EasyChair Preprint

№ 15330

HCGN: Hierarchical Convolution and Graph Network for Predicting Knee Osteoarthritis

Xionghui Yang, Pengju Tang, Kai Zou and Dawei Dai

October 29, 2024

HCGN: Hierarchical Convolution and Graph Network for Predicting Knee Osteoarthritis

Xionghui Yang
Chongqing Key Laboratory of Computational Intelligence (Chongqing University of Posts and Telecommunications)
Chongqing, China
2308767924@qq.com

Pengju Tang
Chongqing Key Laboratory of Computational Intelligence (Chongqing University of Posts and Telecommunications)
Chongqing, China
s220201087@stu.cqupt.edu.cn

Kai Zou
Chongqing Key Laboratory of Computational Intelligence (Chongqing University of Posts and Telecommunications)
Chongqing, China
zk618209@outlook.com

Dawei Dai*
Chongqing Key Laboratory of Computational Intelligence (Chongqing University of Posts and Telecommunications)
Chongqing, China
dw_dai@163.com

***Abstract*—Knee osteoarthritis (KOA) is a common joint disease that severely affects the normal lives of patients. Typically, in clinical practice, the severity of KOA is evaluated by observing X-ray images of the knee joint. This approach is highly dependent on the subjective experience of the doctor and may vary among doctors. In this study, we propose a deep convolutional neural network (CNN) model that integrates structural information processing to predict KOA severity automatically based on the Kellgren-Lawrence (KL) grading system. Specifically, (1) The knee joint regions of the original X-ray images are segmented using automatic detection for subsequent model predictions; (2) We employed popular pre-trained deep CNN models to perform feature extraction, obtain their multi-scale feature maps, and construct their corresponding graph representations; (3) A graph attention network (GAT) was designed as a fine-tuning module to build a KOA prediction model. In our experiments, we tested various pretrained models combined with a GAT fine-tuning module to evaluate their performance on the Osteoarthritis Initiative (OAI) dataset. The results show that our proposed method significantly improves the predictive performance in multiple aspects compared to the original model. In addition, our proposed method has good decision interpretability.(https://github.com/ddw2AIGROUP2CQUPT/HCGN)**

***Index Terms*—knee osteoarthritis, feature space, graph structure, image classification**

I. Introduction

Knee osteoarthritis (KOA) refers to the gradual damage, thinning, and erosion of the cartilage tissue in the knee joint, leading to joint pain, stiffness, swelling, and functional impairment. It is a leading cause of disability in the United States and worldwide [1]. Statistical data show that 392 million people worldwide had KOA in 2019. The number of people with KOA is predicted to reach 583 million [2] by 2040. Several factors may contribute to KOA, including injury, obesity, and age. Among them, obesity is a major risk factor [3]. KOA worsens with age, causing inflammation and affecting patients' daily lives. Early diagnosis and treatment of osteoarthritis can help effectively to reduce chronic pain and improve joint function [4]. Joint space narrowing (JSN) and osteophyte formation are key pathological features of KOA [5]. In terms of medical imaging, KOA primarily uses X-ray and MRI technologies to display the bone and soft tissue structures of the knee joint. Radiography is the most common and cost-effective imaging technique, through which doctors make diagnoses by observing the bone structure of the knee joint, including the patella, femur, and tibia. The Kellgren-Lawrence (KL) grading system [6] is a standardized method for assessing the severity of KOA. This system categorizes KOA into grades 0-4, as shown in Fig 1, with higher grades indicating greater severity. Grade 0 indicates no signs of KOA, whereas grade 1 suggests possible joint space narrowing and osteophytes. Grade 2 indicates definite osteophytes and possible joint space narrowing, whereas grade 3 indicates moderate joint space narrowing, multiple osteophytes, sclerosis, and possible bone deformity. Grade 4 is characterized by severe sclerosis, large osteophytes, and definite bone deformity.

Typically, doctors analyze the severity of KOA based on features such as the patient's knee joint space and the number of bone spurs observed in X-ray images. However, the subjective experience and knowledge level of doctors may lead to differences in analysis results, potentially causing misdiagnosis or missed diagnosis. In contrast, artificial intelligence technology is not affected by these factors, and therefore, offers a means to improve the accuracy and efficiency of KOA diagnosis. The current assessment of the severity of KOA using deep learning methods is generally divided into two stages: first, segmenting the region of interest (ROI) in knee X-ray images to enable the model to focus on specific areas and reduce the intervention of feature engineering, and second, designing a deep learning model to predict the severity of KOA. Therefore, more accurate localization of the ROI in X-ray images can affect the performance of a model.

In the second stage, there is a common method that entails designing deep CNN models with different structures and optimization objectives to improve their predictive performance.

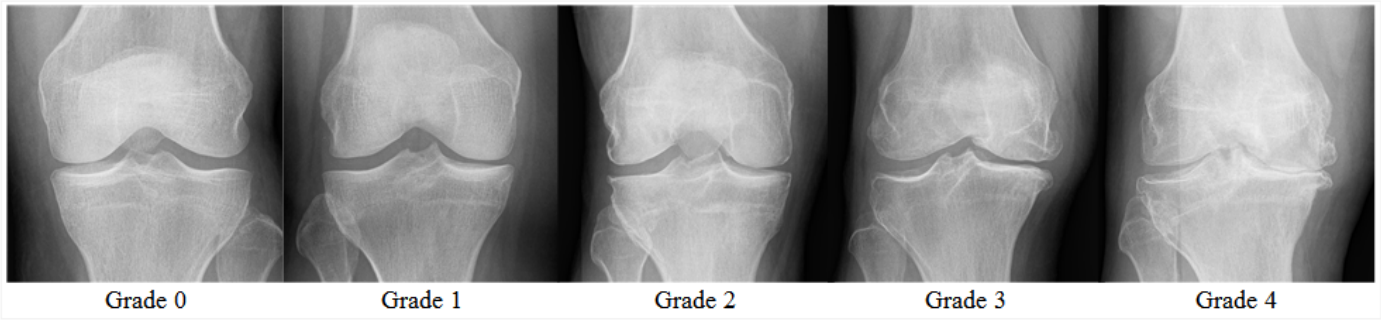

Fig. 1. Illustrations of KL grade samples of the knee joint from grade 0 to grade 4.

Current CNN models can be divided into two parts: feature extraction and classification. Owing to the limitations of its mechanism, CNN can extract only detailed local information within the receptive field, and lacks the ability to perceive structural features of the image content. Consequently, traditional CNN models based on pretrained models for designing new tasks often suffer from domain-shift problems. Clinicians can usually observe structural information in X-ray images to determine the severity of KOA. To this end, we propose a new framework that integrates the structural information of X-ray images into deep neural models to predict the severity. Particularly, (1) we segmented the ROI in X-ray images, with manual localization using the BoneFinder® software [7] and automatic detection using the YOLOv7 [8] model. (2) We used mainstream pretrained models to extract feature maps from knee joint region images and build graph structure representations based on channels. (3) We designed a graph attention network (GAT) as a fine-tuning module to process the structural information in the feature maps, and predict the KL grade of KOA. In summary, our contributions are as follows:

(1) We fine-tuned the YOLOv7 model to segment the knee region from X-ray images automatically, thereby reducing the need for human intervention in feature engineering.

(2) We propose a new framework that integrates a deep CNN and graph convolutional network (GCN) to consider the structural information in X-ray images effectively and predict the severity of KOA automatically based on the KL grading system. Experiments demonstrate that our proposed method can improve the performance of various mainstream pretrained models.

II. Related Work

A. Image Classification

In recent years, deep neural network models have achieved notable success in the field of image recognition and have been applied widely in many scenarios. LeNet [9], proposed by LeCun et al., is one of the earliest CNN models. AlexNet [10], proposed by Krizhevsky et al., consists of five convolutional layers and three fully connected layers, which significantly improves the accuracy of image classification. Zeiler and Fergus proposed ZFNet [11], which improves on AlexNet by optimizing the network structure using visualization techniques. Karen et al. proposed the VGG [12] model which uses small convolutional kernels and repetitive convolutional and pooling layers to construct a deep neural network. He et al. proposed ResNet [13]., which uses residual connections to solve the problem of degradation in deep neural networks, improving their trainability and accuracy. Szegedy et al. [14]–[17] proposed the Inception series of models, which introduced techniques such as batch normalization [15], decomposed convolution operations [16], and residual connections [17]. Hu et al. proposed SENet [18], which used a novel attention mechanism that learns the correlations between different feature channels adaptively. Huang et al. proposed DenseNet [19], which introduces dense connections into the layers of a deep model. Tan and Le proposed EfficientNet [20], which significantly improves computational efficiency and performance using a compound scaling technique and an automatically searched network structure. Zhuang et al. proposed ConvNext [59], which rethinks the design of convolutional neural networks and uses advanced training strategies. Alexey et al. proposed the VIT [21] model, which uses a self-attention mechanism to extract feature information between different blocks.

Recently, graph neural networks (GNN) [22] have made significant progress in the processing of graph-structured data. Kipf and Welling [23] proposed GCN, which generalize the concept of convolutional neural networks to learn node representations by aggregating information from local neighborhoods. Hamilton et al. proposed GraphSAGE [24], an unsupervised framework for generating node embeddings that generates node representations by sampling and aggregating information from adjacent nodes and is capable of efficiently processing large-scale graph data. Veličković et al. proposed the GAT [25], which introduces self-attention mechanisms into a GCN and calculates the attention coefficients between nodes to weight the neighbor node information. Han et al. proposed the VIG [26] model, which divides images into equally sized grids of image blocks and uses a GNN for classification tasks. Chen et al. proposed Patch-GCN [27] for survival prediction of patients from whole slide images in medical imaging, which achieved significant improvements compared to weakly supervised approaches. Yi et al. proposed

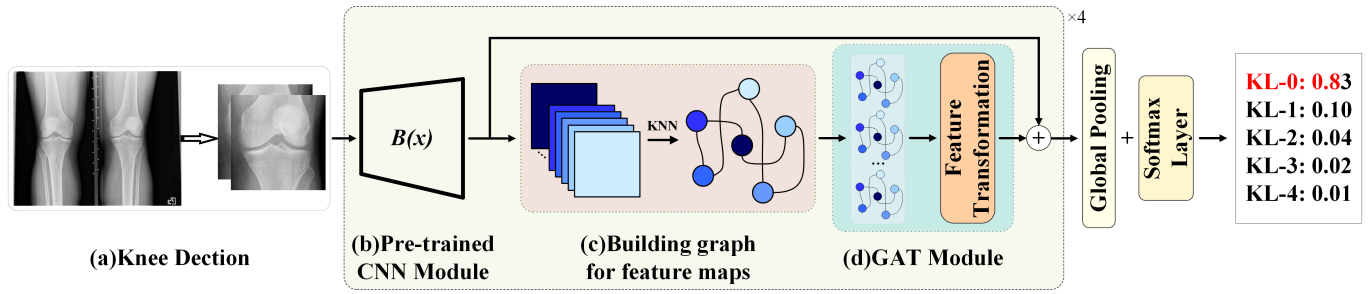

Fig. 2. Overview of our proposed, which consists of 4 parts. (a) Segmentation of the ROI of the knee joint. (b) Feature extraction module of pre-trained deep neural model. (c) Graph representation of the feature space. (d) GAT module.

the Graph-Transformer model [28] for the classification of WSIs, which is an interpretable and effective framework for WSI-level classification.

B. Joint Detection

ROI detection in X-ray images can be performed either manually or automatically. For the manual method, Woloszynski et al. developed a similarity measurement [32] to locate the tibial position in X-ray images for segmenting the ROI. Marijnissen et al. [33] developed the Knee Image Digital Analysis (KIDA) software, which can measure continuous variables such as joint space width, osteophyte area, subchondral bone density, joint angles, and tibial tuberosity height from knee joint images. For the automatic method, Shamir et al. [34] used a moving window to calculate the Euclidean distance between the window and the predefined joint area while scanning X-ray images of knees. Antony et al. [5] used a support vector machine (SVM) and Sobel horizontal image gradient to detect the center of the knee joint and segment the ROI along the center. In recent studies, deep CNNs have been applied to ROI detection. Antony et al. [35] used a fully convolutional neural network to detect knee joint ROI and achieved state-of-the-art results. Berk et al. [36] used a manual template matching method to extract knee joint bounding boxes and then fine-tuned a U-Net [29]. Chen et al. [37] manually annotated the boundary boxes in X-ray images, trained with the YOLOv2 [38] model, and used a CNN to rate the severity of the knee joint. Albert et al. [39] employed a fast region-based CNN (R-CNN) [30] to detect the posterior-anterior and lateral regions in knee X-ray images, and to evaluate the severity of KOA. Yang et al. [40] used RefineDet [31] to localize the knee joint and predict the severity of KOA.

C. KOA Severity Grading

Antony et al. [5] used a pretrained deep CNN to classify knee KL grades after fine tuning, and achieved an accuracy of 59.6%. Suresha et al. [41] used a neural network model pretrained on ImageNet [42] for severity classification of knee joints. Tiulpin et al. [43] proposed a KOA severity classification method based on a Siamese CNN, which helped the model to learn the symmetry of the X-ray images. Berk et al. [36] applied dense modules to their network model to classify the severity of KOA. Brahim et al. [44] used the power spectral density in different directions of the image as features and applied a logistic regression classifier to classify the KL grade. Chen et al. [37] employed transfer learning and assigned higher penalties to misclassifications with larger distances between predicted and actual KL grades. Liu et al. [45] used two networks, a region proposal network (RPN) and Fast R-CNN [56], where the RPN was trained to generate the knee joint ROI, and the Fast R-CNN was used for KL grade classification. Bayramoglu et al. [4] used a small CNN model combined with a joint gap feature and bone texture to detect the presence of KOA. Thomas et al. [46] used a pretrained ResNet169 model on ImageNet to evaluate the severity of KOA. Wang et al. [47] introduced a self-attention mechanism in their CNN to explore the correlation between the different regions of a joint. Albert et al. [39] developed an automated deep learning method that jointly uses the posterior-anterior and lateral views of knee joint X-ray images to assess the severity of KOA. Feng et al. [48] introduced channel and spatial attention modules in their model to improve the effective use of information. Ahmed et al. [49] proposed the Deep Hybrid Learning (DHL)framework, in which the first pretrained CNN was used for feature extraction, then principal component analysis (PCA) was used to reduce the dimensionality of features, and finally, a support vector machine (SVM) was used for classification. Abdullah et al. [50] trained Faster R-CNN [58] to locate the ROI in X-ray images and extracted features using a pretrained ResNet-50. Finally, another pretrained AlexNet [10] was used to classify the severity of KOA. Gu et al. [51] used a deep CNN for the initial evaluation of the severity of KOA, calculated the JSN, and then combined the JSN and initial evaluation to determine the KL grade. Dharmani et al. [52] used a pre-trained EfficientNet [20] to assess the severity of KOA in knee joint X-ray.

III. Method

A. Overview of Our Proposed

Our approach fully exploits the feature extraction capabilities of deep neural networks, while enhancing the generalization ability of the model by representing feature maps as graphs. The overall framework of our approach is illustrated in Fig 2: (1) The knee joint ROIs are segmented from the original

X-ray images through automatic detection to reduce the impact of irrelevant areas. (2) A pre-trained deep neural model is applied as the feature extraction module to extract feature maps of the knee joint region. (3) The multi-scale feature maps are represented in the graph structure, and a graph neural network classifier is fine-tuned using a cross-entropy loss function.

B. Knee Joint Segmentation

After pre-processing the Osteoarthritis Initiative (OAI) dataset, we further segmented the knee joint ROI's in the images. Considering computational complexity and detection accuracy, we adopted YOLOv7 as our detection model. Following the method of Chen et al. [37], we used a portion of the original dataset as the knee joint detection dataset. The higher the intersection over union (IoU) score, the better is the convergence effect of the model. The IoU represents the ratio of the intersection area to the union area between the predicted bounding box and ground truth (See (1)).

$$\text{IoU}_{(A,B)} = \frac{A \cap B}{A \cup B}. \tag{1}$$

Where, A represents the predicted bounding box and B represents the ground truth. To improve the accuracy of knee joint detection, we fine-tuned the YOLOv7-based model pretrained on the COCO dataset [57]. Particularly, we removed the classification loss and retained the objectness and location losses . This reduces noise and complexity during training and improves the accuracy and efficiency of the model. To evaluate the performance of the detection model, we calculated the proportion of knees that achieved an IoU ≥ 0.75.

C. Approach

1) Pre-trained CNN Module: Owing to the lack of high-quality annotated KOA X-ray image training data, pre-trained models can help mitigate the insufficient training data by exploiting the learned generic features to assist in learning new tasks, accelerating the training process, improving performance, and combating overfitting [53]. We adopted classic deep models that well trained on the ImageNet dataset as the pre-trained CNN module, including VGGNet, ResNet, and ConvNext [12] [13] [59]. We describe this process as $z = B(x)$, where x is the input image and $z \in \mathbb{R}^{C \times H \times W}$, C is the number of channels in the feature map, and H and W represent the height and width of the feature maps.

2) Building Graph for Feature Maps: We extract feature maps from x using convolutional blocks. The feature of each channel in the feature maps is treated as a node, and we use the K-Nearest Neighbor (KNN) [60] algorithm to build edges from these nodes to construct an undirected, fully connected graph $G = (V, E)$. Here, V is the set of nodes, and E is the set of edges connecting the nodes. We consider each channel in z as a node, with a total of $k = C$ nodes. For a set of edges E, each node v_i is connected to its k nearest neighboring nodes. By establishing the edge relationships between nodes, performing graph convolution helps to propagate information throughout the graph, enabling the modeling of relationships at the channel level.

3) GAT Module: A GAT [25] enables information propagation between nodes by computing the weights of the relationships between each node and its neighbors adaptively. The GAT module includes multiple heads, each of which captures a different relational pattern. Within each head, the GAT module computes the representation of each node by multiplying its node features by attention weights, and then summing them based on the weighted contributions. The process can be simply expressed as $z' = GAT(G, h)$, where G is the graph constructed from the feature map z, h represents the number of heads in the GAT module. As shown in Fig 3, in order to better integrate GAT modules with convolutional neural networks and obtain richer features, we introduce ConvNext block as our Feature Transformation block.

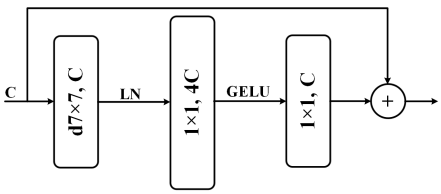

Fig. 3. Feature Transformation block.

D. Evaluation Metrics

We introduce four metrics to evaluate the performance of the models: Accuracy, Recall, Specificity, and F1-score. Accuracy represents the proportion of correctly predicted samples to the total number of samples (i.e., traditional Top-1 accuracy), which measures the overall prediction ability of the model. Recall measures the recognition ability of a model for each specific KL grade of KOA X-ray images, whereas Precision refers to the measure of how many of the samples predicted as positive actually belong to the positive class. F1-score is a metric that takes into account both the precision and recall of the proposed model. Calculations were shown in Eq (2)(3)(4)(5).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{2}$$

$$Recall = \frac{TP}{TP + FN}. \tag{3}$$

$$Precision = \frac{TP}{TP + FP}. \tag{4}$$

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{5}$$

For a given KL grading (where the samples defining this grading are considered to be positive examples and the remaining samples are considered to be negative examples), TP refers to the number of positive samples that are correctly classified, FN refers to the number of positive samples that are incorrectly classified as negative, FP refers to the number of negative samples that are incorrectly classified as positive, and TN refers to the number of negative samples that are correctly classified as negative, as illustrated in the Fig 4.

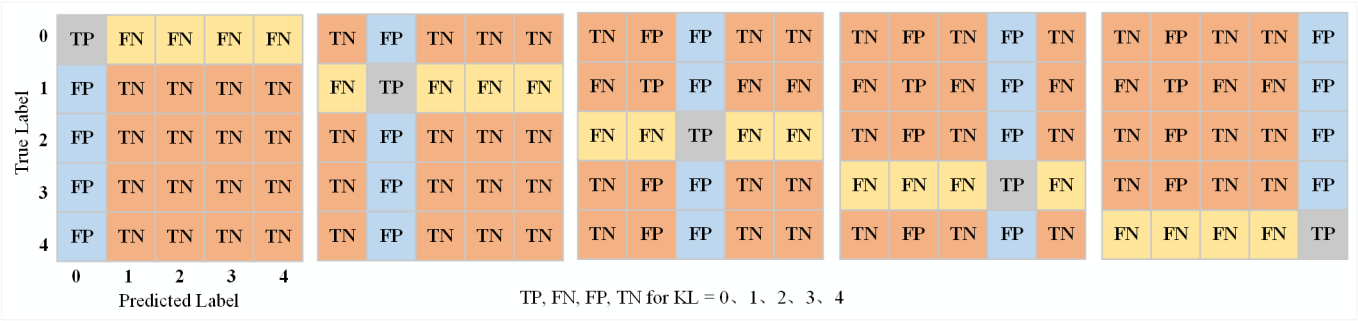

Fig. 4. TP, FN, FP, and TN displayed for KL-0, 1, 2, 3, and 4.

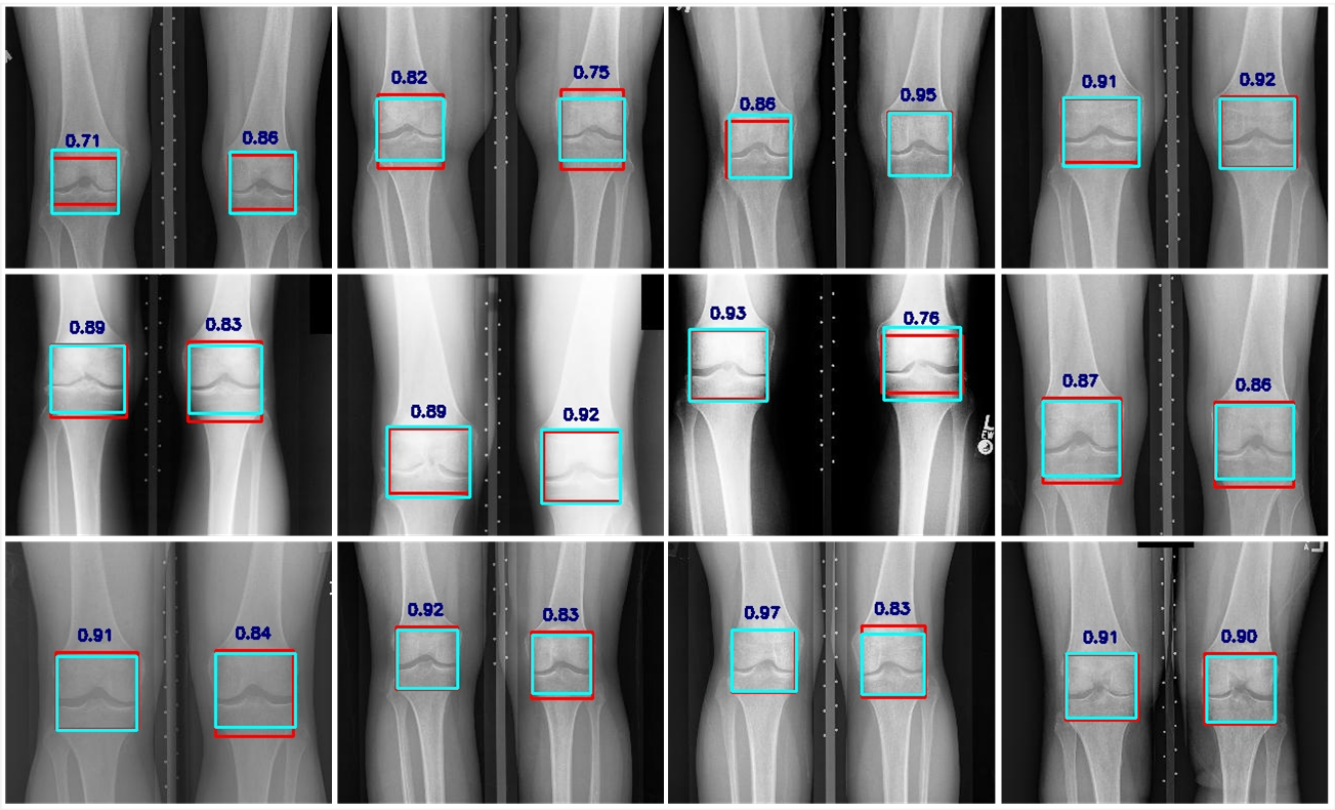

Fig. 5. ROI of the left and right knee detected by YOLOv7, along with their respective IoU scores. The red boxes represent the manually annotated bounding boxes, while the blue boxes represent the detection boxes identified by YOLOv7.

IV. Experiment

A. Experimental Settings

1) Data Preprocessing: We used the OAI knee joint X-ray image dataset (http://www.oai.ucsf.edu/). We selected baseline knee images consisting of 16-bit DICOM X-ray files and evaluation information for 4,796 samples for patients with ages ranging from 45 to 79 years. We first converted all images to standard 8-bit grayscale format and then excluded samples with unclear label information, resulting in 4,506 images in the final dataset.

2) Implementation Details: Owing to the inconsistent sizes of the image slices, we resized all images to 224×224 for ResNet, VggNet, and to both 224×224 and 384×384 for ConvNext. To prevent overfitting, we applied data augmentation by flipping the images horizontally and vertically with a probability of 0.5, adjusting the brightness and saturation by a factor of 0.33 and using random affine. We used PyTorch 1.8 as the software framework. For hardware, we used $2\times$ Nvidia A100 GPU with 40 GB of memory for each experiment. During training, the feature extractor had an initial learning rate of 0.0001, whereas the feature transformation blocks and graph neural network layers had an initial learning rate of 0.001. The batch size was set to 8 and the learning rate was reduced by a factor of 10 if the test accuracy did not improve after every 7500 batches, for a total of 100,000 iterations. The optimizer used was stochastic gradient descent (SGD), with a

weight decay of 0.001 and momentum of 0.9.

B. Knee Joint Detection

We calculated the recall of the model using a threshold of 0.75, with samples having IoU $\geq$ 0.75 considered to be positive samples, and samples with IoU $<$ 0.75 considered as negative samples. We also computed the average IoU for knee detection. Among all knee test samples, all knees were detected, with 98.25% of the samples having IoU $\geq$ 0.75 and an average IoU of 89.01%, which is significantly higher than previous object detection models. The detection metrics are listed in TABLE I. The automatic detection results, as shown in Fig 5, demonstrate a high degree of similarity with the manually annotated bounding boxes. In subsequent prediction model training, we cropped the ROI of the OAI dataset images using automatic detection methods. Particularly, we extended the YOLOv7 detection boxes by a scale factor of 1.4 and randomly split the dataset into training and testing sets with a ratio of 8:2. The results of the dataset splitting are presented in TABLE II.

TABLE I
RESULTS OF KNEE JOINT DETECTION BASED ON YOLOV7 MODEL. L REPRESENTS THE LEFT KNEE, AND R REPRESENTS THE RIGHT KNEE.

Methods	Recall	Mean IoU
FCN [5]	0.892	0.83
HOG-SVM [55]	-	0.84
YOLOv2 [37]	0.922	0.859
YOLOv7(ours)	**0.983** (L:0.972 R:0.993)	**0.890** (L:0.884 R:0.896)

TABLE II
DATA PARTITIONING OF CROPPED ROI REGIONS.

	KL 0	KL 1	KL 2	KL 3	KL 4	Total
Train set	2759	1278	1900	992	236	7165
Test set	689	319	474	247	59	1788
Total	3448	1597	2374	1239	295	8953

TABLE III
COMPARISON OF OUR METHOD WITH DIRECTLY USING DEEP LEARNING MODELS (RN, VGG AND CN FOR COMPREHENSIVE PREDICTION PERFORMANCE IN AUTOMATICALLY DETECTING KNEE JOINT ROI USING YOLOV7.

Models	Accuracy	Recall	Precision	F1-score
RN-50/_HCGN_224	69.4/**71.4**	66.2/**68.2**	67.8/**71.2**	65.3/**69.0**
RN-101/_HCGN_224	70.2/**70.5**	**69.2**/68.2	**71.7**/71.1	**70.3**/69.2
VGG16/_HCGN_224	70.5/**71.7**	66.8/**68.1**	71.6/**72.2**	68.5/**69.0**
VGG19/_HCGN_224	70.5/**71.3**	67.3/**68.9**	70.9/**72.5**	68.6/**70.2**
CN-Ti/_HCGN_224	71.6/**72.2**	68.2/**70.0**	72.1/**72.6**	68.5/**70.2**
CN-Ti/_HCGN_384	72.3/**73.8**	69.0/**72.6**	73.1/**73.9**	70.1/**72.8**

C. Comprehensive Performance and Analysis

We evaluated the performance of the current popular deep neural models (VGGNet, ResNet, and ConvNext) for automatic detection on the OAI dataset. As shown in TABLE III, we can conclude that (1) our proposed method can improve the predictive performance of most deep learning models. (2) For automatic detection, HCGN_CN-Ti_384 achieved the best classification accuracy of 73.8%, the best Recall of 72.6%, the best Precision of 73.9%, and the best F1-score of 72.8%. We conclude that the overall performance of the models improved almost uniformly.

As shown in TABLE IV, our method achieved the highest prediction accuracy for each of the fine-grained metrics.

TABLE IV
COMPARISON OF CLASSIFICATION ACCURACY, RECALL, PRECISION AND F1-SCORE BETWEEN THE EXISTING END-TO-END TRAINING METHODS ON THE OAI DATASET AND OUR PROPOSED METHOD (HCGN_CN-TI_384).

Methods	Accuracy	Recall	Precision	F1-score
VGG-19 [5] (2016)	53.4	-	-	-
CNN [35] (2017)	61.9	62.0	57.0	56.0
VGG-19-Ordinal loss [37] (2019)	70.4	-	-	-
DesNet169(TL) [46] (2020)	71.0	71.0	72.0	71.0
CNN with self-attention [47] (2021)	69.2	-	-	-
DenseNet121-DRS [61] (2021)	71.1	71.0	68.0	68.0
CNN with attention [48] (2021)	70.2	68.2	71.0	68.0
DenseNet161-FOL [62] (2022)	67.43	66.0	66.0	65.0
ViT [63] (2022)	71.20	-	-	-
OsteoHRNET [64] (2023)	71.74	71.0	73.0	72.0
Ours(HCGN_CN-Ti_384)	**73.8**	**72.6**	**73.9**	**72.8**

D. Ablation experiment of the GAT module

In order to better integrate the GAT block with convolutional neural networks, we introduced the Feature Transformation (FT) block after the GAT block. We evaluated the influence of different components of the GAT module on the experimental results. As shown in Table V, we observed that combining the GAT block with the Feature Transformation block yields better results than using the GAT block alone.

TABLE V
COMPARISON OF OUR METHOD(HCGN_CN-TI_384) WITH USING DIFFERENT COMPONENTS(GAT AND FEATURE TRANSFORMATION BLOCK) OF THE GAT MODULE ON THE OAI DATASET.

GAT	FT	Accuracy	Recall	Precision	F1-score
✗	✗	72.32	68.97	73.11	70.13
✓	✗	73.26	71.07	73.40	71.73
✗	✓	72.31	69.48	70.89	69.59
✓	✓	**73.76**	**72.61**	**73.92**	**72.79**

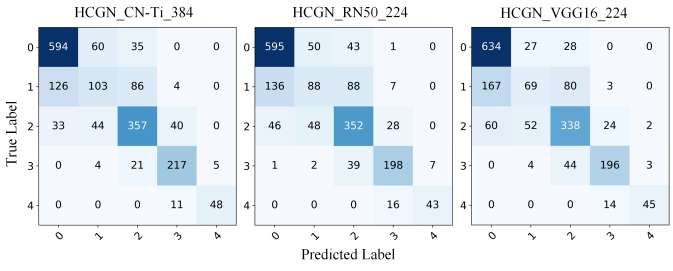

Fig. 6. Comparison of our methods under automatic segmentation.

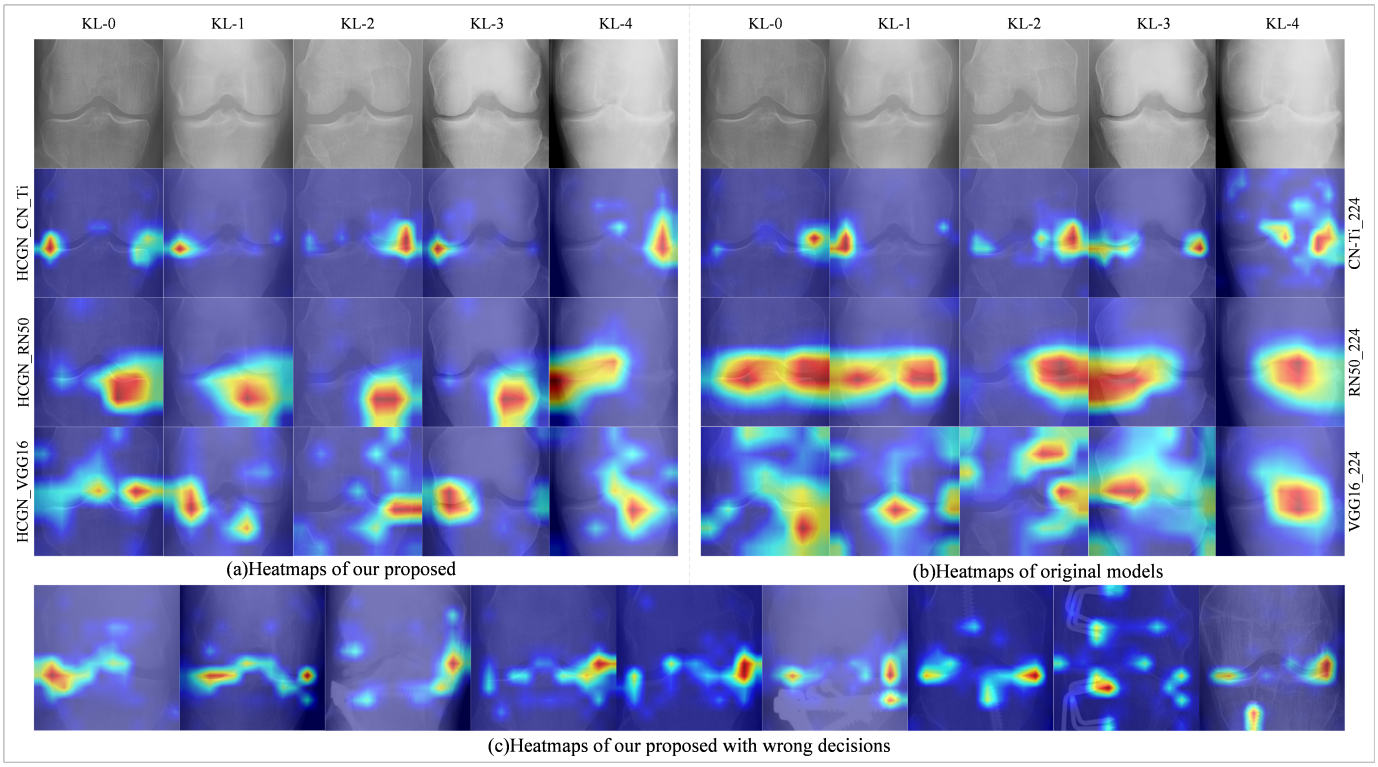

Fig. 7. Comparison of our proposed (DFCR_CN-Ti_384, DFCR_RN50_224, and DFCR_VGG16_224) and original models (CN-Ti_384, RN50_224, and VGG16_224) using the same input.

E. Prediction Result Analysis

In the clinical diagnosis of KOA, there are cases where the differences between X-ray images of adjacent grades are subtle, leading to different doctors diagnosing the same image with different grades. However, these diagnostic differences typically do not cause significant harm. Fig 6 shows the confusion matrix of our model's prediction results (i.e., the statistics of all test sample prediction results), which reveals that (1) most of the prediction errors occur between adjacent grades, and (2) the model misclassifies most severity grade 1 samples as grades 0 and 2, with the lowest accuracy among the five grades, indicating that the distinction between grade 1 and grades 0 and 2 is the least apparent. (3) The prediction differences among the different models were not significant.

F. Credibility Analysis

In clinical diagnosis, the main issue with artificial intelligence models is not accuracy but the credibility of their decisions. To verify whether the decision basis of our model conforms to clinical cognition, we display the regions to which our model pays the most attention during the decision-making process. We calculated the importance of each node in the graph and mapped it to the corresponding positions in the original X-ray image to obtain an heat map.

Fig 7(a) shows the heat maps for KL grade classification from 0 to 4 (from left to right) using our proposed method. The color of the regions, from blue to red, indicates that the region is becoming increasingly important in the decision-making process of the model. For correctly predicted examples, as judged by clinicians, the focus areas of the model's decision-making process are generally located in arthritis-affected regions, indicating that the model's decisions are credible. We also generated heatmaps of the original models(CN-Ti, RN50, and VGG16). Specifically, we calculated the average value of the model's last layer feature map across channels and then mapped it back to the original X-Ray image to obtain the heatmap. As shown in Fig 7(b), we can observe that the original ones can localize the lesion region but with a relatively wide range. We can note that our proposed demonstrates better understandability. In Fig 7(c), we visualize an example of incorrect decision-making by HCGN_CN-Ti_384, where the focus regions of its decision deviate from the arthritis-affected region, and the model pays attention not only to the knee joint but also to the screw. This indicates that these predictions lack credibility. In practice, such erroneous decisions can be identified easily by doctors.

In summary, our method transforms image classification in the feature space into graph classification, in which each node in the graph has a corresponding relationship with the feature in the original image.

G. Comparison with Doctors

We invited an orthopedic clinician (Doctor1), a radiologist (Doctor2), and a medical imaging graduate student (Doctor3) to diagnose all 1788 images in the test set independently. The annotations for this dataset were derived from doctors'

diagnoses based on ***X-ray images combined with clinical symptoms***. In our study, we could not obtain clinical diagnoses corresponding to these X-rays, and the invited participants' diagnoses were made without any clinical symptom data. The confusion matrices for the diagnoses from the three participants are shown in Fig 8. Although there are significant differences and low accuracy among the different doctors' diagnoses, most errors occur in the diagnosis of adjacent KL grades, which causes relatively minor harm in clinical practice. TABLE VI presents the statistical indicators used in this study.We observed that owing to the lack of patients' clinical symptoms and the large workload completed in a short time, the various indicators of the independent diagnoses of the three invited participants were far lower than those of the AI prediction model. In summary, the proposed end-to-end model for predicting KOA severity achieved satisfactory results.

TABLE VI
STATISTICAL INDICATORS OF THE DIAGNOSES MADE BY THE THREE INVITED DOCTORS ON THE OAI TEST DATASET.

	Accuracy	Recall	Precision	F1-score
Doctor1 (orthopedic clinician)	39.88	53.35	50.75	52.02
Doctor2 (radiologist)	44.13	49.10	48.75	48.92
Doctor3 (graduate student)	47.32	49.96	49.16	49.56
Ours(HCGN_CN-Ti_384)	**73.37**	**72.60**	**73.92**	**72.79**

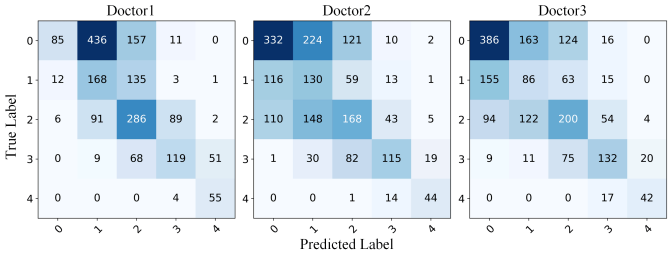

Fig. 8. Confusion matrices of the diagnoses made by the three invited doctors on the OAI test dataset.

V. CONCLUSION

KOA is a widespread joint disease that significantly affects the normal lives of patients. Early prediction can substantially reduce the incidence of KOA and improve treatment outcomes. Typically, doctors assess the severity of KOA by observing knee X-ray images and clinical symptoms, a process that relies heavily on the doctors' subjective experience and that varies among doctors. In this study, we propose an end-to-end reliable prediction model for KOA severity based on a feature map representation. Experiments show that compared to the original model, our proposed method demonstrates significant improvements in multiple predictive performance metrics and exhibits good decision understandability.However, in this paper, we merely explored the role of the graph representation module in ConvNets, rather than its role in Transformers.

VI. DATA AVAILABILITY

Data used in this article were obtained from the Osteoarthritis Initiative (OAI) database, which is available for public access at http://www.oai.ucsf.edu/. datasets used are baseline knee images with 0.C.2 and 0.E.1.

VII. CODE AVAILABILITY

The code used for preprocessing the data, training the YOLOv7 and HCGN model, and the parameters of the HCGN model have been made publicly available on Github at the following link : https://github.com/ddw2AIGROUP2CQUPT/HCGN.

REFERENCES

[1] K. Zhou, Y.-J. Li, E. J. Soderblom, A. Reed, V. Jain, S. Sun, M. A. Moseley, and V. B. Kraus, "A "best-in-class" systemic biomarker predictor of clinically relevant knee osteoarthritis structural and pain progression," *Science Advances*, vol. 9, no. 4, p. eabq5095, 2023.

[2] M. Cross, E. Smith, D. Hoy, S. Nolte, I. Ackerman, M. Fransen, L. Bridgett, S. Williams, F. Guillemin, C. L. Hill *et al.*, "The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study," *Annals of the rheumatic diseases*, vol. 73, no. 7, pp. 1323–1330, 2014.

[3] E. M. Roos and N. K. Arden, "Strategies for the prevention of knee osteoarthritis," *Nature Reviews Rheumatology*, vol. 12, no. 2, pp. 92–101, 2016.

[4] N. Bayramoglu, M. T. Nieminen, and S. Saarakkala, "A lightweight cnn and joint shape-joint space () descriptor for radiological osteoarthritis detection," in *Medical Image Understanding and Analysis: 24th Annual Conference, MIUA 2020, Oxford, UK, July 15-17, 2020, Proceedings*. Springer, 2020, pp. 331–345.

[5] J. Antony, K. McGuinness, N. E. O'Connor, and K. Moran, "Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 1195–1200.

[6] J. H. Kellgren and J. Lawrence, "Radiological assessment of osteo-arthrosis," *Annals of the rheumatic diseases*, vol. 16, no. 4, p. 494, 1957.

[7] C. Lindner, S. Thiagarajah, J. M. Wilkinson, G. A. Wallis, T. F. Cootes, arcOGEN Consortium *et al.*, "Fully automatic segmentation of the proximal femur using random forest regression voting," *IEEE transactions on medical imaging*, vol. 32, no. 8, pp. 1462–1472, 2013.

[8] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv preprint arXiv:2207.02696*, 2022.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[11] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*. Springer, 2014, pp. 818–833.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[15] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. pmlr, 2015, pp. 448–456.

[16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[17] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[20] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[22] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE transactions on neural networks*, vol. 20, no. 1, pp. 61–80, 2008.

[23] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[24] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

[25] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[26] K. Han, Y. Wang, J. Guo, Y. Tang, and E. Wu, "Vision gnn: An image is worth graph of nodes," *arXiv preprint arXiv:2206.00272*, 2022.

[27] R. J. Chen, M. Y. Lu, M. Shaban, C. Chen, T. Y. Chen, D. F. Williamson, and F. Mahmood, "Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*. Springer, 2021, pp. 339–349.

[28] Y. Zheng, R. H. Gindra, E. J. Green, E. J. Burks, M. Betke, J. E. Beane, and V. B. Kolachalama, "A graph-transformer for whole slide image classification," *IEEE transactions on medical imaging*, vol. 41, no. 11, pp. 3003–3015, 2022.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[30] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[31] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, "Single-shot refinement neural network for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4203–4212.

[32] T. Woloszynski, P. Podsiadlo, G. Stachowiak, and M. Kurzynski, "A signature dissimilarity measure for trabecular bone texture in knee radiographs," *Medical physics*, vol. 37, no. 5, pp. 2030–2042, 2010.

[33] A. C. Marijnissen, K. L. Vincken, P. A. Vos, D. Saris, M. Viergever, J. Bijlsma, L. Bartels, and F. Lafeber, "Knee images digital analysis (kida): a novel method to quantify individual radiographic features of knee osteoarthritis in detail," *Osteoarthritis and cartilage*, vol. 16, no. 2, pp. 234–243, 2008.

[34] L. Shamir, S. M. Ling, W. W. Scott, A. Bos, N. Orlov, T. J. Macura, D. M. Eckley, L. Ferrucci, and I. G. Goldberg, "Knee x-ray image analysis method for automated detection of osteoarthritis," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, pp. 407–415, 2008.

[35] J. Antony, K. McGuinness, K. Moran, and N. E. O'Connor, "Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks," in *Machine Learning and Data Mining in Pattern Recognition: 13th International Conference, MLDM 2017, New York, NY, USA, July 15-20, 2017, Proceedings 13*. Springer, 2017, pp. 376–390.

[36] B. Norman, V. Pedoia, A. Noworolski, T. M. Link, and S. Majumdar, "Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs," *Journal of digital imaging*, vol. 32, no. 3, pp. 471–477, 2019.

[37] P. Chen, L. Gao, X. Shi, K. Allen, and L. Yang, "Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 84–92, 2019.

[38] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[39] A. Swiecicki, N. Li, J. O'Donnell, N. Said, J. Yang, R. C. Mather, W. A. Jiranek, and M. A. Mazurowski, "Deep learning-based algorithm for assessment of knee osteoarthritis severity in radiographs matches performance of radiologists," *Computers in biology and medicine*, vol. 133, p. 104334, 2021.

[40] J. Yang, Q. Ji, M. Ni, G. Zhang, and Y. Wang, "Automatic assessment of knee osteoarthritis severity in portable devices based on deep learning," *Journal of Orthopaedic Surgery and Research*, vol. 17, no. 1, pp. 1–8, 2022.

[41] S. Suresha, L. Kidziński, E. Halilaj, G. Gold, and S. Delp, "Automated staging of knee osteoarthritis severity using deep neural networks," *Osteoarthritis and Cartilage*, vol. 26, p. S441, 2018.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[43] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala, "Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach," *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.

[44] A. Brahim, R. Riad, and R. Jennane, "Knee osteoarthritis detection using power spectral density: Data from the osteoarthritis initiative," in *Computer Analysis of Images and Patterns: 18th International Conference, CAIP 2019, Salerno, Italy, September 3–5, 2019, Proceedings, Part II 18*. Springer, 2019, pp. 480–487.

[45] B. Liu, J. Luo, and H. Huang, "Toward automatic quantification of knee osteoarthritis severity using improved faster r-cnn," *International journal of computer assisted radiology and surgery*, vol. 15, pp. 457–466, 2020.

[46] K. A. Thomas, Ł. Kidziński, E. Halilaj, S. L. Fleming, G. R. Venkataraman, E. H. Oei, G. E. Gold, and S. L. Delp, "Automated classification of radiographic knee osteoarthritis severity using deep neural networks," *Radiology: Artificial Intelligence*, vol. 2, no. 2, p. e190065, 2020.

[47] Y. Wang, X. Wang, T. Gao, L. Du, and W. Liu, "An automatic knee osteoarthritis diagnosis method based on deep learning: data from the osteoarthritis initiative," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–10, 2021.

[48] Y. Feng, J. Liu, H. Zhang, and D. Qiu, "Automated grading of knee osteoarthritis x-ray images based on attention mechanism," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021, pp. 1927–1932.

[49] S. M. Ahmed and R. J. Mstafa, "Identifying severity grading of knee osteoarthritis from x-ray images using an efficient mixture of deep learning and machine learning models," *Diagnostics*, vol. 12, no. 12, p. 2939, 2022.

[50] S. S. Abdullah and M. P. Rajasekaran, "Automatic detection and classification of knee osteoarthritis using deep learning approach," *La radiologia medica*, vol. 127, no. 4, pp. 398–406, 2022.

[51] H. Gu, K. Li, R. J. Colglazier, J. Yang, M. Lebhar, J. O'Donnell, W. A. Jiranek, R. C. Mather, R. J. French, N. Said *et al.*, "Automated grading of radiographic knee osteoarthritis severity combined with joint space narrowing," *arXiv preprint arXiv:2203.08914*, 2022.

[52] B. C. Dharmani and K. Khatri, "Deep learning for knee osteoarthritis severity stage detection using x-ray images," in *2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2023, pp. 78–83.

[53] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[54] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, "Gated graph sequence neural networks," *arXiv preprint arXiv:1511.05493*, 2015.

[55] A. Tiulpin, J. Thevenot, E. Rahtu, and S. Saarakkala, "A novel method for automatic localization of joint area on knee plain radiographs," in

Image Analysis: 20th Scandinavian Conference, SCIA 2017, Tromsø, Norway, June 12–14, 2017, Proceedings, Part II 20. Springer, 2017, pp. 290–301.

[56] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[57] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 740–755.

[58] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[59] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," 2022.

[60] T. Abeywickrama, M. A. Cheema, and D. Taniar, "K-nearest neighbors on road networks: a journey in experimentation and in-memory implementation," *arXiv preprint arXiv:1601.01549*, 2016.

[61] A. Mikhaylichenko and Y. Demyanenko, "Automatic grading of knee osteoarthritis from plain radiographs using densely connected convolutional networks," in *Recent Trends in Analysis of Images, Social Networks and Texts*, W. M. P. van der Aalst, V. Batagelj, A. Buzmakov, D. I. Ignatov, A. Kalenkova, M. Khachay, O. Koltsova, A. Kutuzov, S. O. Kuznetsov, I. A. Lomazova, N. Loukachevitch, I. Makarov, A. Napoli, A. Panchenko, P. M. Pardalos, M. Pelillo, A. V. Savchenko, and E. Tutubalina, Eds. Cham: Springer International Publishing, 2021, pp. 149–161.

[62] W. Liu, T. Ge, L. Luo, H. Peng, X. Xu, Y. Chen, and Z. Zhuang, "A novel focal ordinal loss for assessment of knee osteoarthritis severity," *Neural Processing Letters*, vol. 54, no. 6, pp. 5199–5224, 2022.

[63] E. A. Alshareef, F. O. Ebrahim, Y. Lamami, M. B. Milad, M. S. Eswani, S. A. Bashir, S. A. Bshina, A. Jakdoum, A. Abourqeeqah, M. O. Elbasir *et al.*, "Knee osteoarthritis severity grading using vision transformer," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 6, pp. 8303–8313, 2022.

[64] R. K. Jain, P. K. Sharma, S. Gaj, A. Sur, and P. Ghosh, "Knee osteoarthritis severity prediction using an attentive multi-scale deep convolutional neural network," *Multimedia Tools and Applications*, vol. 83, no. 3, pp. 6925–6942, 2024.