



## OPEN ACCESS

## EDITED BY

Xiaodu Wang,  
University of Texas at San Antonio, United States

## REVIEWED BY

Łukasz A. Poniąkowski,  
Medical University of Warsaw, Poland  
Yupeng Niu,  
Sichuan Agricultural University, China

## \*CORRESPONDENCE

Ting Li,  
✉ 181141305@qq.com  
Yongjing Cheng,  
✉ chengyongjing3427@bjhmoh.cn

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 03 June 2024

ACCEPTED 12 December 2024

PUBLISHED 03 January 2025

## CITATION

Ren X, Hou L, Liu S, Wu P, Liang S, Fu H, Li C, Li T and Cheng Y (2025) OA-MEN: a fusion deep learning approach for enhanced accuracy in knee osteoarthritis detection and classification using X-Ray imaging.  
*Front. Bioeng. Biotechnol.* 12:1437188.  
doi: 10.3389/fbioe.2024.1437188

## COPYRIGHT

© 2025 Ren, Hou, Liu, Wu, Liang, Fu, Li, Li and Cheng. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# OA-MEN: a fusion deep learning approach for enhanced accuracy in knee osteoarthritis detection and classification using X-Ray imaging

Xiaolu Ren<sup>1,2†</sup>, Lingxuan Hou<sup>3†</sup>, Shan Liu<sup>1</sup>, Peng Wu<sup>4</sup>, Siming Liang<sup>4</sup>, Haitian Fu<sup>5</sup>, Chengquan Li<sup>5</sup>, Ting Li<sup>1\*</sup> and Yongjing Cheng<sup>6\*</sup>

<sup>1</sup>Department of Radiology, General Hospital of Ningxia Medical University, Yinchuan, China, <sup>2</sup>School of Health Sciences, Universiti Sains Malaysia, Kota Bharu, Kelantan, Malaysia, <sup>3</sup>College of Biomedical Engineering, Sichuan University, Chengdu, Sichuan, China, <sup>4</sup>Department of Orthopedics, General Hospital of Ningxia Medical University, Yinchuan, China, <sup>5</sup>School of Clinical Medicine, Tsinghua University, Beijing, China, <sup>6</sup>Department of Rheumatology and Immunology, Beijing Hospital, National Centre of Gerontology, Beijing, China

**Background:** Knee osteoarthritis (KOA) constitutes the prevailing manifestation of arthritis. Radiographs function as a common modality for primary screening; however, traditional X-ray evaluation of osteoarthritis confronts challenges such as reduced sensitivity, subjective interpretation, and heightened misdiagnosis rates. The objective of this investigation is to enhance the validation and optimization of accuracy and efficiency in KOA assessment by utilizing fusion deep learning techniques.

**Methods:** This study aims to develop a highly accurate and lightweight model for automatically predicting and classifying KOA through knee X-ray imaging. We propose a deep learning model named OA-MEN, which integrates a hybrid model combining ResNet and MobileNet feature extraction with multi-scale feature fusion. This approach ensures enhanced extraction of semantic information without losing the advantages of large feature maps provided by high image resolution in lower layers of the network. This effectively expands the model's receptive field and strengthens its understanding capability. Additionally, we conducted unseen-data tests and compared our model with widely used baseline models to highlight its superiority over conventional approaches.

**Results:** The OA-MEN model demonstrated exceptional performance in tests. In the unseen-data test, our model achieved an average accuracy (ACC) of 84.88% and an Area Under the Curve (AUC) of 89.11%, marking improvements over the best-performing baseline models. These results showcase its improved capability in predicting KOA from X-ray images, making it a promising tool for assisting radiologists in diagnosis and treatment selection in clinical settings.

**Conclusion:** Leveraging deep learning for osteoarthritis classification guarantees heightened efficiency and accuracy. The future goal is to seamlessly integrate deep learning and advanced computational techniques with the expertise of medical professionals.

#### KEYWORDS

knee osteoarthritis, deep learning, decision making, artificial intelligence, convolution natural network

## 1 Introduction

The rising global prevalence of osteoarthritis (OA), especially in the knee joint, poses a significant public health challenge (Latourte et al., 2020). OA is typically triggered by a combination of subtle risk factors present in daily life, including diet, hormonal levels, and genetic predispositions. The incidence of OA significantly increases with age (Jia et al., 2023). A recent Chinese meta-analysis involving 74,908 symptomatic patients revealed an overall Knee osteoarthritis (KOA) prevalence of 14.6% between 2012 and 2016, with higher rates among females (19.1%) compared to males (10.9%) (Li et al., 2020). KOA varies in its progression, with the majority of patients seeing gradual deterioration over decades, while some experience rapid decline. Moreover, the onset of KOA often coincides with other comorbidities, exacerbating its impact and ultimately leading to the necessity for knee replacement surgery in some cases (Kotela et al., 2019; Wojdasiewicz et al., 2020; Wu et al., 2022). Early and accurate diagnosis is vital for effective management and timely intervention in KOA cases (Wojdasiewicz et al., 2020).

X-ray imaging is indispensable for the diagnosis of KOA, providing a non-invasive and cost-effective means to assess joint damage and monitor disease progression (Hirvasniemi et al., 2019; Jansen et al., 2021). In clinical practice, physicians rely on the Kellgren-Lawrence (K-L) grading system (Kellgren and Lawrence, 1957), designed for visual inspection of X-ray images, to measure the severity of KOA. The K-L system categorizes KOA severity into five grades, ranging from grade 0 (normal) to grade 4 (severe). While K-L grading methods are valuable tools for clinicians, the clinical assessment of conditions like KOA faces inherent challenges marked by subjectivity, the lack of quantitative data, and the absence of standardization (Abdullah and Rajasekaran, 2022; Cueva et al., 2022). This subjectivity leads to varying interpretations among healthcare practitioners, resulting in inconsistent diagnoses and treatment recommendations. The absence of quantitative data, combined with the lack of standardization in assessment techniques and protocols across healthcare providers, exacerbates these disparities, making it difficult to effectively monitor changes over time or evaluate the effectiveness of different treatment approaches. Consequently, the development of a fully automatic and efficient auxiliary classification method becomes imperative.

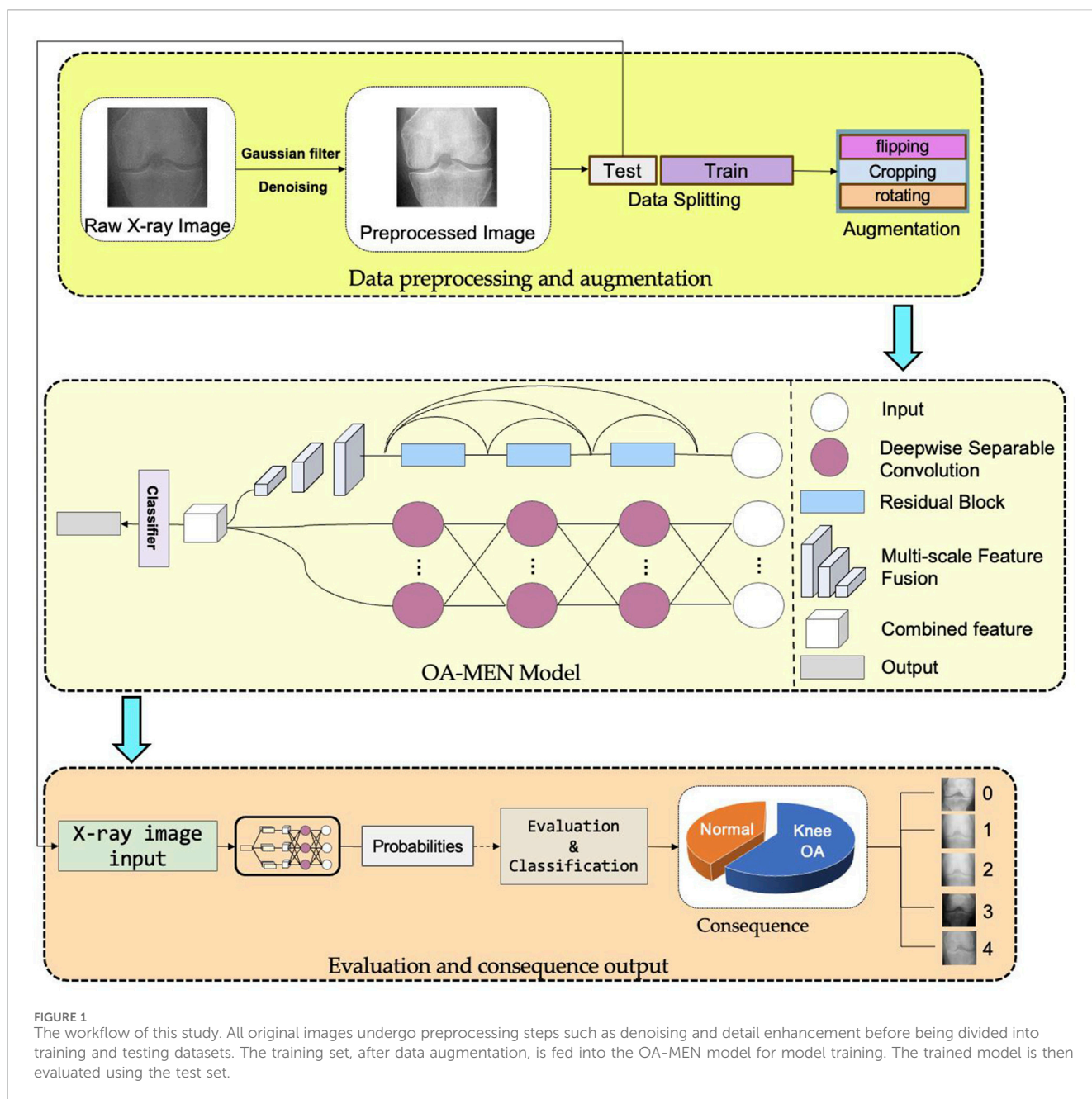
The rise of artificial intelligence technology has prompted extensive research into the application of deep learning in arthritis diagnosis. Deep learning, particularly convolutional neural networks (CNNs), excels at automatic feature extraction from intricate data, enabling the identification of subtle differences that may challenge human perception. This capability positions deep learning as a potent tool for tasks like image recognition and classification (Panwar et al., 2020). Deep learning

has proven effective in various knee-related applications, including the diagnosis of bone tumors (Do et al., 2021), identification of knee joint injuries (Kim et al., 2022; Qu et al., 2022; Wang et al., 2023), and the recognition and segmentation of anatomical structures (Gan et al., 2021; Quinsten et al., 2022). Concurrently, numerous studies concentrate on employing deep learning for knee grading, showcasing diverse techniques for automated KOA severity grading from X-ray images. Chen et al. (2019) utilized YOLO2 and fine-tuned VGG-19, achieving 69.7% accuracy with a Mean Absolute Error of 0.344. Thomas et al. (2020) developed a DenseNet with an accuracy of 0.71 and an F1 score of 0.70 for KOA grading. Wang et al. (2021) integrated YOLO with a visual transformer, achieving 69.18% accuracy in diagnosis. Cueva et al. (2022) present a CADx model employing Deep Siamese CNNs and fine-tuned ResNet-34 for simultaneous detection of OA lesions in both knees (KL scale), achieving 61% average multi-class accuracy. Mohammed et al. (2023) attained peak classification accuracies of 69%, 83%, and 89%, respectively, utilizing the ResNet101 deep neural network (DNN) model. However, the diagnosis of KOA primarily relies on texture features, and general models exhibit limited capability in recognizing these features, resulting in lower accuracy rates that fail to meet clinical application standards. Hence, there is a pressing need to develop a highly accurate model for the automatic diagnosis and grading of KOA. Such a model would assist radiologists in making more accurate diagnoses, reduce the rate of misdiagnosis, and consequently minimize patient suffering.

In this study, we developed a deep learning model named OAMEN, which is a fusion model obtained by integrating multi-scale feature fusion with parallel ResNet and MobileNet architectures. The multi-scale feature fusion within the ResNet structure effectively increases the network depth, combining the high-level semantic expression capabilities of the upper network layers with the geometric information representation of the lower layers. This integration expands the model's receptive field and utilizes residual blocks to prevent gradient vanishing, thereby avoiding ineffective training. MobileNet enhances the model's comprehension ability by extracting more effective features without significantly increasing model complexity. The final prediction and output are conducted through a fully connected layer. Tests with unseen data and comparisons with baseline models demonstrate this method's effectiveness in enhancing the model's performance and understanding ability for multi-classification of osteoarthritis.

## 2 Materials and methods

The workflow of this study is shown in Figure 1.



**FIGURE 1** The workflow of this study. All original images undergo preprocessing steps such as denoising and detail enhancement before being divided into training and testing datasets. The training set, after data augmentation, is fed into the OA-MEN model for model training. The trained model is then evaluated using the test set.

**TABLE 1** The knee K-L grading system.

Grade	Description
0	Healthy knee image
1 (Doubtful)	Doubtful joint narrowing with possible osteophytic lipping
2 (Minimal)	Definite presence of osteophytes and possible joint space narrowing
3 (Moderate)	Multiple osteophytes, definite joint space narrowing, with mild sclerosis
4 (Severe)	Large osteophytes, significant joint narrowing, and severe sclerosis

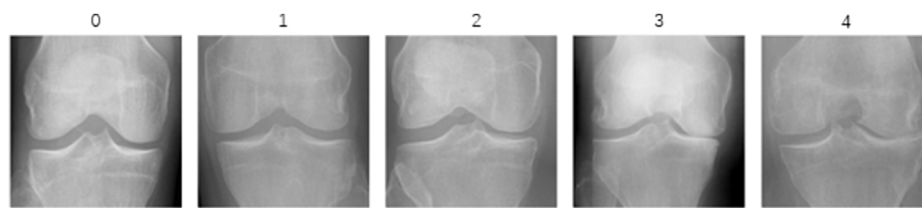


FIGURE 2

The diagram of KOA Severity Grading Dataset. The image demonstrates increasing severity from left to right, starting with a normal joint in Grade 0 and advancing through slight changes in Grade 1, evident osteophytes in Grade 2, significant joint space narrowing and sclerosis in Grade 3, to severe joint space loss and bone deformity in Grade 4.

## 2.1 Database

In this study, we utilized the KOA Severity Grading Dataset from Kaggle (Chen, 2018) as our training and testing data source. This dataset comprises 9786 high-quality knee X-ray images, each labeled by clinical experts according to the knee Kellgren-Lawrence (KL) grading standards (Kellgren and Lawrence, 1957). The KL grading system is a widely used radiological classification system for assessing the severity of KOA. Initially proposed by Kellgren and Lawrence in 1957, it primarily focuses on X-ray features to evaluate the progression and severity of KOA. The KL grading system categorizes KOA into levels as outlined in Table 1. The diagram of KOA Severity Grading Dataset is shown in Figure 2, which illustrates the progression of knee osteoarthritis through the KL grading system. Grade 0 shows no signs of osteoarthritis with a normal joint appearance. Grade 1 reveals slight narrowing of the joint space and possible osteophytic lipping. Grade 2 displays definite osteophytes and possible narrowing of the joint space. Grade 3 shows moderate multiple osteophytes, definite narrowing of the joint space, some sclerosis, and possible deformity of bone contour. Grade 4 exhibits large osteophytes, marked narrowing of the joint space, severe sclerosis, and definite deformity of bone contour.

## 2.2 Data preprocessing

In the initial stages of the experiment, we conducted preprocessing operations on all training images. Firstly, we applied a Gaussian filter to the training set images to remove high-frequency noise, followed by edge detection using the Laplacian operator for edge feature extraction, which was then combined with the original images to enhance the representation of texture features. Additionally, histogram equalization was performed in the brightness space to augment the display of image brightness and details. Moreover, the original dataset was divided into training and testing sets in an 8:2 ratio for conducting unseen-data tests.

Given the presence of data imbalance in the original dataset and to enhance the model's ability to recognize anomalous inputs, we employed data augmentation techniques for the training set. For each category, we increased the number of images through cropping, flipping, and translation, equalizing the image quantity across categories to prevent model overfitting due to data imbalance.

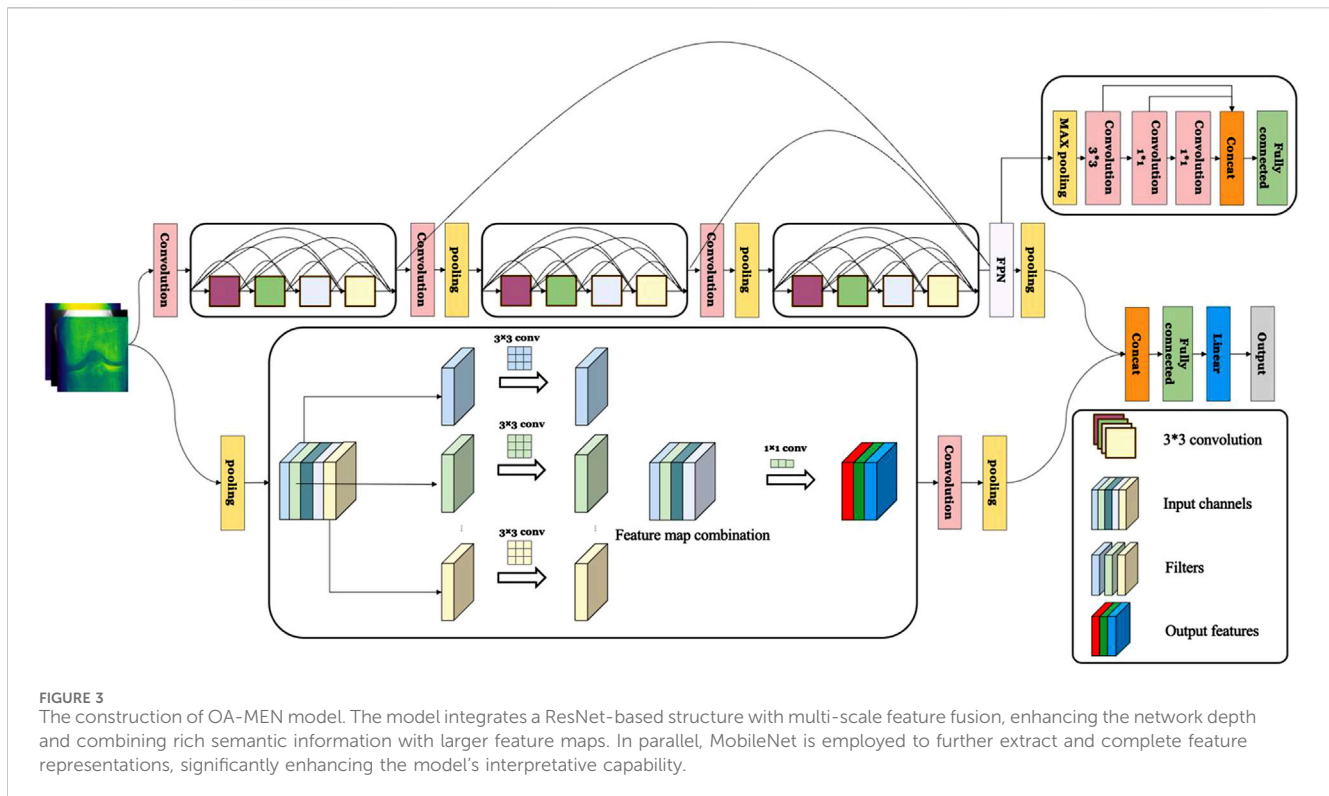
For fair comparison, all baseline models used datasets that had undergone similar data augmentation and preprocessing. Tests showed that such preprocessing successfully improved the model's predictive performance.

## 2.3 Model construction

### 2.3.1 OA-MEN model

In our preliminary experiments, it was observed that most deep learning models lacked the capability to effectively extract texture information in KOA recognition, thereby limiting their performance. To address this issue, we propose the OA-MEN model, a fusion of multi-scale feature-integrated ResNet and MobileNet. ResNet, introduced by He et al. (2016) in 2015, achieves this goal through the incorporation of "residual blocks." Within these blocks, the input is not only passed to the next layer but also directly through skip connections to deeper layers. This architecture allows the network to learn residuals between inputs and outputs, rather than learning the outputs directly, facilitating the construction of deeper models without encountering the problem of gradient vanishing. MobileNet, developed by Howard et al. (2017) from Google in 2017, is centered around depthwise separable convolutions, which decompose traditional convolution into depthwise and pointwise convolutions. Depthwise convolution applies individual filters to each input channel, while pointwise convolution, a  $1 \times 1$  convolution, combines the output of depthwise convolutions. This approach significantly reduces the model's complexity and computational demands while maintaining robust performance. As another branch of our model, MobileNet enhances the extraction of detailed features without substantially increasing computational load. To enhance the recognition of critical texture features in KOA, the model utilizes multi-scale feature fusion, capturing fine surface details (acquired in shallower network layers) and more abstract texture patterns (captured in deeper layers).

In summary, the ResNet module in the OA-MEN model allows for the construction of deeper learning architectures and avoids gradient vanishing, laying the groundwork for multi-scale feature fusion to extract information at various levels. The multi-scale feature fusion module effectively extracts both fine surface details and elusive texture features across different layers and scales. Furthermore, MobileNet, as a complementary branch, captures detailed features without significantly increasing computational



load. These modules work in synergy to not only extract surface details from Knee Joint X-ray images but also effectively capture texture features, enhancing the model's ability to diagnose and classify KOA. The model construction of OA-MEN is shown in Figure 3.

### 2.3.2 Comparison models

To objectively evaluate the performance of the OA-MEN model proposed in this study, we employed traditional deep learning models (Xception, MobileNet, ResNet, DenseNet, and NasNet) for comparison. These models have been widely applied by researchers and have demonstrated their superior performance and reliability, serving as benchmarks and future directions for our research. The Xception model, an evolution of the Inception (Szegedy et al., 2015) architecture, was introduced by Chollet in 2016. It utilizes depthwise separable convolutions within the Inception module, enabling the network to learn spatial hierarchies of features more efficiently while reducing the model's parameters and computational complexity. Xception has shown excellent performance in various image recognition tasks and has been increasingly applied in KOA recognition in recent years (Sindhu et al., 2022; Abd El-Ghany et al., 2023).

DenseNet, proposed by Huang et al., in 2016, features a densely connected pattern where each layer is directly connected to all preceding layers, effectively improving information and gradient flow and enhancing parameter efficiency. DenseNet has been extensively proven effective in various image classification and recognition tasks, including medical imaging and KOA recognition (Chaugule and Malemath, 2022; Alexopoulos et al., 2023).

NASNet, a neural network architecture designed using automated machine learning techniques, was introduced by

Google Brain researchers Zoph et al. (2018) in 2018. NASNet is a product of Neural Architecture Search (NAS), wherein machine learning autonomously discovers high-performance network architectures. Exhibiting outstanding performance in multiple standard image recognition benchmarks, NASNet represents a trend in discovering neural network architectures through automated methods and has also been applied in medical imaging and KOA recognition (Sultan et al., 2021; Yoon et al., 2023).

### 2.4 Model evaluation

To evaluate our OA-MEN model, we used several key performance metrics: Accuracy (ACC), Precision (PRE), Recall (REC), and Area Under the ROC Curve (AUC). These metrics are derived from the counts of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), which reflect correct or incorrect predictions by the model. The formulas for calculating ACC, PRE, REC, and F1 are as follows (Equations 1–4):

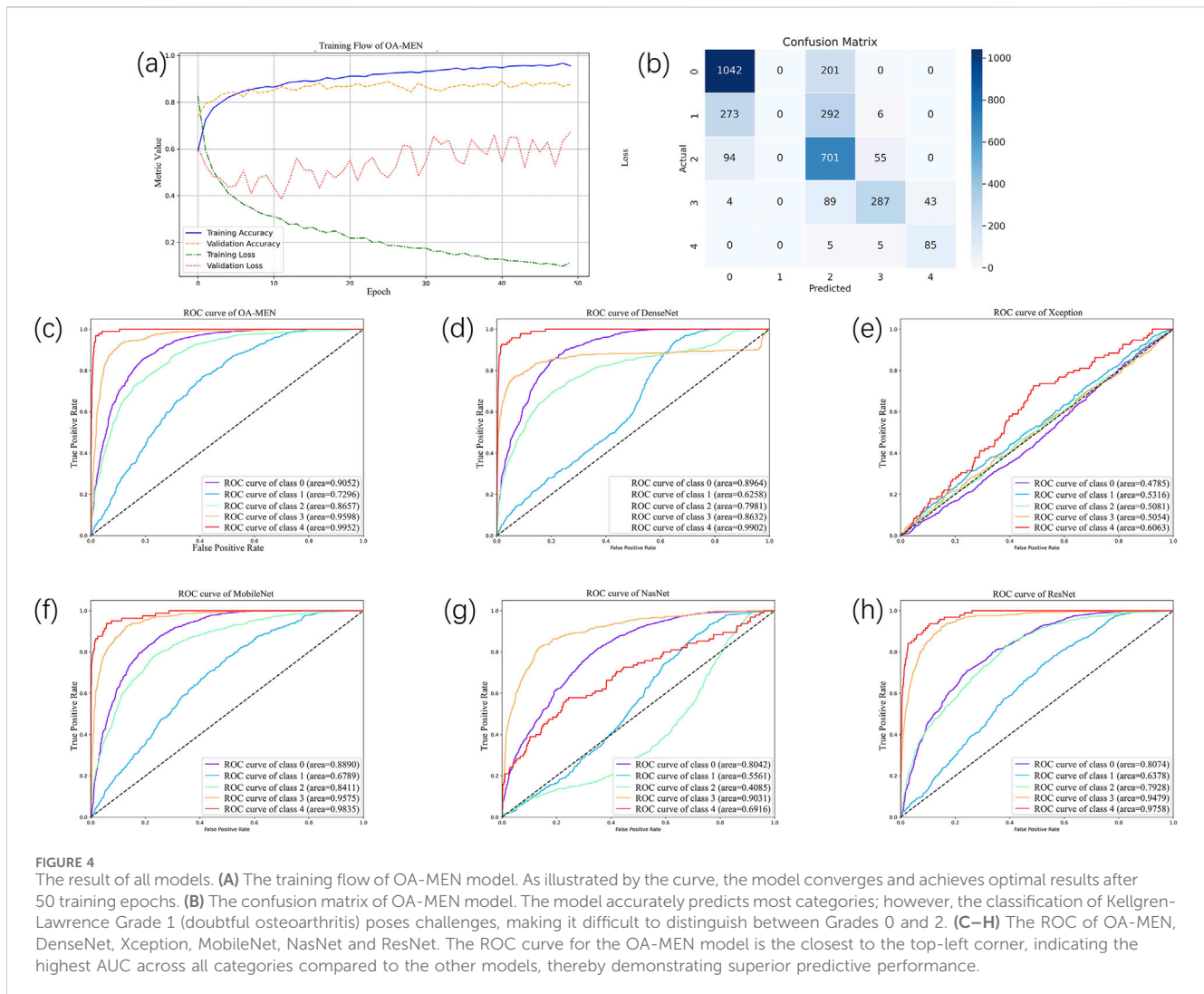
$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$PRE = \frac{TP}{TP + FP} \quad (2)$$

$$REC = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 \times PRE \times REC}{PRE + REC} \quad (4)$$

The ROC curve plots the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) at various thresholds, with



the AUC indicating the model's ability to distinguish between classes. An AUC close to 1 suggests high effectiveness, while an AUC near 0.5 implies random performance.

## 3 Results

### 3.1 Experimental setup

In this study, we implemented uniform parameter optimization across all models to ensure an objective evaluation. We employed the gradient threshold method for tuning all parameters. After extensive experimentation and testing, we chose the suitable number of epochs to make the models to fully converge without exhibiting signs of overfitting. The learning rate was set correctly to facilitate appropriate convergence speed and mitigated the risk of settling at local optima.

The experiments in this study were conducted on a system operating with Windows 11 Professional, utilizing Python 3.10.9 as the programming environment. In terms of software libraries, we employed Pytorch 2.0.1 + cu117, Scikit-learn, Sklearn 0.0.post1, and scipy 1.10.0, along with other

mathematical libraries, to facilitate the development of the model's architecture and the validation of its results. The hardware setup encompassed an Intel Core i7 10750H processor, featuring a base frequency of 2.6 GHz and a turbo frequency of up to 5 GHz, with 6 cores and 12 threads. Additionally, the system was equipped with an NVIDIA GeForce GTX 1080Ti graphics card, which has an 8 GB memory capacity and a 128-bit memory bus width.

### 3.2 Model results

To conduct a more comprehensive and objective evaluation of the predictive performance of the model proposed in this paper, we divided the dataset into training and testing sets in an 8:2 ratio. The testing set, serving as unseen data, was utilized to assess the model's predictive capabilities on datasets it had not previously encountered. Through experimental validation, we ultimately determined that training the model for 50 epochs with a learning rate of 0.00001 allowed for complete convergence without signs of overfitting. The iteration graph of the OA-MEN model is depicted in Figure 4A.

TABLE 2 The performance of all models.

	Mean ACC	Mean PRE	Mean REC	Mean AUC
OA-MEN	84.88	63.57	64.03	89.11
MobileNet	83.50	65.07	61.60	87.00
Xception	80.69	51.66	60.52	83.47
ResNet	83.86	60.13	62.71	83.23
DenseNet	83.62	59.09	60.68	87.56
NasNet	64.78	40.09	24.33	67.27

In the unseen-data test, the OA-MEN model exhibited outstanding performance. Specifically, for mild KOA (class 1), the model proposed in this study achieved a classification accuracy of 75.87%, the highest compared to all traditional models. In terms of average ACC and AUC, the OA-MEN model reached 84.88% and 89.11%, respectively, as shown in Table 2. Such high-performance metrics position the OA-MEN model as a promising tool for deployment in clinical settings, assisting radiologists in diagnosis. The confusion matrix and ROC curve of the OA-MEN model are illustrated in Figures 4B, C.

Among the comparison models, ResNet demonstrated the best performance, with an average ACC and AUC of 83.86% and 83.23%, respectively. However, there was a significant gap between ResNet and the OA-MEN model, particularly in the classification accuracy for mild KOA, where ResNet achieved only 73.88%. This discrepancy could potentially mislead radiologists in selecting appropriate treatment methods and impose unnecessary burdens on patients. The ROC curves for the comparison models are depicted in Figures 4D–H.

### 3.3 Model visualization and clinical interpretability

Most grade 0 Grad-CAM heatmaps do not exhibit a distinct gradient color heatmap compared to cases of higher-grade osteoarthritis, suggesting that the model's focus may not be primarily on specific areas associated with pathological features in this non-osteoarthritis category. Instead, accurate predictions of grade 0 may be attributed to a more extensive and precise assessment of normal anatomy, aligning with clinical expectations of a lack of significant pathological manifestations. The gradient color heatmaps detected in a limited number of level 0 images suggest subtle changes or features capturing the model's attention. This prompts the intriguing hypothesis that these images may not be genuinely normal and could potentially exhibit some pathological alterations. The distribution of most heatmaps for grades 1 to 4 aligns with the distribution of narrow joint spaces. However, the eye-catching appearance of the gradient color heatmaps of the subarticular area, combined with previous studies, leads to the proposal of a second interesting hypothesis that altered remodeling of subchondral trabecular bone may attract the attention of deep learning models. All of this necessitates future

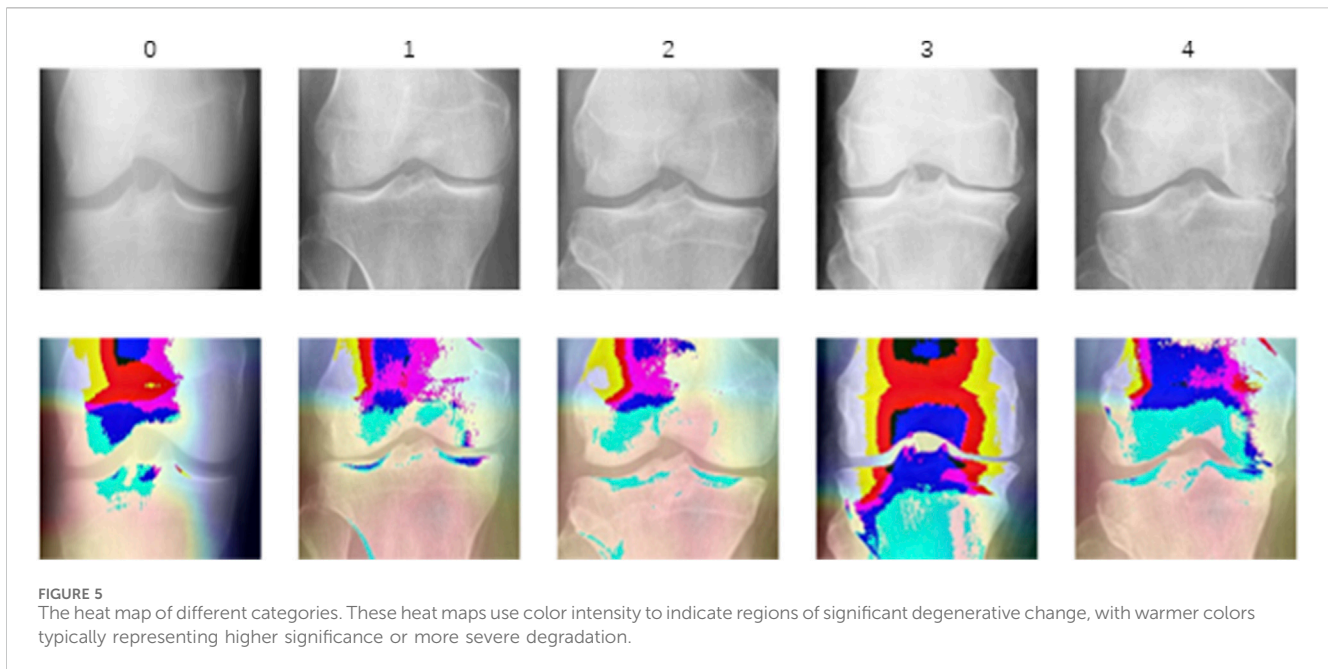
research to incorporate this potential hypothesis into the study design for further investigation and validation.

Visualization of heat maps enhances the interpretability of deep learning models but also presents certain challenges, especially compared to traditional imaging biomarkers. The heat map of five categories is shown in Figure 5. Heat maps capture static snapshots of decisions made at a given moment, providing a partial view of the model's attention and visualizing the features the model learns. Grade 0 displays minimal color variation, indicating an absence of significant osteoarthritic changes, which corroborates with the absence of OA. Grade 1 shows slight color changes around the joint space, suggesting the model's early detection of potential osteoarthritic alterations. Grade 2 is characterized by increased color intensity near the joint edges, highlighting early degenerative changes such as minor joint space narrowing and possible osteophyte formation. Grade 3 exhibits pronounced warm colors in the joint space and surrounding bone, signifying moderate reductions in joint space, osteophyte growth, and potential sclerosis. Grade 4 focuses on the bone itself, where the model pays particular attention to changes in bone morphology and density as well as the formation of osteophytes. However, their display may not be consistent with conventional imaging biomarkers that clinicians are familiar with. Deep learning models, including those that use heat maps, often have black-box properties, making the internal decision-making process not completely transparent. Additionally, these models may be inherently uncertain, and heat maps may not clearly convey the confidence or uncertainty of the model's predictions. These limitations may affect the interpretation process and results, necessitating increased adoption of these novel visualization techniques by radiologists and clinicians and further exploration of interpretability.

## 4 Discussion

Aiming to enhance the detection accuracy of KOA and reduce the misdiagnosis and underdiagnosis of mild cases, this paper introduces a novel deep learning model, named the OA-MEN model. It is a fusion model, combining ResNet with multi-scale feature integration and MobileNet in parallel. This model achieves accurate detection of KOA by synthesizing surface information and detailed texture features from feature maps of various network depths, thus increasing the detection accuracy for mild KOA and overall.

The OA-MEN model leverages the ResNet architecture to increase network depth without encountering issues like gradient vanishing. It integrates the semantic expressiveness of higher-level network structures with the geometric representational ability of lower-level structures through multi-scale feature fusion, thereby expanding the model's receptive field. In parallel, MobileNet extracts more effective features without significantly increasing the model's complexity, enhancing its understanding capability. The model culminates in a fully connected layer for prediction and output. Therefore, it effectively captures texture features and surface details of KOA, enhancing the model's performance and understanding of the multi-classification of osteoarthritis.



Despite positive outcomes in existing studies, this investigation in the field of precise and automatic osteoarthritis grading aims to enhance, validate, and extend the applicability of findings. Acknowledging positive outcomes from previous studies, the research maintains a rigorous approach, emphasizing the necessity for validation across diverse datasets and external test data. This work not only advances the current understanding of osteoarthritis grading but also sets the stage for future refinements, multi-center validations, and the development of a deployable medical tool, demonstrating an ongoing innovation in automated medical grading systems.

Despite the promising implications of our study, our grading system has limitations. Limited external test data from a specific cohort impacts generalizability. Future research should include diverse data from multiple centers. Furthermore, the KL grading system presents difficulties in differentiating between KL grade 1 (doubtful OA) and other grades due to its definition and the limited data available, an issue our model has not yet resolved. Future research should aim to increase the training samples and enhance the model to address this challenge (Zhao et al., 2024). Also, our system is not yet deployable for clinical use, it necessitates additional development. Moreover, our study lacks external validation, which is critical for ensuring the robustness and generalizability of our proposed model. It is essential to conduct further testing on more diverse datasets and to acquire clinical samples of KOA to demonstrate the superiority and clinical applicability of our model. Future research can refine our findings, contributing to a sophisticated and clinically applicable osteoarthritis grading system.

## 5 Conclusion

This study aims to precisely predict and grade KOA, introducing the OA-MEN model, a hybrid deep learning model that leverages ResNet and a multi-scale feature fusion strategy. This approach

enhances the model's ability to extract texture features while capturing rich information from high-resolution feature maps. Additionally, the integration of a parallel MobileNet allows the model to extract more effective features without substantially increasing its complexity, thus augmenting the model's overall understanding of KOA characteristics. Through unseen-data testing and comparison with traditional models, the OA-MEN model has demonstrated superior performance. Using the Knee KL grading system for classification and grading, the model achieved average ACC and AUC of 84.88% and 89.11%, respectively, fulfilling the objective of accurately predicting KOA. Further validation and optimization of this study will enhance the translation of these promising results into practical applications. Future research should encompass a more diverse array of external test data, employ more sophisticated methodologies, undergo further development and validation, and transition from a research setting to a clinically applicable tool. Looking forward, this study is expected to be deployed in clinical settings to assist physicians in diagnosis, reduce misdiagnosis rates, and minimize the patient harm caused by misdiagnoses and underdiagnoses.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: P. J. M. D. Chen, Knee osteoarthritis severity grading dataset, 1 (2018) 21-23.

## Ethics statement

The studies involving humans were approved by Ethics Committee of Medical Research, General Hospital of Ningxia Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. Written



informed consent for participation was not required from the participants or the participants'; legal guardians/next of kin because Research conducted on open data from public databases.

## Author contributions

XR: Conceptualization, Data curation, Investigation, Software, Visualization, Writing—original draft, Writing—review and editing. LH: Conceptualization, Investigation, Software, Visualization, Writing—review and editing. ShL: Data curation, Methodology, Supervision, Writing—review and editing. PW: Formal Analysis, Project administration, Validation, Writing—review and editing. SiL: Formal Analysis, Project administration, Validation, Writing—review and editing. HF: Resources, Software, Visualization, Writing—review and editing. CL: Resources, Software, Visualization, Writing—review and editing. TL: Investigation, Supervision, Validation, Writing—review and editing. YC: Investigation, Software, Supervision, Visualization, Writing—review and editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the fund of National High Level Hospital

## References

- Abd El-Ghany, S., Elmogy, M., and Abd El-Aziz, A. J. E. I. J. (2023). A fully automatic fine tuned deep learning model for knee osteoarthritis detection and progression analysis. *24*, 229–240. doi:10.1016/j.eij.2023.03.005
- Abdullah, S. S., and Rajasekaran, M. P. J. L. r.m. (2022). Automatic detection and classification of knee osteoarthritis using deep learning approach. *Radiol. Med.* *127* (4), 398–406. doi:10.1007/s11547-022-01476-7
- Alexopoulos, A., Hirvasniemi, J., Klein, S., Donkervoort, C., Oei, E., and Tümer, N. J. O. I. (2023). Early detection of knee osteoarthritis using deep learning on knee mri. *Osteoarthr. Imaging* *3*, 100112. doi:10.1016/j.ostima.2023.100112
- Chaugule, S. V., and Malemath, V. J. S. C. S. (2022). Knee osteoarthritis grading using DenseNet and radiographic images. *4*(1), 63. doi:10.1007/s42979-022-01468-4
- Chen, P., Gao, L., Shi, X., Allen, K., Yang, L. J. C. M. I., and Graphics (2019). Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. *Comput. Med. Imaging Graph.* *75*, 84–92. doi:10.1016/j.compmedimag.2019.06.002
- Chen, P. J. M. D. (2018). Knee osteoarthritis severity grading dataset. *1*, 21–23.
- Cueva, J. H., Castillo, D., Espinós-Morató, H., Durán, D., Díaz, P., and Lakshminarayanan, V. J. D. (2022). Detection and classification of knee osteoarthritis. *Diagn. (Basel)*. *12* (10), 2362. doi:10.3390/diagnostics12102362
- Do, N.-T., Jung, S.-T., and Kim, S.-H. J. C. C. R. (2021). Abstract PO-024: boosting up knee bone tumor detection from radiology and magnetic resonance imaging by using deep learning techniques. *Clin. Cancer Res.* *27* (5), PO-024. doi:10.1158/1557-3265.a021-po-024
- Gan, H.-S., Ramlee, M. H., Wahab, A. A., Lee, Y.-S., and Shimizu, A. J. A. I. R. (2021). From classical to deep learning: review on cartilage and bone segmentation techniques in knee osteoarthritis research. *Artif. Intell. Rev.* *54* (4), 2445–2494. doi:10.1007/s10462-020-09924-4
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Hirvasniemi, J., Niinimäki, J., Thevenot, J., and Saarakkala, S. J. A. o.b.e. (2019). Bone density and texture from minimally post-processed knee radiographs in subjects with knee osteoarthritis. *Ann. Biomed. Eng.* *47*, 1181–1190. doi:10.1007/s10439-019-02227-y

Clinical Research Funding (BJ-2023-089) and in part by the Ningxia Medical University 2024 University-Level Key Scientific Research Project (Project No. XZ2024034).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2024.1437188/full#supplementary-material>

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: efficient convolutional neural networks for mobile vision applications.

Jansen, M. P., Welsing, P. M., Vincken, K., Mastbergen, S. J. O., and Cartilage (2021). Performance of knee image digital analysis of radiographs of patients with end-stage knee osteoarthritis. *Osteoarthr. Cartil.* *29* (11), 1530–1539. doi:10.1016/j.joca.2021.07.013

Jia, S., Yu, Z., and Bai, L. (2023). Exerkines and osteoarthritis. *Front. Physiology* *14*, 1302769. doi:10.3389/fphys.2023.1302769

Kellgren, J. H., and Lawrence, J. J. A. o.t.r.d. (1957). Radiological assessment of osteoarthrosis. *Ann. Rheum. Dis.* *16* (4), 494–502. doi:10.1136/ard.16.4.494

Kim, D. H., Chai, J. W., Kang, J. H., Lee, J. H., Kim, H. J., Seo, J., et al. (2022). Ensemble deep learning model for predicting anterior cruciate ligament tear from lateral knee radiograph. *Skelet. Radiol.* *51* (12), 2269–2279. doi:10.1007/s00256-022-04081-x

Kotela, A., Wojdasiewicz, P., Łęgosz, P., Sarzyńska, S., Dreła, K., Pulik, Ł., et al. (2019). Increased serum levels of progranulin (PGRN) in patients with haemophilic arthropathy. *Clin. Exp. Pharmacol. Physiology* *46* (4), 373–379. doi:10.1111/1440-1681.13054

Latourte, A., Kloppenburg, M., and Richette, P. J. N. R. R. (2020). Emerging pharmaceutical therapies for osteoarthritis. *Nat. Rev. Rheumatol.* *16* (12), 673–688. doi:10.1038/s41584-020-00518-6

Li, D., Li, S., Chen, Q., and Xie, X. J. F. i.m. (2020). The prevalence of symptomatic knee osteoarthritis in relation to age, sex, area, region, and body mass index in China: a systematic review and meta-analysis. *Front. Med. (Lausanne)*. *7*, 304. doi:10.3389/fmed.2020.00304

Mohammed, A. S., Hasanaath, A. A., Latif, G., and Bashar, A. J. D. (2023). Knee osteoarthritis detection and severity classification using residual neural networks on preprocessed X-ray images. *Diagn. (Basel)*. *13* (8), 1380. doi:10.3390/diagnostics13081380

Panwar, H., Gupta, P., Siddiqui, M. K., Morales-Menendez, R., Singh, V. J. C., and Solitons, F. (2020). Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. *Chaos Solit. Fractals* *138*, 109944. doi:10.1016/j.chaos.2020.109944

- Qu, C., Yang, H., Wang, C., Wang, C., Ying, M., Chen, Z., et al. (2022). A deep learning approach for anterior cruciate ligament rupture localization on knee MR images. *Front. Bioeng. Biotechnol.* 10, 1024527. doi:10.3389/fbioe.2022.1024527
- Quinsten, A. S., Umutlu, L., Forsting, M., Nassenstein, K., and Demircioğlu, A. J. S. R. (2022). Determining the anatomical site in knee radiographs using deep learning. *Sci. Rep.* 12 (1), 3995. doi:10.1038/s41598-022-08020-7
- Sindhu, N., Mishra, S., Gowrishankar, S., Anushka, S., Snehananda, H., and Veena, A. (2022). "Prediction of knee-replacement using deep-learning approach," in 2022 international Conference on edge Computing and applications (ICECAA): *ieee*, 1213–1218.
- Sultan, H., Owais, M., Park, C., Mahmood, T., Haider, A., and Park, K. R. J. J. o.P. M. (2021). Artificial intelligence-based recognition of different types of shoulder implants in X-ray scans based on dense residual ensemble-network for personalized medicine. *medicine* 11 (6), 482. doi:10.3390/jpm11060482
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9.
- Thomas, K. A., Kidziński, L., Halilaj, E., Fleming, S. L., Venkataraman, G. R., Oei, E. H., et al. (2020). Automated classification of radiographic knee osteoarthritis severity using deep neural networks. *neural Netw.* 2 (2), e190065. doi:10.1148/ryai.2020190065
- Wang, Y., Li, Y., Lin, G., Zhang, Q., Zhong, J., Zhang, Y., et al. (2023). Lower-extremity fatigue fracture detection and grading based on deep learning models of radiographs. *Eur. Radiol.* 33 (1), 555–565. doi:10.1007/s00330-022-08950-w
- Wang, Y., Wang, X., Gao, T., Du, L., and Liu, W. J. J. o.H. E. (2021). An automatic knee osteoarthritis diagnosis method based on deep learning: data from the osteoarthritis initiative. *J. Healthc. Eng.* 2021, 1–10. doi:10.1155/2021/5586529
- Wojdasiewicz, P., Poniatowski, L. A., Kotela, A., Skoda, M., Pyzlak, M., Stangret, A., et al. (2020). Comparative analysis of the occurrence and role of CX3CL1 (fractalkine) and its receptor CX3CR1 in hemophilic arthropathy and osteoarthritis. *J. Immunol. Res.* 2020 (1), 1–12. doi:10.1155/2020/2932696
- Wu, R., Ma, Y., Yang, Y., Li, M., Zheng, Q., and Fu, G. J. C. r. (2022). *A clinical model for predicting knee replacement in early-stage knee osteoarthritis: data from osteoarthritis initiative*, 1–12.
- Yoon, J. S., Yon, C.-J., Lee, D., Lee, J. J., Kang, C. H., Kang, S.-B., et al. (2023). Assessment of a novel deep learning-based software developed for automatic feature extraction and grading of radiographic knee osteoarthritis. *BMC Musculoskelet. Disord.* 24 (1), 869. doi:10.1186/s12891-023-06951-4
- Zhao, H., Ou, L., Zhang, Z., Zhang, L., Liu, K., and Kuang, J. (2024). The value of deep learning-based X-ray techniques in detecting and classifying KL grades of knee osteoarthritis: a systematic review and meta-analysis. *Eur. Radiol.* 35, 327–340. doi:10.1007/s00330-024-10928-9
- Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V. (2018). "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8697–8710.