

# The application of machine learning in early diagnosis of osteoarthritis: a narrative review

Anran Xuan<sup>\*</sup>, Haowei Chen<sup>\*</sup>, Tianyu Chen, Jia Li, Shilong Lu, Tianxiang Fan, Dong Zeng, Zhibo Wen, Jianhua Ma, David Hunter<sup>\*</sup>, Changhai Ding<sup>\*</sup> and Zhaozhua Zhu

**Abstract:** Osteoarthritis (OA) is the commonest musculoskeletal disease worldwide, with an increasing prevalence due to aging. It causes joint pain and disability, decreased quality of life, and a huge burden on healthcare services for society. However, the current main diagnostic methods are not suitable for early diagnosing patients of OA. The use of machine learning (ML) in OA diagnosis has increased dramatically in the past few years. Hence, in this review article, we describe the research progress in the application of ML in the early diagnosis of OA, discuss the current trends and limitations of ML approaches, and propose future research priorities to apply the tools in the field of OA. Accurate ML-based predictive models with imaging techniques that are sensitive to early changes in OA ahead of the emergence of clinical features are expected to address the current dilemma. The diagnostic ability of the fusion model that combines multidimensional information makes patient-specific early diagnosis and prognosis estimation of OA possible in the future.

**Keywords:** early diagnosis, machine learning, osteoarthritis, prediction model

Received: 2 September 2022; revised manuscript accepted: 1 February 2023.

## Introduction

Osteoarthritis (OA), one of the commonest forms of rheumatic diseases, is an age-related joint disorder. Pain and dysfunction are major symptoms of OA, and the knee joint is the most affected site.<sup>1,2</sup> OA affects more than 520 million people globally in 2019,<sup>3,4</sup> and the aging global population is driving a much greater increase in absolute number of new cases of OA, causing a growing burden on health services and societies worldwide. In the United States, the estimated medical costs and lost earnings caused by OA is 303 billion dollars annually.<sup>5</sup>

To date, there are no disease-modifying drugs that can delay the progression of OA. Early identifying patients with OA or at a higher risk of progression could facilitate clinical decision-making and allow the design of more effective and specific therapeutic interventions. To date, X-rays are the 'gold standard' for radiographic osteoarthritis

(ROA) diagnosis. However, these X-rays lack sensitivity and specificity in changes of bone and cartilage, which leads to a delay in timely clinical intervention.<sup>6</sup> Magnetic resonance imaging (MRI) can detect pre-radiographic abnormalities in bone and soft tissues.<sup>7</sup> There are multiple evaluation methods to assess OA abnormalities using MRI; however, the commonly accepted diagnostic standard for OA using MRI is still absent. In addition, clinical symptoms in many patients with OA are not consistent with imaging findings.<sup>8</sup>

Research on OA biomarkers has attracted much attention, with the latest development of sequencing and imaging technologies.<sup>9,10</sup> Currently, biochemical markers are mainly derived from blood, urine, and joint synovial fluid.<sup>11</sup> In the last few years, a growing number of investigators have been working on establishing quantitative imaging biomarkers as reproducible and reliable indicators for early diagnosis of OA.<sup>12</sup> Biochemical

*Ther Adv Musculoskelet Dis*

2023, Vol. 15: 1–17

DOI: 10.1177/  
1759720X231158198

© The Author(s), 2023.  
Article reuse guidelines:  
sagepub.com/journals-  
permissions

Correspondence to:

**Zhaozhua Zhu**  
Clinical Research Centre,  
Zhujiang Hospital,  
Southern Medical  
University, Guangzhou  
510280, China  
[zhaozhua.zhu@uts.edu.au](mailto:zhaozhua.zhu@uts.edu.au)

**Changhai Ding**  
Clinical Research Centre,  
Zhujiang Hospital,  
Southern Medical  
University, 261 Industry  
Road, Guangzhou, 510280,  
China.

Department of  
Rheumatology, Guangzhou  
First People's Hospital,  
School of Medicine,  
South China University of  
Technology, Guangzhou,  
China

Department of  
Orthopaedics, Affiliated  
Hospital of Youjiang  
Medical University for  
Nationalities, Baise, China  
Menzius Institute for  
Medical Research,  
University of Tasmania,  
Hobart, TAS, Australia  
[changhai.ding@uts.edu.au](mailto:changhai.ding@uts.edu.au)

**Anran Xuan**  
The Second Clinical  
Medical School, Zhujiang  
Hospital, Southern Medical  
University, Guangzhou,  
China

**Haowei Chen**  
The Second Clinical  
Medical School, Zhujiang  
Hospital, Southern Medical  
University, Guangzhou,  
China

Clinical Research Centre,  
Zhujiang Hospital,  
Southern Medical  
University, Guangzhou,  
China

**Tianyu Chen**  
**Tianxiang Fan**  
Clinical Research Centre,  
Zhujiang Hospital,  
Southern Medical  
University, Guangzhou,  
China

**Jia Li**  
Division of Orthopaedic  
Surgery, Department of  
Orthopaedics, Nafang  
Hospital, Southern  
Medical University,  
Guangzhou, China

**Shilong Lu**  
**Zhibo Wen**  
Department of Radiology,  
Zhujiang Hospital,  
Southern Medical  
University, Guangzhou,  
China

**Dong Zeng**  
College of Automation  
Science and Engineering,  
South China University of  
Technology, Guangzhou,  
China

**Jianhua Ma**  
School of Biomedical  
Engineering, Southern  
Medical University,  
Guangzhou, China

**David Hunter**  
Clinical Research Centre,  
Zhujiang Hospital,  
Southern Medical  
University, Guangzhou,  
China

Department of  
Rheumatology, Royal  
North Shore Hospital  
and Institute of Bone and  
Joint Research, Kolling  
Institute, University of  
Sydney, Sydney, NSW,  
Australia

\*These authors  
contributed equally.

and radiographic factors are the major categories of intense research for identifying OA biomarkers.<sup>13,14</sup>

Machine learning (ML) is a type of artificial intelligence (AI) that applies algorithmic methods and enables machines to solve problems without specific computer programming.<sup>15</sup> ML can be mainly divided into supervised learning and unsupervised learning. The most common classifications of supervised learning include classification and regression. Typical examples of unsupervised learning include clustering, including k-means clustering, Hierarchical clustering and dendrograms. ML was first promoted in the 1950s and has recently shown impressive performance in a variety of domains.<sup>16</sup> Automatic diagnosis and evaluation of disease prognosis based on ML models have shown promising potential in fields such as cancer, sepsis, and ankylosing spondylitis.<sup>17–19</sup> The use of ML in OA diagnosis has also increased dramatically in recent years. Therefore, we aimed to review the research progress on ML in the early diagnosis of OA when patients have few or no symptoms, but there may be early sub-clinical structural changes.<sup>20</sup> We also discuss the current trends and limitations of ML approaches and propose future research priorities for this approach in the field of OA.

## Materials and methods

A literature search was conducted using different databases such as PubMed, Web of Science, Embase, Google Scholar, and Cochrane Library for all types of articles on the application of ML in the early diagnosis of OA. Additional references were identified by searching the bibliographies of retrieved articles. We performed a keyword search restricted to the abstract of the articles, using terms relating to ML and early diagnosis of OA. The date of publication was not used as an entry selection criterion. Articles were reviewed according to their titles and abstracts, with all article types taken into consideration if they provided data relevant to the research questions.

## ML application in conventional radiographs and computed tomography scans

X-ray imaging is frequently used to define ROA, because it is cost-effective, easy, and rapid.<sup>6</sup> The Kellgren–Lawrence (KL) grading system and Osteoarthritis Research Society International (OARSI) atlas are both widely used, yet both

suffer from the subjectivity of the readers.<sup>21</sup> ML could be a promising technique to improve this situation by automating the two grading systems.<sup>21</sup> One study comparing the reliability of ROA assessment between non-clinical readers and an experienced radiologist, suggested that in cost-constrained settings, ML models may be potentially suitable readers of ROA for increased throughput in image inspection and accuracy, that can achieve ‘super-human performances’.<sup>22</sup> Antony<sup>23</sup> proposed an approach based on a convolutional neural network (CNN) for the synchronous analysis of KL and OARSI grades using two large public data sets, the Osteoarthritis Initiative (OAI) study and the Multi-center Osteoarthritis (MOST) study. Nevertheless, the agreements between the predictions and the labels of the test set for KL and OARSI grades were lower than the human observers’ inter-rater agreements.<sup>21</sup> In the study by Tiulpin and Saarakkala,<sup>21</sup> a method based on deep learning (DL) was developed to automatically perform KL and OARSI grading simultaneously from knee X-ray images using transfer learning, and demonstrated an excellent agreement with the test set labels. This was the first study of OA that used an independent test set for automatic OARSI grading from plain radiographs. Brahim *et al.*<sup>24</sup> developed a computer-aided diagnosis (CAD) system to detect early knee OA (KOA) using knee radiographs and ML algorithms. This approach was applied to the classification using naive Bayes and random forest (RF) classifiers on 1024 knee radiographs from the OAI database, which showed excellent diagnostic values for OA (82.98% accuracy, 87.15% sensitivity, and up to 80.65% specificity). However, it does not account for the variations in the acquisition parameters or the impact of soft tissue on X-ray absorption and diffusion. To build a clinically robust applicable model, another study produced the most advanced methods for KL grade, trained solely with the MOST data set and tested with the OAI data set. This method yielded a quadratic kappa coefficient of 0.83. The area under the receiver operating characteristic (ROC) curve was 0.93, and the average multiclass accuracy was 66.71%. This research indicates that the ability of the model to learn features related to OA is robust toward different data settings,<sup>25</sup> being practical in clinical settings. Bayramoglu *et al.*<sup>26</sup> used a landmark detection tool (Bone Finder) to automatically detect the patella region of interest (ROI) on the lateral view of 5507 knee X-ray images from the MOST database. They trained an ML model

using a gradient boosting machine for the radiographic patellofemoral OA (PFOA) for texture detection, using local binary pattern (LBP) features. Furthermore, a deep neural network (DNN) method, that trained end-to-end, was used to directly detect PFOA on the texture patches. The results showed that textural ROI classification using DNN had good prediction performance, demonstrating the potential of using texture features of the patella in knee radiographs for PFOA prediction.<sup>27</sup>

To automatically diagnose hip OA, a DNN method was trained and validated on 420 hip plain radiographs from Beijing Chaoyang Hospital, China.<sup>6</sup> The study applied a verified CNN model and compared it with diagnoses from physicians with different experience levels. The CNN model's diagnostic performance was comparable to that of an attending physician with 10 years of experience.<sup>6</sup> Another study adopted a model based on end-to-end DL, namely, faster R-CNN (region-CNN), that can directly locate the knee joint in plain radiographs and quantify KOA severity automatically. To fit the real data of the radiographs, a study by Bin *et al.*<sup>28</sup> made some adjustments, which indicated that the adjusted model outperformed the faster R-CNN. It yielded a mean precision of nearly 0.82, a sensitivity of 78%, and a specificity of 94%, and only took 0.33 s to test each image, and thus achieved a better balance between accuracy and speed.

An early stage of KOA detection technique using ML for feature extraction and classification was presented by Mahum *et al.*<sup>29</sup> The input radiographs taken from Mendeley Data set VI were pre-processed, where features of ROA were extracted by using hybrid feature descriptors such as CNN through LBP, and CNN with a histogram of oriented gradient (HOG). Multi-class classifiers, including a support vector machine (SVM), RF, and K-nearest neighbor (KNN), were used for KOA classification that performed five-fold validation and cross-validation. The HOG feature descriptor yielded an accuracy of 97.14% in cross-validation and 98% in five-fold validation for the early KOA detection and classification. The proposed algorithm yielded 93% average accuracy within 1870 iterations.

CT has a high resolution to reflect pathological conditions more clearly and a high detection rate for KOA.<sup>30</sup> A novel approach using phase-contrast imaging (PCI) with CT for the

knee cartilage matrix visualization enables direct examination of chondrocyte patterns and their subsequent correlation with OA.<sup>14</sup> Nagarajan *et al.*<sup>14</sup> extracted statistical features using gray-level co-occurrence matrices (GLCMs) and geometric features using scaling index method (SIM) from 404 ROIs that annotated on PCI images of normal and osteoarthritic specimens. Both feature extractions achieved great performance with an area under the curve of ROC (AUC) value of 0.98 in classification between normal and OA, suggesting these quantitative knee cartilage matrix analyses can highly accurately distinguish between normal and diseased tissue patterns. Abidin *et al.*<sup>31</sup> used two different CNN models, CaffeNet and Inception-v3, to extract features for characterizing healthy and diseased tissues. They used a classification task measured by the ROC curve to quantitatively evaluate the features and visualize them by *t*-distributed stochastic neighbor embedding (t-SNE) a dimension reduction approach. Using features from the last convolutional layer and fully connected layer for CaffeNet showed the best classification performance of an AUC up to 0.91. Simultaneously, the approach using Inception-v3 yielded similar classification performance of an AUC up to 0.95. The visualized features from these layers showed adequate characterization of chondrocyte patterns in classification between normal and osteoarthritic tissues. These techniques can potentially be used to detect the presence of the early osteoarthritic changes in the articular cartilage, which was previously impossible on CT images. For early diagnosis of OA, imaging changes could be earlier than the emergence of clinical symptoms. The application of ML in radiological imaging makes the early diagnosis of OA possible. However, studies have not yet been carried out on a large-scale clinical application, and its clinical benefits need to be further explored.

### ML application in MRI

MRI plays an increasingly essential role in OA diagnosis and related research, and has been regarded as a hot spot in combination with computer-aided diagnostic technology.<sup>32</sup> According to previous studies, cartilage degeneration is not a geometrically regular process.<sup>33</sup> Tibial trabecular bone structural markers detected in MRI could be effective predictors of articular cartilage loss.<sup>34</sup> A community-based cohort study, including 159 participants with MRI images of both the left and right knees, presented a framework using ML and

texture analysis to diagnose ROA. They used an ML approach to apply feature selection and extraction to identify the linear combination of the features that best related to the volume of cartilage loss. However, due to the limited sample size and selection bias, the results can only be partially applied to patients at the early stages of ROA.<sup>34</sup> Chang *et al.*<sup>35</sup> used a methodology based on bone flattening to estimate the presence of osteophytes and ROA ( $KL \geq 2$ ). The AUC for osteophytes was 0.85. However, the test set size was relatively small, suggesting that the results require further validation.

Kundu *et al.*<sup>36</sup> developed a model to distinguish between progressors and non-progressors from baseline cartilage texture maps that enabled the early detection of OA using pre-symptomatic cartilage texture maps derived from transport-based learning. Their work in detecting future symptomatic OA progression 3 years prior to its occurrence, achieved a robust test accuracy of 0.78 and also demonstrated that early OA diagnosis was possible at a potentially reversible phase. In another study, 313 knee scans were included, and six different ML classifiers were used to analyze trabecular bone structures and quantify OA risks. They found that the optimal approaches were linear discriminant analysis (LDA) (with a ROC of 0.82) and weighted KNN (with a ROC of 0.81). The results demonstrated the feasibility of using ML to separate healthy and OA knees based on trabecular bone structures. Moreover, due to the fully automatic framework, the markers were well-suited for longitudinal clinical studies.<sup>37</sup> In a study consisting of 68 participants from the OAI database, an ML algorithm—weighted neighbor distance using a compound hierarchy of algorithms representing morphology (WND-CHRM) was applied to T2 mapping of the central medial femoral condyle. This study found that WND-CHRM could classify T2 maps of cartilage, which has 75% accuracy in predicting the development of OA.<sup>12</sup> Another study used fast imaging and employed a steady-state acquisition sequence to quantitatively analyze bone. A total of 665 right knees from the Rotterdam Study, scanned with 1.5 T MRI and elastic net models, were used to differentiate knees with and without MRI-defined OA. The image feature (17 shape features, 3 orientation features, 12 histogram-based features, and 271 texture features) model had an AUC under 0.80. These results suggest that radiomic features from the subchondral bone are useful imaging biomarkers for the diagnosis of OA.<sup>38</sup>

The quality of infrapatellar fat pad (IPFP) is considered as a marker to predict KOA.<sup>39</sup> However, current semiquantitative scoring system is largely constrained to reader-dependent reliability, and signal alterations of the IPFP are considered non-specific findings.<sup>40</sup> Jia Li *et al.*<sup>41</sup> investigated 690 participants with KOA risk from the Pivotal OAI MRI Analyses incident OA cohort, being distributed to development ( $n = 500$ , 340 women; mean age, 60 years) and test ( $n = 190$ , 120 women; mean age, 61 years) cohorts. All knees had a KL grade of 1 or less at baseline. IPFP texture features were extracted by three-dimensional (3D) texture analysis based on MRI. In both cohorts, IPFP texture scores ( $AUC \geq 0.75$ ) showed greater discrimination than clinical scores ( $AUC \leq 0.6$ ) at baseline, P-1 (1 year before P0), and P0 (the incident time point of KOA), with significant differences in pairwise comparisons ( $p \leq 0.002$ ). Greater predictive and concurrent validities of IPFP texture scores ( $AUC \geq 0.75$ ) compared with MRI Osteoarthritis Knee Scores ( $AUC \leq 0.66$ ) were also demonstrated ( $p < 0.001$ ). Therefore, MRI-based 3D texture of the IPFP could help predict incident radiographic KOA and the development of incident KOA.

The low resolution of MRI of the knee joint seriously affects the diagnosis of KOA. In one study, a high resolution ( $\sim 0.3$  mm in plane pixel size) was applied to make an accurate early diagnosis.<sup>42</sup> Another study compared the efficient medical image super-resolution (EMISR) method with the sparse coding-based network (SCN), super-resolution convolutional neural network (SRCNN), and efficient sub-pixel convolution neural network (ESPCN) methods. They found that the reconstruction quality of EMISR was higher, and the reconstruction speed of EMISR was faster than its counterparts.<sup>43</sup> However, T1 and T2 alone are not representative of cartilage changes; thus, combining them may improve the sensitivity and specificity of OA diagnosis. Pedoia *et al.*<sup>44</sup> mapped 178 participants by integrating demographic and clinical information, gait kinematics, kinetics, cartilage compositional  $T_{1\rho}$  and  $T_2$ , and  $R_2-R_{1\rho}$  ( $1/T_2 - 1/T_{1\rho}$ ) acquired at 3 T and whole-organ magnetic resonance imaging score (WORMS) morphological grading. This study suggested  $R_2-R_{1\rho}$  as an early predictor of cartilage lesion progression in KOA, reflecting the advantage of data-driven topological data analysis (TDA). However, an unsupervised learning method was used in this study. In the future, a supervised learning method should be adopted to



improve predictive ability.<sup>44</sup> Although current developments in MRI hardware, such as compositional and functional assessments, can provide helpful information for clinicians, challenges remain in selecting the best features from a large number of variables to integrate into prediction models. Different types of MRI scanning may have an impact on the sensitivity and accuracy of the prediction model. Therefore, further research should make full use of MRI data and explore the best features of MRI to reflect the imaging changes of OA for early diagnosis of OA and make a more accurate diagnosis.

### ML application in biochemical biomarkers

As OA is a multifaceted chronic joint disease, it is not very persuasive to study OA from an imaging perspective only. Biomechanical biomarkers that have the potential to diagnose OA have also been identified, with the development of sequencing technology and application of ML approaches.

Diagnostic algorithms were applied in a study comprising 46 patients with early OA in the training data set. Glycated, oxidized, and nitrated proteins and amino acids were accessed in the synovial fluid and plasma of patients with early OA.<sup>45</sup> In the algorithm-based model, feature selection by the ML process produced algorithms with multiple oxidation, nitration, and glycation adducts, suggesting that assessment of these adducts increase the diagnostic ability of discriminating early-stage disease from control.<sup>45</sup> However, due to the complexity and heterogeneity of OA, a single biomarker is unlikely to achieve a comprehensive classification. Thus, the investigation of aggregate biomarkers based on both biochemical and imaging assessments is complementary and could be a superior strategy for the early diagnosis of OA. Using publicly available data from the Foundation for the National Institutes of Health (FNIH) cohort study, Nelson *et al.*<sup>46</sup> incorporated data on biomechanical, demographic and clinical variables, and MRI assessment, to simultaneously distinguish different KOA phenotypes. This was achieved through the application of novel ML methodologies, including the distance-weighted discrimination (DWD), t-SNE, and principal components analysis (PCA), which have identified various variables that are mostly related to progressor status (e.g. bone marrow lesions, osteophytes, medial meniscal extrusion, and urine CTX-II). These innovative ML algorithms provide a way to

estimate numerous variables of multiple types and scalings simultaneously in relation to KOA progression, and could be applied to other fields in the future. However, the authors admitted that this study was a preliminary step, requiring further internal validation using other methodologies in this cohort, and external validation in other data sets, to test the validity of the results.

Temporomandibular joint (TMJ) disease is the second most common musculoskeletal disorder. A study enrolled 46 TMJ-OA patients and 46 healthy controls and used four ML models (logistic regression (LR), RF, LightGBM, and XGBoost) to analyze 52 features, including 5 clinical variables, 20 radiomics features from high-resolution cone-beam tomography imaging (HR-CBCT), 25 biomolecular features (13 from serum and 12 from saliva), and 2 demographic variables (age and gender). Results showed that XGBoost + LightGBM model achieved accuracy of 0.82 (SD 0.03), and an AUC of 0.87. Composed of clinical, biological, and radiomics features, this study is expected to boost future studies by integrating multidimensional information for the diagnosis of OA.<sup>47</sup> Another study by Celia Le *et al.*<sup>48</sup> included 46 TMJ OA patients and 46 healthy controls. 52 features were obtained from ROI, which was the lower part of the TMJ. Meanwhile, 32 mandibular fossa radiomic features from the upper part of the TMJ were tested. A total of 3828 features and interactions features are obtained in the training data set. Results showed LightGBM and XGBoost were the most efficient diagnostic methods because they had the highest AUCs (0.831) and F1-scores (0.761). However, this study only had 92 participants, so it can still be improved by expanding sample size to improve the robustness of predictive models. OA is a complex disease with various clinical manifestations and imaging features in different populations and stages of the disease. Therefore, it is necessary to combine multidimensional information, including a variety of biomarkers, to establish an early diagnosis model of the disease. In addition, the internal and external validation of the model needs to be further explored to enhance the credibility of the model.

### ML application in the multi-omics study

Omics include but are not limited to, genomics, epigenomics, transcriptomics, proteomics, and metabolomics. It has been increasingly used for the identification of novel biomarkers, pathways,

and mechanisms for OA,<sup>49</sup> and significant progress has been made by integrating multiple algorithm techniques and multi-omics data sets.<sup>50</sup> Swan *et al.*<sup>13</sup> used high throughput proteomics to identify potential biomarkers in OA articular cartilage and found that interleukin 1 $\beta$  (IL-1 $\beta$ ) played a significant role in the diagnosis of OA. As a rule-based ML technique, bioinformatics-oriented hierarchical evolutionary learning (BioHEL) proved to be excellent in classifying proteomics results, which indicated that some proteins, including matrix metalloproteinase-3 (MMP-3), interleukin 8 (IL-8), hyaluronan and proteoglycan link protein 1 (HPLN1), matrix gla protein (MGP), and apolipoprotein E (APOE), were associated with the occurrence of OA and others had little effect on OA. Another study developed an ML heuristic based on BioHEL and rule-guided iterative feature elimination (RGIFE), whose diagnosis accuracy for OA was better in proteomics and transcriptomics data sets compared with other ML methods.<sup>51</sup>

The application of ML has significantly improved the efficiency of research in the context of the Big Data era. A primary challenge faced by multi-omics studies is the under-representation of the lower abundance of potentially valuable proteins. Therefore, ML sensitivity plays an important role in resolving this issue.<sup>13</sup> Further analyses should use larger sample sizes and more replicate populations to investigate the individual proteins in the networks and to provide patients with individualized treatment guidance. Moreover, building more publicly available data sets, developing more transparent and reproducible computational analyses, and validating biomarkers across populations and time points are of crucial importance for early diagnostic OA biomarkers derived from multi-omics studies.

### ML application in other studies

In addition to the above methods, many other studies have applied ML approaches for the early diagnosis of OA, such as gait analysis, vibroarthrography, and cartilage texture analysis.

Kotti *et al.*<sup>52</sup> built a computer system that used body kinetics as input and produced an output of an estimation of the KOA presence that filled the interpretability gap between engineering and medical approaches. They used RF regressors to map the parameters by rule induction for KOA grading that led to an accuracy of quintuple

cross-validation of  $72.60\% \pm 4.20\%$ . Vibroarthrography can evaluate the cartilage damage of KOA during extension and flexion movements that is cheap and radiation-free.<sup>53</sup> Using ML with a linear SVM, a satisfactory classification performance could be achieved that produced a specificity of nearly 0.80 and a sensitivity of 0.75.<sup>54</sup> Lim *et al.*<sup>55</sup> employed a DNN method for OA detection using 5749 patients' medical utilization data and their health behavior information. They used PCA with quantile transformer scaling to generate features from background medical information of the participants. This study yielded an AUC of 0.77, and the proposed method minimized the effort for features generation.

One study included 407 middle-aged women with a mean body mass index (BMI) of 27 kg/m<sup>2</sup> and without symptomatic KOA, and built an analytics pipeline with the RGIFE method to predict the early incidence of KOA using clinical variables, food and pain questionnaires, and biochemical and imaging information.<sup>56</sup> Five highly predictive models were proposed and could possibly be adopted for early diagnosis of the disease. All models achieved the outstanding performance of an AUC > 0.70 when only a few variables were used. The study demonstrated the potential of applying the ML approach to generate predictive models integrating all types of information for KOA diagnosis. The method proposed in this study is universal and can be extended to other diseases. The influence of BMI on the development of OA is recognized. Therefore, subsequent researches can comprehensively consider various information, including BMI to increase the applicability of diagnosis model.

### Perspectives of ML in OA

In this review, we summarized various early predictive models based on ML, and different perspectives that can be applied in patients with OA, including radiographs, MRI, biochemical biomarkers, and multi-omics (Table 1 and Figure 1). The commonest method for radiographic definition is the KL grading system and atlas, which has been used for more than 60 years and contains common limitations of reader-dependent reliability.<sup>21</sup> The widely used American College of Rheumatology (ACR) classification criteria for clinical OA diagnosis was first proposed in 1995.<sup>57</sup> Comprehensive predictive models combining multidimensional information are needed to early diagnose OA and further understand the etiology

**Table 1.** Extracted data for OA early diagnosis using machine-learning techniques.

| Author                            | Year | Information source | Data set   | Population characteristics   | Inputs data  | Feature/ biomarkers   | Model                                 | Joint | Performance  |
|-----------------------------------|------|--------------------|--|--|--|---|---------------------------------------|-------|--|
| Yanping Xue <sup>6</sup>          | 2017 | X-ray              | 420 hip X-ray images between March 2016 and April 2016 from Beijing Chaoyang Hospital, China       | NA   | Anteroposterior view hip X-ray images  | The positions of osteophytes and altered shape of the joint ends                            | CNN                                   | Hip   | A sensitivity of 95.00%, a specificity of 90.70% and an accuracy of 92.80% compared to the chief physicians for hip OA diagnosis   |
| Aleksei Tiulpin <sup>25</sup>     | 2018 | X-ray              | Training set from the MOST; validation set from the OAI  | Both datasets include data from men and women aged between 50–79 years old and 45–79, respectively | Plain radiographs annotated the knee joint areas   | Class-discriminating “attention maps”   | DNN                                   | Knee  | A quadratic Kappa coefficient of 0.83 and average multiclass accuracy of 66.71% compared to the annotations given by a committee of clinical experts; a radiological OA diagnosis AUC 0.93   |
| Bin Liu <sup>28</sup>             | 2020 | X-ray              | Images from a hospital in Shanghai   | NA   | Normalized knee X-ray images processed and converted into JPG format                               | Clinically relevant features  | Faster R-CNN                          | Knee  | A mean average precision nearly 0.82; a sensitivity above 78.00%; a specificity above 94%; 0.33s to test each image for automatic diagnosis of KOA   |
| Aleksei Tiulpin <sup>21</sup>     | 2020 | X-ray              | Training set: 4796 participants from the OAI dataset; testing set: 3021 participants from the MOST | 45–79 and 50–79 years old for OAI and MOST, respectively   | Bilateral posterior–anterior fixed-flexion knee radiographs performed knee joint area localization | FL, TL, FM and TM compartments as well as the JSN grades in lateral and medial compartments | DNN                                   | Knee  | Cohen’s kappa coefficients of 0.82 for KL-grade and 0.79, 0.84, 0.94, 0.83, 0.84 and 0.90 for femoral osteophytes, tibial osteophytes and joint space narrowing for lateral and medial compartments, respectively; AUC 0.98; average precision of 0.98 for detecting the presence of radiographic OA |
| Abdelbasset Brahim <sup>24</sup>  | 2019 | X-ray              | 1024 knees from the OAI dataset  | 45–79 years old  | Intensity normalization knee X-ray images  | Tibial trabecular bone ROIs   | Naive Bayes and RF classifiers        | Knee  | An accuracy of 82.98%, a sensitivity of 87.15% and a specificity of 80.65% for KOA detection   |
| Neslihan Bayramoglu <sup>26</sup> | 2021 | X-ray              | 5507 knees from the MOST dataset   | 50–79 years old  | Hand-crafted features based on Local Binary Patterns   | Patellar region ROIs  | Textural ROI classification using CNN | Knee  | AUC 0.90, AP 0.71 for detecting patellofemoral OA  |
| Neslihan Bayramoglu <sup>27</sup> | 2021 | X-ray              | 18436 knees from the MOST dataset  | 50–79 years old  | Normalized knee lateral view radiographs   | Patellar ROI  | Textural ROI classification using CNN | Knee  | AUC 0.96, AP 0.86 for detecting patellofemoral OA  |

(Continued)

Table 1. (Continued)

| Author                            | Year | Information source             | Data set  | Population characteristics  | Inputs data  | Feature/ biomarkers   | Model                               | Joint | Performance  |
|-----------------------------------|------|--------------------------------|---|---|--|---|-------------------------------------|-------|--|
| Rabbia Mahum <sup>29</sup>        | 2021 | X-ray                          | 500 images from the Mendeley Dataset V1   | NA  | Knee X-ray images processed and converted into TIFF format   | Tibiofemoral joint ROI  | CNN, SVM, RF, KNN                   | Knee  | An accuracy on cross-validation of 97.14% and accuracy on five-fold validation of 98%  |
| Mahesh B. Nagarajan <sup>14</sup> | 2013 | PCI-CT                         | Two osteochondral cylindrical samples of diameter 7 mm and height 12 mm were extracted from the lateral facet of the patella of 2 patients, one healthy and one osteoarthritic, within 48 hours post-mortem. These samples were subject to PCI-CT and annotated 404 ROIs totally. | NA  | Axial images of dimensions 1124x1124 and in-plane resolution of 8 µm x 8 µm were reconstructed by using a direct Hamming filtered-back projection algorithm. | 404 ROIs of dimensions 51x51 pixels, capturing chondrocyte patterns   | KNN                                 | Knee  | GLCM and geometric features derived from SIM were extracted from 404 ROIs. The highest AUC value was observed with the SIM feature vector (0.98 ± 0.01). Similar performance was achieved by GLCM Correlation features |
| Anas Z. Abidin <sup>31</sup>      | 2018 | PCI-CT                         | Five ex-vivo specimens of the cartilage extracted from the retropatellar joint  | Age < 40 years old  | Cross-sectional view of single slice of articular cartilage as obtained using PCI-CT   | 842 square ROIs capturing chondrocyte patterns in the radial zone of the cartilage matrix                         | Two CNNs: CaffeNet and Inception-v3 | Knee  | CaffeNet (AUCs > 0.91); Inception-v3 (AUC > 0.95) of classifying healthy or KOA  |
| Gary H Chang <sup>35</sup>        | 2021 | 3D DESS sagittal MRI sequences | 9,434 knees from the OAI  | Age from 45-79; minorities made up 20.9% of the population of whom 87.3% were African Americans, 4.5% Asians, and 8.2% fell into the "Other" category | Sagittal MRI slices subsequently underwent histogram equalization  | ROIs comprised of all the subchondral compartments containing the manually annotated femoral and tibia cartilages | DNN                                 | Knee  | The odds ratios between lowest and highest quartiles corresponding to SBL changes for future KR were 5.68 and 7.19   |

(Continued)



**Table 1.** (Continued)

| Author                          | Year | Information source  | Data set   | Population characteristics   | Inputs data   | Feature/biomarkers   | Model  | Joint | Performance  |
|---------------------------------|------|---|--|--|---|--|--|-------|--|
| Joselene Marques <sup>37</sup>  | 2012 | Low-field Turbo 3D T1 MRI sequence  | 313 knee scans from the CCBR   | Age $56 \pm 16$ , BMI $26 \pm 4\%$ and 47% female  | Sagittal view of low-field knee MRI scan  | ROI extracted from the subchondral region of trabecular bone, the tibial cartilage and the femoral cartilage; a single value for each ROI, which is defined as the bone structure marker | LDA, QDA, NN, KNN, wNN, wkNN                               | Knee  | LDA: AUC 0.82, KNN: AUC 0.81 for analyzing trabecular bone structures and quantifying OA risks   |
| Joselene Marques <sup>34</sup>  | 2013 | Turbo 3D T1 MRI sequence  | 268 knees MRI from the CCBR  | Age $56 \pm 16$ , BMI $26 \pm 4\%$ and 47% female  | Sagittal slices of knee MRI   | Trabecular bone defined as the ROI   | A ML approach for feature selection and feature extraction | Knee  | The ability to separate ROA ( $KL > 1$ ) and healthy knees of the bone structure marker reached a generalization AUC of 0.92 prior to normalization for the confounding factors and AUC of 0.86 after correction |
| Beth G. Ashinsky <sup>12</sup>  | 2017 | Multi-slice T2-weighted MRI   | 68 participants from the OAI   | Mean age 56 and mean BMI 25 in non-progression group; mean age 59 and mean BMI 29 in symptomatic progression of OA group | Knee MRI registered using a nonlinear image registration algorithm                    | A generic set of numerical image content descriptors, including textures   | WND-CHRM   | Knee  | WND-CHRM classified the isolated T2 maps for the progression to symptomatic OA with 75.00% accuracy, 74.00% sensitivity and 76.00% specificity   |
| Jukka Hirvasniemi <sup>38</sup> | 2021 | MRI scanning protocol consisted of Dual echo FSE PD, FSE T2-weighted FS, SGE FS, FIESTA sequences | Right knees of 665 females from the population-based Rotterdam Study | Age $54.6 \pm 3.7$ , BMI $26.8 \pm 4.6$  | Semi-automatically extracted MRI-based radiomic features from tibial subchondral bone | Radiomic features related to the shape and texture calculated from six VOIs in the proximal tibia, they are SBM, SBL, MidM, MidL TBM and TBL   | Elastic Net  | Knee  | Covariate model: AUC 0.68, image features model: AUC 0.80, Model combined image features from all VOIs and covariates: AUC 0.80  |

(Continued)

Table 1. (Continued)

| Author                         | Year | Information source   | Data set  | Population characteristics   | Inputs data  | Feature/ biomarkers   | Model  | Joint | Performance   |
|--------------------------------|------|--|---|--|--|---|--|-------|---|
| Jia Li <sup>41</sup>           | 2022 | Sagittal, intermediate-weighted, fat-suppressed, two-dimensional fast spin-echo MRI sequence   | 690 participants with KOA risk from the Pivotal OAI MRI Analyses incident osteoarthritis cohort   | Mean age 60 and female 68% in development cohort; mean age 61 and female 83.16% in test cohort | IPFP masks   | IPFP texture features   | 3D texture analysis  | Knee  | In both cohorts, IPFP texture scores [AUC $\geq 0.75$ ] showed greater discrimination than clinical scores [AUC $\leq 0.6$ ] at baseline, P-1 (1 year before P0), and P0 (the incident time point of KOA), with significant differences in pairwise comparisons ( $P \leq 0.002$ ). |
| Valentina Pedita <sup>44</sup> | 2018 | Demographic, clinical information, gait kinematics and kinetics, cartilage compositional T1q, T2 and R2-R1p(1/T2-1/T1p) acquired at 3T | 178 participants in a multidimensional space integrating  | Age > 35 years old   | The observed compositional MRI, MRI morphological grading, and biomechanics variables                      | The combined compositional metric $R_2-R_1p$ ( $1/T_2-1/T_1p$ )   | A multidimensional platform used TDA and Kolmogorov Smirnov test | Knee  | Considering T2 LF alone: AUC 91.06%; using a threshold of 23.97 msec (sensitivity 86.76% and specificity 86.84%)  |
| Usman Ahmed <sup>45</sup>      | 2016 | Protein oxidation, nitration and glycation biomarkers  | 225 participants from Orthopedic Clinics, University Hospital Coventry & Warwickshire (UHCW), Coventry, Department of Rheumatology, Ipswich Hospital NHS Trust, Department of Rheumatology, Royal Devon and Exeter NHS Foundation Trust, Exeter and Rapid Access Rheumatology Clinic, City Hospital, Birmingham, UK | NA   | Oxidized, nitrated and glycated protein and oxidation, nitration adduct in plasma/serum and synovial fluid | Protein oxidation, nitration and glycation adducts in protein from synovial fluid and plasma, and related oxidized, nitrated and glycated amino acids | RF, multiclass LR, multi-class sparse LR, SVM                    | Knee  | For eOA, eRA and non-RA, AUC values were in the range 0.68–0.98, 0.77–0.93 and 0.62–0.91 for training set and test set cross-validations and test set validation, respectively. Related sensitivities/specificities for eOA, eRA and non-RA were 0.92/0.90, 0.80/0.78 and 0.70/0.65 |

(Continued)

Table 1. (Continued)

| Author                        | Year | Information source   | Data set  | Population characteristics  | Inputs data  | Feature/ biomarkers  | Model                            | Joint        | Performance   |
|-------------------------------|------|--|---|---|--|--|----------------------------------|--------------|---|
| Jonas Bianchi <sup>47</sup>   | 2020 | Biomolecular features (proteins of serum and saliva), imaging features (trabecular bone radiomics and morphometry) and clinical features | 92 patients, 46 TMJ OA patients and 46 controls from Michigan   | Age 38.50 ± 13.26 and female 84.78% in control group; age 38.00 ± 12.34 and female 84.78% in TMJ OA group | 52 clinical, biological and high-resolution CBCT (radiomics) markers                     | TGF-β1, PA1, VE-cadherin and angiogenin in saliva, VE-cadherin in serum  | LR, RF, LightGBM, XGBoost        | TMJ          | XGBoost + LightGBM: an accuracy of 0.82, AUC 0.87   |
| Celia Le <sup>48</sup>        | 2021 | A total of 3828 features and interactions features   | 92 subjects, 46 TMJ OA and 46 healthy controls  | NA  | 52 clinical, biological and jaw condyle radiomic markers                                 | 2 demographic values, 13 protein level values from serum, 12 protein level values from saliva, 5 clinical features evaluating the pain and 20 imaging features representing the gray-level values of the ROI | RF, XGBoost, LightGBM, RR and LR | TMJ          | LightGBM and XGBoost are the most efficient diagnostic methods with AUCs (0.831) and F1-scores (0.761); the inclusion of 32 additional markers from the mandibular fossa of the joint improved the AUC prediction performance from 0.83 to 0.88 |
| Anna L Swan <sup>13</sup>     | 2013 | Proteomics and transcriptomics   | Male German Shepherd army dogs, over 5 years of age   | NA  | EmpAI scores and ProteinProphet scores   | ECM proteins, including MMP-3, IL-8 and MGP  | BioHEL                           | Elbow        | BioHEL correctly classified eighteen out of twenty-three samples, giving a classification accuracy of 78.3% for the dataset   |
| Anna L Swan <sup>51</sup>     | 2015 | Proteomics and transcriptomics   | Uniprot database  | NA  | Proteomics database  | MMP-3, IL-8, thrombospondin-1, hyaluronan and proteoglycan link protein 1, clusterin and fetuin-A  | BioHEL                           | Knee and hip | Classification accuracy of 73.90%, with sensitivity and specificity of 96.00% and 99.00% respectively   |
| Margarita Kotti <sup>52</sup> | 2017 | Vertical, anterior-posterior, and medio-lateral ground reaction forces, such as mean value, push-off time, and slope                     | 47 subjects with KOA and 47 healthy subjects from hospital clinics and GP surgeries, local community in London, United States | Age 54.40 ± 13.30 and female 53.19% in control group; age 58.10 ± 12.70 and female 53.19% in KOA group    | Parameters of the vertical, anterior-posterior, and medio-lateral ground reaction forces | Ground reaction force  | RF                               | Knee         | An accuracy of 72.61% ± 4.24% for automatically estimating the degree of KOA  |

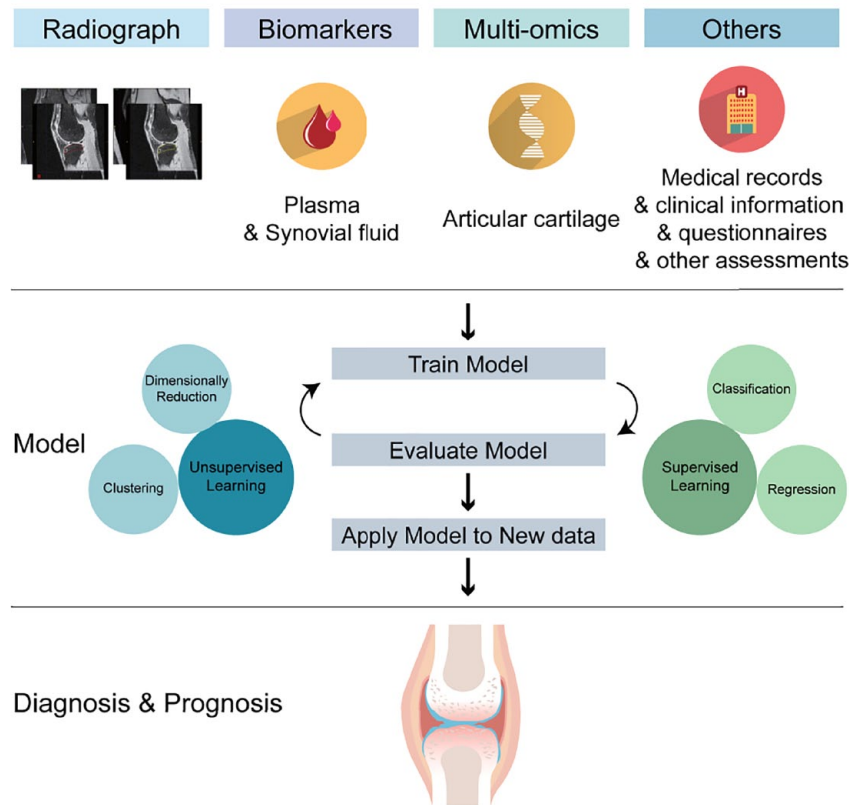
(Continued)

Table 1. (Continued)

| Author                     | Year | Information source  | Data set   | Population characteristics   | Inputs data   | Feature/ biomarkers  | Model                                | Joint              | Performance  |
|----------------------------|------|---|--|--|---|--|--------------------------------------|--------------------|--|
| Nima Bafui <sup>54</sup>   | 2018 | Vibroarthro-graphic signals   | 59 asymptomatic knees and 40 knees with OA from Munich, Germany            | Age $26.90 \pm 4.18$ and female $46.67\%$ in healthy group; age $55.20 \pm 13.5$ and female $61.54\%$ in patient group | Semi-automatic segmentation of the structure-borne noise acoustic signals, frequency features were generated for the extension as well as the flexion phase | Normalized nature of the frequency features, such as attenuation by fat tissue and variance in acoustic coupling | A linear SVM                         | Knee               | A classification specificity of approximately 0.80; a sensitivity of 0.75  |
| Jihye Lim <sup>55</sup>    | 2019 | Medical utilization and health behavior information   | 5749 participants from 2015–2016 KNHANCES                                  | Age > 50 years old   | Medical utilization and health behavior information   | NA   | DNN, PCA                             | Hand, hip and knee | An accuracy of 71.97% with sensitivity and specificity of 66.67% and 73.35% respectively, and AUC 76.80% for automatic OA prediction |
| N. Lazzarini <sup>56</sup> | 2017 | Clinical variables, food and pain questionnaires, biochemical markers and imaging-based information | 365 middle-aged overweight women without clinical KOA from the PROOF study | Age $55.70 \pm 3.20$ , BMI $32.30 \pm 4.30$ , female 100% and western ethnicity 96%                                    | Clinical variables, food and pain questionnaires, biochemical markers and imaging-based information   | Early signs (e.g. pain while jumping) as well as risk factors (e.g. BMI or waist circumference)                  | An analytics pipeline based on RGIFE | Knee               | KL incidence model: AUC 0.823; the Knee pain model: AUC 0.755  |

3D, three-dimensional; AP, average precision; AUC, area under curve; BioHEL, bioinformatics-oriented Hierarchical Evolutionary Learning; BMI, body mass index; CCB, Center for Clinical and Basic Research; CNN, convolutional neural network; DESS, dual-echo in steady state; DNN, deep neural network; ECM, extracellular matrix; eOA, early-stage OA; eRA, early rheumatoid arthritis; FIESTA, fast imaging employing steady-state acquisition; FL, femoral lateral; FM, femoral medial; FS, fat suppression; FSE, fast spin echo; GLCM, gray-level co-occurrence matrices; GP, General Practitioner; IL, interleukin; IPFP, infrapatellar fat pad; JSN, joint space narrowing; KL, Kellgren-Lawrence; KNN, k-nearest neighbor; KOA, knee OA; KR, knee replacement; LDA, linear discriminant analysis; LR, logistic regression; MGP, matrix Gla protein; MidL, lateral mid-part; MidM, medial mid-part; ML, machine learning; MMP, matrix metalloproteinase; MOST, Multicenter Osteoarthritis Study; MRI, Magnetic resonance imaging; NA, not available; NN, nearest-neighbor; OA, osteoarthritis; OAI, Osteoarthritis Initiative; PA, prealbumin; PCA, principal components analysis; PCI-CT, Phase-contrast X-ray computed tomography; PD, proton density; QDA, quadratic discriminant analysis; R-CNN, Region-CNN; RF, random forest; RGIFE, Ranked Guided Iterative Feature Elimination; ROA, radiographic osteoarthritis; ROI, region of interest; RRI, regions of interest; RR, ridge regression; SBL, lateral subchondral bone; SBL, medial subchondral bone; SGE, spoiled gradient echo; SIM, Scaling Index Method; SVM, support vector machine; TBL, lateral trabecular bone; TBM, medial trabecular bone; TDA, topological data analysis; ToF- $\beta$ , transforming growth factor- $\beta$ ; TL, tibial-lateral; TM, tibial-medial; TMJ, temporomandibular joint; VOIs, volumes of interests; wkNN, weighted k-nearest-neighbor; WND-CHRM, weighted neighbor distance using compound hierarchy of algorithms representing morphology; wNN, weighted nearest-neighbor.





**Figure 1.** Machine-learning model working flow chart.

and progression of joint degeneration. Effective prediction models could pave the way for developing efficient and personalized therapies for OA patients.

Recent studies have regarded MRI as an important imaging tool in clinical and research applications, especially high-resolution MRI. However, early diagnostic criteria for OA using MRI have yet to be established. It is accepted that high-resolution MRI reflects cartilage changes earlier than radiographs, however, the course of OA is typically diverse and in an individual OA patient is largely unpredictable.<sup>58</sup> In the future, supervised learning methods could significantly improve the predictive ability of high-resolution MRI. Moreover, although aging has always been considered as an important etiology for OA, the mechanisms are still not fully understood. Current ML methods cannot distinguish the osteoarthritic structural changes from normal aging manifestations.<sup>59</sup>

Biomarkers are essential for the early diagnosis of OA. However, the development of biochemical biomarkers that can accurately and early diagnose

OA is far from trivial, since changes in the disease are subtle. Recent advances in multi-omics technology, including genomics, transcriptomics, proteomics, and metabolomics, have promoted the development of new methods to identify disease biomarkers. Bioinformatics tools have detected several OA-related proteins and have shown the potential to investigate the individual early diagnosis of OA.

Due to complexity and heterogeneity of OA, it is unlikely that a single marker will provide an accurate early diagnosis of the disease. The challenges on KOA using ML remain because of high heterogeneity among different OA subtypes or localizations, which calls for diversifying the studies. Moreover, how to best subset OA into phenotypes and endotypes and whether certain subsets are of any clinical value is an important issue that requires both internal and external validation both in discovery cohort and in external populations. In this regard, the availability of high-quality and easy-to-obtain patient data, ideally longitudinal and from different populations, is key for OA phenotyping and/or endotyping

research.<sup>60–62</sup> Therefore, additional research combining measurements from MRI, biomarkers, clinical manifestations, and multi-omics are needed to build a more accurate and generalizable model for early diagnosis and prediction of OA. Along with the development of related techniques, a growing number of public and private data sets will provide extremely large amounts of statistics for OA research. Thus, patient-specific prognosis estimation and personalized prevention and treatment of OA can be achieved. Further studies are needed to verify whether the current ML algorithms are also approximate to other data sets and to identify their generalizability.

One major challenge that needs to be highlighted is consistent quality and standardization to extract OA features between studies in respect to acquisition techniques. Although the data sources appeared to be larger and larger, the lack of high-quality labeling may reduce their effectiveness in the ML model training. Until now, the principle of the feature map or the target location of the most important features extracted by ML is unknown. Therefore, standardized protocols and acceptance testing should be established to ensure ML studies can meet certain criteria before routine clinical use. Moreover, most studies applying ML to OA tested or trained in the same data set, which often have high comparability but limited generalizability. Besides, most of the cohorts currently available predate the interest of ML approaches in the field. This situation reinforces the need for purpose-built cohorts for ML analyses and future studies should aim to test ML models on broad data to demonstrate the robustness and generalizability.<sup>63–65</sup>

Although ML could be an important tool to aid the work of researchers, and eventually clinicians, in the early identification of OA, acceptable quality standards for real-world applications have not been fully met. It is hoped that the heightened pace of external validation added to the current pace of model development and internal validation will accelerate the translation to the accurate, rapid, and cost-effective detection of OA in the near future.<sup>66</sup>

The main limitation of ML in OA is that the performance of the algorithm may be biased. Although the method could perform well on the entire data set, it is still unsatisfactory when the test set is stratified according to the stages of OA development. Therefore, future studies should focus on

improving the performance when OA is not severe. In addition, most of the current studies were associated with imaging and lack external validation and real-world clinical applications, and few articles merged the use of clinical and biological data. Therefore, it is necessary to use larger scale and multifaceted data to make the diagnosis model more credible and reproducible. Most of the applied ML algorithms are based on Supervised Learning, which may limit the ability to identify new phenotypes. Moreover, developing interpretable ML models that can determine which features contribute to a specific model decision is also a challenge.<sup>65</sup> Despite these limitations, we believe that investigational methodologies are emerging in clinical trials, and could provide better quantitative information about the joints of patients who already have OA in a systematic way.<sup>21</sup>

In conclusion, we have demonstrated current predictive models based on ML for the early diagnosis of OA. An accurate prediction model that can detect early changes in OA before the emergence of clinical features is expected to alter the current state of OA diagnosis. Although there are neither cluster of imaging nor biological markers allowing routine early diagnosis of OA in clinical practice, the diagnostic performance of the fusion model combining multidimensional information could enable patient-specific prognosis estimation and personalized therapy for OA in the future. In addition, to standardize and improve the quality of such studies, it is necessary to develop ML checklists, consensus, and training for the scientific community of OA.

## Declarations

*Ethics approval and consent to participate*  
Not applicable.

*Consent for publication*  
Not applicable.

## Author contributions

**Anran Xuan:** Writing – original draft.

**Haowei Chen:** Writing – original draft.

**Tianyu Chen:** Writing – review & editing.

**Jia Li:** Writing – review & editing.

**Shilong Lu:** Writing – review & editing.

**Tianxiang Fan:** Writing – review & editing.

**Dong Zeng:** Writing – review & editing.

**Zhibo Wen:** Writing – review & editing.

**Jianhua Ma:** Writing – review & editing.

**David Hunter:** Writing – review & editing.

**Changhai Ding:** Writing – review & editing.

**Zhaohua Zhu:** Writing – review & editing.

### Acknowledgements

None.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The present study was supported by the National Natural Science Foundation of China (grant no. 32000925), Guangzhou Science and Technology Program (grant no. 202002030481) and Guangdong Basic and Applied Basic Research Foundation (grant no. 2019A1515111169). The funding body had no role in the performance of this review or writing the manuscript.

### Competing interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Availability of data and material

Not applicable.

### ORCID iDs

Anran Xuan  <https://orcid.org/0000-0003-3631-8961>

David Hunter  <https://orcid.org/0000-0002-4065-7395>

Changhai Ding  <https://orcid.org/0000-0002-9479-730X>

### References

1. Bitton R. The economic burden of osteoarthritis. *Am J Manag Care* 2009; 15: 230–235.
2. Martel-Pelletier J, Boileau C, Pelletier JP, *et al.* Cartilage in normal and osteoarthritis conditions. *Best Pract Res Clin Rheumatol* 2008; 22: 351–384.
3. Katz JN, Arant KR and Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis: a review. *JAMA* 2021; 325: 568–578.
4. Long H, Liu Q, Yin H, *et al.* Prevalence trends of site-specific osteoarthritis from 1990 to 2019: findings from the global burden of disease study 2019. *Arthritis Rheumatol* 2022; 74: 1172–1183.
5. Abramoff B and Caldera FE. Osteoarthritis: pathology, diagnosis, and treatment options. *Med Clin North Am* 2020; 104: 293–311.
6. Xue Y, Zhang R, Deng Y, *et al.* A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PLoS ONE* 2017; 12: e0178992.
7. van Oudenaarde K, Jobke B, Oostveen AC, *et al.* Predictive value of MRI features for development of radiographic osteoarthritis in a cohort of participants with pre-radiographic knee osteoarthritis-the CHECK study. *Rheumatology* 2017; 56: 113–120.
8. Jamshidi A, Pelletier JP and Martel-Pelletier J. Machine-learning-based patient-specific prediction models for knee osteoarthritis. *Nature Reviews Rheumatology* 2019; 15: 49–60.
9. Richette P and Latourte A. [Osteoarthritis: value of imaging and biomarkers]. *Rev Prat* 2019; 69: 507–509.
10. Henrotin Y. Osteoarthritis in year 2021: biochemical markers. *Osteoarthritis Cartilage* 2022; 30: 237–248.
11. Nagy EE, Nagy-Finna C, Popoviciu H, *et al.* Soluble biomarkers of osteoporosis and osteoarthritis, from pathway mapping to clinical trials: an update. *Clin Interv Aging* 2020; 15: 501–518.
12. Ashinsky BG, Bouhrara M, Coletta CE, *et al.* Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative. *J Orthop Res* 2017; 35: 2243–2250.
13. Swan AL, Hillier KL, Smith JR, *et al.* Analysis of mass spectrometry data from the secretome of an explant model of articular cartilage exposed to pro-inflammatory and anti-inflammatory stimuli using machine learning. *BMC Musculoskelet Disord* 2013; 14: 349.
14. Nagarajan MB, Coan P, Huber MB, *et al.* Characterizing healthy and osteoarthritic knee cartilage on phase contrast CT with geometric texture features. *Proc SPIE Int Soc Opt Eng* 2013; 8672: 86721J.
15. Ngiam KY and Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* 2019; 20: e262–e73.

16. Rauschert S, Raubenheimer K, Melton PE, *et al.* Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification. *Clin Epigenetics* 2020; 12: 51.
17. Delahanty RJ, Alvarez J, Flynn LM, *et al.* Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. *Ann Emerg Med* 2019; 73: 334–344.
18. Levine AB, Schlosser C, Grewal J, *et al.* Rise of the machines: advances in deep learning for cancer diagnosis. *Trends Cancer* 2019; 5: 157–169.
19. Walsh JA, Rozycki M, Yi E, *et al.* Application of machine learning in the diagnosis of axial spondyloarthritis. *Curr Opin Rheumatol* 2019; 31: 362–367.
20. Glyn-Jones S, Palmer AJ, Agricola R, *et al.* Osteoarthritis. *Lancet* 2015; 386: 376–387.
21. Tiulpin A and Saarakkala S. Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks. *Diagnostics* 2020; 10: 932.
22. Pedoia V and Majumdar S. Translation of morphological and functional musculoskeletal imaging. *J Orthop Res* 2019; 37: 23–34.
23. Antony AJ. *Automatic quantification of radiographic knee osteoarthritis severity and associated diagnostic features using deep convolutional neural networks*. PhD Thesis, Dublin City University, Dublin, Ireland 2018.
24. Brahim A, Jennane R, Riad R, *et al.* A decision support tool for early detection of knee osteoarthritis using X-ray imaging and machine learning: data from the osteoarthritis initiative. *Comput Med Imaging Graph* 2019; 73: 11–18.
25. Tiulpin A, Thevenot J, Rahtu E, *et al.* Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. *Sci Rep* 2018; 8: 1727.
26. Bayramoglu N, Nieminen MT and Saarakkala S. Machine learning based texture analysis of patella from X-rays for detecting patellofemoral osteoarthritis. *Int J Med Inform* 2022; 157: 104627.
27. Bayramoglu N, Nieminen MT and Saarakkala S. Automated detection of patellofemoral osteoarthritis from knee lateral view radiographs using deep learning: data from the Multicenter Osteoarthritis Study (MOST). *Osteoarthritis Cartilage* 2021; 29: 1432–1447.
28. Liu B, Luo J and Huang H. Toward automatic quantification of knee osteoarthritis severity using improved Faster R-CNN. *Int J Comput Assist Radiol Surg* 2020; 15: 457–466.
29. Mahum R, Rehman SU, Meraj T, *et al.* A novel hybrid approach based on deep CNN features to detect knee osteoarthritis. *Sensors* 2021; 21: 6189.
30. Khalid H, Hussain M, Al Ghamdi MA, *et al.* A comparative systematic literature review on knee bone reports from MRI, X-rays and CT scans using deep learning and machine learning methodologies. *Diagnostics* 2020; 10: 518.
31. Abidin AZ, Deng B, DSouza AM, *et al.* Deep transfer learning for characterizing chondrocyte patterns in phase contrast X-Ray computed tomography images of the human patellar cartilage. *Comput Biol Med* 2018; 95: 24–33.
32. Juras V, Chang G and Regatte RR. Current status of functional MRI of osteoarthritis for diagnosis and prognosis. *Curr Opin Rheumatol* 2020; 32: 102–109.
33. Łuczkiwicz P, Daszkiewicz K, Chróścielewski J, *et al.* The influence of articular cartilage thickness reduction on meniscus biomechanics. *PLoS ONE* 2016; 11: e0167733.
34. Marques J, Genant HK, Lillholm M, *et al.* Diagnosis of osteoarthritis and prognosis of tibial cartilage loss by quantification of tibia trabecular bone from MRI. *Magn Reson Med* 2013; 70: 568–575.
35. Chang GH, Park LK, Le NA, *et al.* Subchondral bone length in knee osteoarthritis: a deep learning derived imaging measure and its association with radiographic and clinical outcomes. *Arthritis Rheumatol* 2021; 73: 2240–2248.
36. Kundu S, Ashinsky BG, Bouhrara M, *et al.* Enabling early detection of osteoarthritis from presymptomatic cartilage texture maps via transport-based learning. *Proc Natl Acad Sci U S A* 2020; 117: 24709–24719.
37. Marques J, Granlund R, Lillholm M, *et al.* Automatic analysis of trabecular bone structure from knee MRI. *Comput Biol Med* 2012; 42: 735–742.
38. Hirvasniemi J, Klein S, Bierma-Zeinstra S, *et al.* A machine learning approach to distinguish between knees without and with osteoarthritis using MRI-based radiomic features from tibial bone. *Eur Radiol* 2021; 31: 8513–8521.
39. Hunter DJ, Guermazi A, Lo GH, *et al.* Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). *Osteoarthritis Cartilage* 2011; 19: 990–1002.
40. Lu M, Chen Z, Han W, *et al.* A novel method for assessing signal intensity within infrapatellar fat pad on MR images in patients with knee osteoarthritis. *Osteoarthritis Cartilage* 2016; 24: 1883–1889.
41. Li J, Fu S, Gong Z, *et al.* MRI-based texture analysis of infrapatellar fat pad to predict knee osteoarthritis incidence. *Radiology* 2022; 212009.



42. Jerban S, Chang EY and Du J. Magnetic resonance imaging (MRI) studies of knee joint under mechanical loading: review. *Magn Reson Imaging* 2020; 65: 27–36.
43. Qiu D, Zhang S, Liu Y, *et al.* Super-resolution reconstruction of knee magnetic resonance imaging based on deep learning. *Comput Methods Programs Biomed* 2020; 187: 105059.
44. Pedoia V, Haefeli J, Morioka K, *et al.* MRI and biomechanics multidimensional data analysis reveals R2 -R1rho as an early predictor of cartilage lesion progression in knee osteoarthritis. *J Magn Reson Imaging* 2018; 47: 78–90.
45. Ahmed U, Anwar A, Savage RS, *et al.* Protein oxidation, nitration and glycation biomarkers for early-stage diagnosis of osteoarthritis of the knee and typing and progression of arthritic disease. *Arthritis Res Ther* 2016; 18: 250.
46. Nelson AE, Fang F, Arbeeve L, *et al.* A machine learning approach to knee osteoarthritis phenotyping: data from the FNIH Biomarkers Consortium. *Osteoarthritis Cartilage* 2019; 27: 994–1001.
47. Bianchi J, de Oliveira Ruellas AC, Gonçalves JR, *et al.* Osteoarthritis of the temporomandibular joint can be diagnosed earlier using biomarkers and machine learning. *Sci Rep* 2020; 10: 8012.
48. Le C, Deleat-Besson R, Turkestani NA, *et al.* TMJOAI: an artificial web-based intelligence tool for early diagnosis of the temporomandibular joint osteoarthritis. *Clin Image Based Proced Distrib Collab Learn Artif Intell Combat COVID 19 Secur Priv Preserv Mach Learn* 2021; 12969: 78–87.
49. Ruiz-Romero C and Blanco FJ. Proteomics role in the search for improved diagnosis, prognosis and treatment of osteoarthritis. *Osteoarthritis Cartilage* 2010; 18: 500–509.
50. Ratneswaran A and Kapoor M. Osteoarthritis year in review: genetics, genomics, epigenetics. *Osteoarthritis Cartilage* 2021; 29: 151–160.
51. Swan AL, Stekel DJ, Hodgman C, *et al.* A machine learning heuristic to identify biologically relevant and minimal biomarker panels from omics data. *BMC Genomics* 2015; 16(Suppl. 1): S2.
52. Kotti M, Duffell LD, Faisal AA, *et al.* Detecting knee osteoarthritis and its discriminating parameters using random forests. *Med Eng Phys* 2017; 43: 19–29.
53. Krishnan S, Rangayyan RM, Bell GD, *et al.* Adaptive time-frequency analysis of knee joint vibroarthrographic signals for noninvasive screening of articular cartilage pathology. *IEEE Trans Biomed Eng* 2000; 47: 773–783.
54. Befrui N, Elsner J, Flesser A, *et al.* Vibroarthrography for early detection of knee osteoarthritis using normalized frequency features. *Med Biol Eng Comput* 2018; 56: 1499–1514.
55. Lim J, Kim J and Cheon S. A deep neural network-based method for early detection of osteoarthritis using statistical data. *Int J Environ Res Public Health* 2019; 16: 1281.
56. Lazzarini N, Runhaar J, Bay-Jensen AC, *et al.* A machine learning approach for the identification of new biomarkers for knee osteoarthritis development in overweight and obese women. *Osteoarthritis Cartilage* 2017; 25: 2014–2021.
57. Jang S, Lee K and Ju JH. Recent updates of diagnosis, pathophysiology, and treatment on osteoarthritis of the knee. *Int J Mol Sci* 2021; 22: 2619.
58. Mahmoudian A, Lohmander LS, Mobasheri A, *et al.* Early-stage symptomatic osteoarthritis of the knee – time for action. *Nat Rev Rheumatol* 2021; 17: 621–632.
59. Rahmati M, Nalesso G, Mobasheri A, *et al.* Aging and osteoarthritis: central role of the extracellular matrix. *Ageing Res Rev* 2017; 40: 20–30.
60. Mobasheri A, van Spil WE, Budd E, *et al.* Molecular taxonomy of osteoarthritis for patient stratification, disease management and drug development: biochemical markers associated with emerging clinical phenotypes and molecular endotypes. *Curr Opin Rheumatol* 2019; 31: 80–89.
61. Devez LA, Nelson AE and Loeser RF. Phenotypes of osteoarthritis: current state and future implications. *Clin Exp Rheumatol* 2019; 37(Suppl 120): 64–72.
62. Angelini F, Widera P, Mobasheri A, *et al.* Osteoarthritis endotype discovery via clustering of biochemical marker data. *Ann Rheum Dis* 2022; 81: 666–675.
63. Shen D, Wu G and Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng* 2017; 19: 221–248.
64. Schwendicke F, Chaurasia A, Arsiwala L, *et al.* Deep learning for cephalometric landmark detection: systematic review and meta-analysis. *Clin Oral Investig* 2021; 25: 4299–4309.
65. Binvignat M, Pedoia V, Butte AJ, *et al.* Use of machine learning in osteoarthritis research: a systematic literature review. *RMD Open* 2022; 8: e001998.
66. Nwanosike EM, Conway BR, Merchant HA, *et al.* Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review. *Int J Med Inform* 2022; 159: 104679.