

Explainable Machine Learning for Knee Osteoarthritis Diagnosis Based on a Novel Fuzzy Feature Selection Methodology

Christos Kokkotis (✉ chkokkotis@gmail.com)

CERTH: Ethniko Kentro Ereunas & Technologikes Anaptyxes

Charis Ntakolia

National Technical University Of Athens School of Naval Architecture and Marine Engineering: Ethniko Metsobio Polytechneio Schole Naupegon Mechanologon Mechanikon

Serafeim Moustakidis

AIDEAS OU

Giannis Giakas

Department of Physical Education and Sport Science, University of Thessaly

Dimitrios Tsaopoulos

CERTH: Ethniko Kentro Ereunas & Technologikes Anaptyxes

Research Article

Keywords: KOA Diagnosis, Machine Learning, Clinical data, Explainability, Feature selection

Posted Date: August 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-777000/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Physical and Engineering Sciences in Medicine on January 31st, 2022. See the published version at <https://doi.org/10.1007/s13246-022-01106-6>.

Abstract

Knee Osteoarthritis (KOA) is a degenerative joint disease of the knee that results from the progressive loss of cartilage. Due to KOA's multifactorial nature and the poor understanding of its pathophysiology, there is a need for reliable tools that will reduce diagnostic errors made by clinicians. The existence of public databases has facilitated the advent of advanced analytics in KOA research however the heterogeneity of the available data along with the observed high feature dimensionality make this diagnosis task difficult. The objective of the present study is to provide a robust Feature Selection (FS) methodology that could: (i) handle the multidimensional nature of the available datasets and (ii) alleviate the defectiveness of existing feature selection techniques towards the identification of important risk factors which contribute to KOA diagnosis. For this aim, we used multidisciplinary data obtained from the Osteoarthritis Initiative database for individuals without or with KOA. The proposed fuzzy ensemble feature selection methodology aggregates the results of several FS algorithms (filter, wrapper and embedded ones) based on fuzzy logic. The effectiveness of the proposed methodology was evaluated using an extensive experimental setup that involved multiple competing FS algorithms and several well-known ML models. A 73.55 % classification accuracy was achieved by the best performing model (Random Forest classifier) on a group of twenty-one selected risk factors. Explainability analysis was finally performed to quantify the impact of the selected features on the model's output thus enhancing our understanding of the rationale behind the decision-making mechanism of the best model.

Introduction

Knee Osteoarthritis (KOA) is one of the most common types of osteoarthritis and musculoskeletal disorder. Being the 11th highest global disability, KOA is a multifactorial disease that results from mechanical and constitutional factors [1]. Obesity, age, gender, knee injuries and lifestyle are likely risk factors of KOA as they have been highlighted in the relevant recent literature [2]. In addition, swelling, pain and stiffness have been characterized as typical symptoms of the disease with irreversible cartilage damage being KOA's main consequence [3–5]. KOA is closely associated with a huge economic burden for the healthcare system and an unbearable health burden of the patients and their families. Significant consequences of KOA are the social isolation of the individual and the low quality of life [6, 7]. Furthermore, the quantification of KOA is performed with the Kellgren–Lawrence (KL) severity grading scale, which is the most commonly grading system (current gold standard) and consists of five severity grades, from 0 to 4 [8].

Despite the fact that the scientific community has put a lot of effort into KOA research, a major challenge remains with respect to early diagnosis, long-term diagnosis and treatment of KOA. The parallel increase in computing power along with the collection of big datasets combined with the need to address the above challenges has led many research teams to use artificial intelligence (AI) techniques in the field of KOA [9]. In light of the above, several AI enabled studies have been proposed in the recent literature with the objective to diagnose or predict KOA. Yoo et al. used data from the Fifth Korea National Health and Nutrition Examination Surveys (KNHANES V-1) and the Osteoarthritis Initiative (OAI) to build an artificial

neural network (ANN)-based a scoring system for the identification of KOA severity [10]. The proposed ANN model achieved an area under the curve (AUC) of 0.76 for the symptomatic KOA in an external validation with OAI data. In another study, Lim et al. proposed a method for early diagnosis of KOA based on clinical data from Korean National Health and Nutrition Examination Survey (KNHANES) [11]. They achieved an 76.8% AUC by using a deep neural network with scaled principal component analysis. In 2019, Christodoulou et al. investigated the deep learning capabilities in KOA diagnosis [12]. They used clinical data from OAI database and they achieved an 86.95% accuracy working on an aged subgroup (70+).

In another study, Moustakidis et al. worked on self-reported clinical data (OAI) and proposed a deep learning methodology for the recognition of participants being at high risk of developing KOA in at least one knee and participants with symptomatic KOA [13]. They achieved accuracies up to 86.95%. Furthermore, Kwon et al. proposed an automatic classification of KOA severity that made use of gait analysis data and radiographic imaging (from Seoul National University Hospital) [14]. They employed Inception-ResNet-v2 for feature extraction from X-rays and a support vector machine for KOA diagnosis achieving accuracies of 0.93, 0.82, 0.83, 0.88, and 0.97 for the KL grades 0–4, respectively. In addition, Moustakidis et al. proposed a KOA classification approach with a focus on both accuracy and fairness [15]. They worked on different subgroups of participants from self-reported clinical data (OAI) and the dense neural networks methodology improved the accuracy up to 79.6% with fairness measured by balanced equalized odds (~ 92%) and demographic parity (98.5%) in the KOA case study.

Given that medical data and features can be subjective or difficult to interpret, medical decision making has a great potential to benefit by the use of fuzzy logic (FL). FL has been used to diagnose or facilitate decision making systems tackling many diseases, including OA. Hardi et al. proposed an expert system based on the fuzzy Tsukamoto method for OA diagnosis [16]. They treated symptoms of OA as fuzzy values that were further converted into firm value by using a weighted average demonstrating a 90% accuracy in the task of diagnosis of osteoarthritis disease. In general, various feature selection methods have integrated fuzzy logic in their internal mechanisms in order to handle the observed fuzziness and therefore improve the way that features are treated and combined. For instance, with emphasis to medical applications, mutual information method combined with FL was used: (i) to select miRNAs in cancer [17]; (ii) to classify tumors [18]; and to select features for multilabel learning [19]. Similar studies include fuzzy entropy by using thresholds [20] for feature selection in various medical datasets and fuzzy rough sets [21, 22] for dimensionality reduction of feature space to prevent samples from misclassification.

It is well known that each one of the existing FS algorithms comes with its own advantages and disadvantages introducing a certain level of bias. To handle the multidimensional nature of the OAI dataset and to avoid bias and alleviate the defectiveness of single feature selection results, a fuzzy ensemble FS methodology is proposed in this paper that aggregates the results of several FS algorithms (filter, wrapper and embedded ones). Fuzzy logic is employed to combine multiple feature importance scores thus leading to a more robust selection of informative features. The proposed method contributes

to the significant reduction of the initial OAI feature dimensionality and to a decrease of the computational complexity of the classification models employed. To prove the effectiveness of the proposed methodology, an extensive experimental setup was designed involving multiple competing FS algorithms and several well-known ML models. As a post-hoc explainability, SHAP model was finally employed to identify the contribution of the selected features and the rationale behind the decision-making mechanism of best performing model.

Methods

Dataset Description

For the purpose of this study, data were employed from the osteoarthritis initiative (OAI) database (available on <https://nda.nih.gov/oai/>). OAI is a prospective observational, multi-center and longitudinal study of KOA. OAI has enrolled 4796 women and men, aged 45-79 years. The present study used clinical evaluation data (643 features in total) from the baseline visit from all participants with or without KOA. The features of clinical dataset were divided into seven categories as shown in Table 1. Furthermore, in the present study, Kellgren and Lawrence (KL) grades were used as the outcome for the classification task.

Table 1. Main categories of the clinical evaluation data considered in this study.

| Category | Description |
|-------------------------|---------------------------------------------------------------------------------------------------------|
| Medical history | Medications and health histories based on questionnaire results (not included medical imaging outcomes) |
| Symptoms | Arthritis symptoms or health-related disability and function based on questionnaire data |
| Subject characteristics | Includes variables which describe anthropometric parameters and personal information |
| Nutrition | Questionnaire based on Block Food Frequency |
| Physical exam | Includes performance measures and knee and hand exams |
| Physical activity | Questionnaire results regarding living and leisure activities |
| Behavioral | Consists of variables which quantify the social behavior and the quality level of daily routine |

Methodology

The proposed AI methodology for KOA diagnosis consists of five processing steps: i) data pre-processing, ii) application of FS techniques, iii) learning process, iv) evaluation of the classification results and v) explainability analysis, as illustrated in Figure 1. An extensive presentation of the steps of the proposed methodology is given in the following subsections.

Problem Definition

In this study, we defined the KL-grade prediction task as a binary class classification problem. Specifically, the subjects of the study (3872 subjects in total) were divided into two equal groups:

- i) KOA - participants who have KL ≥ 2 at baseline. Participant in the group had KL grades equal (early diagnosis) or higher than 2 in at least one of the two knees or in both at baseline;
- ii) non-KOA - participants who had KL0 or KL1 grade at baseline. Especially, this group of participants do not have KOA in any of their knees.

Data Pre-processing

Mode imputation was employed to handle categorical and continuous missing values [23]. In our study, data were normalised to [0, 1] to build a common basis for the FS algorithms and learning techniques that follow [24]. Furthermore, to cope with the imbalance data problem a stratified strategy for data resampling was applied. In particular, the number of the subjects in the majority class was reduced in order to become equal with the number of samples on the minority class [25].

Proposed FS methodology

The proposed Fuzzy logic enhanced Feature Selection method (FSFL) combines the outputs of six well-known feature selection methods from three feature selection categories (Filter, Wrapper and Embedded). Specifically, from the filter category, the mutual information [26] and the f-ANOVA [27] techniques were applied. From the wrapper category, we employed a recursive feature elimination (RFE) based on logistic regression [28] and an RFE based on support vector machine [29] techniques, respectively. Furthermore, from the embedded category, a LightGBM [30] and a random forest technique [31] were applied. To calculate the importance of a feature for each category, the scores of the associated FS techniques were used as input to the Fuzzy Inference System (FIS) 1 that was implemented with Mamdani inference methodology [32]. The output of the FIS 1 was the defuzzification value that represents the feature importance score for the specific feature selection category. Then, the defuzzification score of each category was used as input to the FIS 2 where the output defuzzification value represents the overall feature importance. Figure 2 illustrates the FSFL flowchart with the defined fuzzy rules for each FIS and the selected feature selection methods for this study. Figure 3 shows the fuzzy sets used in the presented methodology for the input variables for FIS 1 and FIS 2, while Figure 4 shows the fuzzy sets of output variable for FIS 1 and 2.

Learning

In order to handle the demanding task of KOA classification, we investigated various ML models for their suitability and behavior in this problem. Specifically, random forest (RF) [33], multilayer perceptron (MLP) [34], logistic regression (LR) [35], support-vector machines (SVMs) [36], and k-nearest neighbors (KNN) [37] classifiers were tested. Furthermore, to avoid overfitting, and to optimize the performance of our models hyperparameter selection was applied individually per model.

Validation

For the experimental evaluation, a repeated stratified 5-fold cross validation was used [38]. Furthermore, the performance of the classifiers was also evaluated in terms of the recall, f1-score and precision as additional evaluation criteria [39]. A brief description of these metrics is given below. Initially, the accuracy is the ratio of correctly predicted observations to the total observations and can be characterized as the most intuitive performance measure. Recall (or Sensitivity) is the ratio of correctly predicted positive observations to all observations in the actual class. Moreover, the ratio of correctly predicted positive observations to the total predicted positive observations is called precision or positive predictive value. F1-score is the weighted average of Precision and Recall.

Explainability

In the present work, we also examine how the risk factors have contributed to the final decision of KOA diagnosis. In order to achieve this, we used SHapley Additive exPlanations (SHAP), which is an approach to explain individual predictions based on Shapley Values of game theory and local explanations [40,41]. In particular, we employed SHAP to rank features in terms of their impact on the final ML (Random Forest) outputs and to build a mini explainer model, which contributes to understanding the behavioral and the contribution of the risk factors in KOA diagnosis.

Results

In this section, we demonstrate the overall diagnosis performance of the models in relation to the first 100 selected features, and the highest metrics of the best models are also presented. Then, reference is made in the most important risk factors as they have been selected by the proposed Fuzzy FS methodology. Moreover, a comparative analysis is presented to prove the superiority of the proposed FS methodology compared to a number of well-known FS techniques. For the interpretation of the best model, an explainability analysis is employed to enhance our understanding of the reasoning behind its decision-making mechanism.

Diagnosis Performance

This subsection presents the results of a comparative analysis over a number of well-known ML models on the diagnosis classification task by using the first 100 selected risk factors. Figure 5 shows the testing accuracy performance (%) of the competing ML models with respect to the number of selected features. Specifically, KNN failed in diagnosis task, recording low testing accuracy performances. The rest of the

ML models had an upward trend in the range of the first 15 risk factors. Overall, the best overall performance was achieved by RF with a maximum of 73.55% at 21 features.

Furthermore, the classification performance of the best performing ML models was further evaluated with respect to various validation metrics including class precision, recall, and f1-score. Table 2 demonstrates the best performance metrics of RF, MLP, LR, SVMs, and KNN models on the diagnosis task. In particular, RF achieved the best overall performance (73.55% accuracy) on the group of the twenty-one (21) risk factors. SVMs achieved the second-highest accuracy (73.36%). The rest of the ML models achieved lower accuracies.

Table 2. Summary of best metrics per model and number of selected features.

| Models | Accuracy | Precision | Recall | F1-Score | Num. of Features |
|--------|----------|-----------|--------|----------|------------------|
| RF | 73.55 | 73.82 | 73.64 | 73.59 | 21 |
| MLP | 73.20 | 73.48 | 73.20 | 73.13 | 17 |
| LR | 73.27 | 73.38 | 73.27 | 73.24 | 17 |
| SVMs | 73.36 | 73.68 | 73.36 | 73.27 | 18 |
| KNN | 71.55 | 71.74 | 71.55 | 71.49 | 12 |

Features Selected

Figure 6 reveals more information about the origin of the 21 risk factors as selected by the chosen Fuzzy FS approach. As observed in Figure 6, six features describing subject characteristics were among the selected risk factors e.g., the age of the participants, the body mass index (BMI), and the diastolic blood pressure. Moreover, five out of the 21 selected risk factors come from the symptom's category, representing clinical parameters related to stiffness, knee difficulty, swelling, and pain, demonstrating the indication of the existence of KOA. Four of the risk factors are related to physical exams, whereas another two medical history and two physical activity parameters were selected as relevant to KOA occurrence. A behavioural risk factor and a nutrition risk factor were also selected by the proposed Fuzzy FS approach.

Comparative Analysis

The performance of the proposed FSFL methodology was compared with each one of the six FS techniques that were also implemented independently. Finally, another recently published FS technique was also selected as comparative in which the final feature ranking, is decided on the basis of a majority vote scheme [42, 43].

Table 3 shows the maximum achieved accuracy in the first selected 100 features of OAI dataset and the number of features where the maximum accuracy was reached for each feature selection method used in

the experimental evaluation with the best performed model (RF). The last row in table 3 shows the dimensionality reduction achieved with the proposed FS method compared to other competitive methods. Specifically, the metric DR was defined to quantify the difference (%) in dimensionality reduction compared to FSFL:

$$DR = 1 - \frac{\text{Max number of features (FSFL)}}{\text{Max number of features (Compared method)}} \quad (1)$$

The proposed FSFL method achieved the best trade-off between performance and dimensionality reduction being capable of reducing significantly the feature set dimensionality while achieving slightly higher or comparable prediction performance with the rest competing algorithms. Specifically, the proposed FSFL technique reaches the highest accuracy (73.55%) at 21 selected features while the second-best accuracy (73.51%) was achieved by LGBM Emb at 87 features. This shows that the proposed FSFL technique results to a 76% smaller set of selected features compared to the second-best performing technique. On the other hand, the second-best performer with respect to dimensionality reduction was RF Emb with 73.36% accuracy achieved on a considerably larger feature subset with more than double features (43) compared to FSFL (21).

Table 3. Comparative analysis of FS methods.

| | FSFL | Vote FS | RF Emb FS | LGBM Emb FS | SVM RFE FS | LR RFE FS | Filter MI FS | Filter f-ANOVA FS |
|------------------------------------|-------|---------|-----------|-------------|------------|-----------|--------------|-------------------|
| Maximum Accuracy (%) | 73.55 | 72.99 | 73.36 | 73.51 | 70.53 | 73.50 | 72.75 | 73.44 |
| Number of Selected Features | 21 | 76 | 43 | 87 | 96 | 60 | 91 | 53 |
| DR (%) | - | +72% | +51% | +76% | +78% | +65% | +77% | +60% |

Explainability Results

Figure 7a depicts how the features' impact shapes the output of the final model (RF) on the testing dataset. The features are sorted by the sum of SHAP value magnitudes over all testing subjects. Furthermore, the SHAP values are used to demonstrate the contribution of each risk factor (negative or positive) on the model's output. Specifically, blue color represents low feature values, whereas red color represents high values, respectively. In particular, a high value of PO2ELGRISK (knee symptoms, risk factors, or both status) increases the probability of the subjects to be assigned to class KOA. Similarly to PO2ELGRISK, the higher the values of risk factors V00AGE, P02KSRG, P01BM1, V00RKFHDEG, P01WEIGHT, V00LKFHDEG, V00WTMACKG, V00BRDIAS, V00KPLKN1, and P02PA1, the more probable for

subjects to belong to class KOA. The rest of the selected risk factors in Figure 7 have the opposite effect pushing the prediction output of the model to the class of healthy subjects. Figure 7b presents the SHAP global feature importance. The risk factors are sorted by the mean $[|SHAP\ value|]$, which is the average impact on model output magnitude.

Figure 8 interprets locally the behavior of the model for the prediction output in a subject that suffers by KOA. P02ELGRISK (with a value of 2) and P01BMI (with a value of 29.8) push the predictions towards the class of KOA patients. Therefore, a high value of the aforementioned risk factors results to the increase of the output probability of the subject to be classified as KOA patient. On the contrary, increase of the risk factors P02KSURG, V00RKFHDEG, V00KOOSQOL, and V00KOOSKPR lowers the probability of a subject to be classified as KOA. Since, our prediction score = 0.51 > base value = 0.49, this subject has been positively classified, i.e., class KOA status.

Discussion

Handling the multidimensional nature of the OAI dataset, a novel fuzzy ensemble FS methodology was designed, implemented and tested in this paper. Its main novelty lies on the combination of several well-known FS algorithms based on a properly designed fuzzy inference mechanism that effectively aggregates their outputs. The superiority of the proposed FS technique was demonstrated through a thorough comparative investigation that included several state-of-the-art algorithms coming from different FS families (filter, wrapper, embedded and hybrid).

The proposed fuzzy FS methodology outperformed the aforementioned FS techniques achieving the best trade-off between dimensionality reduction and prediction accuracy. Working on a high-dimensional dataset of 643 features, twenty-one risk factors were selected for the objective of KOA diagnosis. Observing the nature of the selected risk factors, it was found that subject characteristics, symptoms, and physical exams are the most important risk factors contributing considerably to the KOA diagnosis. Overall, it was concluded that a combination of heterogeneous risk factors coming from different feature categories is needed for the effective diagnosis of KOA.

To sanity check the AI models beyond mere performance and further quantify the relevance of the selected risk factors, a post hoc explainability analysis was also conducted using SHAP. As observed by SHAP, P02ELGRISK, P02KSURG, V00AGE, P01BMI and V00KOOSQOL are five risk factors that have a major impact to the prediction output, which are in line with the existing literature. Specifically, P02ELGRISK, that represents knee symptoms, is an important risk factor in the diagnosis of KOA, as it has been identified by Lespasio et al. [44]. The history of knee surgery (P02KSURG) has been recognised as an important risk factor of KOA by Katz et al. [45], whereas the age of the subjects was also characterized as crucial in the occurrence of KOA and therefore was considered in the development of a predictive model for KOA diagnosis [13]. The knee injury and osteoarthritis outcome (KOOS) is a well-known knee-specific instrument that has been widely employed to evaluate quality of life in patients with knee injuries and identify patients who are at risk of developing OA [46]. Moreover, high BMI is suggested

to be a high-risk factor in the development of KOA. High BMI values lead to the increment of knee joint mechanical loading [47].

Although the proposed FSFL technique selects a subset of risk factors with a significant dimensionality reduction compared to popular FS techniques, the application of a post-hoc explainability is still important in order to identify the contribution of the selected features to prediction output of the model. The use of explainability analysis algorithms for the interpretation of the ML models increases the understanding of the principle of operation of each ML model and reveal the interactions that shape the diagnosis outcome.

The proposed methodology can be considered as computationally intensive; however, FS is considered here as an offline process and therefore the execution time does not play a crucial role. Future work will focus on the identification of easily measurable biomarkers and biomechanical parameters derived from musculoskeletal models, in combination with the already selected risk factors for the early diagnosis of KOA in the general population. Hence, to achieve this goal more advanced AI analytics tools in combination with the FSFL algorithm will be employed.

Conclusion

To enforce the development of more reliable, powerful, and non-invasive diagnostic tools, this study focuses on the identification and interpretation of the risk factors that contribute on the diagnosis of KOA. The proposed methodology is based on a novel fuzzy logic-based feature selection followed by learning algorithms and subsequently a post-hoc explainability analysis. The proposed technique aggregates the results of several FS algorithms (filter, wrapper and embedded ones), whereas fuzzy logic was employed to combine multiple feature importance scores thus leading to a more robust selection of informative features. The results showed that the presented methodology was capable to select a subset of risk factors that increase the performance accuracy of various ML models, compared to popular FS techniques. This was achieved with a significant decrease on the feature dimensionality (up to 78%). SHAP was finally applied to enhance our understanding of the rationale behind the decision-making mechanism of the selected ML model and the impact of the used risk factors on the prediction output.

Declarations

Funding

This research was funded by the European Community's H2020 Programme, under grant agreement No. 777159 (OACTIVE).

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data from the osteoarthritis initiative (OAI) database (available upon request at <https://nda.nih.gov/oai/>).

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Malanga G, Niazi F, Kidd VD, Lau E, Kurtz SM, Ong KL, Concoff AL (2020) Knee Osteoarthritis Treatment Costs in the Medicare Patient Population. *American health & drug benefits* 13(4) 144.
- [2] Johnson VL, Hunter DJ (2014) The epidemiology of osteoarthritis. *Best practice & research Clinical rheumatology* 28(1) 5-15. 10.1016/j.berh.2014.01.004
- [3] Silverwood V, Blagojevic-Bucknall M, Jinks C, Jordan J, Protheroe J, Jordan K (2015) Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. *Osteoarthritis Cartilage* 23(4) 507-515. 10.1016/j.joca.2014.11.019
- [4] Ackerman IN, Kemp JL, Crossley KM, Culvenor AG, Hinman, RS (2017) Hip and Knee Osteoarthritis Affects Younger People, Too. *J Orthop Sports Phys Ther* 47(2) 67-79. 10.2519/jospt.2017.7286
- [5] Toivanen AT, Heliövaara M, Impivaara O, Arokoski JP, Knekt P, Lauren H, Kröger H (2010) Obesity, physically demanding work and traumatic knee injury are major risk factors for knee osteoarthritis—a population-based study with a follow-up of 22 years. *Rheumatology* 49(2) 308-314. 10.1093/rheumatology/kep388
- [6] Mahir L, Belhaj K, Zahi S, Azanmasso H, Lmidmani F, El Fatimi A (2016) Impact of knee osteoarthritis on the quality of life. *Annals of physical and rehabilitation medicine* 59 e159. <https://doi.org/10.1016/j.rehab.2016.07.355>
- [7] Jack Farr I, Miller LE, Block JE (2013) Quality of life in patients with knee osteoarthritis: a commentary on nonsurgical and surgical treatments. *The open orthopaedics journal* 7 619. 10.2174/1874325001307010619
- [8] Kohn MD, Sassoon AA, Fernando ND (2016) Classifications in brief: Kellgren-Lawrence classification of osteoarthritis. *Clinical Orthopaedics and Related Research®* 474(8) 1886-1893. 10.1007/s11999-016-4732-4

- [9] Kokkotis C, Moustakidis S, Papageorgiou E, Giakas G, Tsaopoulos D (2020) Machine Learning in Knee Osteoarthritis: A Review. *Osteoarthritis and Cartilage Open* 100069. <https://doi.org/10.1016/j.ocarto.2020.100069>
- [10] Yoo TK, Kim DW, Choi SB, Oh E, Park JS (2016) Simple Scoring System and Artificial Neural Network for Knee Osteoarthritis Risk Prediction: A Cross-Sectional Study. *PLoS One* 11(2) e0148724. 10.1371/journal.pone.0148724
- [11] Lim J, Kim J, Cheon S (2019) A Deep Neural Network-Based Method for Early Detection of Osteoarthritis Using Statistical Data. *International journal of environmental research and public health* 16(7) 1281. 10.3390/ijerph16071281
- [12] Christodoulou E, Moustakidis S, Papandrianos N, Tsaopoulos D, Papageorgiou E (2019) *Exploring deep learning capabilities in knee osteoarthritis case study for classification*. Paper presented at the 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA). 10.1109/IISA.2019.8900714
- [13] Moustakidis S, Christodoulou E, Papageorgiou E, Kokkotis C, Papandrianos N, Tsaopoulos D (2019) Application of machine intelligence for osteoarthritis classification: a classical implementation and a quantum perspective. *Quantum Machine Intelligence*. 10.1007/s42484-019-00008-3
- [14] Kwon SB, Han HS, Lee MC, Kim H C, Ku Y (2020) Machine learning-based automatic classification of knee osteoarthritis severity using gait data and radiographic images. *IEEE Access* 8 120597-120603. 10.1109/ACCESS.2020.3006335
- [15] Moustakidis S, Papandrianos N, Christodolou E, Papageorgiou E, Tsaopoulos D (2020) Dense neural networks in knee osteoarthritis classification: a study on accuracy and fairness. *Neural Computing and Applications* 1-13. <https://doi.org/10.1007/s00521-020-05459-5>
- [16] Hardi S, Triwiyono A (2020) *Expert System for Diagnosing Osteoarthritis with Fuzzy Tsukamoto Method*. Paper presented at the Journal of Physics: Conference Series. 10.1088/1742-6596/1641/1/012107
- [17] Pal JK, Ray SS, Pal SK (2017) Fuzzy mutual information based grouping and new fitness function for PSO in selection of miRNAs in cancer. *Computers in biology and medicine* 89 540-548. 10.1016/j.combiomed.2017.08.013
- [18] Dai J, Chen J (2020) Feature selection via normative fuzzy information weight with application into tumor classification. *Applied Soft Computing* 92 106299. <https://doi.org/10.1016/j.asoc.2020.106299>
- [19] Lin Y, Hu Q, Liu J, Li J, Wu X (2017) Streaming feature selection for multilabel learning based on fuzzy mutual information. *IEEE Transactions on Fuzzy Systems* 25(6) 1491-1507. 10.1109/TFUZZ.2017.2735947

- [20] Jaganathan P, Kuppuchamy R (2013) A threshold fuzzy entropy based feature selection for medical database classification. *Computers in biology and medicine* 43(12) 2222-2229. <https://doi.org/10.1016/j.compbiomed.2013.10.016>
- [21] Wang C, Qi Y, Shao M, Hu Q, Chen D, Qian Y, Lin Y (2016) A fitting model for feature selection with fuzzy rough sets. *IEEE Transactions on Fuzzy Systems* 25(4) 741-753. 10.1109/TFUZZ.2016.2574918
- [22] Qian Y, Wang Q, Cheng H, Liang J, Dang C (2015) Fuzzy-rough feature selection accelerator. *Fuzzy Sets and Systems* 258 61-78. <https://doi.org/10.1016/j.fss.2014.04.029>
- [23] Silva-Ramírez EL, Pino-Mejías R, López-Coello M, Cubiles-de-la-Vega MD (2011) Missing value imputation on missing completely at random data using multilayer perceptrons. *Neural Networks* 24(1) 121-129. <https://doi.org/10.1016/j.neunet.2010.09.008>
- [24] Pihera J, Musliu N (2014) *Application of machine learning to algorithm selection for TSP*. Paper presented at the 2014 IEEE 26th International Conference on Tools with Artificial Intelligence. 10.1109/ICTAI.2014.18
- [25] Japkowicz N (2000) *Learning from imbalanced data sets: a comparison of various strategies*. Paper presented at the AAAI workshop on learning from imbalanced data sets.
- [26] Vergara JR, Estévez PA (2014) A review of feature selection methods based on mutual information. *Neural Computing and Applications* 24(1) 175-186. <https://doi.org/10.1007/s00521-013-1368-0>
- [27] Suchwalko A, Buzalewicz I, Podbielska H (2013). *Identification of bacteria species by using morphological and textural properties of bacterial colonies diffraction patterns*. Paper presented at the Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection. <https://doi.org/10.1117/12.2020337>
- [28] Zhao J, Karimzadeh M, Masjedi A, Wang T, Zhang X, Crawford MM, Ebert DS (2019) *FeatureExplorer: Interactive feature selection and exploration of regression models for hyperspectral images*. Paper presented at the 2019 IEEE Visualization Conference (VIS). 10.1109/VISUAL.2019.8933619
- [29] Ding J, Shi J, Wu FX (2011) SVM-RFE based feature selection for tandem mass spectrum quality assessment. *International journal of data mining and bioinformatics* 5(1) 73-88. 10.1504/ijdmb.2011.038578
- [30] Ye Y, Liu C, Zemiti N, Yang C (2019) *Optimal feature selection for EMG-based finger force estimation using lightGBM model*. Paper presented at the 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). 10.1109/RO-MAN46459.2019.8956453
- [31] Mate Y, Somai N (2021) *Hybrid Feature Selection and Bayesian Optimization with Machine Learning for Breast Cancer Prediction*. Paper presented at the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). 10.1109/ICACCS51430.2021.9441914

- [32] Gayathri B, Sumathi C (2015). *Mamdani fuzzy inference system for breast cancer risk detection*. Paper presented at the 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICIC). 10.1109/ICIC.2015.7435670
- [33] Ram M, Najafi A, Shakeri MT (2017) Classification and biomarker genes selection for cancer gene expression data using random forest. *Iranian journal of pathology* 12(4) 339. PMID: 29563929 PMCID: PMC5844678
- [34] Parisi L, Biggs PR, Whatling GM, Holt CA (2015) *A novel comparison of artificial intelligence methods for diagnosing knee osteoarthritis*. Paper presented at the XXV congress of the international society of biomechanics. 10.13140/RG.2.1.4197.6163
- [35] Ntakolia C, Kokkotis C, Moustakidis S, Tsaopoulos D (2020) *A machine learning pipeline for predicting joint space narrowing in knee osteoarthritis patients*. Paper presented at the 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE). 10.1109/BIBE50027.2020.00158
- [36] Kubkaddi S, Ravikumar K (2017) Early detection of Knee Osteoarthritis using SVM Classifier. *IJSEAT* 5(3) 259-262.
- [37] Long NP, Park S, Anh NH, Min JE, Yoon SJ, Kim HM, Lim J (2019) Efficacy of integrating a novel 16-gene biomarker panel and intelligence classifiers for differential diagnosis of rheumatoid arthritis and osteoarthritis. *Journal of clinical medicine* 8(1) 50. 10.3390/jcm8010050
- [38] Wong TT, Yeh PY (2019) Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering* 32(8) 1586-1594. 10.1109/TKDE.2019.2912815
- [39] Ghosh M, Sanyal G (2018) An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning. *Journal of Big Data* 5(1) 1-25.
<https://doi.org/10.1186/s40537-018-0152-5>
- [40] Lundberg SM, Lee SI (2017) *A unified approach to interpreting model predictions*. Paper presented at the Advances in neural information processing systems.
<https://dl.acm.org/doi/10.5555/3295222.3295230>
- [41] Nohara Y, Matsumoto K, Soejima H, Nakashima N (2019) *Explanation of machine learning models using improved Shapley Additive Explanation*. Paper presented at the Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. 10.1145/3307339.3343255
- [42] Kokkotis C, Moustakidis S, Giakas G, Tsaopoulos D (2020) Identification of Risk Factors and Machine Learning-Based Prediction Models for Knee Osteoarthritis Patients. *Applied Sciences* 10(19) 6797.
<https://doi.org/10.3390/app10196797>

- [43] Ntakolia C, Kokkotis C, Moustakidis S, Tsaopoulos D (2021) Prediction of Joint Space Narrowing Progression in Knee Osteoarthritis Patients. *Diagnostics* 11(2) 285. <https://doi.org/10.3390/diagnostics11020285>
- [44] Lespasio MJ, Piuzzi NS, Husni ME, Muschler GF, Guarino A, Mont MA (2017) Knee osteoarthritis: a primer. *The Permanente Journal* 21. 10.7812/TPP/16-183
- [45] Katz JN, Arant KR, Loeser RF (2021) Diagnosis and treatment of hip and knee osteoarthritis: a review. *Jama* 325(6) 568-578. 10.1001/jama.2020.22171
- [46] Roos EM, Lohmander, LS (2003) The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health and quality of life outcomes* 1(1) 1-8. 10.1186/1477-7525-1-64
- [47] Cooper C, Snow S, McAlindon TE, Kellingray S, Stuart B, Coggon D, Dieppe PA (2000) Risk factors for the incidence and progression of radiographic knee osteoarthritis. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* 43(5) 995-1000. 10.1002/1529-0131(200005)43:5<995::AID-ANR6>3.0.CO;2-1

Figures

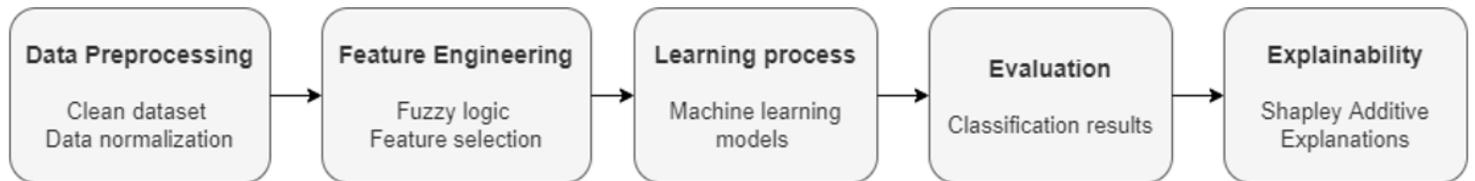


Figure 1

The proposed AI methodology for KOA diagnosis.

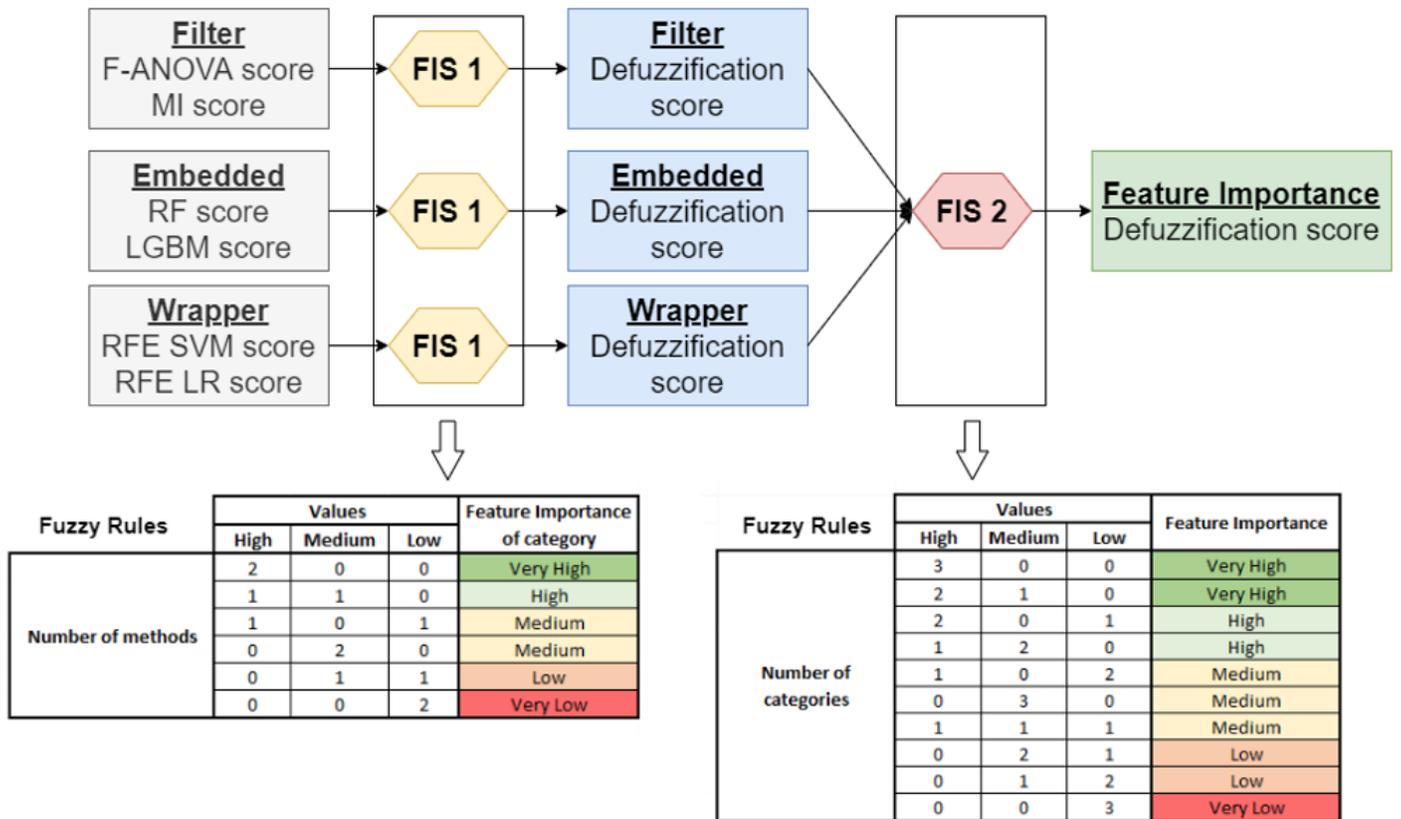


Figure 2

Feature Selection method based on Fuzzy Logic flowchart.

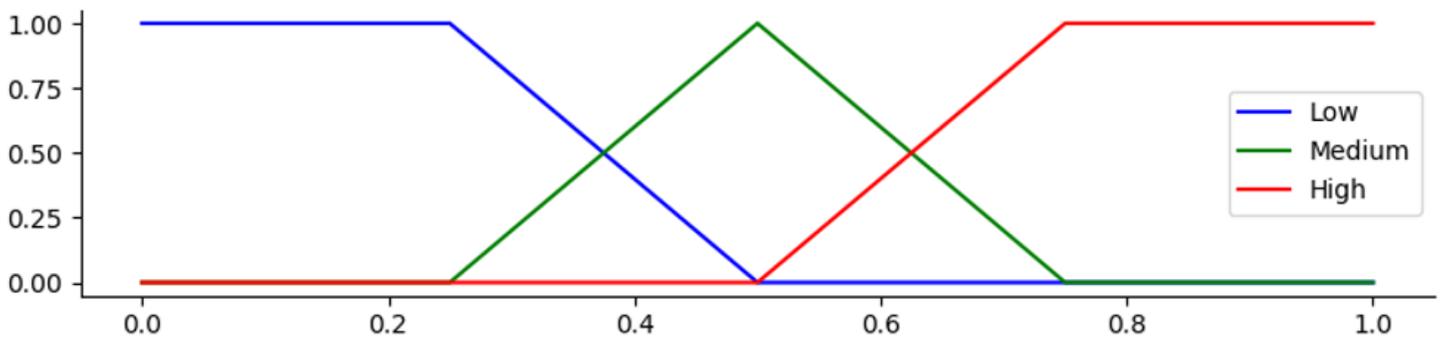


Figure 3

Fuzzy set of input variables for FIS 1 and 2.

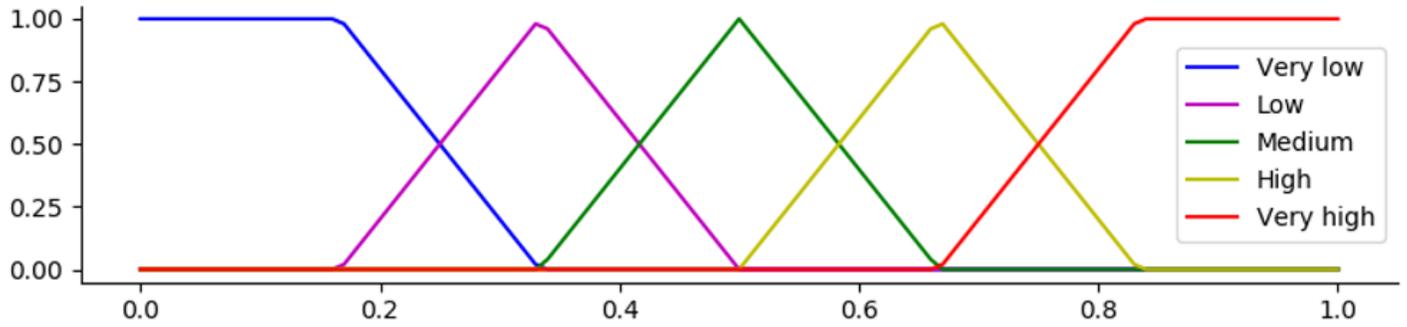


Figure 4

Fuzzy set of output variable for FIS 1 and 2.

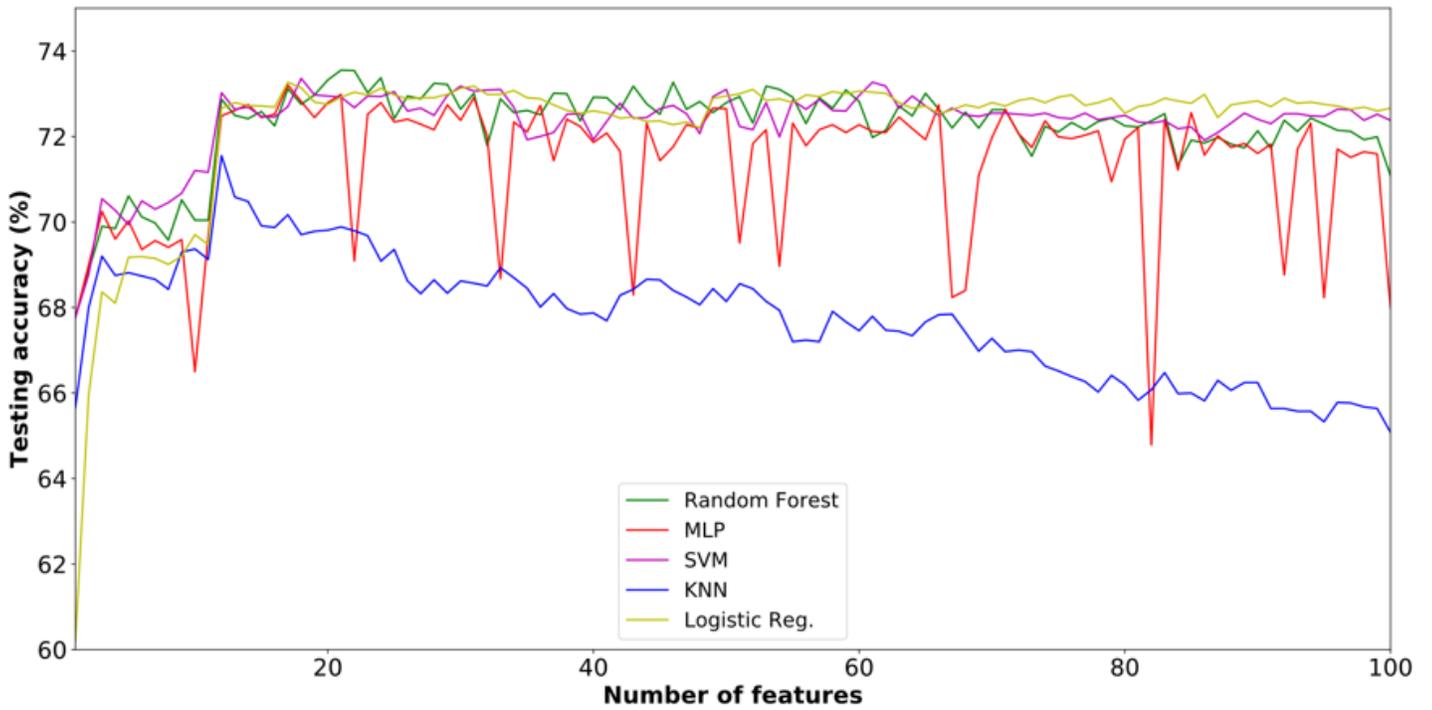


Figure 5

Curves with testing accuracy scores with respect to the number of selected features for different ML models.

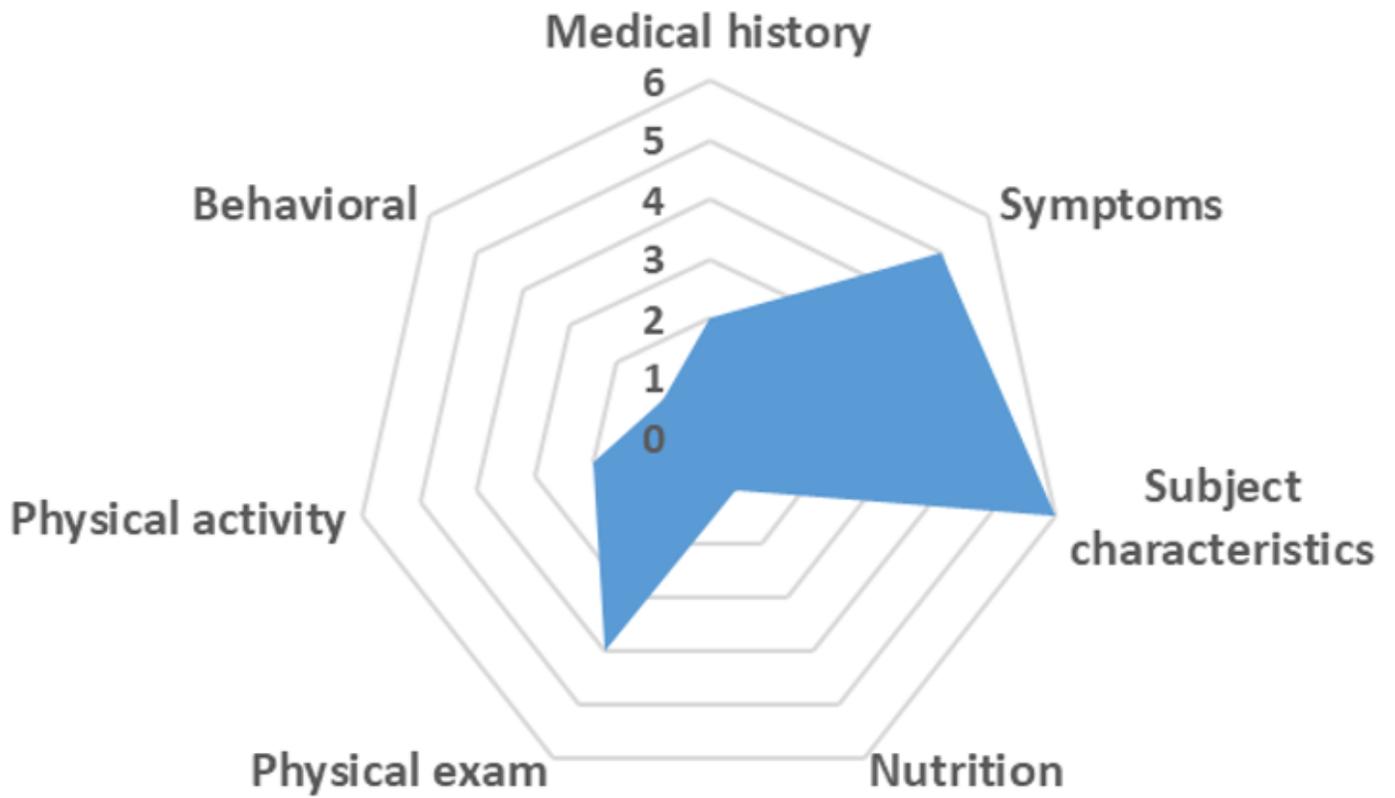


Figure 6

The 21 most informative selected risk factors per category.

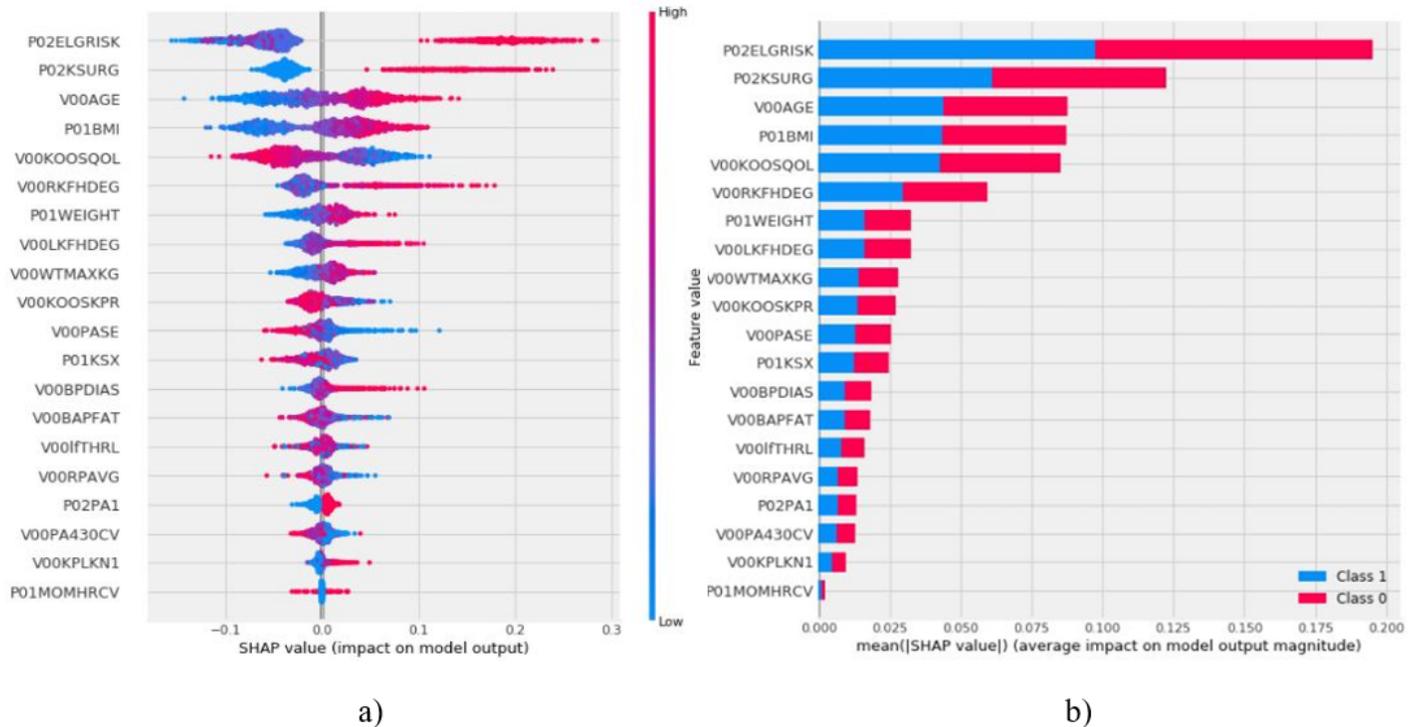


Figure 7

a) Features' impact on Random Forest (21F) model output for the testing set of OAI dataset. b) Features' average impact magnitude for testing instances.

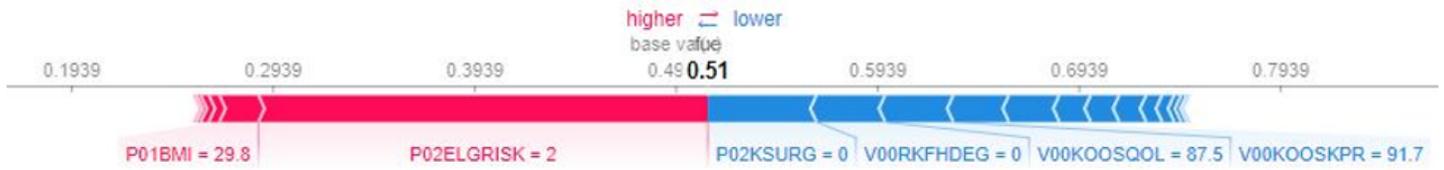


Figure 8

Risk factors contributions to ML model output for a KOA status subject

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixA.docx](#)