



## Research article

# A joint multi-modal learning method for early-stage knee osteoarthritis disease classification

Liangliang Liu, Jing Chang, Pei Zhang, Qingzhi Ma, Hui Zhang, Tong Sun, Hongbo Qiao\*

College of Information and Management Science, Henan Agricultural University, Zhengzhou, Henan, 450046, PR China



## ARTICLE INFO

**Keywords:**

Classification  
Intercorrelation  
Multi-modal learning  
Osteoarthritis  
Regularization term

## ABSTRACT

Osteoarthritis (OA) is a progressive and chronic disease. Identifying the early stages of OA disease is important for the treatment and care of patients. However, most state-of-the-art methods only use single-modal data to predict disease status, so that these methods usually ignore complementary information in multi-modal data. In this study, we develop an integrated multi-modal learning method (MMLM) that uses an interpretable strategy to select and fuse clinical, imaging, and demographic features to classify the grade of early-stage knee OA disease. MMLM applies XGboost and ResNet50 to extract two heterogeneous features from the clinical data and imaging data, respectively. And then we integrate these extracted features with demographic data. To avoid the negative effects of redundant features in a direct integration of multiple features, we propose a L1-norm-based optimization method (MMLM) to regularize the inter-correlations among the multiple features. MMLM was assessed using the Osteoarthritis Initiative (OAI) data set with machine learning classifiers. Extensive experiments demonstrate that MMLM improves the performance of the classifiers. Furthermore, a visual analysis of the important features in the multimodal data verified the relations among the modalities when classifying the grade of knee OA disease.

## 1. Introduction

Osteoarthritis (OA), a progressive and chronic disease, is among the most frequently occurring and disabling chronic diseases worldwide [1–3]. The World Health Organization reported that OA disease affects up to 9.6% of men and 18.0% of women among middle-aged and elderly people worldwide [4–6]. OA disease was the fourth leading cause of disability globally in 2020 [7,8]. OA disease frequently occurs at the joints of the body, especially the knees, hips, and hands, of which the knee is most often affected [9,10]. Experienced clinicians usually base a diagnosis of OA disease on an X-ray image of the joints. They use the Kellgren and Lawrence (KL) radiographic grade to classify the severity of the disease for individual joints [11,12]. The KL grade ranges from 0 representing normal to 4 being severe OA disease. In clinical diagnosis, grades 1 and 2 are considered as early-stage OA disease, and grades 3 and 4 are considered as the severe stage [13]. Early classification of grades 0, 1, and 2 not only can alleviate the pain of patients but also correct the disease and prevent further deterioration of the joint. However, the early symptoms of OA disease are not obvious, and an evaluation depends on the subjective experience of clinicians. Therefore, to improve clinical diagnoses and clinical drug trials, it would

\* Corresponding author.

E-mail address: [qiaohb@henau.edu.cn](mailto:qiaohb@henau.edu.cn) (H. Qiao).

<https://doi.org/10.1016/j.heliyon.2023.e15461>

Received 11 November 2022; Received in revised form 5 April 2023; Accepted 10 April 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Table 1**  
Details of entire cohort.

Type	Detail
Male or female	The recruitment goal is for approximately equal numbers of men and women overall and in each subcohort.
Ages 45-79	Enrollment goals will be specified for each decade of age within each gender and subcohort.
All ethnic groups	The recruitment goal is for approximately 23% of the cohort from ethnic minority groups.
Knee Radiographs	Time points: baseline, 12 m, 24 m, 36 m, 48 m, 72 m, 96 m.
Examination Measures	Blood collection, fasting, Urine collection, Height, standing, Weight, ...
Demographics	Age, gender, ethnicity, education, marital status, residency, income.

be useful to develop an auxiliary diagnostic model for assessing OA. Accordingly, many studies have been carried out from the perspective of assisting the diagnosis of early-stage knee OA disease [14–17].

Many researchers have explored how to predict the progression of knee OA disease and determine the grade of OA disease-based on various types of medical data [18–20]. Among these methods, machine learning approaches are the most common. For example, knee OA risk prediction is hot topic [17,21]. Liu et al. [22] proposed a supervised method (XGBoost model) to predict the side effects of analgesics on knee OA patients. The proposed method could not only accurately distinguish the severity of knee OA but also identify risk features from electronic medical records, which will help to assist patients with medication. Furthermore, Zhang et al. [23] adopted a logistic regression model for knee OA risk prediction.

Despite this progress, as far as we know, most previous OA works have focused on only one type of medical data and have not considered the inherent correlations among different modalities of data. In clinical diagnosis, the severity of the disease is embodied in the diagnostic information. Moreover, severity is highly correlated to the patient's clinical outcomes, biomedical markers, and examination images [24]. Some current methods have shown moderate success in predicting the KL grade, primarily using clinical data, X-ray images, or magnetic resonance imaging (MRI) as the input to the prediction model [11,12]. Compared with clinical data-based methods, the imaging-based diagnostic methods have improved the performance of OA risk-assessment models, especially the advanced deep learning methods [25–27]. However, owing to the lack of complementarity among different types of features, the performance of deep learning methods is also limited to a single data type or modality. In addition, the “black box” problem of deep learning makes it unconvincing.

Previous studies have shown that an integrated analysis of clinical modality and imaging modality is a promising strategy for identifying knee OA disease and predicting risk. For example, Guan et al. [16] developed a neural network for evaluating OA risk. They combined demographic and examination imaging features to predict the risk level of knee OA patients. In addition, a multi-modal prediction model was developed for incident knee OA [28]. This model can be used to predict the proportions of different modalities that produce adverse outcomes. These multi-modal integration strategies can improve the performance of prediction models. Tiulpin et al. [29] proposed an OA progression prediction model based on CNN and manual screening features, but this model was a semi-automatic method with low efficiency and did not pay attention to the differences between different modes. At present, most multi-modal-based approaches usually splice different modality vectors into a mix-vector and use traditional machine learning for the prediction [30–33]. The combination of multi-modal features is key for the subsequent classifiers. However, using a rough combination of multimodal features can easily lead to an increase in the number of features (dimensions), resulting in an exponential increase in the amount of calculation required.

Inspired by multi-modal learning, we developed an integrated multi-modal learning method (MMLM) that learns and fuses multi-feature from multi-modal (clinical, imaging, and demographic modalities). Specifically, based on a multi-modal eigenvalue relation-learning framework, MMLM automatically optimizes the features used from the input modalities. Furthermore, MMLM captures the correlations among the selected demographic, clinical, and imaging features by adding regularization terms.

There are four main contributions of our work.

- 1) We proposed a multi-modal feature integration method (MMLM) that fuses clinical, imaging, and demographic features relating to knee OA.
- 2) We used L1-norm-based optimization method (MMLM) to optimize the expressiveness of the different type feature before constructing the multi-modal tensor.
- 3) On the basis of MMLM output, the performance of machine learning classifiers were improved.
- 4) We provided the in-depth explanation of the relations among features from different modalities in the osteoarthritis disease.

The rest of this paper is organized as follows. The materials and the proposed method are introduced in Section 2. Extensive experimental results are presented in Section 3. In Section 4, we discuss the significance of our study from the perspectives of both method strategy and visual analysis. Finally, we draw our conclusions in Section 5.

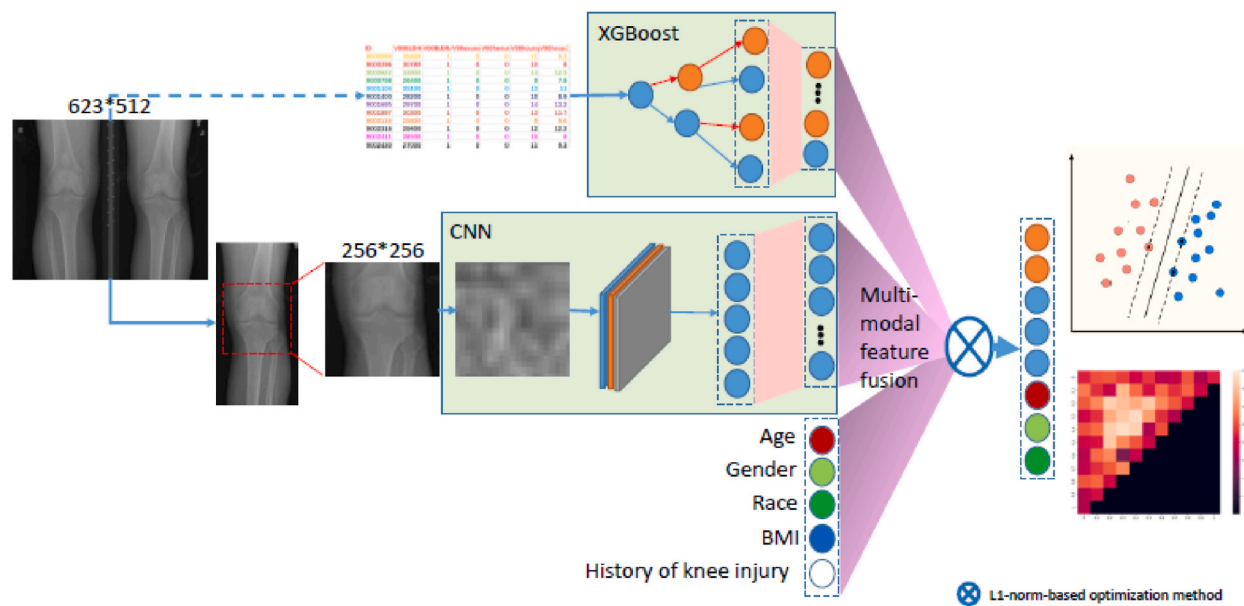
## 2. Materials and methods

### 2.1. Data set and preprocessing

Our experimental data came from the Osteoarthritis Initiative (OAI) database. OAI is a public database on OA disease that includes

**Table 2**  
Distribution of baseline KL grades for all knees in our study.

KL Grade	Participant number	Knee number
0	400	800
1	260	520
2	300	600
1 & 2	240	240&240



**Fig. 1.** The pipeline for the proposed MMLM.

demographic data, clinical records, X-ray images, and biospecimens from 4796 men and women aged from 45 to 79 who were enrolled between February 2004 and May 2006. The details of the entire cohort is shown in Table 1. The OAI database can be accessed at <https://nda.nih.gov/oai/>. The subjects in the OAI database can be divided into two types: 1) individuals who have been diagnosed with knee OA and have disease progression and 2) individuals who are at high risk of developing knee OA. The status of each knee for each person is described in clinical records and includes the KL grade, the central longitudinal measurements of joint space width (JSW), Western Ontario MacMaster (WOMAC) score, Osteoarthritis Research Society International (OARSI) score, as well as variables giving the case/control status for each knee [34]. The OAI database uses the KL grade to label the status of individual joints. In our study, the KL grade was used to judge whether a patient was assigned to the healthy control group (HC; KL grade 0) or the abnormal group. Note that each subject has two knees, and each knee has an independent KL grade.

In clinical diagnosis, KL grades 1 and 2 are considered as early-stage OA, whereas grades 3 and 4 are considered as severe-stage OA [13]. Our goal is to distinguish between grade 0 and the early stages of OA (grades 1 and 2) for a knee. Owing to the high similarity between grade 0 and grades 1 and 2, this is a very challenging task.

Of the 4796 patients in the OAI database, clinical and imaging data were available for 2301 people (4602 knees). We excluded patients with a KL grade of 3 or 4. The resulting data set had 2147 patients (4294 knees). To avoid any statistical bias due to an imbalance in the sizes of the data sets, we set the number of each knee type to be the same: 800 knees from 400 HC subjects (grade 0), 800 knees from 400 OA subjects at grade 1, and 800 knees from 400 OA subjects at grade 2. Thus, we had data for 1200 patients (2400 knees). The distribution of baseline KL grades for all knees is shown in Table 2. The demographic record, clinical record, and image data of each individual were linked by their sample ID.

In addition, we preprocessed the image provided by OAI. We manually cut the image with size of 623\*512 to 256\*256. The purpose is as follows: 1. Reducing the image size helps reduce the difficulty and time of model training. 2. We focus on the research of the knee joint. Too much background will interfere with the attention of the model to the knee joint.

## 2.2. Overview of our method

Given the correlations among the clinical data, imaging data, and demographic data with OA outcomes, we proposed MMLM to fuse three modalities to improve the classification of early-stage knee OA data compared to using single-modality data.

MMLM is a novel approach to integrate clinical data, imaging data, and demographic data. As shown in Fig. 1, in the MMLM

**Table 3**  
The main architecture of ResNet50.

Layer name	50-layer
Conv1	7*7,64, stride 2
Conv2	3*3 max pool, stride 2 [1 * 1; 64; 3 *3; 64; 1 * 1; 256] * 3
Conv3	[1 * 1; 128; 3 * 3; 128; 1 * 1; 512] * 4
Conv4	[1 * 1; 256; 3 * 3; 256; 1 * 1; 1024] * 6
Conv5	[1 * 1; 512; 3 * 3; 512; 1 * 1; 2048] * 6
Conv6	average pool, 1000-d fc, softmax

pipeline, clinical features are extracted by using XGBoost, knee radiography features using a convolutional neural network (CNN), and demographic features by a manual method. We adopt L1-norm regularization terms to control the expressiveness of each type of feature. The L1-norm regularized terms are used to regularize the relations among these correlated features. Then, these features are fused and optimized into a multi-modal tensor, which can capture all possible interaction spaces across the features in the three modalities. MMLM provides high-quality fusion features for a subsequent classifier. Compared with demographic features and clinical features, the imaging features can directly reflect the status of knee OA. The demographic features are used as statistical analysis terms.

### 2.3. Learning clinical features from clinical data

The clinical data reveals the status of knee OA from an intuitive view. In our study, we used data for 1200 people with 443 features, including various types of clinical information: medical history, physical measurements, nutrition, physical examinations, and so on. We removed those feature variables with >50% missing rates. Some missing values were imputed by the hot-deck method. Other missing values were imputed from the same variable in prior or later observations of the specific patient in the data set. In addition, we removed the features that were directly related to KL grade, such as WOMAC score,<sup>1</sup> OARSI score,<sup>2</sup> and so on. Finally, we removed 105 features and got 338 features from clinical data for each sample. This resulted in a 1200\*338 feature matrix. We used the eXtreme Gradient Boosting (XGBoost) method to extract the important features from the clinical feature matrix. In XGBoost, we employed the split node weight value as the importance of each feature in a tree. The important feature measure is based on the number of times that a node is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees, the importance of each splitting node defined as

$$\hat{I}_j^2(T) = \sum_{t=1}^{J-1} \hat{i}_t^2 1(v_t = j), \quad (1)$$

where the summation is over the non-terminal nodes  $t$  of the  $J$ -terminal node of tree  $T$ ,  $1(\cdot)$  represents the indicator function which is associated with squared-influence.  $v_t$  is the splitting variable associated with node  $t$ , and  $\hat{i}_t^2$  is the corresponding empirical improvement in squared error as a result of the split,  $\hat{i}_t^2$  is defined as

$$\hat{i}_t^2 = i^2(R_l, R_r) = \frac{w_l w_r}{w_l + w_r} (\bar{y}_l - \bar{y}_r)^2 \quad (2)$$

where the mean weight of the left and right children nodes of  $t$  are expressed  $\bar{y}_l \bar{y}_r$ , and  $w_l, w_r$  are the corresponding sums of the weights. For a collection of decision trees  $\{T_M\}_1^M$ , obtained through boosting, Eq. (2) can be generalized by its average over all of the trees in the sequence. Eq. (1) can be redefined as Eq. (3):

$$\hat{I}_j^2 = \frac{1}{M} \sum_{m=1}^M \hat{I}_j^2(T_M). \quad (3)$$

Finally, we extracted 19 features from the clinical modality, as described in more detail in Section 3.1.

### 2.4. Learning imaging features from knee radiographs

Knee radiographs are commonly obtained from patients. They can reveal the severity of OA. In the past century, owing to the availability and accessibility of plain X-ray films, knee radiographs have been the most popular standard technology for diagnosing knee OA [35–37]. For example, subchondral bone changes, osteophyte formation, and joint space stenosis can be detected from X-ray images. These features are considered to have clinical significance in OA grading and prediction.

As knee radiographs can contain a large region of irrelevant tissue and the acquisition parameters (resolution, contrast, exposure,

<sup>1</sup> The WOMAC score is a scoring system developed by Bellamy and colleagues specifically for hip arthritis and knee arthritis. WOMAC score ∈ [0–4].

<sup>2</sup> The OARSI score is the combined value of grade and stage. OARSI score ∈ [0–24].

etc.) of knee radiographs vary greatly, we used the semi-automatic segmentation method proposed by Ref. [37] to extract the region near the knee joint as the region of interest (ROI). In our experiments, the ROIs (1200\*2) of each patient's left and right knees were used as the representative area for knee OA to train a CNN. We used ResNet50 with batch normalization as the architecture of the CNN model. The architecture of ResNet50 was shown in Table 3. Finally, a feature vector ( $g_{out} \in R^{n \times 1}$ ) was extracted from the last hidden layer of the CNN.  $g_{out} \in R^{n \times 1}$  are image features that form one part of the multi-modal tensor, where n is the length of the feature vector. In our study,  $n = 128$ . The length of the imaging feature vector is 128.

## 2.5. Multi-modal feature fusion method

Multi-modal information is highly correlated with the patient's clinical outcomes. Thus, we proposed an integrated method (MMLM) to mine the hidden relationship among these modalities to boost classifier performance. In addition to the clinical and imaging modalities, OAI database also provides demographic information, including gender, age, race, history of knee injury, and body mass index (BMI) [36,38,39]. These data have been confirmed to be correlated with the grade of knee OA. We applied these demographic features as one of the modalities in our framework.

The fusion process works as follows. Let X be a vector, and N the number of samples.  $X^C$ ,  $X^G$ , and  $X^D$  correspond to the clinical modality, imaging modality, and demographic modality. Eq. (4) represents the multi-modal feature fusion vector:

$$X = [x_1, x_2, \dots, x_n]^T = [X^C, X^G, X^D] \in R^{N \times (c+g+d)}, \quad (4)$$

where c, g, and d are the numbers of output features for the clinical modality  $X^C$ , the imaging modality  $X^G$ , and the demographic modality  $X^D$ .

In classifying the OA grade of a knee,  $x_i$  denotes the i-th sample and  $y_i$  denotes the label (KL grade) of  $x_i$  ( $y_i \in \{0, 1, 2\}$ ), where  $x_i \in R^{c+g+d}$ , the sample feature vector. The objective function of the MMLM framework is

$$\min \sum_{i=1}^N l(x_i, w_i) + \gamma \text{tr}(WM^{-1}W^T), \quad (5)$$

where  $l(x_i) = \|x_i w - y_i\|_2^2$  indicates the empirical loss for the classification task.  $\gamma \text{tr}(WM^{-1}W^T)$  can capture the inherent correlation among the features of the three modalities. Here,  $\gamma$  is the Lagrange multiplier, and  $W = [w_c, w_G, w_D]$ ,  $w_c, w_G, w_D$  is the feature of different modalities, respectively.  $M^{-1}$  is the inverse matrix of M, which is a task covariance matrix that helps to learn W by automatically inducing the correct relation M among the features of the three modalities. In our study, M is a constraint term. We set  $M \geq 0$  to ensure that it is a positive semi-definite matrix. In addition, to penalize the complexity of M, we set  $\text{tr}(M) = 1$ .

However, a simple concatenation of modality features may be affected by the "curse of dimensionality". To solve this problem, optimized multi-modal methods have been widely used in biomarker recognition tasks and deep learning methods [32,33]. In our study, we proposed the MMLM to integrate features of multiple modalities in classifying early-stage knee OA. We express the three types of feature vector ( $X^C$ ,  $X^G$ , and  $X^D$ ) as linear functions of three features, namely  $w_c, w_G, w_D$ . The objective function Eq. (5) can be extended to:

$$\min \sum_{i=1}^N l(x_i) + \alpha \left( \|X^C w^C - X^G w^G\|_2^2 + \|X^C w^C - X^D w^D\|_2^2 + \|X^G w^G - X^D w^D\|_2^2 \right) + \sum_{j \in \{C, G, D\}} \beta^j \|w^j\|_1 + \gamma \text{tr}(WM^{-1}W^T), \quad (6)$$

where the first term is the empirical loss for the classification task. The second term ensures that the projections of each modality are as close as possible, so that the inter-correlations across different modalities can be captured. The third term ( $\beta^j \|w^j\|_1$ ) includes the L1-norm regularized terms, which is proportional the absolute value of the values in  $w^j$ . In our study, L1-norm regularized terms are used to select a small number of features from the multi-modal data. Here,  $\beta^j \geq 0$   $j \in \{C, G, D\}$  and  $\sum_{j=1}^3 \beta^j = 1$  is used as the regularization parameter. A larger  $\beta^j$  implies that more features in the j-th modality will be used for classification.

## 2.6. L1-norm-based optimization method

The second and third terms of Eq. (6) are used to capture and constrain the inter-correlations across different modalities. We use an iterative approach to optimize W and M. Let M be a fixed term. Then, we optimize W as follows:

$$\min_w \|X^C w^C + X w^G + X^D w^D - Y\|_2^2 + \sum_{j \in \{C, G, D\}} (\beta^j \|w^j\|_1 + \gamma M (w^j)^T w^j), \quad (7)$$

where  $Y = [y_1, y_2, y_3, \dots, y_N]^T \in R^n$  denotes the label of the sample.

In Eq. (7),  $w^C, w^G$ , and  $w^D$  are variables. To ensure these three variables have a minimum value, Eq. (7) works like a similar convex function for each  $w^j \in W(j \in \{C, G, D\})$ . Any of these three variables can be obtained by giving and fixing another variable. For example, the objective function for finding  $w^C$  can be expressed as:

$$\min_w \|X^C w^C + X^G w^G + X^D w^D - Y\|_2^2 \beta^C \|w^C\|_1 + \gamma M (w^C)^T w^C, \quad (8)$$

where variables  $w^G$  and  $w^D$  can also be derived in a manner like that for  $w^C$ .

The L1-norm can regularize the number of multi-modal features from the different modalities. Note that the L1-norm cannot be non-differentiable at zero. We guarantee that the L1-norm term leads to a smooth approximation by giving it a very small value. Specifically, we do this by taking the derivative with respect to  $w^C$  and setting it to zero. The expression for  $w^C$  can then be redefined as follows:

$$w^C = \frac{2(X^G)^T (X^G w^G + X^D w^D + 2(X^C)^T) Y}{[2(X^C)^T X^C + \beta^C + 2\gamma MI]} \quad (9)$$

In Eq. (9), the variables  $w^G$  and  $w^D$  can also be derived in a manner like that for  $w^C$ . After each  $w$  in  $W$  has been determined, a closed-form solution for  $M$  can be obtained by following the method proposed by Liu et al. [40], which can be stated as follows:

$$M = \frac{(W^T W)^{\frac{1}{2}}}{\text{tr}(W^T W)^{\frac{1}{2}}} \quad (10)$$

We update the weight of each sample iteratively. The above optimization process continues until  $W$  and  $M$  converge. The optimization algorithm for MMLM is shown in Algorithm 1.

## 2.7. Evaluation metrics

To verify the proposed model, the accuracy (Acc), precision (Prec), recall (Rec), and F1-score are used as metrics to evaluate the performance of all the comparative experiments, these metrics are list in Eq. (11):

$$\text{Acc} = \frac{TP + TN}{TP + FP + TN + FN}, P_{rec} = \frac{TP}{TP + FP}, R_{ec} = \frac{TP}{TP + FN}, F_{1-score} = 2 \times \frac{P_{rec} \times R_{ec}}{P_{rec} + R_{ec}} \quad (11)$$

where TP denotes the number of true positives, FP the number of false positives, TN the number of true negatives, and FN the number of false negatives.

---

Algorithm 1 Optimization pseudocode for the proposed MMLM.

---

Input:  
 $X^C$   
 $Y \in \mathbb{R}^N$   
Output:  
 $w^C$   
1:  $w^C = 0$   
2:  $M = [1, 0; 0, 1]$   
3: While not converged do:  
4: Update  $w^C$  in turn according to Eq. (9);  
5: Update  $M$  according to Eq. (10);  
6: End While  
7: Output:  
8:  $w^C$

---

## 3. Experiments and results

### 3.1. Experimental settings

We developed MMLM as a novel way to improve the classification accuracy for early-stage knee OA. MMLM generates a multi-modal feature tensor by extracting and optimizing clinical, imaging, and demographic features. We used 1200 subjects with 2400 knee images in our study. The data for the individuals were split into a training and validation set and a testing set in the ratio 80%–20%. The training and validation set was used to train and ne-tune XGBoost, ResNet50, and MMLM. The testing set was used to evaluate the performance of the model. To avoid bias, the extracted features from the clinical and demo-graphic modalities were at subject level. The extracted features from the knee imaging modality were at knee level. In MMLM method, clinical modality data was used as the input XGBoost, X-ray image was used as the input of CNN, and manual entry of gender, age, sex etc as a demographic feature.

For the clinical modality, we used XGBoost [41] to select the important features from the training set (960 subjects with clinical modality) in a 10-fold cross-validation strategy. Then, we used a statistical method to select features with the highest importance

**Table 4**  
Comparison of feature selection algorithms with SVM or DT.

		SVM				DT			
Method	Modality(ies)	Acc(%)	Prec(%)	Rec(%)	F1-Score(%)	Acc(%)	Prec(%)	Rec(%)	F1-Score(%)
sLASSO	Imaging	45.32	47.44	48.63	47.95	45.01	45.79	46.94	47.01
sLASSO	Clinical	64.83	67.05	66.93	63.47	62.09	64.59	65.73	61.84
mLASSO	Imaging & clinical	68.27	67.95	69.34	67.28	67.90	69.23	68.76	67.84
spCCA	Imaging & clinical	72.89	74.20	72.13	75.45	73.05	72.64	71.72	73.06
MMLM*	Clinical data & demographic	71.10	72.42	71.37	71.58	71.84	71.24	72.68	73.38
MMLM*	Imaging & demographic	73.79	73.21	75.89	75.87	74.12	75.46	73.13	75.48
MMLM*	Imaging & clinical data	76.54	77.18	75.43	78.75	76.01	77.69	74.72	78.58
MMLM	Imaging & clinical & demographic	83.45	79.27	82.04	84.83	81.27	78.06	81.54	84.13

Note that we did not use the demographic modality as a single modality with sLASSO because the simple linear relations between the demographic features and OA disease is insufficient to predict reliably the grade of OA. This modality can be used only as an auxiliary subset of multi-modal features, as it enhances the statistical analysis.

value. We recorded the F1-score of the features that were higher than 0 in each iteration. Finally, we extracted 19 clinical features for each individual (as described in Section 2.3). For the knee imaging modality, first, we used ResNet50 as the image extraction model and fine-tuned it using preexisting weights trained on ImageNet. Then, we replaced the original classification layer of ResNet50 with a new classification layer suitable for the.

OAI data set. Finally, the pre-trained network was retrained for the knee imaging diagnostic task (1920 knee images for 960 subjects). Thus, we obtained 128 imaging features for each knee (as described in Section 2.4). For the demographic modality, features such as race, gender, BMI, age, and history of knee injury play a significant role in predicting the grade of knee OA. These data are for an individual.

In our experiments, we optimized the fusion features following the strategy introduced in Section 2.6. Specifically, for each knee, we concatenated the 19 features selected from the clinical modality, the 128 features selected from the imaging modality, and the 5 features selected from the demographic modality into a vector with 152 units. These features were normalized to have zero mean and unit standard deviation. In this process, we used MMLM to fuse and optimize the features of different modalities. For the optimized multi-modal features, we applied two machine learning methods, namely a support vector machine (SVM) and a decision tree (DT), to verify the performance of the proposed MMLM in classifying early-stage knee OA (into grade 0 or grades 1 or 2).

Our experiments were implemented in Python (Tensorow package [42]). The training and testing processes were implemented in Python on a PC with an Nvidia Titan X Pascal CUDA GPU processor.

### 3.2. Comparison results

We compared the performance of MMLM using three modalities (the imaging, clinical, and demographic modalities) with the performance of strategies using a single modality (i.e., the imaging or clinical modality) or two modalities (the imaging and clinical modalities or the clinical and demographic modalities). We did this comparison with four feature selection methods (sLASSO, mLASSO, spCCA, and MMLM\*). sLASSO is based on LASSO and uses a single modality to predict the grade of OA. mLASSO is also based on LASSO and concatenates data for two modalities to predict the grade of OA. spCCA is a sparse version of CCA and concatenates the data for two modalities [43] to predict the grade of OA. MMLM\* is a variant of MMLM that uses all the different combinations of two modalities to predict the grade of OA. For fairness of comparison, all methods used the SVM and DT classifiers. We present the comparison results in Table 4. We evaluated MMLM and the other methods on accuracy, precision, recall, and F1-score.

As shown in Table 4, we can draw the following conclusions.

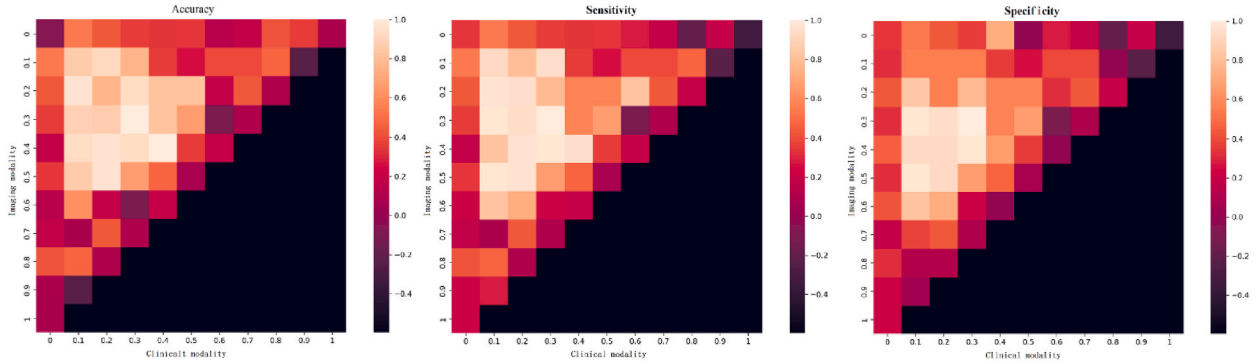
1. The two-modality algorithms (mLASSO, spCCA, and MMLM\*) achieved better classification performance on the testing set than the single-modality method (sLASSO with imaging or clinical data). Compared with the classification accuracy of mLASSO, the classification accuracy of the other three models improved by at least 8% (mLASSO improved by at least 3%), which demonstrates the advantages of comprehensively fusing imaging and clinical features in OA grade classification.
2. The MMLM and MMLM\* classifiers performed better than the other methods when MMLM\* used clinical and demographic data. This verifies that the proposed MMLM could improve the performance of the classifier since imaging data are an important modality in our method.
3. The classification results for the three MMLM\* classifiers were different, which means that there is heterogeneity between the three modalities.
4. Compared with the MMLM\* classifiers, MMLM (with either SVM or DT) had better classification performance, which verifies that considering the correlation between all three modalities is beneficial for feature selection.

### 3.3. Comparison with CNN methods

Next, we compared MMLM-based SVM with five CNNs: VGG16 [44], ResNet [45], CNN [46], DConvNet [47], and deep learning (DL) [16]. All these CNNs have been used for medical image classification. These other methods were trained and tested only using

**Table 5**  
Comparison of performance of different methods.

Method	Material	Acc(%)	Prec(%)	Rec(%)	F1-Score(%)	Time(hour)
VGG16 [44]	Imaging	70.80	69.74	71.28	71.19	5.2
ResNet [45]	Imaging	72.87	72.24	71.18	73.49	6.0
CNN [46]	Imaging	75.46	76.27	77.72	77.38	6.4
2DConvNet [47]	Imaging	76.26	77.08	77.49	78.52	6.5
DL [16]	Demographic & Imaging	79.65	78.28	79.81	80.14	7.0
SVM	MMLM(Our proposed)	83.45	79.27	82.04	84.83	6.6



**Fig. 2.** Influence of different weights for grade 1 knees (the squares at the top left are for the imaging features, those at the top right for the demographic features, and those at bottom left for the clinical features).

imaging data. We used the codes released by the authors. To ensure that the comparison was fair, we adopted the experimental settings applied in Section 3.1.

The results are shown in Table 5. The classification accuracy of the image-based CNN methods was higher than 70.00%, which shows the advantage of such methods in image processing. The performance of image-based SVM methods (Table 4) is worse than that of the CNN methods. Moreover, the performance of signal-based methods is worse than that of multi-modal methods. Compared with the other five CNN methods, our proposed MMLM with SVM achieves good classification performance. The accuracy of SVM is 3.80% higher (better) than that of the best of the other methods, DL. There are two main reasons for this. First, information relevant for classification is hidden in multi-modal data, but the image-based CNNs cannot obtain features other than from image information. Second, we use L1-norm regulated terms to control the expressiveness of each modality feature before constructing the multi-modal tensor. Thus, our method has a higher quality representation of knee OA. As a result, we suggest that MMLM can help SVM to classify knee grades more accurately.

In addition, we compare the time complexity of these methods in Table 5. Among these methods, the time complexity of MMLM method is 6.6 h, second only to DL [16]. This means that the MMLM method needs more training time. Although SVM plays as a simple machine learning classification in MMLM, MMLM method also includes CNN branch. Extracting features from multi-modal data is a time-consuming process. This is the reason why the time complexity of our proposed method is high. This is also the limitation of our model.

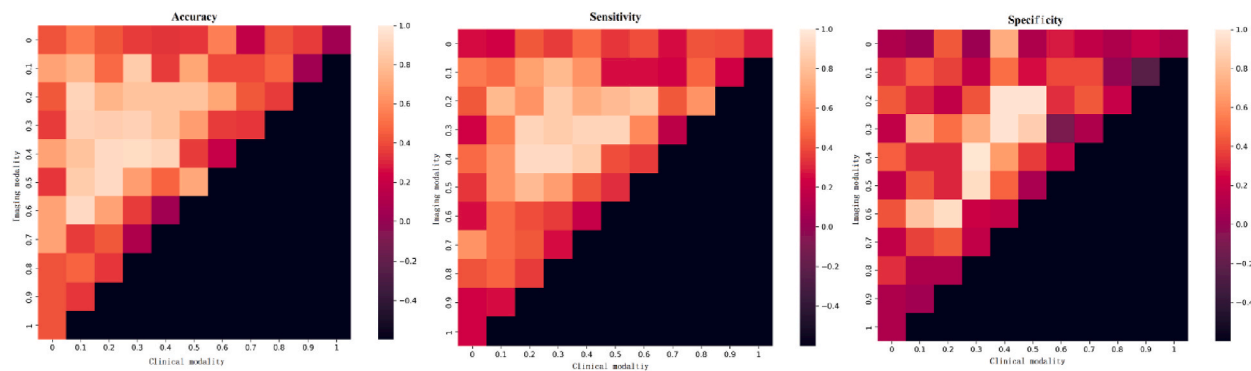
### 3.4. Influence of different combination schemes

To analyze the effects of different weights for the different modalities on the performance of MMLM, we adopted three strategies for the regularization parameters  $\beta^j$ , where  $\beta^1$  is the weighting factor on the subset of features of the clinical modality,  $\beta^2$  is the weighting factor on the subset of features of the imaging modality, and  $\beta^3$  is the weighting factor on the subset of features of the demographic modality. Their values lie in the range [0, 1]. We verified the possible values of  $\beta^j$  under the constraint  $\sum_{j=1}^3 \beta^j = 1$  on the SVM classifier.

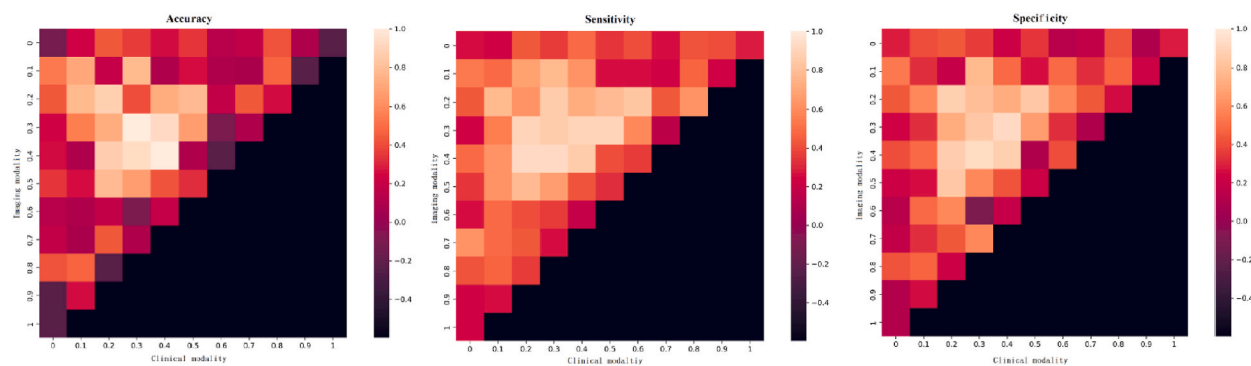
Figs. 2 and 3 are heat maps that represent the relations among these different combinations of weight for OA patients. Fig. 2 shows the results for grade 1 knees and Fig. 3 for grade 2 knees. We use three metrics (accuracy, sensitivity, and specificity) to visualize the effects of different weights for the clinical, imaging, and demographic features. In these plots, the squares at the top left are for the imaging features, those at the top right for the demographic features, and those at bottom left for the clinical features. The inner squares in the upper triangle denote the effectiveness of combining the three modalities. The brighter the color of a square, the more important the feature weights (the better the classification result). Under the constraint  $\sum_{j=1}^3 \beta^j = 1$ , only the squares in the upper triangles have valid values.

In Figs. 2 and 3, the inner squares in the upper triangles have much brighter colors than those near the three vertices. These brighter





**Fig. 3.** Influence of different weights for grade 2 knees (the squares at the top left are for the imaging features, those at the top right for the demographic features, and those at bottom left for the clinical features.).



**Fig. 4.** Influence of different weights for healthy controls (the squares at the top left are for the imaging features, those at the top right for the demographic features, and those at bottom left for the clinical features.).

colors indicate the effectiveness of combining features from the three modality features in classifying the grade of knee OA. The most brightly colored squares indicate the highest classification accuracy. These are mainly on the left or in the middle of each upper triangle, instead of in the inner or boundary regions. In Figs. 2 and 3, the squares corresponding to demographic features are darker, which indicates that these features have limited influence. This means that among the three modalities, the clinical and imaging features are more important in classifying the grade of knee OA than demographic features, which are an auxiliary factor. This does not mean that demographic features are not important. On the contrary, the features from the demographic modality have an important role in judging whether someone is suffering from knee OA.

We compare the effects of different weights for the features from the three modalities for HCs. In Fig. 4, the squares with high accuracy are close to the right or middle of each upper triangle, indicating that demographic features have a more significant role in distinguishing HCs among the three modalities. When classifying grade 1 knees (Fig. 2), the higher accuracy squares were mainly distributed in the middle of the lower left area of the upper triangle, indicating that the clinical and imaging modalities have almost the same influence for grade 1. When classifying grade 2 knees (Fig. 3), the higher accuracy squares tended to be on the boundaries of the imaging features, indicating that the imaging modality has a more important role in classifying a grade 2 knee.

### 3.5. Comparison of HCs and knee OA patients

The KL score is used as an index of the grade of knee OA in clinical examinations. In this study, we use the KL grade to label the status of knee OA. Our goal is to distinguish between grade 0 and the early stages of OA (grades 1 and 2). Owing to the high similarity of the knee joint in these conditions, it is a challenging task to judge the OA grade through a texture analysis of knee OA. Therefore, we group grade 1 and 2 knees together and compared them statistically with grade 0 knees. We use the features of the imaging and clinical modalities to locate and quantify the JSW index of each knee joint, and use demographic information to obtain the specific value of the corresponding JSW and mark it on the X-ray. The samples in the two comparison groups are from 365 individuals of very similar age and BMI ( $p = 4.77 \times 10^{-8}$  and  $p = 1.2 \times 10^{-7}$ , respectively).

In clinical diagnosis, a quantitative measurement of the JSW between the tibia and femur is one of the criteria used to assess the status of the knee. Therefore, we statistically analyze the tibiofemoral JSW at the specific anatomical position in the knee joint. The specific steps are as follows. First, we adopt a customized software tool to measure JSW from radiographs of a serial fixed flexion knee

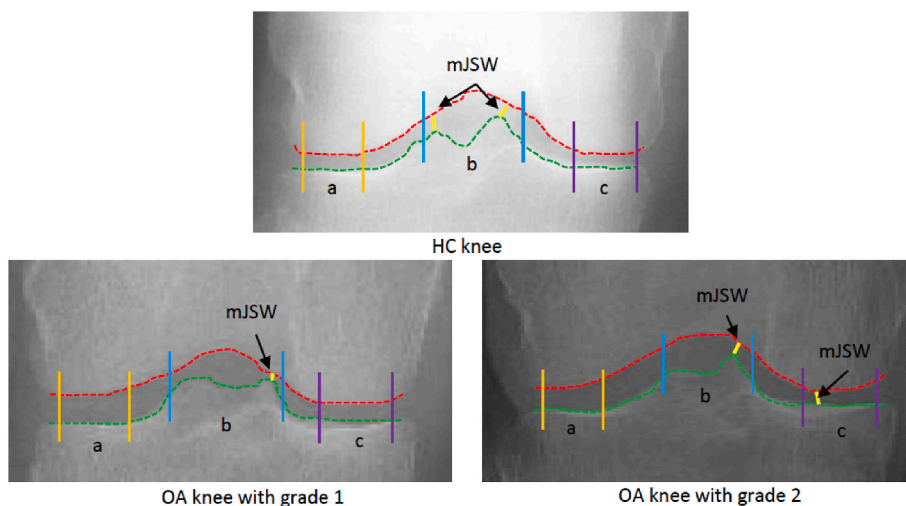


Fig. 5. Comparison of HC knee (top) and OA knees (bottom).

from the OAI [48]. As shown in Fig. 5, the software can automatically annotate the edges of the femoral and tibial condyles. We manually correct the contours generated by the software according to the recommendations of an orthopedic doctor. Then, we count the clinical features of HCs and OA patients that are extracted by MMLM. We statistically analyzed the feature values related to the typical JSW. Various representative minimum JSW (mJSW) values are marked with yellow solid lines in the figure and visualized in the internal measurement chamber of the joint in the X-ray image. Finally, we select three regions of the 380 medial chamber of each joint (a, b, and c) and use these in a comparative analysis.

The image in the first row of Fig. 5 is for a HC (grade 0), whereas the images in the second row are for two OA patients (grades 1 and 2). As shown in the first row for the HC, the JSW values for regions a, b, and c are relatively similar. The mJSW values in region b are marked since there are larger differences in this region. The annotation here indicates that the mJSW values in region b are almost the same, although the differences are smaller for regions a and c. In the second row, the variations in the JSW for regions a, b, and c are relatively large. In particular, the mJSW values at the two peaks in region b are obviously different. This is mainly due to the uneven narrowing of the knee joint space in arthritis. This is also consistent with the fact that the early onset of knee OA is not obvious, which makes it more difficult to distinguish between grades 1 and 2 clinically. Although the differences between HCs and OA patients can be confirmed by a clinical diagnosis [17], one limitation that cannot be ignored in our study is that a diagnosis depends on how the X-ray image was acquired.

#### 4. Discussion

In this study, MMLM was proposed as a novel method for learning multi-modal features from three OA modalities. We used two machine learning classifiers to verify the performance of MMLM in classifying early-stage knee OA. The experimental results show that MMLM can improve the performance of machine learning classifiers.

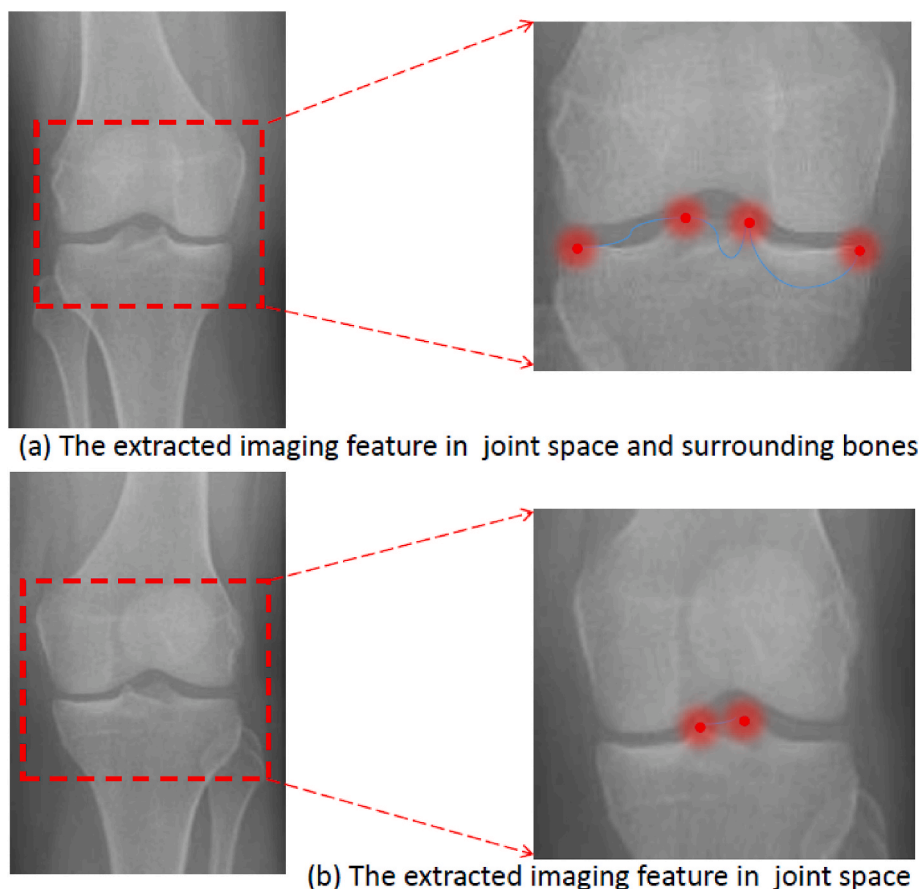
##### 4.1. Multi-modal data fusion

Various studies have confirmed that a combination of multi-modal data can improve the performance of predictive models [33,39,49–52]. Most of these studies concatenated the feature vectors into a single vector for prediction or classification. However, the disadvantage of these methods was that they do not optimize the features from different modalities. Thus, we developed the integrated MMLM. Compared with signal-based methods or direct feature concatenation methods, MMLM has two advantages: 1) It constrains the projection of each modality to make each modality as close as possible, which ensures that the inter correlations across different modalities can be captured. 2) It uses L1-norm regularized terms to balance the feature weights of different modalities. MMLM is a reasonable way to optimize the three types of modality (clinical data, imaging data, and demographic data) to guide classification.

In this study, we focused on classifying early-stage knee OA. We combined not only the clinical and imaging modalities but also the demographic modality. As shown in Table 4, a combination with each modality was essential in classification. In addition, compared with the other feature extraction methods (sLASSO, mLASSO, and spCCA), MMLM optimized the three different types of feature. Therefore, compared with sLASSO, mLASSO, and spCCA, the MMLM classifiers achieved better performance. In addition, according to the results in Figs. 2–4, when distinguishing between HC and OA subjects, the three modalities are equally important, whereas when distinguishing between grade 1 and grade 2 knees, the clinical and imaging modalities are more important.

**Table 6**  
Selected important features for the clinical modality.

Feature Type	Selected Features
Bone attributes	BMD of lateral tibial ROI, Bone Volume Fraction, and Trabecular thickness.
Clinical symptoms	Medial minimum JSW, medial/lateral JSW, and BMI.



**Fig. 6.** Visualization of the extracted imaging features on X-ray images.

#### 4.2. Comparison of the important features of the imaging and clinical modalities

The clinical and imaging modalities are very important in classifying potential factors associated with the grade of OA. In our study, we ran several experiments to evaluate the effectiveness of MMLM on three different types of feature (i.e., clinical, imaging, and demographic modalities). Here, we used cross-validation with the training and validation set to assess the efficiency of MMLM. Different features were selected in each fold, so we took the common overlapping features of the clinical and imaging modalities that appear in at least eight folds of the 10-fold cross-validation as the most important features. These important features of the clinical and imaging modalities are compared in Table 6 and Fig. 6, respectively.

Table 6 shows that 6 out of the 19 selected clinical features are correlated with features from the imaging modality. These can be grouped into two types: bone attributes and clinical symptoms. Bone attributes include clinical readings such as bone mass density (BMD) and quantitative values for the left and right knees. These are the BMD of the lateral tibial ROI, bone volume fraction, and trabecular thickness. Clinical symptoms include central longitudinal measurements of the JSW, as well as variables indicating the case/control status for each knee, for example, the medial minimum JSW, medial/lateral JSW, and BMI. As efficient early-stage case finding and diagnosis in primary care would enable health-care providers to proactively and substantially reduce the burden of disease through proper management including structured education, exercise and weight management (when needed) and addressing lifestyle-related risk factors for disease progression.

The bone attributes are regarded as measures of the ‘roughness’ of a bone surface by clinicians and are used to analyze and quantify the complex shape or structure of bone. Such attributes have also been used in OA-related studies. For example, both Lynch

et al. [53] and Eckstein et al. [54] used several bone attributes to quantify trabecular bone texture. Both research teams confirmed that knee OA is correlated with spacing, cross-connections between trabecular structures, variations in thickness, and variations in orientation. In MMLM, these extracted imaging features are correlated with clinical features, which mainly focus on JSW. In knee OA classification, the variation in JSW in X-ray images can be used to diagnose whether the person has OA [17,55].

To illustrate the relations among these features and knee imaging, we visualized the extracted imaging features and marked the corresponding regions of these features on an X-ray saliency diagram (Fig. 6). These regions are mainly composed of the joint space and surrounding bones. Some of the marked regions correspond to the features of clinical symptoms, especially the region of the joint space. These results show that clinical and imaging features are related and consistent in the diagnosis of knee OA.

Our research has two limitations: 1). The number of samples is limited and there is an imbalance problem. In future research, we need to collect more data to alleviate this problem. 2). MMLM is a complex integrated model that requires a lot of training time. The MMLM method needs further optimization.

## 5. Conclusions

In this study, we presented MMLM, a features optimization learning method for multi-modal. By fusing clinical, imaging, and demographic modalities, it can assist the classifier to assign the grade of early-stage knee OA. MMLM can select effective features from the three modalities and can improve overall classification performance. The main advantage of MMLM is that it uses the internal correlations among different modalities to enhance the efficiency of feature extraction and fusion, which improves grade classification of early-stage knee OA. In the future, we plan to fuse genetic information into deep neural networks so that they can learn more advanced features to enhance knee OA diagnosis further. We hope to provide efficient prediction tools for medical science and assist clinical diagnosis based on computational science.

## Author contribution statement

Liangliang Liu and Hongbo Qiao conceived and designed the experiments.

Liangliang Liu and Jing Chang performed the experiments.

Liangliang Liu and Hui Zhang analyzed and interpreted the data.

Pei Zhang, Qingzhi Ma, Tong Sun contributed reagents, materials, analysis tools or data.

Liangliang Liu and Jing Chang wrote the paper.

## Funding statement

Prof. Liangliang Liu was supported by Natural Science Foundation of Henan [222102310085].

## Data availability statement

No data was used for the research described in the article.

## Declaration of interest's statement

The authors declare no conflict of interest.

## Acknowledgements

The work described in this paper was supported by the Natural Science Foundation of Henan (Grant No. 222102310085) and the Key R&D Projects during the 14th Five Year Plan of China (Grant No. 2022YFD1400302).

## References

- [1] D.T. Felson, Y. Zhang, An update on the epidemiology of knee and hip osteoarthritis with a view to prevention, *Arthritis Rheum.: Official Journal of the American College of Rheumatology* 41 (8) (1998) 1343–1355.
- [2] D.T. Felson, An update on the pathogenesis and epidemiology of osteoarthritis, *Radiol. Clin.* 42 (1) (2004) 1–9.
- [3] L. Sharma, Osteoarthritis of the knee, *N. Engl. J. Med.* 384 (1) (2021) 51–59.
- [4] G. Peat, R. McCarney, P. Croft, Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care, *Ann. Rheum. Dis.* 60 (2) (2001) 91–97.
- [5] A. Mahmoudian, L.S. Lohmander, A. Mobasheri, M. Englund, F.P. Luyten, Early-stage symptomatic osteoarthritis of the knee|time for action, *Nat. Rev. Rheumatol.* 17 (10) (2021) 621–632.
- [6] S. Ravalli, F. Roggio, G. Lauretta, M. Di Rosa, A.G. D'Amico, V. D'agata, G. Maugeri, G. Musumeci, Exploiting Real-World Data to Monitor Physical Activity in Patients with Osteoarthritis: the Opportunity of Digital Epidemiology, *Heliyon*, 2022, e08991.
- [7] A.D. Woolf, B. Peger, Burden of major musculoskeletal conditions, *Bull. World Health Organ.* 81 (2003) 646–656.
- [8] K. Lim, C.S. Lau, Perception is everything: oa is exciting, *Int. J. Rheumatic Diseases* 14 (2) (2011) 111–112.
- [9] D.T. Felson, R. Hodgson, Identifying and treating pre-clinical and early osteoarthritis, *Rheum. Dis. Clin. N. Am.* 40 (4) (2014) 699.
- [10] A. Eitner, A.G. Culvenor, W. Wirth, H.-G. Schaible, F. Eckstein, Impact of diabetes mellitus on knee osteoarthritis pain and physical and mental status: data from the osteoarthritis initiative, *Arthritis Care Res.* 73 (4) (2021) 540–548.

- [11] Y.T. Keun, K.D. Won, C.S. Beom, O. Ein, P.J. Soo, C. Jose, Simple scoring system and artificial neural network for knee osteoarthritis risk prediction: a cross-sectional study, *PLoS One* 11 (2) (2016), e0148724.
- [12] N. Lazzarini, J. Runhaar, A.C. Bay-Jensen, C.S. Thudium, S.M.A. Bierma-Zeinstra, Y. Henrotin, J. Bacardit, A machine learning approach for the identification of new biomarkers for knee osteoarthritis development in overweight and obese women, *Osteoarthritis Cartilage* 25 (12) (2017) 2014–2021.
- [13] N. Arden, M.C. Nevitt, Osteoarthritis: epidemiology, *Best Pract. Res. Clin. Rheumatol.* 20 (1) (2006) 3–25.
- [14] L. Liu, Y. Yu, Z. Fei, M. Li, F.-X. Wu, H.-D. Li, Y. Pan, J. Wang, An interpretable boosting model to predict side effects of analgesics for osteoarthritis, *BMC Syst. Biol.* 12 (6) (2018) 29–38.
- [15] A. Brahim, R. Jennane, R. Riad, T. Janvier, L. Khedher, H. Toumi, E. Lespessailles, A decision support tool for early detection of knee osteoarthritis using x-ray imaging and machine learning: data from the osteoarthritis initiative, *Comput. Med. Imag. Graph.* 73 (2019) 11–18.
- [16] B. Guan, F. Liu, A. Haj-Mirzaian, S. Demehri, A. Samsonov, T. Neogi, A. Guermazi, R. Kijowski, Deep learning risk assessment models for predicting progression of radiographic medial joint space loss over a 48-month follow-up period, *Osteoarthritis Cartilage* 28 (4) (2020) 428–437.
- [17] Y. Nasser, R. Jennane, A. Chetouani, E. Lespessailles, M. El Hassouni, Discriminative regularized auto-encoder for early detection of knee osteoarthritis: data from the osteoarthritis initiative, *IEEE Trans. Med. Imag.* 39 (9) (2020) 2976–2984.
- [18] A. Soto-Hermida, M. Fernandez-Moreno, N. Oreiro, C. Fernandez-Lopez, S. Pertega, E. Cortes-Pereira, I. Rego-Perez, F.J. Blanco, Mitochondrial dna (mtdna) haplogroups influence the progression of knee osteoarthritis. Data from the osteoarthritis initiative (oai), *PLoS One* 9 (2014).
- [19] S. Heisinger, W. Hitzl, G.M. Hobusch, R. Windhager, S. Cotozana, Predicting total knee replacement from symptomatology and radiographic structural change using artificial neural networks data from the osteoarthritis initiative (oai), *J. Clin. Med.* 9 (5) (2014).
- [20] C.K. Kwok, H. Guehring, A. A. M. K. F. Eckstein, M.C. Hochberg, Predicting knee replacement in participants eligible for disease-modifying osteoarthritis drug treatment with structural endpoints - sciencedirect, *Osteoarthritis Cartilage* 28 (6) (2020) 782–791.
- [21] K.L. Urish, M.G. Keffalas, J.R. Durkin, D.J. Miller, C.R. Chu, T.J. Mosher, T2 texture index of cartilage can predict early symptomatic oa progression: data from the osteoarthritis initiative, *Osteoarthritis Cartilage* 21 (10) (2013) 1550–1557.
- [22] L. Liu, Y. Ying, Z. Fei, L. Min, J. Wang, An interpretable boosting model to predict side effects of analgesics for osteoarthritis, *BMC Syst. Biol.* 12 (S6) (2006).
- [23] W. Zhang, D.F. McWilliams, S.L. Ingham, S.A. Doherty, S. Muthuri, K.R. Muir, M. Doherty, Extended report: nottingham knee osteoarthritis risk prediction models, *Ann. Rheum. Dis.* 70 (2011) 1599–1604.
- [24] A. Scarpa, W. Mantovani, P. Capelli, S. Beghelli, L. Boninsegna, R. Bettini, F. Panzuto, P. Pederzoli, G.D. Fave, M. Falconi, Pancreatic endocrine tumors: improved tmn staging and histopathological grading permit a clinically efficient prognostic stratification of patients, *Mod. Pathol.* 23 (6) (2010) 824–833.
- [25] G.B. Joseph, C.E. Mcculloch, M.C. Nevitt, J. Neumann, A.S. Gersing, M. Kretzschmar, B.J. Schwaiger, J.A. Lynch, U. Heilmeier, N.E. Lane, Tool for osteoarthritis risk prediction (toarp) over 8 years using baseline clinical data, x-ray, and mri: data from the osteoarthritis initiative, *J. Magn. Reson. Imag.* 47 (6) (2018) 1517–1526.
- [26] A. Tulpin, J. Thevenot, E. Rahtu, P. Lehenkari, S. Saarakkala, Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach, *Sci. Rep.* 8 (1) (2018) 1–10.
- [27] P.S.Q. Yeoh, K.W. Lai, S.L. Goh, K. Hasikin, Y.C. Hum, Y.K. Tee, S. Dhanalakshmi, Emergence of deep learning in knee osteoarthritis diagnosis, *Comput. Intell. Neurosci.* 2021 (2021) 1–20.
- [28] H.J. Kerkhof, S.M. Bierma-Zeinstra, A. Hofman, A.G. Uitterlinden, C.A. Janssens, J. Meurs, Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors, *Osteoarthritis Cartilage* 20 (1) (2012) S183.
- [29] A. Tulpin, S. Klein, S.M. Bierma-Zeinstra, J. Thevenot, E. Rahtu, J.v. Meurs, E.H. Oei, S. Saarakkala, Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data, *Sci. Rep.* 9 (1) (2019), 20038.
- [30] F. Liu, C.-Y. Wee, H. Chen, D. Shen, Inter-modality relationship constrained multi-modality multi-task feature selection for alzheimer's disease and mild cognitive impairment identification, *Neuroimage* 84 (2014) 466–475.
- [31] J. Cheng, J. Zhang, Y. Han, X. Wang, X. Ye, Y. Meng, A. Parwani, Z. Han, Q. Feng, K. Huang, Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis, *Cancer Res.* 77 (21) (2017) 91–100.
- [32] A.-R. Mohammadi-Nejad, G.-A. Hossein-Zadeh, H. Soltanian-Zadeh, Structured and sparse canonical correlation analysis as a brain-wide multi-modal data fusion approach, *IEEE Trans. Med. Imag.* 36 (7) (2017) 1438–1448.
- [33] L. Liu, S. Tang, F. Wu, Y.-P. Wang, J. Wang, An ensemble hybrid feature selection method for neuropsychiatric disorder classification, *IEEE ACM Trans. Comput. Biol. Bioinf* 19 (3) (2022) 1459–1471.
- [34] G. Lester, The osteoarthritis initiative: a nih public-private partnership, *Hss Journal the Musculoskeletal Journal of Hospital for Special Surgery* 8 (12) (2012) 2047–2054.
- [35] T. Woloszynski, P. Podsiadlo, G. Stachowiak, M. Kurzynski, L. Lohmander, M. Englund, Prediction of progression of radiographic knee osteoarthritis using tibial trabecular bone texture, *Arthritis Rheum.* 64 (3) (2012) 688–695.
- [36] N. Lazzarini, J. Runhaar, A. Bay-Jensen, C. Thudium, S. Bierma-Zeinstra, Y. Henrotin, J. Bacardit, A machine learning approach for the identification of new biomarkers for knee osteoarthritis development in overweight and obese women, *Osteoarthritis Cartilage* 25 (12) (2017) 2014–2021.
- [37] T. Janvier, R. Jennane, A. Valery, K. Harrar, E. Lespessailles, Subchondral tibial bone texture analysis predicts knee osteoarthritis progression: data from the osteoarthritis initiative, *Osteoarthritis Cartilage* 25 (2) (2017) 259.
- [38] H. Kerkhof, S. Bierma-Zeinstra, N. Arden, S. Metrustry, M. Castano-Betancourt, D. Hart, A. Hofman, F. Rivadeneira, E. Oei, T.D. Spector, et al., Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors, *Ann. Rheum. Dis.* 73 (12) (2014) 2116–2121.
- [39] E. Halilaj, Y. Le, J.L. Hicks, T.J. Hastie, S.L. Delp, Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative, *Osteoarthritis Cartilage* 26 (12) (2018) 1643–1650.
- [40] M. Liu, D. Zhang, S. Chen, H. Xue, Joint binary classifier learning for ecoc-based multi-class classification, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (11) (2015) 2335–2341.
- [41] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
- [42] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: a system for large-scale machine learning, in: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, pp. 265–283.
- [43] J.R. Kettenring, Canonical analysis of several sets of variables, *Biometrika* 58 (3) (1971) 433–451.
- [44] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv preprint arXiv:1409.1556.
- [45] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [46] Z. Li, Y. Wang, J. Yu, Y. Guo, W. Cao, Deep learning based radiomics (dlr) and its usage in noninvasive idh1 prediction for low grade glioma, *Sci. Rep.* 7 (1) (2017) 1–11.
- [47] Y. Zhuge, H. Ning, P. Mathen, J.Y. Cheng, A.V. Krauze, K. Camphausen, R.W. Miller, Automated glioma grading on conventional mri images using deep convolutional neural networks, *Med. Phys.* 47 (7) (2020) 3044–3053.
- [48] J. Duryea, J. Li, C.G. Peterfy, C. Gordon, H.K. Genant, Trainable rule-based algorithm for the measurement of joint space width in digital radiographic images of the knee, *Med. Phys.* 27 (3) (2020).
- [49] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, A.D.N. Initiative, et al., Multimodal classification of alzheimer's disease and mild cognitive impairment, *Neuroimage* 55 (3) (2011) 856–867.
- [50] W. Hu, B. Cai, A. Zhang, V.D. Calhoun, Y.-P. Wang, Deep collaborative learning with application to the study of multimodal brain development, *IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng.* 66 (12) (2019) 3346–3359.
- [51] J. Liu, G. Tan, W. Lan, J. Wang, Identification of early mild cognitive impairment using multi-modal data and graph convolutional networks, *BMC Bioinf.* 21 (6) (2020) 1–12.

- [52] W. Hu, X. Meng, Y. Bai, A. Zhang, G. Qu, B. Cai, G. Zhang, T.W. Wilson, J.M. Stephen, V.D. Calhoun, et al., Interpretable multimodal fusion networks reveal mechanisms of brain cognition, *IEEE Trans. Med. Imag.* 40 (5) (2021) 1474–1483.
- [53] J. Lynch, D. Hawkes, J. Buckland-Wright, Analysis of texture in macro-radiographs of osteoarthritic knees, using the fractal signature, *Phys. Med. Biol.* 36 (6) (1991) 709.
- [54] F. Eckstein, D. Burstein, T.M. Link, Quantitative mri of cartilage and bone: degenerative changes in osteoarthritis, *NMR in Biomedicine, An Int. J. Devoted to the Development and Application of Magnetic Resonance In vivo* 19 (7) (2006) 822–854.
- [55] C. Mercier, M. Piperno, E. Vignon, K. Brandt, M. Hochberg, M.-P.H. Le Graverand, In normal knees, joint space width (jsw) is correlated with the intermargin distance (imd), a measure of medial tibial plateau alignment. variations in imd explain variability in jsw in serial radiographs, *Joint Bone Spine* 80 (2) (2013) 183–187.