An Integrated Multi-modal Learning Method for Early-stage Knee Osteoarthritis Disease Classification

Liangliang Liu (liangliu@henau.edu.cn)  
Henan Agricultural University

Jing Chang  
Henan Agricultural University

Qingzhi Ma  
Henan Agricultural University

Research Article

Keywords: classification, multi-modal learning, osteoarthritis, regularization term

Posted Date: December 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2329021/v1

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

Additional Declarations: No competing interests reported.
An Integrated Multi-modal Learning Method for Early-stage Knee Osteoarthritis Disease Classification

Liangliang Liu*, Jing Chang and Qingzhi Ma

Abstract

Background: 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.

Methods: 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 $L_1$-norm-based optimization method (MMLM) to regularize the inter-correlations among the multiple features.

Results: 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. Based on MMLM, the accuracy of SVM and DT are 83.45% and 81.27%, respectively, which are much higher than other feature fusion method. In addition, the visual analysis of the important features in the multi-modal data verified the relations among the different modalities.

Conclusion: MMLM 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.

Keywords: classification; multi-modal learning; osteoarthritis; regularization term

1 Introduction

Osteoarthritis (OA), a progressive and chronic disease, is among the most frequently occurring and disabling chronic diseases worldwide [1, 2]. 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 [3]. OA disease was the fourth leading cause of disability globally in 2020 [4, 5]. OA disease frequently occurs at the joints of the body, especially the knees, hips, and hands, of which the knee is most often affected [6, 7]. 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 [8, 9].
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 [10]. 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 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 [11, 12, 13, 14].

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 [15]. Among these methods, machine learning approaches are the most common. For example, Liu et al. [16] 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. Also based on an analysis of electronic medical records, Zhang et al. [17] adopted a logistic regression model for knee OA risk prediction. The model estimated the risk reduction resulting from the modification of potential risk factors. Urish et al. [18] combined a linear discriminant and feature reduction to identify the texture metric, which accurately predicted the onset of OA symptoms. Furthermore, Nasser et al. [14] proposed a discriminative regularized autoencoder method, which could learn both relevant and discriminative properties from X-ray images of knee OA that improved classification performance.

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 [19]. 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 [8, 9]. Compared with methods using clinical data, the imaging-based diagnostic methods have improved the performance of OA risk-assessment models, especially the advanced deep learning methods [20, 21, 22]. 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. [13] 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 [23]. 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. At present,
most multi-modality approaches usually splice different modality vectors into a mix-vector and use traditional machine learning for the prediction [24, 25, 26, 27]. The combination of multi-modal features is key for the subsequent classifiers. However, using a rough combination of multi-modal 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 combines 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 regularized terms 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 are 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 [1]. OAI is a public database on OA disease that includes 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 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), WOMAC scores, OARSI scores, as well as variables giving the case/control status for each knee [28]. The OAI database uses the KL grade to label the status of individual joints. In our study, the KL grade is used to judge whether a patient is 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 [10]. Our goal was 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 1. The demographic record, clinical record, and image data
of each individual were linked by their sample ID.

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 combine these three modalities
to improve the classification of early-stage knee OA data compared to using single-
modality data.

MMLM is a novel approach to integrating clinical data, imaging data, and dem-
ographic data. In the MMLM pipeline, as shown in Fig. 1, clinical features are
extracted using XGBoost, knee radiography features using a convolutional neural
network (CNN), and demographic features by a manual method. We next adopt
$L_1$-norm regularization terms to control the expressiveness of each type of feature.
The $L_1$-norm regularized terms are used to regularize the relations among these
correlated features. These features are then 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, 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. We removed the features that were directly related to KL grade, such as
WOMAC scores, OARSI scores, and so on. Finally, we got 338 features from clinical
data for each sample. This resulted in a $1200 \times 338$ 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. 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 [20, 29, 21]. 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, etc.) of knee radiographs vary greatly, we use the semi-automatic segmentation method proposed by [21] to extract the region near the knee joint as the region of interest (ROI). In our experiments, the ROIs (1200 × 2) of the left and right knees of each patient are used as the representative area for knee OA to train a CNN. We use ResNet50 with batch normalization as the architecture of the CNN model. Finally, a feature vector ($g_{out} \in \mathbb{R}^{n \times 1}$) is extracted from the last hidden layer of the CNN.

2.5 Multi-modal Feature Fusion

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) [30, 29, 31]. 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. The multi-modal feature fusion vector is

$$X = [x_1, x_2, \ldots, x_N]^T = [X^C, X^G, X^D] \in \mathbb{R}^{N \times (c+g+d)}, \quad (1)$$

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 $x_i \in \mathbb{R}^{c+g+d}$, the sample feature vector. The objective function of the MMLM framework is

$$\min \sum_{i=1}^{N} l(x_i, w_k) + \gamma \text{tr}(WM^{-1}W^T), \quad (2)$$

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 $W = [w_C, w_G, w_D]$. $M^{-1}$ 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 [26, 27]. 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 $w^C$, $w^G$, and $w^D$. The objective function Eq. (2) can be extended to:

$$\min_{W} \sum_{i=1}^{N} l(x_i) + \alpha \left( \|X^C w^C - X^G w^G\|^2 + \|X^C w^C - X^D w^D\|^2 + \|X^G w^G - X^D w^D\|^2 \right) + \sum_{j \in \{C, G, D\}} \beta_j \|w^j\|_1 + \gamma \text{tr}(WM^{-1}W^T), \quad (3)$$

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 intercorrelations across different modalities can be captured. The third term includes the $L_1$-norm regularized terms, which 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 Optimization

The second and third terms of Eq. (3) are used to capture and constrain the intercorrelations 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:

$$\min_{W} \|X^C w^C + X^G w^G + X^D w^D - Y\|^2 + \sum_{j \in \{C, G, D\}} (\beta_j \|w^j\|_1 + \gamma M(w^j)^T w^j), \quad (4)$$

where $Y = [y_1, y_2, y_3, \ldots, y_N]^T \in \mathbb{R}^N$ denotes the label of the sample.

In Eq. (4), $w^C$, $w^G$, and $w^D$ are variables. To ensure these three variables have a minimum value, Eq. (4) 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:

$$\min_{W} \|X^C w^C + X^G w^G + X^D w^D - Y\|^2 + \beta^C \|w^C\|_1 + \gamma M(w^C)^T w^C), \quad (5)$$

where variables $w^G$ and $w^D$ can also be derived in a manner like that for $w^C$. 
The $L_1$-norm can regularize the number of multi-modal features from the different modalities. Note that the $L_1$-norm cannot be non-differentiable at zero. We guarantee that the $L_1$-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:

$$w^C = \frac{2(X_G)^T(X_GW_G + X_DW_D + 2(X_H)^TY)}{2(X_C)^TX_C + \beta^C + 2\gamma MI}.$$ (6)

After each $w^j (j \in \{C, G, D\})$ in $W$ has been determined, a closed-form solution for $M$ can be obtained by following the method proposed by Liu et al. [32], which can be stated as follows:

$$M = \frac{(W^TW)^{1/2}}{\text{tr}(W^TW)^{1/2}}.$$ (7)

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 $F_1$-score are used as metrics to evaluate the performance of all the comparative experiments:

$$\text{Acc} = \frac{TP + TN}{TP + FP + TN + FN},$$ (8)

$$\text{Prec} = \frac{TP}{TP + FP},$$ (9)

$$\text{Rec} = \frac{TP}{TP + FN},$$ (10)

$$F_1\text{-score} = 2 \times \frac{\text{Prec} \times \text{Rec}}{\text{Prec} + \text{Rec}},$$ (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.
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 can generate a multi-modal feature tensor by extracting and optimizing clinical, imaging, and demographic features. We used data for 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% to 20%. The training and validation set was used to train and fine-tune XGBoost, ResNet50, and MMLM. The testing set was used to evaluate the performance of the model. To avoid bias, the features extracted from the clinical and demographic modalities were at subject level. The features extracted from the knee imaging modality were at knee level.

For the clinical modality, we used XGBoost [33] 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 value. We recorded the $F_1$-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 pretrained network was retrained for the knee imaging diagnostic task (1920 knee images for 960 subjects). We, thus, 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 (Tensorflow package [34]). 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 [35] 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 2. We evaluated MMLM and the other methods on accuracy, precision, recall, and \( F_1 \)-score.

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.

From the results shown in Table 2, 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 SVM using the multi-modal tensor from MMLM with five other CNNs: VGG16 [36], ResNet [37], CNN [38], 2DConvNet [39], and deep learning (DL) [13]. All these CNNs have been used for medical image classification. These other methods were trained and tested using only 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 3. 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 2) was worse than that of the CNN methods. Moreover, the performance of signal-based methods was 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 was 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.

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 used 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 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. 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 only 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 compared 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 used the KL grade to label the status of knee OA. Our goal was 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 grouped grade 1 and 2 knees together and compared them statistically with grade 0 knees. We used the features of the imaging and clinical modalities to locate and quantify the JSW index of each knee joint, and used demographic information to obtain the specific value of the corresponding JSW and marked it on the X-ray. The samples in the two comparison groups were from 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 analyzed 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 from the OAI [40]. As shown in Fig. 5, the software can automatically annotate the edges of the femoral and tibial condyles. We manually corrected the contours generated by the software according to the recommendations of an orthopedic doctor. Then, we counted the clinical features of HCs and OA patients that were 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 selected three regions of the medial chamber of each joint (a, b, and c) and used 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 [14], 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 [41, 42, 31, 43, 27, 44]. Most of these studies concatenated the feature vectors into a single vector for prediction or classification. However, the disadvantage of these methods is 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 intercorrelations across different modalities can be captured. 2) It uses $L_1$-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 2, 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, 3, and 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.

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 use 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 4 and Fig. 6, respectively.

Table 4 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.

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. [45] and Eckstein et al. [46] 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 [47, 14].

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.

5 Conclusions
In this study, we presented MMLM, a multi-modal method for learning features. 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.

Acknowledgements
We gratefully acknowledge the support of Xuchang Zhongyuan Artificial Intelligence Cluster Computing for the algorithm implementation and training of this project, we also want to use MindSpore in future projects, which is a new deep learning computing framework. These problems are left for future work.

Funding
The work described in this paper was supported by the Natural Science Foundation of Henan (Grant No. 222102310085).

Availability of data and materials
The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Authors’ contributions
Liangliang Liu and Jing Chang designed the study, performed the research, analysed data, and wrote the paper. Qingzhi Ma analysed data.

Author details
College of Information and Management Science., Henan Agricultural University, Zhengzhou, Henan 450046, P.R. China.
References


neuropsychiatric disorder classification. IEEE/ACM Transactions on Computational Biology and Bioinformatics (2021)


**Figures**

**Tables**

**Table 1** Distribution of baseline KL grades for all knees in our study

<table>
<thead>
<tr>
<th>KL Grade</th>
<th>Participant number</th>
<th>Knee number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>400</td>
<td>800</td>
</tr>
<tr>
<td>1</td>
<td>260</td>
<td>520</td>
</tr>
<tr>
<td>2</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td>1 &amp; 2</td>
<td>240</td>
<td>240 &amp; 240</td>
</tr>
</tbody>
</table>
Figure 1  The pipeline for the proposed MMLM.

Figure 2  Influence of different weights for grade 1 knees.

Figure 3  Influence of different weights for grade 2 knees.

Table 2  Comparison of feature selection algorithms with SVM or DT

<table>
<thead>
<tr>
<th>Method</th>
<th>Modality(ies)</th>
<th>Acc(%)</th>
<th>Prec(%)</th>
<th>Rec(%)</th>
<th>F1-Score(%)</th>
<th>Acc(%)</th>
<th>Prec(%)</th>
<th>Rec(%)</th>
<th>F1-Score(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sLASSO Imaging</td>
<td>45.32 47.44 48.81 47.95</td>
<td>50.00</td>
<td>45.79</td>
<td>46.94</td>
<td>47.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sLASSO Clinical</td>
<td>64.83 67.05 66.93 63.47</td>
<td>67.28</td>
<td>64.59</td>
<td>65.73</td>
<td>61.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mLASSO Imaging &amp; clinical</td>
<td>68.27 67.95 69.34 67.28</td>
<td>67.90</td>
<td>67.23</td>
<td>67.16</td>
<td>67.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>spCCA Imaging &amp; clinical</td>
<td>72.89 74.20 72.13 75.45</td>
<td>72.06</td>
<td>72.64</td>
<td>72.72</td>
<td>73.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMLM</td>
<td>Clinical data &amp; demographic</td>
<td>73.79 74.24 73.92 73.87</td>
<td>74.14</td>
<td>75.45</td>
<td>75.84</td>
<td>73.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMLM</td>
<td>Imaging &amp; clinical &amp; demographic</td>
<td>83.45 79.27 82.04 84.83</td>
<td>81.27</td>
<td>78.06</td>
<td>81.54</td>
<td>84.13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4 Influence of different weights for healthy controls.

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

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Comparison of performance of different methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Material</td>
</tr>
<tr>
<td>VGG16 [36]</td>
<td>Imaging</td>
</tr>
<tr>
<td>ResNet [37]</td>
<td>Imaging</td>
</tr>
<tr>
<td>CNN [38]</td>
<td>Imaging</td>
</tr>
<tr>
<td>2DConvNet [39]</td>
<td>Imaging</td>
</tr>
<tr>
<td>DL [13]</td>
<td>Demographic &amp; Imaging</td>
</tr>
<tr>
<td>SVM</td>
<td>MMLM(Our proposed)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Selected important features for the clinical modality.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Type</td>
<td>Selected Features</td>
</tr>
<tr>
<td>Bone attributes</td>
<td>BMD of lateral tibial ROI, Bone Volume Fraction, and Trabecular thickness.</td>
</tr>
<tr>
<td>Clinical symptoms</td>
<td>medial minimum JSW, medial/lateral JSW, and BMI.</td>
</tr>
</tbody>
</table>
Figure 6 Visualization of the extracted imaging features on X-ray images.