using Machine Learning: Model Development and Validation Study

Zuzanna Nowinka[1], BSc, MBChB; M Abdulhadi Alagha[1,2], MD; Khadija Mahmoud[1], BSc; Gareth G Jones[1], PhD

[1]MSk Lab, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, United Kingdom
[2]Data Science Institute, London School of Economics and Political Science, London, United Kingdom

Corresponding Author:
M Abdulhadi Alagha, MD
MSk Lab
Department of Surgery and Cancer, Faculty of Medicine
Imperial College London
South Kensington Campus
London, SW7 2AZ
United Kingdom
Phone: 44 020 7589 5111
Email: h.alagha@imperial.ac.uk

Abstract

Background: Knee osteoarthritis (OA) is the most common form of OA and a leading cause of disability worldwide. Chronic pain and functional loss secondary to knee OA put patients at risk of developing depression, which can also impair their treatment response. However, no tools exist to assist clinicians in identifying patients at risk. Machine learning (ML) predictive models may offer a solution. We investigated whether ML models could predict the development of depression in patients with knee OA and examined which features are the most predictive.

Objective: The primary aim of this study was to develop and test an ML model to predict depression in patients with knee OA at 2 years and to validate the models using an external data set. The secondary aim was to identify the most important predictive features used by the ML algorithms.

Methods: Osteoarthritis Initiative Study (OAI) data were used for model development and external validation was performed using Multicenter Osteoarthritis Study (MOST) data. Forty-two features were selected, which denoted routinely collected demographic and clinical data such as patient demographics, past medical history, knee OA history, baseline examination findings, and patient-reported outcome measures. Six different ML classification models were trained (logistic regression, least absolute shrinkage and selection operator [LASSO], ridge regression, decision tree, random forest, and gradient boosting machine). The primary outcome was to predict depression at 2 years following study enrollment. The presence of depression was defined using the Center for Epidemiological Studies Depression Scale. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC) and F1 score. The most important features were extracted from the best-performing model on external validation.

Results: A total of 5947 patients were included in this study, with 2969 in the training set, 742 in the test set, and 2236 in the external validation set. For the test set, the AUC ranged from 0.673 (95% CI 0.604-0.742) to 0.869 (95% CI 0.824-0.913), with an F1 score of 0.435 to 0.490. On external validation, the AUC varied from 0.720 (95% CI 0.685-0.755) to 0.876 (95% CI 0.853-0.899), with an F1 score of 0.456 to 0.563. LASSO modeling offered the highest predictive performance. Blood pressure, baseline depression score, knee pain and stiffness, and quality of life were the most predictive features.

Conclusions: To our knowledge, this is the first study to apply ML classification models to predict depression in patients with knee OA. Our study showed that ML models can deliver a clinically acceptable level of performance (AUC>0.7) in predicting the development of depression using routinely available demographic and clinical data. Further work is required to address the class imbalance in the training data and to evaluate the clinical utility of the models in facilitating early intervention and improved outcomes.

XSL•FO
RenderX

Introduction

Knee osteoarthritis (OA) is the most common form of OA and a leading cause of disability worldwide, with global prevalence estimated at 16% for individuals aged 15 years and over [1]. Knee OA is a chronic, progressive condition characterized by structural damage to the cartilage [2]. Knee OA results in chronic pain and impaired joint function, significantly limiting the activities of daily living [1,3]. Consequently, these patients experience a poorer health-related quality of life and are at higher risk of developing depression compared to the general population [4]. It has been estimated that up to 20% of patients with knee OA may be suffering from depression [3].

Several studies suggest that depression has an adverse impact on OA prognosis, quality of life, pain levels, as well as treatment effectiveness [5-7]. A longitudinal study conducted by Rathbun et al [8] found that depressive symptoms affected the physical functioning and pain severity of patients with knee OA. Another study showed that a persistently depressed mood significantly increases the severity of pain [9]. Additionally, a bidirectional relationship between pain and depression in patients with knee OA has been described, where concurrent depression increases pain perception and, reciprocally, higher pain levels may lead to a more depressed state [9-11]. It is therefore essential to recognize and address the vicious pain-depression cycle early.

Unsurprisingly, patients with knee OA and comorbid depression report lower coping ability, which translates into more frequent medical help-seeking and reduced satisfaction from treatment, including surgical interventions such as knee arthroplasty [3,10,12,13]. Ultimately, this accounts for a substantial rise in the health care cost burden [14,15]. Agarwal et al [16] estimated that the health care costs per year increase by US $4400 (US $13,684 vs US $9284) for every patient with concurrent OA and depression. The economic cost associated with knee OA is likely to rise in the upcoming years due to increasing life expectancy and thus the proportion of patients with knee OA [2]. With no curative treatment in sight, emphasis should be made on preventative and nonoperative strategies to manage the disease symptoms and reduce worsening factors such as depression [1,12].

Obtaining adequate mental health support should be of primary importance, as the presence of depressive symptoms is a significant predictor of worsening outcomes [17]. At the same time, appropriate therapy with antidepressants and counseling has been shown to significantly lower the perceived severity of pain [18]. However, less than half of all patients affected by knee OA and concurrent depression actively seek support or receive adequate treatment [19,20]. Unfortunately, poor mental health is frequently overlooked by clinicians, who focus primarily on the physical aspects of knee OA and so fail to recognize depression or its role in contributing to persisting knee symptoms [12,21]. Being able to predict which patients are at risk of experiencing depression would facilitate a targeted, preventative strategy against worsening outcomes such as pain and declining physical function [17].

Identifying patients with depression early would be helpful; however, no such tools currently exist. Although one previous study has tried to predict depression in this patient population, the model was based on conventional statistical methods, had low accuracy (area under the receiver operating characteristic curve [AUC]=0.742, 95% CI 0.622-0.862), and lacked external validation [22]. This represents a significant gap in care. The solution may lie in machine learning (ML) models. The ability of ML algorithms to handle large data sets, and evaluate complex and nonlinear relationships between variables theoretically makes them better suited for predictive tasks than standard statistical methods [23,24]. To date, no previous study has attempted to build an ML prediction model to detect the development of depression in patients with knee OA.

The primary objective of this study was to apply ML models to predict depression in patients with knee OA, using routinely available clinical data. We hypothesized that ML models can deliver a clinically acceptable level of performance, defined as an AUC greater than 0.7. Our secondary objective was to identify the most important predictive features used by the ML algorithms to make this prediction.

Methods

Data Sources and Study Cohort

We used data from the Osteoarthritis Initiative (OAI) database for model development and data from the Multicenter Osteoarthritis Study (MOST) for external validation. Both are publicly available, prospective cohort studies investigating knee OA progression in the US population [25,26]. The OAI study included adults aged 45-79 years, enrolled between February 2004 and May 2006, and the MOST included adults aged 50-79 years, recruited in 2003.

We included patients who attended the baseline and 15-month/24-month follow-ups, with preexisting knee OA (defined as the presence of symptoms and radiographic evidence of OA) or at high risk of developing knee OA (symptoms of pain, stiffness, and swelling). Patients with a history of rheumatoid arthritis, missing data for the depression scale scores at either consultation, missing radiographic data, missing baseline examination findings, or missing patient-reported outcome measures were excluded.

Ethics Considerations

No ethical approval was required for this study owing to the open access nature of the OAI and MOST databases.

Prediction Outcome

Our primary outcome was the development of depression at 2 years following enrollment in the database. Depression was defined using the Center for Epidemiological Studies Depression Scale (CES-D), which is based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition formulation of

depression, containing 20 questions evaluating the severity of psychosomatic symptoms [27]. The score ranges from 0 to 60, with higher values indicating greater symptom severity. A score of 16 points or more has previously been linked to clinical depression and as such was used in this study to dichotomize patients as either depressed or not depressed [27].

In the MOST, follow-up visits were scheduled at different time points compared with those used in the OAI study, and therefore CES-D scores captured during the 15-month visit were used for external validation.

Variable Selection

Variable selection was guided by the literature and clinical relevance as judged by the senior author who is a specialist in the field. To facilitate external validation, equivalent variables had to be available in both the OAI and MOST data sets. In total, there were 2532 baseline variables in the OAI database and 1842 baseline variables in the MOST database; 70 and 66 variables were selected from the respective databases for model development. Variables included information on patient demographics, past medical history, knee OA history, baseline examination findings, and baseline patient-reported outcome measures.

Patient demographics included age, sex, ethnicity, BMI, marital status, living arrangements, current employment, education, and smoking status. Past medical history encompassed the history of heart attack, heart failure, stroke, asthma, chronic obstructive pulmonary disease, peptic ulcer disease, diabetes, kidney disease, and osteoporosis medication. Variables relating to knee OA history consisted of past knee injury, past knee surgery, steroid knee injections, analgesic medication for knee pain, as well as other arthritis medication. Baseline examination findings covered systolic and diastolic blood pressure, medial and lateral tibiofemoral, Kellgren-Lawrence grade, the 20-meter-walk test, the five-times-sit-to-stand test, and baseline CES-D score. Patient-reported outcome measures were the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Physical Activity Scale for the Elderly (PASE), and 12-item Short-Form Health Survey (SF-12).

Data Preprocessing

Binning the Features

Smoking status was stratified according to smoking intensity into light (1-5 pack-year history of smoking), moderate (10-20 pack-years), or severe (>20 pack-years). BMI was grouped into underweight (BMI<18.5 kg/m^2), normal weight (BMI 18.5-24.9 kg/m^2), overweight (BMI 25-29.9 kg/m^2), and obese (BMI>30 kg/m^2), as defined by the World Health Organization [28]. Patients were categorized according to the American Heart Association Hypertension Guidelines to denote the stage of hypertension using variables for systolic and diastolic blood pressures [29]. Results of the five-times-sit-to-stand test were dichotomized, given that ≥10 seconds is the optimal cutoff for predicting the development of disability [30].

Feature Engineering

Feature engineering involves the combination of separate variables into a new, "engineered" feature, based on domain expertise and literature evidence. This action decreases the number of separate features and has been shown to improve model performance [31]. The "ethnicity" feature was created by merging variables describing race (*white, Black, Hispanic, other*). Variables assessing living arrangements were combined to denote whether the patient lived alone or with someone else. A feature for OA history was created by combining variables denoting the presence of other types of arthritis (*no other arthritis, one or more joints affected by OA, gout, OA and gout*). Variables denoting the use of analgesic medication for knee OA were assigned into a single feature, "analgesic medication" (*no pain relief, topical salicylates, nonsteroidal anti-inflammatory drugs or cyclooxygenase-2 inhibitors, opioid medication, combination of analgesic medication, other*). The "OA medication" feature was created by combining variables with information on OA treatment and supplements (*no medication or vitamin D supplements, bisphosphonates, estrogen/raloxifene, calcitonin/teriparatide, combination of OA medications*). The "arthritis medication" feature was created by merging five variables (*oral corticosteroids, supplements*). The final list of 42 features included in model training is summarized in Table 1.

Table 1. Summary of all features included in the model training.

Feature category	Features
Patient demographics	Age, sex, BMI, ethnicity, employment status, education status, living alone, marital status, smoking status
Past medical history and medication	Heart attack, heart failure, stroke, asthma, chronic obstructive pulmonary disease, peptic ulcer disease, diabetes, kidney disease, osteoporosis medication
Knee osteoarthritis history	Knee arthroscopy, knee meniscectomy, ligament repair, other knee surgery, arthritis of other joints, knee injury, steroid knee injections, analgesic medication for knee osteoarthritis, arthritis medication
Baseline examination findings	Blood pressure, 20-meter-walk test, five-stands-to-sit test, KLG[a,b], CES-D[c] baseline
Patient-reported outcome measures	WOMAC[a,d] (Total, Pain score, Stiffness score); SF-12[e] (Physical components, Mental health component); PASE[f]

[a]Separate feature for the right and left knee.

[b]KLG: Kellgren-Lawrence Grade.

[c]CES-D: Center for Epidemiological Studies Depression Scale.

[d]WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

[e]SF-12: 12-item Short Form Health Survey.

[f]PASE: Physical Activity Scale for the Elderly.

Missing Values

Missing values in the OAI data set were addressed by coding them as "unknown" to match the MOST data set. Following this imputation, only patients with all observations completed were included for analysis.

Model Development

Overview

Figure 1 summarizes the stages of data preprocessing and model development. The OAI data set was randomly divided into training (80% of observations) and test (20% of observations) sets using a computer algorithm, ensuring that each set included an equal proportion of patients with depression. Six common classification ML algorithms (logistic regression, least absolute shrinkage and selection operator [LASSO], ridge, decision tree, random forest, and gradient boosting machine [GBM]) were trained using the same set of 42 features. Classification models are a type of supervised ML where the algorithm calculates a probability of an observation belonging to the "positive" class based on the input data [32]. If the probability is above the threshold, the observation is labeled as "positive" (ie, depressed). The probability threshold is by default set to 0.5 but can be lowered when the cost of missing a "positive" case is high. Therefore, in this study, the threshold was set to 0.2 [33]. For each model, hyperparameter tuning was conducted until the performance on the training set was maximized. All models were developed using RStudio software (version 1.4.1106) [34].

Figure 1. Flowchart summarizing the project timeline and steps of model development. AUC: area under the receiver operating characteristic curve; GBM: gradient boosting machine; LASSO: least absolute shrinkage and selection operator; MOST: Multicenter Osteoarthritis Study; OAI: Osteoarthritis Initiative.
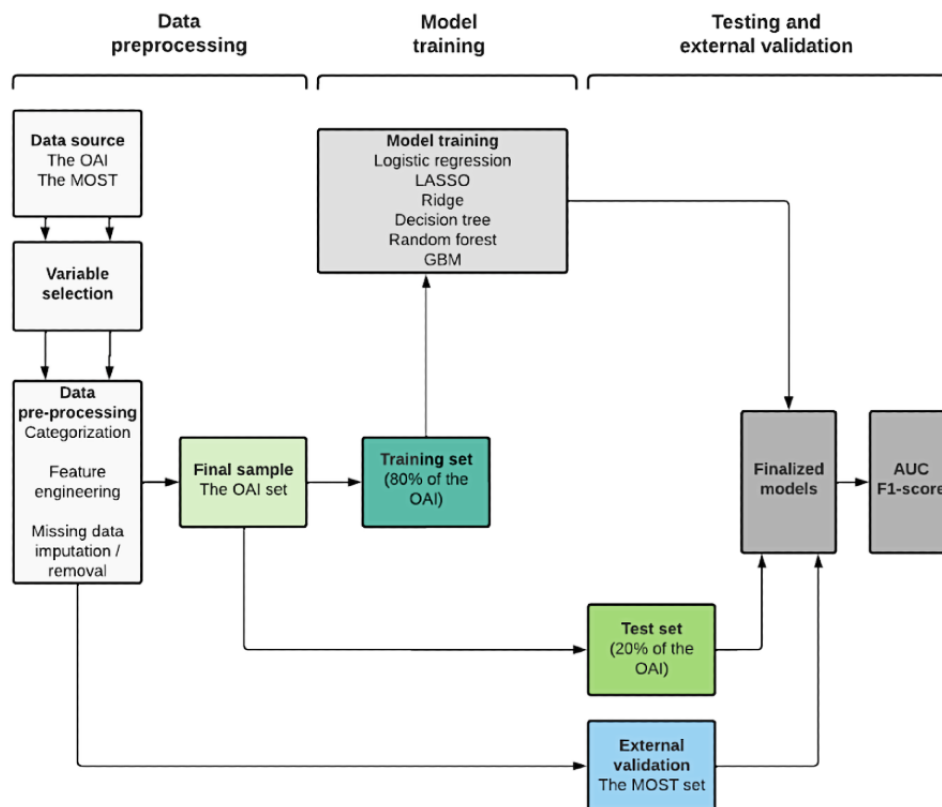
Logistic Regression

Logistic regression is a statistical model that uses a logit function to predict the probability of an observation belonging to the positive class [35]. Logistic regression is well-suited for classification problems such as problems involving describing the risk of developing a disease or the risk of mortality. This model was implemented using the RStudio "stats" package [36].

LASSO and Ridge Regression

LASSO and ridge regression models are based on the logistic regression model [24,32,37]. In LASSO, the algorithm adds a "penalty" to each feature so that features are eliminated if not considered important for the prediction by the algorithm [37]. LASSO shrinks regression coefficients toward 0, and ultimately only top informative features are included. This results in a simpler and more easily interpretable model [37]. In ridge, the algorithm reduces less important features to close to zero but does not eliminate them [32]. In this way, all features are kept in the model, which is beneficial when all features need to be included [32]. LASSO and ridge models were developed using the "glmnet" package with optimal hyperparameters for both algorithms set as follows: nfolds=3, s=lambda.min [38].

Decision Tree and Random Forest

Decision tree is a simple, tree-shaped algorithm, in which each branch of the tree determines a possible decision or course of action [39]. The model was developed with no additional hyperparameters using the "rpart" package [40]. Random forest is an algorithm similar to the decision tree; it operates by building multiple, independently trained decision trees using random subsets of the data [41]. Subsequently, their predictions are combined into a single prediction outcome. Random forest of 500 trees with nodesize=100 and mtry=4 was developed using the "randomForest" package [42].

GBM Model

In GBM, multiple tree-based classifiers are trained to augment each other and to reduce the prediction error [43]. GBM differs from the random forest algorithm in that a new decision tree is trained with the aim to correct errors made by existing trees, rather than training them independently. This model was developed using the "gbm" package and optimum hyperparameters were ntrees=2000, cv.folds=3, interaction.depth=4, and shrinkage=0.1 [44].

Performance Evaluation

The overall model performance was evaluated on the previously unseen OAI test set and externally validated using the MOST data set.

The primary model performance criterion was the AUC, and we considered an AUC greater than 0.7 to indicate clinically acceptable performance [45]. For each model, accuracy, precision, and recall are also reported. In addition, the F1 score, a weighed metric of precision and recall, was calculated according to the formula: $F1=2\times([precision\times recall]/[precision+recall])$. F1 score ranges from 0 (poor performance) to 1 (perfect performance).

While ML may provide a valuable predictive tool, the clinical implementation often raises concerns due to the model's complexity, referred to as the "black-box" problem [46]. One

way of improving model understanding is by extracting the most important predictive features. We therefore identified the most important predictive features from the best-performing model.

Results

Study Participants

The initial OAI data set included 4796 patients (Figure 2). Following exclusion of 1085 patients, the final sample size encompassed 3711 patients. After splitting the sample, the training set included 2969 patients and the test set had 742 observations. In the MOST data set, 790 patients were excluded

from the initial sample of 3026 cases and the final sample included 2236 patients.

Table 2 summarizes the key patient characteristics. The average age was 61.0 years for the OAI sample and 62.1 years for the MOST sample. In both data sets, the majority of patients were female and of white ethnicity. Less than half of the patients had hypertension stage 1 or higher. There were some differences between the OAI and MOST samples. First, the proportion of depressed patients at 2 years was higher in the MOST sample. The MOST population also had higher average WOMAC scores for both the right and left knees, and a greater proportion of patients using analgesic medication for knee OA.

Figure 2. Summary of patient flow for both databases. CES-D: Center for Epidemiological Studies Depression Scale.
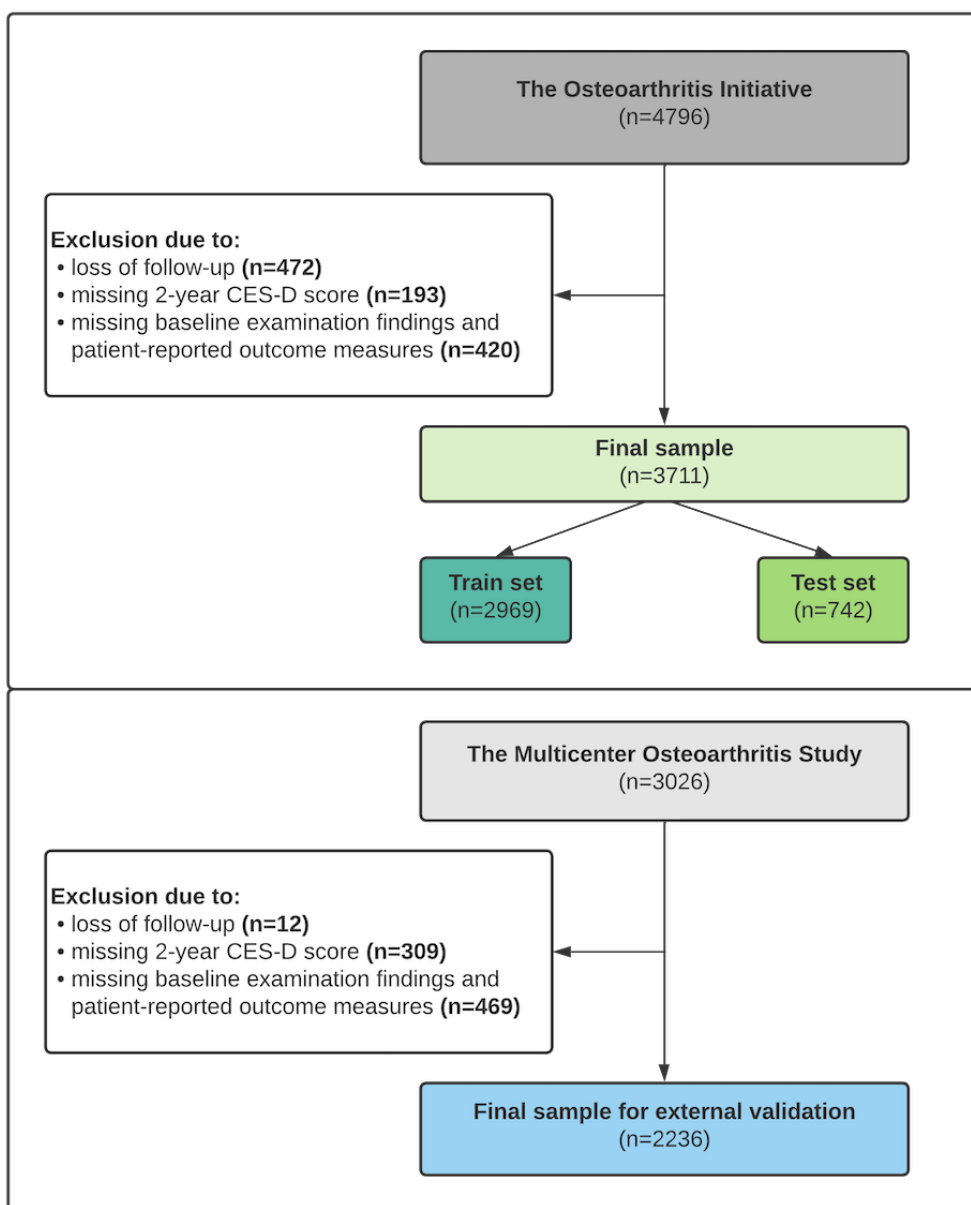
Table 2. Key patient demographic and clinical data.

Characteristic	OAI[a] (n=3711)	MOST[b] (n=2236)
Age, mean (SD)	61.0 (9.1)	62.1 (8.1)
BMI, mean (SD)	28.4 (4.8)	30.4 (5.9)
Sex (female), n (%)	2149 (57.91)	1297 (58.01)
Ethnicity (white), n (%)	3082 (83.05)	1932 (86.40)
Blood pressure (hypertension stage≥1), n (%)	1847 (49.77)	1008 (45.08)
Other arthritis, n (%)	1454 (39.18)	1071 (47.90)
Analgesic medication for knee OA[c] (any), n (%)	845 (22.77)	1804 (80.68)
KLG[d], n (%)		
Right knee, grade 1 or higher	2294 (61.82)	1180 (52.77)
Left knee, grade 1 or higher	2206 (59.44)	1264 (56.53)
WOMAC[e]-total, mean (SD)		
Right knee	10.7 (10.3)	18.6 (17.5)
Left knee	10.7 (10.4)	18.3 (17.5)
Baseline CES-D[f], mean (SD)	6.3 (6.0)	6.7 (6.2)
Depression at 2-year visit, n (%)	342 (9.22)	265 (11.85)

[a]OAI: Osteoarthritis Initiative.

[b]MOST: Multicenter Osteoarthritis Study.

[c]OA: osteoarthritis.

[d]KLG: Kellgren-Lawrence Grade.

[e]WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

[f]CES-D: Center for Epidemiological Studies Depression Scale.

Model Performance

In total, six classification models were trained using all 42 features. The results for each model are summarized in Table 3. Figure 3 and Figure 4 present the AUC plots for the internal test set and the external validation set, respectively. The AUC ranged from 0.673 to 0.869 for the internal test set and from 0.720 to 0.876 for the external validation set. Except for the decision tree algorithm, all models yielded an AUC>0.7, suggesting clinically acceptable discrimination between depressed and nondepressed patients [45]. LASSO was the model with the highest AUC on both the internal test set and external validation set.

The accuracy, precision, recall, and F1 scores for the test and validation sets are summarized in Table 4 and Table 5, respectively. The accuracy on the OAI test set varied from 0.895 (decision tree) to 0.923 (random forest). The performance on this metric was lower for the MOST data set, ranging from 0.865 (GBM) to 0.895 (ridge). Despite high accuracy, the proportion of correctly classified positive cases was relatively low. For the internal test set, the F1 scores varied from 0.435 (decision tree) to 0.490 (LASSO), and from 0.456 (ridge) to 0.536 (LASSO) on external validation. LASSO had a consistently high performance for the AUC and F1 score in comparison to the other models, ranking first on both the internal test and external validation sets.

Table 3. Model performance for the internal test set and external validation set.

Rank[a]	Model	Test set (OAI[b]), AUC[c] (95% CI)	External validation set (MOST[d]), AUC (95% CI)
1	LASSO[e]	0.869 (0.824-0.913)	0.876 (0.853-0.899)
2	GBM[f]	0.858 (0.813-0.903)	0.872 (0.849-0.895)
3	Ridge	0.864 (0.818-0.910)	0.852 (0.827-0.878)
4	Random forest	0.808 (0.741-0.874)	0.822 (0.790-0.853)
5	Logistic regression	0.837 (0.786-0.888)	0.808 (0.775-0.840)
6	Decision tree	0.673 (0.604-0.742)	0.720 (0.685-0.755)

[a]Models are ranked by their performance on the external validation data set.

[b]OAI: Osteoarthritis Initiative.

[c]AUC: area under the receiver operating characteristic curve.

[d]MOST: Multicenter Osteoarthritis Study.

[e]LASSO: least absolute shrinkage and selection operator.

[f]GBM: gradient boosting machine.

Figure 3. AUC plot of all models tested on the OAI test set (20% of the initial OAI data set). The test set was not used at any stage of model training. AUC: area under the receiver operating characteristic curve; GBM: gradient boosting machine; LASSO: least absolute shrinkage and selection operator; MOST: Multicenter Osteoarthritis Study; OAI: Osteoarthritis Initiative.
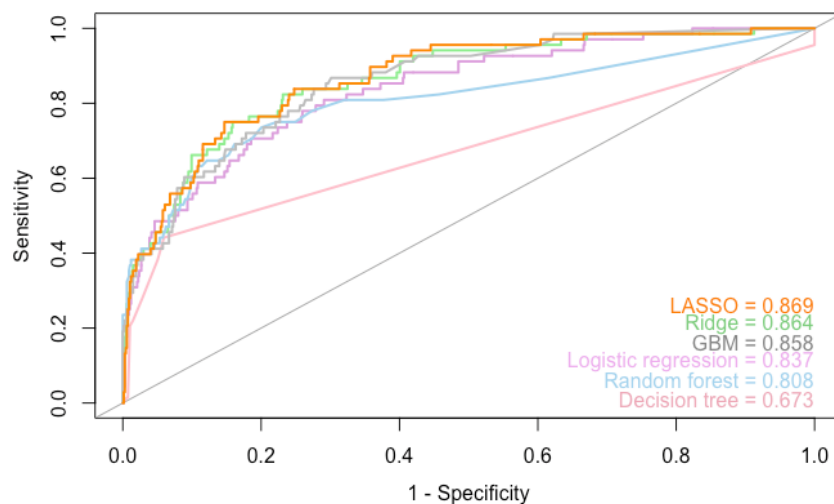
Figure 4. AUC plot of all models externally validated on the MOST data set. AUC: area under the receiver operating characteristic curve; GBM: gradient boosting machine; LASSO: least absolute shrinkage and selection operator; MOST: Multicenter Osteoarthritis Study; OAI: Osteoarthritis Initiative.
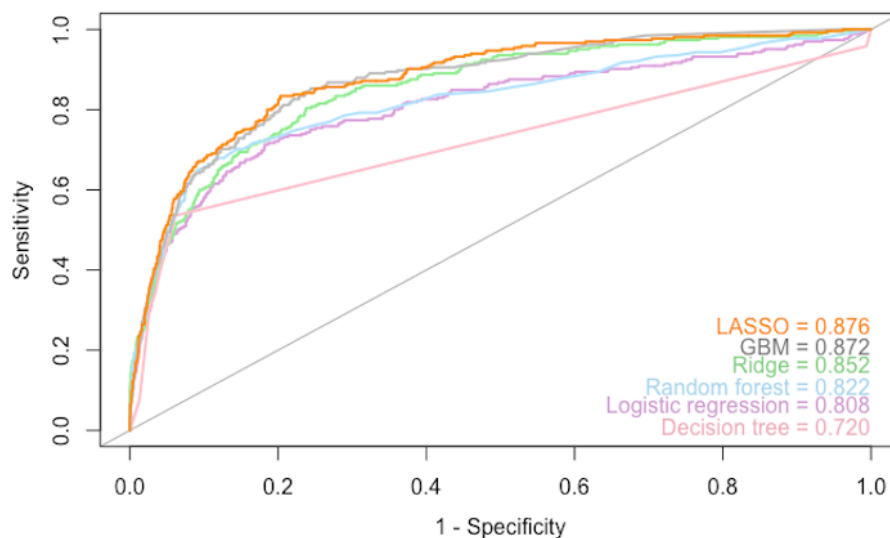
Table 4. Accuracy, precision, recall, and F1 scores for the test set, ranked by the F1 score.

Rank	Model	Accuracy	Precision	Recall	F1
1	LASSO[a]	0.902	0.467	0.515	0.490
2	Random forest	0.923	0.628	0.397	0.486
3	Logistic regression	0.906	0.485	0.485	0.485
4	GBM[b]	0.901	0.466	0.500	0.482
5	Decision tree	0.895	0.429	0.441	0.435
6	Ridge	0.908	0.500	0.426	0.460

[a]LASSO: least absolute shrinkage and selection operator.
[b]GBM: gradient boosting machine.

Table 5. Accuracy, precision, recall, and F1 scores for the validation set, ranked by the F1 score.

Rank	Model	Accuracy	Precision	Recall	F1
1	LASSO[a]	0.889	0.528	0.604	0.563
2	Decision tree	0.890	0.538	0.536	0.537
3	GBM[b]	0.865	0.453	0.657	0.536
4	Random forest	0.894	0.556	0.506	0.530
5	Logistic regression	0.886	0.344	0.698	0.461
6	Ridge	0.895	0.593	0.370	0.456

[a]LASSO: least absolute shrinkage and selection operator.
[b]GBM: gradient boosting machine.

Most Important Predictive Features

The most important predictive features identified by LASSO were blood pressure, CES-D score at baseline, total WOMAC score for both knees, and mental and physical components of the SF-12 survey. Blood pressure had the highest coefficient (0.173), followed by the baseline CES-D score (0.126), WOMAC total for the right knee (0.004), and WOMAC total for the left knee (0.003). The mental and physical components of SF-12 had negative coefficients (–0.032 and –0.009, respectively).

Discussion

Principal Findings

The results of this study demonstrate that it is possible, with high accuracy, to predict depression in patients with knee OA using a variety of routinely collected data such as patient demographics, medical history, examination findings, and patient-reported outcome measures. The developed ML models achieved clinically relevant discrimination between depressed and nondepressed patients, with LASSO identified as the best-performing model, yielding an AUC of 0.876 (95% CI 0.853-0.899) on external validation. The accuracies for external validation were high, ranging from 0.865 (GBM) to 0.895 (ridge), meaning that between 86.5% and 89.5% of all patients were correctly classified. However, the F1 scores ranged from 0.456 (ridge) to 0.563 (LASSO). Low F1 scores despite high accuracy implies that the models can identify patients without

depression more accurately than those with depression. This is likely due to class imbalance in the data set, which is a common problem in medical research that results in predictive modeling bias toward the majority [47].

While ML may provide a valuable predictive tool, the clinical implementation often raises concerns due to model complexity, often referred to as the "black-box" problem [46]. One way of improving model understanding is to extract the most important features [48]. In this study, blood pressure, the baseline CES-D, the total WOMAC, as well as mental and physical components for SF-12 were identified as being the most informative measures for prediction. Although this does not imply a statistically significant correlation between the features and the prediction outcome, it is reassuring that the input features identified by LASSO have previously been highlighted as factors associated with an increased risk of developing depression in patients with OA [8,9,49]. Surprisingly, blood pressure was identified as being the most informative factor for prediction. The presence of multiple comorbidities can further increase the risk of depression development in patients with knee OA, regardless of their pathophysiology [49]. Notably, the radiographic severity of OA was not highlighted as a predictive feature for depression development. This is consistent with previous research showing that depression and pain are independent from the extent of radiographic degenerative changes [50]. This known discrepancy between knee OA symptoms and radiographic severity highlights the complex nature of the disease and the need for more objective assessment tools. The association between depression, chronic conditions,

and pain is complex. The temporality of the relationship between depression and pain has been poorly researched, but it appears that both factors potentiate each other, with higher pain severity increasing the persistence of depressed mood and the presence of pain increasing the incidence of depression [5,7,28,51,52]. This highlights the essential role of appropriate, interdisciplinary mental health support for patients with knee OA.

ML predictive models have an important role in augmenting clinical judgment, and when compared with standard predictions, they produce more accurate and less variable risk estimates [53]. The best-performing model in our study, LASSO, could be potentially used to aid in identification of patients at risk of future depression. Since the CES-D score has been designed as a screening tool, the patients identified as "positive" by our model would have to undergo further, more specialist mental health assessment. Depending on that outcome, the patients could be offered either a self-help aid, or potentially, a specialist referral. This would be more economical and time-efficient than assessing every patient attending with knee pain. However, further research is required since the implementation of predictive models is often difficult due to lack of clear clinical guidance on how to act upon the predicted outcome [54].

The advantage of our models lies in their simplicity as they rely on easily accessible clinical information. In addition, LASSO identified only 6 features to be crucial for prediction, making the model more practical. Blood pressure is routinely measured by primary health care practitioners, and WOMAC, SF-12, and CES-D scores are commonly used patient-reported outcome measures [55-57]. The aforementioned questionnaires are brief and require minimal training. Currently, there is no proven strategy to prevent or cure knee OA, and the therapy is focused on alleviating pain and addressing functional limitations [9]. Since depression is a potentially modifiable risk factor for worsening pain and function in knee OA, our prediction model could offer a targeted, preventative strategy. Diagnosing depression in patients with concurrent chronic pain conditions is challenging and having such information would facilitate discussions around the patient's mental health, even at times when the patient is not yet aware of their symptoms. While further research is required to evaluate the practical aspects of the clinical application, the findings of our study represent an important step toward developing a potential diagnostic aid, addressing a significant gap in knee OA care.

Comparison With Prior Work

To the best of our knowledge, this is the first study applying ML to predict depression in patients with knee OA. One previous study attempted to develop a prediction model based on logistic regression using conventional statistical methods [22]. Although the model achieved a clinically acceptable performance with an AUC of 0.742 (95% CI 0.622-0.862), it was built using a small sample of patients and was not tested on an independent sample or externally validated [22].

Diagnosis of depression is challenging in clinical practice, and ML models have been previously applied to predict illness in different patient populations [58-62]. Clinically relevant predictive performance of common ML classification algorithms was shown in two studies predicting postpartum depression [58,59]. Cvetkovic [60] used a deep-learning approach to predict depression in breast cancer patients, achieving high internal accuracy. However, the study methodology was poorly reported, with information lacking on data preprocessing and model testing [60]. In another study, depression and anxiety in college students were estimated using GBM, with satisfactory performance yielding an AUC of 0.730 [61]. When applied to community-residing older adults, a logistic regression model achieved variable accuracy, ranging from 58.33% for severe depression to 90.44% for mild depression [62]. The variation in model performance achieved by these studies could be attributed to the use of different algorithms, different evaluation tools for detection of depressive symptoms, as well as the use of different predictive features.

Strengths

Our study is strengthened by the use of a large patient cohort for model development, testing, and validation. The list of input features was carefully curated, with selection based on literature evidence, domain expertise, and data completeness. In addition, our predictive models were externally validated and performed well in an independent cohort, demonstrating their generalizability and potential for clinical application. Notably, LASSO identified only six features to be crucial for prediction, which showcases the simplicity of our method and the ease with which this tool could be used in a clinical setting.

Limitations

Several limitations should be addressed in future research. First, the study sample used for model development might not be representative of a general population of patients with knee OA. The prevalence of depressed patients in the training set was 9.2%, which is much lower than the 20% rate previously suggested by the literature [63]. The OAI study excluded patients with end-stage OA, morbid obesity, or those with terminal diseases, whereas these factors are associated with an even higher risk of depression [25,49]. Second, both the OAI and the MOST data sets were based in the United States with patients from a predominantly white ethnic background [25,26]. Further validation of our prediction model in a more ethnically and socioeconomically diverse population would help to detect any potential discrimination. Third, due to differences in the OAI and MOST protocols, follow-up times differed by 15 months between the training and external validation sets. Nevertheless, the models were able to predict on the external data set with similar performance. Lastly, the presence of depression at 2 years was defined using the CES-D scale; although this tool has been validated for use in patients with chronic illness and OA, it is not considered a gold standard for the diagnosis of depression [27]. However, the CES-D questionnaire has the advantage of being brief, easy to understand, and requiring minimal training for the assessor [27].

Conclusions

This is the first study to apply ML classification models to predict depression in patients with knee OA using routinely collected patient data. The LASSO model offered the highest quality of prediction, with an AUC of 0.876 (95% CI 0.853-0.899) on external validation. The advantages of our

method include the use of a large patient cohort and routinely collected data, as well as external validation on an independent data set. This tool offers a potential opportunity to assess a patient's risk of future depression, facilitating early intervention. Further research is required to establish where such a tool would fit within the care pathway, and while the harmful effects of depression on knee OA are well documented, it will be necessary to confirm that early detection and management of depression in this population leads to the expected improvement in outcomes.

Acknowledgments

Authors' Contributions

ZN, MAA, KM, and GGJ were involved in setting out the project aim and methodology. ZN conducted the literature search and wrote the original draft. ZN, MAA, and KM contributed to data curation and analysis. MAA and GGJ contributed to study design. GGJ supervised the conduction of the study, and reviewed and edited the manuscript. All authors had access to the raw data and have approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Cui A, Li H, Wang D, Zhong J, Chen Y, Lu H. Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. EClinicalMedicine 2020 Dec;29-30:100587 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100587] [Medline: 34505846]

2. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. The Lancet 2019 Apr 27;393(10182):1745-1759. [doi: 10.1016/S0140-6736(19)30417-9] [Medline: 31034380]

3. Marks R. Depression and osteoarthritis: impact on disability. Aging Sci 2014;02(03):1000126. [doi: 10.4172/2329-8847.1000126]

4. van 't Land H, Verdurmen J, Ten Have M, van Dorsselaer S, Beekman A, de Graaf R. The association between arthritis and psychiatric disorders; results from a longitudinal population-based study. J Psychosom Res 2010 Feb;68(2):187-193. [doi: 10.1016/j.jpsychores.2009.05.011] [Medline: 20105702]

5. Axford J, Heron C, Ross F, Victor CR. Management of knee osteoarthritis in primary care: pain and depression are the major obstacles. J Psychosom Res 2008 May;64(5):461-467. [doi: 10.1016/j.jpsychores.2007.11.009] [Medline: 18440398]

6. Wang S, Ni G. Depression in osteoarthritis: current understanding. Neuropsychiatr Dis Treat 2022;18:375-389. [doi: 10.2147/NDT.S346183] [Medline: 35237034]

7. Previtali D, Andriolo L, Di Laura Frattura G, Boffa A, Candrian C, Zaffagnini S, et al. Pain trajectories in knee osteoarthritis-a systematic review and best evidence synthesis on pain predictors. J Clin Med 2020 Sep 01;9(9):2828 [FREE Full text] [doi: 10.3390/jcm9092828] [Medline: 32882828]

8. Rathbun A, Shardell M, Stuart E, Yau M, Gallo J, Schuler M, et al. Pain severity as a mediator of the association between depressive symptoms and physical performance in knee osteoarthritis. Osteoarthritis Cartilage 2018 Nov;26(11):1453-1460 [FREE Full text] [doi: 10.1016/j.joca.2018.07.016] [Medline: 30092262]

9. Rathbun AM, Stuart EA, Shardell M, Yau MS, Baumgarten M, Hochberg MC. Dynamic effects of depressive symptoms on osteoarthritis knee pain. Arthritis Care Res 2018 Jan 06;70(1):80-88 [FREE Full text] [doi: 10.1002/acr.23239] [Medline: 28320048]

10. White DK, Neogi T, Nguyen UDT, Niu J, Zhang Y. Trajectories of functional decline in knee osteoarthritis: the Osteoarthritis Initiative. Rheumatology 2016 May;55(5):801-808 [FREE Full text] [doi: 10.1093/rheumatology/kev419] [Medline: 26705330]

11. Kroenke K, Wu J, Bair MJ, Krebs EE, Damush TM, Tu W. Reciprocal relationship between pain and depression: a 12-month longitudinal analysis in primary care. J Pain 2011 Sep;12(9):964-973 [FREE Full text] [doi: 10.1016/j.jpain.2011.03.003] [Medline: 21680251]

12. Sharma A, Kudesia P, Shi Q, Gandhi R. Anxiety and depression in patients with osteoarthritis: impact and management challenges. Open Access Rheumatol 2016;8:103-113. [doi: 10.2147/OARRR.S93516] [Medline: 27843376]

13. Perruccio AV, Power JD, Evans HMK, Mahomed SR, Gandhi R, Mahomed NN, et al. Multiple joint involvement in total knee replacement for osteoarthritis: effects on patient-reported outcomes. Arthritis Care Res 2012 Jun;64(6):838-846. [doi: 10.1002/acr.21629] [Medline: 22570306]

14. Rosemann T, Gensichen J, Sauer N, Laux G, Szecsenyi J. The impact of concomitant depression on quality of life and health service utilisation in patients with osteoarthritis. Rheumatol Int 2007 Jul 23;27(9):859-863. [doi: 10.1007/s00296-007-0309-6] [Medline: 17242902]

XSL•FO

RenderX

15. Gong L, Chen H. Descriptive analysis of the cost-effectiveness of depressed patients undergoing total knee arthroplasty: an economic decision analysis. J Orthop Sci 2014 Sep;19(5):820-826. [doi: 10.1007/s00776-014-0599-y] [Medline: 24996623]

16. Agarwal P, Sambamoorthi U. Healthcare expenditures associated with depression among individuals with osteoarthritis: post-regression linear decomposition approach. J Gen Intern Med 2015 Dec 20;30(12):1803-1811 [FREE Full text] [doi: 10.1007/s11606-015-3393-4] [Medline: 25990191]

17. Riddle DL, Kong X, Fitzgerald GK. Psychological health impact on 2-year changes in pain and function in persons with knee pain: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2011 Sep;19(9):1095-1101 [FREE Full text] [doi: 10.1016/j.joca.2011.06.003] [Medline: 21723400]

18. Lin E, Tang L, Katon W, Hegel M, Sullivan M, Unützer J. Arthritis pain and disability: response to collaborative depression care. Gen Hosp Psychiatry 2006;28(6):482-486. [doi: 10.1016/j.genhosppsych.2006.08.006] [Medline: 17088163]

19. Gleicher Y, Croxford R, Hochman J, Hawker G. A prospective study of mental health care for comorbid depressed mood in older adults with painful osteoarthritis. BMC Psychiatry 2011 Sep 12;11:147 [FREE Full text] [doi: 10.1186/1471-244X-11-147] [Medline: 21910895]

20. Agarwal P, Pan X, Sambamoorthi U. Depression treatment patterns among individuals with osteoarthritis: a cross sectional study. BMC Psychiatry 2013 Apr 22;13:121 [FREE Full text] [doi: 10.1186/1471-244X-13-121] [Medline: 23607696]

21. Cohen E, Lee YC. A mechanism-based approach to the management of osteoarthritis pain. Curr Osteoporos Rep 2015 Dec 30;13(6):399-406 [FREE Full text] [doi: 10.1007/s11914-015-0291-y] [Medline: 26419467]

22. Sayre E, Esdaile J, Kopec J, Singer J, Wong H, Thorne A, et al. Specific manifestations of knee osteoarthritis predict depression and anxiety years in the future: Vancouver Longitudinal Study of Early Knee Osteoarthritis. BMC Musculoskelet Disord 2020 Jul 16;21(1):467 [FREE Full text] [doi: 10.1186/s12891-020-03496-8] [Medline: 32677938]

23. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina 2020 Sep 08;56(9):455 [FREE Full text] [doi: 10.3390/medicina56090455] [Medline: 32911665]

24. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019 Mar 19;19(1):64 [FREE Full text] [doi: 10.1186/s12874-019-0681-4] [Medline: 30890124]

25. Nevitt MC, Felson DT, Lester G. The osteoarthritis initiative: protocol for the cohort study. National Institute of Mental Health Data Archive. 2006. URL: https://nda.nih.gov/static/docs/StudyDesignProtocolAndAppendices.pdf [accessed 2021-04-15]

26. Segal NA, Nevitt MC, Gross KD, Gross KD, Hietpas J, Glass NA, et al. The Multicenter Osteoarthritis Study: opportunities for rehabilitation research. PM R 2013 Aug;5(8):647-654 [FREE Full text] [doi: 10.1016/j.pmrj.2013.04.014] [Medline: 23953013]

27. Smarr KL, Keefer AL. Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). Arthritis Care Res 2011 Nov;63(Suppl 11):S454-S466. [doi: 10.1002/acr.20556] [Medline: 22588766]

28. WHO Consultation on Obesity. Obesity: preventing and managing the global epidemic: report of a WHO consultation. World Health Organization. 1999. URL: https://apps.who.int/iris/handle/10665/42330 [accessed 2021-05-22]

29. Whelton P, Carey R, Aronow W, Casey D, Collins K, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018 Jun;71(6):1269-1324 [FREE Full text] [doi: 10.1161/HYP.0000000000000066] [Medline: 29133354]

30. Makizako H, Shimada H, Doi T, Tsutsumimoto K, Lee S, Hotta R, et al. Cognitive functioning and walking speed in older adults as predictors of limitations in self-reported instrumental activity of daily living: prospective findings from the Obu Study of Health Promotion for the Elderly. Int J Environ Res Public Health 2015 Mar 11;12(3):3002-3013 [FREE Full text] [doi: 10.3390/ijerph120303002] [Medline: 25768239]

31. Xu Y, Hong K, Tsujii J, Chang EI. Feature engineering combined with machine learning and rule-based methods for structured information extraction from narrative clinical discharge summaries. J Am Med Inform Assoc 2012 Sep 01;19(5):824-832 [FREE Full text] [doi: 10.1136/amiajnl-2011-000776] [Medline: 22586067]

32. Weng W. Machine learning for clinical predictive analytics. In: Celi L, Majumder M, Ordóñez P, Osorio J, Paik K, Somai M, editors. Leveraging data science for global health. Cham: Springer; 2020. URL: https://link.springer.com/chapter/10.1007/978-3-030-47994-7_12

33. Chen JJ, Tsai CA, Moon H, Ahn H, Young JJ. The use of decision threshold adjustment in classification for cancer prediction. Penn State University. 2005. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.5813&rep=rep1&type=pdf [accessed 2021-04-22]

34. RStudio: Integrated Development for R. 2020. URL: https://www.rstudio.com/ [accessed 2021-05-28]

35. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression, 3rd edition. Hoboken, NJ: Wiley; 2013.

XSL•FO
RenderX

36. The R stats package (version 3.6.2). RDocumentation. URL: https://www.rdocumentation.org/packages/stats/versions/3.6.2 [accessed 2021-05-28]

37. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc B 2018 Dec 05;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]

38. Lasso and elastic-net regularized generalized linear models (glmnet package). RDocumentation. URL: https://www.rdocumentation.org/packages/glmnet/versions/4.1-1 [accessed 2021-05-28]

39. Quinlan JR. Induction of decision trees. Mach Learn 1986 Mar;1(1):81-106. [doi: 10.1007/bf00116251]

40. rpart package. RDocumentation. URL: https://www.rdocumentation.org/packages/rpart/versions/4.1-15 [accessed 2021-05-28]

41. Breiman L. Random forests. Mach Learn 2001;45:5-32. [doi: 10.1023/A:1010933404324]

42. randomForest package. RDocumentation. URL: https://www.rdocumentation.org/packages/randomForest/versions/4.6-14 [accessed 2021-05-28]

43. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7:21. [doi: 10.3389/fnbot.2013.00021] [Medline: 24409142]

44. gbm: Generalized Boosted Regression Modeling (GBM). RDocumentation. URL: https://www.rdocumentation.org/packages/gbm/versions/2.1.8/topics/gbm [accessed 2021-05-28]

45. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. Intensive Care Med 2003 Jul;29(7):1043-1051. [doi: 10.1007/s00134-003-1761-8] [Medline: 12734652]

46. Anderson M, Anderson S. How should AI be developed, validated, and implemented in patient care? AMA J Ethics 2019 Feb 01;21(2):E125-E130 [FREE Full text] [doi: 10.1001/amajethics.2019.125] [Medline: 30794121]

47. Japkowicz N, Stephen S. The class imbalance problem: a systematic study. Intell Data Anal 2002 Nov 15;6(5):429-449. [doi: 10.3233/ida-2002-6504]

48. Diprose W, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. J Am Med Inform Assoc 2020 Apr 01;27(4):592-600 [FREE Full text] [doi: 10.1093/jamia/ocz229] [Medline: 32106285]

49. Zheng S, Tu L, Cicuttini F, Zhu Z, Han W, Antony B, et al. Depression in patients with knee osteoarthritis: risk factors and associations with joint symptoms. BMC Musculoskelet Disord 2021 Jan 07;22(1):40 [FREE Full text] [doi: 10.1186/s12891-020-03875-1] [Medline: 33413273]

50. Kim KW, Han JW, Cho HJ, Chang CB, Park JH, Lee JJ, et al. Association between comorbid depression and osteoarthritis symptom severity in patients with knee osteoarthritis. J Bone Joint Surg Am 2011 Mar 16;93(6):556-563. [doi: 10.2106/JBJS.I.01344] [Medline: 21411706]

51. Georgiev T, Angelov AK. Modifiable risk factors in knee osteoarthritis: treatment implications. Rheumatol Int 2019 Jul;39(7):1145-1157. [doi: 10.1007/s00296-019-04290-z] [Medline: 30911813]

52. Monroe SM, Slavich GM, Gotlib IH. Life stress and family history for depression: the moderating role of past depressive episodes. J Psychiatr Res 2014 Feb;49:90-95 [FREE Full text] [doi: 10.1016/j.jpsychires.2013.11.005] [Medline: 24308926]

53. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol 2014 Mar 19;14(1):40 [FREE Full text] [doi: 10.1186/1471-2288-14-40] [Medline: 24645774]

54. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med 2006 Feb 07;144(3):201-209. [doi: 10.7326/0003-4819-144-3-200602070-00009] [Medline: 16461965]

55. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988 Dec;15(12):1833-1840. [Medline: 3068365]

56. Ware J, Kosinski M, Keller S. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]

57. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]

58. Shin D, Lee KJ, Adeluwa T, Hur J. Machine learning-based predictive modeling of postpartum depression. J Clin Med 2020 Sep 08;9(9):2899 [FREE Full text] [doi: 10.3390/jcm9092899] [Medline: 32911726]

59. Wang S, Pathak J, Zhang Y. Using electronic health records and machine learning to predict postpartum depression. Stud Health Technol Inform 2019 Aug 21;264:888-892. [doi: 10.3233/SHTI190351] [Medline: 31438052]

60. Cvetković J. Breast cancer patients' depression prediction by machine learning approach. Cancer Invest 2017 Sep 14;35(8):569-572. [doi: 10.1080/07357907.2017.1363892] [Medline: 28872366]

61. Nemesure MD, Heinz MV, Huang R, Jacobson NC. Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Sci Rep 2021 Jan 21;11(1):1980. [doi: 10.1038/s41598-021-81368-4] [Medline: 33479383]

62. Choi J, Choi J, Choi WJ. Predicting depression among community residing older adults: a use of machine learning approach. Stud Health Technol Inform 2018;250:265. [Medline: 29857458]

63. Stubbs B, Aluko Y, Myint PK, Smith TO. Prevalence of depressive symptoms and anxiety in osteoarthritis: a systematic review and meta-analysis. Age Ageing 2016 Mar;45(2):228-235. [doi: 10.1093/ageing/afw001] [Medline: 26795974]

Abbreviations

AUC: area under the receiver operating characteristic curve
CES-D: Center for Epidemiological Studies Depression Scale
GBM: gradient boosting machine
LASSO: least absolute shrinkage and selection operator
ML: machine learning
MOST: Multicenter Osteoarthritis Study
OA: osteoarthritis
OAI: Osteoarthritis Initiative
PASE: Physical Activity Scale for the Elderly
SF-12: 12-item Short Form Health Survey
WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index

XSL•FO
RenderX