International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Early Osteoarthritis Detection from Knee X-Rays

Harshith V[1], Mohan Gowda L[2], Namish Jain A[3], Swapnil Divate[4]

[1,2,3,4]Department of Computer Science and Engineering in Artificial Intelligence and Machine Learning
Vidyavardhaka College of Engineering Mysuru, India

ABSTRACT:
Millions of people worldwide suffer from osteoarthritis (OA), a degenerative joint disease for which early identification is essential to slowing progression and improving patient outcomes. Even though radiographic imaging—especially X-rays—remains the most affordable and widely available diagnostic method, doctors sometimes struggle to spot minor alterations in the early stages (Kellgren–Lawrence grades 0–1). *Recent advances in deep learning and artificial intelligence (AI) create new opportunities for automated early OA identification and grading on radiographs.* This paper examines cutting-edge methods, stressing both their advantages and disadvantages. *We propose a methodological overview for developing an automated system that integrates preprocessing, region-of-interest localization, and classification to improve sensitivity in early OA detection.* There includes discussion of difficulties such poor generalizability across datasets, class imbalance, and interpretability problems. *Finally, we outline future research initiatives, stressing the necessity of integrating radiomic biomarkers, developing clinically interpretable AI models, and obtaining larger multi-center datasets.*

KEYWORDS: Osteoarthritis, Knee radiograph, Early identification, Kellgren-Lawrence grades, Deep learning, Convolutional neural networks (CNN), Transfer learning, Radiomics, Medical image analysis, Explainable AI, Radiographic images, Machine learning, ROI, classification, and preprocessing.

I. INTRODUCTION:

We propose a methodological overview for developing an automated system that integrates preprocessing, region-of-interest localization, and classification to improve sensitivity in early OA detection. There is discussion of difficulties such as class imbalance, poor generalizability across datasets, and interpretability problems. We wrap up by summarizing possible directions for further research, emphasizing the importance of incorporating radiomic biomarkers, growing multi-center datasets, and creating AI models that are clinically interpretable. *Early detection of OA* is crucial for preventing future disease progression, optimizing treatment programs, and delaying surgeries such as complete knee arthroplasty. OA is typically identified using radiographic imaging, specifically conventional X-rays, employing the *Kellgren–Lawrence (KL) grading system*. However, the earliest radiographic grades, KL 0 and KL 1, show small structural alterations that are unlikely to be visible during a typical clinical examination.
Because of this, a large number of early OA cases go undetected, which restricts treatment choices before it's too late. The automation of radiography image processing has been greatly enhanced by recent advances in artificial intelligence (AI), especially artificial neural networks like deep learning. *According to studies, convolutional neural networks (CNNs) can diagnose and grade OA at a level comparable to musculoskeletal radiologists.* Although the performance of these CNNs in the detection and grading of moderate to severe OA is strong, *the identification of the early stages of OA remains limited due to*

overlapping radiographic features of OA, class imbalance in the training datasets, and lack of sensitivity to assess subtle textural features.

Results in, various further solutions must be created to assess early OA classification (e.g., KL grades 0–1) utilizing AI-based diagnostic frameworks on anteroposterior (AP) knee X-rays. These frameworks can enhance clinician assessments by ensuring accuracy, consistency, and interpretability, resulting in better treatments of early disease and reduced burdens on healthcare systems in the coming decades. *In this paper, we look at developments in automated knee OA identification using radiography and sensor imaging, as well as the problems that have arisen and future research paths that will help to inform clinically applicable solutions.*

II. LITERATURE REVIEW

[1] Kundu et al. (2020) introduced a deep learning-based approach aimed at detecting osteoarthritis at a stage when structural changes are not yet visible through standard radiographic techniques. Their work applied transport-based morphometry to analyze cartilage texture patterns derived from MRI scans, using data obtained from the Osteoarthritis Initiative. The proposed method demonstrated superior predictive performance compared to commonly used clinical risk indicators, showing strong potential for early diagnosis. A notable limitation, however, is that the model depends on MRI imaging, which is costlier and less accessible than traditional X-ray methods, reducing the feasibility of broad clinical deployment.

[2] Thomas et al. (2020) developed a large-scale deep learning system capable of automatically assigning Kellgren–Lawrence (KL) grades directly from knee radiographs. Trained on over 40,000 images from the Osteoarthritis Initiative, the model achieved an accuracy of 0.71 and an F1-score of 0.70, performance levels comparable to experienced musculoskeletal radiologists. Visualization through saliency maps confirmed that the network concentrated on clinically relevant indicators, including osteophytes and reductions in joint space. Although the model is robust and openly accessible via the SimTK platform, further evaluation across diverse clinical datasets is necessary to establish its generalizability.

[3] Lee et al. (2024) proposed an enhanced deep learning strategy for assessing the severity of knee osteoarthritis by introducing plug-in modules (PIMs) designed to strengthen feature extraction within existing models for KL grading. Their approach combined convolutional neural networks with transformer-based architectures, forming an ensemble of four PIM-integrated models. The system was trained on approximately 63,000 radiographs from the Osteoarthritis Initiative and evaluated using more than 17,000 images from the MOST dataset. It achieved an overall accuracy of 75.7%, showing particularly high agreement for severe cases (KL-4 at 96%) and lower performance for early-stage disease (KL-1 at 43%). These results highlight both the method's strong cross-dataset generalization and the inherent challenge of identifying early osteoarthritis.

[4] Pi et al. (2023) developed an ensemble-based deep learning approach for automated grading of knee osteoarthritis using X-ray images. Their method combined multiple convolutional neural networks trained with different input resolutions and merged their predictions through a mixed voting strategy. Using a dataset of 8,260 radiographs from the Osteoarthritis Initiative, the ensemble achieved an accuracy of 76.93% and an F1-score of 0.77, surpassing the performance of OsteoHRNet. Visualizations generated through Grad-CAM confirmed that the networks concentrated on key radiographic markers such as reduced joint space and the presence of osteophytes. Although the ensemble improved robustness overall, its performance remained limited for KL-1 cases, underscoring the continued challenge of reliably detecting early-stage osteoarthritis.

[5] Ren et al. (2025) introduced a hybrid deep learning framework for detecting and classifying knee osteoarthritis from X-ray images. Their approach combined elements of ResNet, MobileNet, and multi-scale feature aggregation to enhance feature representation. After applying preprocessing and augmentation to 9,786 radiographs sourced from Kaggle, the model achieved an accuracy of 84.88% and an AUC of 89.11%. Notably, its performance in identifying KL-1 cases surpassed existing baseline methods. Grad-CAM visualizations showed that the network consistently concentrated on clinically relevant cues such as narrowing of the joint space. Despite its strong performance, the approach is limited by the absence of external validation and potential dataset bias.
[6] Tiulpin and Saarakkala (2019) proposed an ensemble of deep convolutional neural networks—built using ResNet-50 enhanced with squeeze-excitation mechanisms and ResNeXt components—to automatically grade radiographic features related to knee osteoarthritis. Their system evaluated both Kellgren–Lawrence scores and OARSI feature grades. Trained on the Osteoarthritis Initiative dataset and tested on the MOST cohort, the method achieved strong reliability, including a Cohen's kappa of 0.82 for KL grading, up to 0.94 for OARSI joint space narrowing, and an AUC of 0.98 for distinguishing radiographic OA. Although the approach demonstrated impressive cross-dataset generalization, its ensemble design required substantial computational resources.
[7] Tiulpin et al. (2017) introduced a Siamese convolutional neural network designed to diagnose knee osteoarthritis directly from radiographs. The model was trained on 18,376 knee images from the MOST dataset and evaluated using 5,960 images from the Osteoarthritis Initiative. It achieved a quadratic kappa of 0.83, an accuracy of 66.7%, and an AUC of 0.93 for identifying radiographic OA at KL grade 2 or higher. When compared with a standard ResNet-34, the Siamese architecture showed superior capability in focusing on clinically relevant cues such as osteophyte formation and joint space narrowing. Although attention maps improved the interpretability of the predictions, the system continued to underperform for early-stage cases, particularly KL-1 and KL-2.
[8] Alshamrani et al. (2023) explored the use of transfer learning for automated prediction of knee osteoarthritis from X-ray images. Their study compared several pretrained architectures—including a basic CNN, VGG-16, and ResNet-50—using a dataset of 3,836 radiographs sourced from Kaggle. Among the models tested, VGG-16 achieved the highest test accuracy at 92.17%, while ResNet-50 offered faster computation with an accuracy of 90.63%. Grad-CAM visualizations confirmed that the networks consistently focused on clinically relevant features such as joint space narrowing and osteophyte presence. Although the findings support the effectiveness of transfer-learning-based approaches, the authors also pointed out potential risks related to dataset bias and overfitting.
[9] Ahmed and Mstafa (2022) introduced two hybrid deep learning–machine learning frameworks, named DHL-I and DHL-II, for assessing knee osteoarthritis severity from radiographic images. The second framework, which combined fine-tuned convolutional neural networks with support vector machines, delivered stronger performance—achieving around 90.8% accuracy for binary OA classification and 88% accuracy for four-level KL grading. DHL-II consistently surpassed both DHL-I and other benchmark models, although its effectiveness in identifying KL-1 cases remained limited. The study also noted challenges such as imbalanced data distribution and dependence on posteroanterior (PA) radiographs, which constrained overall generalizability.
[10] Tiulpin et al. (2021) presented a deep learning framework for automatically grading tibiofemoral osteoarthritis severity from knee radiographs. Their approach combined Faster R-CNN for accurately localizing the knee joint with a DenseNet-201 model for classification. Using a dataset of 4,334 images,

the system achieved 98.6% accuracy for anteroposterior (AP) views and 92.4% for lateral views. The region-of-interest detection module exceeded 98% IoU, indicating highly precise localization. Although the method reached state-of-the-art performance, its accuracy decreased for lateral images due to overlapping anatomical structures, and the overall training process was computationally intensive.

[11] A prospective research by Podsiadlo et al. (2013) assessed if trabecular bone texture (TBT) measured from standard weight-bearing knee radiographs can predict the likelihood of future joint replacement in individuals with osteoarthritis. Using baseline images from 123 patients, the researchers extracted fractal-based and variance–orientation texture (VOT) descriptors from the medial tibial subchondral region. During the six-year follow-up, 25% of the assessed knees eventually required replacement. Notably, lower medial fractal dimension values were strongly linked to an increased risk of surgery, independent of typical clinical factors such as Kellgren–Lawrence grade, pain severity, body mass index, or demographic variables. The study's strengths include its longitudinal design and its focus on a clinically meaningful outcome. However, the relatively small sample size and reliance on a single imaging modality limit broader applicability, indicating the need for larger, more diverse cohorts for validation.

[12] Almhdie-Imjabbar, Lespessailles, and Toumi (2021) extended trabecular bone texture (TBT)– based prognostic study using information from the Osteoarthritis Initiative to a far bigger cohort. Their study examined 4,382 knees by extracting 64 fractal texture descriptors from 16 defined subregions of the tibial bone and constructing predictive models that incorporated TBT features alongside KL grades and clinical variables. Over a nine-year follow-up period, models that combined TBT with KL scoring achieved an AUC of 0.92, outperforming models relying solely on KL grade, which reached an AUC of 0.86 ($P < 0.001$). The strengths of the work include its large longitudinal dataset and detailed region-specific texture evaluation. However, the analysis was limited to radiographic features without integration of MRI-based biochemical markers, and the findings require external validation across varied imaging protocols to ensure broad applicability.

[13] Kondal et al. (2020) introduced a two-stage deep learning framework for automated assessment of knee osteoarthritis severity from radiographs using the Kellgren–Lawrence scale. The system first localized the knee joint using an object detection model and then applied a regression-based CNN to predict a continuous KL score. When trained on images from the Osteoarthritis Initiative and evaluated on radiographs from an Indian clinical setting, the model showed poor cross-domain performance, with a mean absolute error of about 1.09. However, fine-tuning the model on locally collected images reduced the error substantially to 0.28, highlighting the significant impact of domain shift. The authors also demonstrated that using a regression objective leads to better calibration of severity scores compared to traditional categorical classification.

[14] Touahema et al. (2024) developed a clinician-oriented deep learning system, MedKnee, built on a fine-tuned Xception architecture for automated prediction of radiographic knee osteoarthritis. Trained on approximately 5,000 images from the Osteoarthritis Initiative, the model reached an internal test accuracy of 97.2% and achieved 90–95% agreement when evaluated on external datasets from India and Morocco, following standardized preprocessing and manually selected regions of interest. The software also includes a graphical user interface to facilitate clinical reporting. Key strengths of this work include its high reported accuracy, successful evaluation across different populations, and the availability of a practical interface. However, the system relies on manual ROI selection and fixed-flexion radiograph protocols, and its performance remains sensitive to preprocessing variations.

[15] Sekiya et al. (2023) examined the biological relevance of AI-generated Kellgren–Lawrence grades by comparing them with three-dimensional MRI–derived structural measures in individuals with early knee osteoarthritis. Using MRI data from 469 participants, the study assessed the relationship between AI-assigned KL grades from radiographs and MRI-based evaluations of medial meniscus extrusion (MME) and cartilage thickness across specific regions of the knee. Higher AI KL scores were consistently associated with greater MME and reduced cartilage thickness in medial compartments, supporting the notion that radiograph-based AI grading reflects underlying structural degeneration. A key strength of this work is its multimodal approach, linking X-ray–derived AI predictions with MRI morphometric analyses. However, the cross-sectional design limits causal interpretation, and the study did not consider structural changes in the lateral compartment.

[16]Tiulpin et al. (2020) presented an automated two-step framework for assessing the severity of knee osteoarthritis from radiographs. The system first employed YOLOv2 to accurately localize the knee joint and then applied an ensemble of convolutional neural networks—including ResNet, VGG, DenseNet, and Inception architectures—trained with an ordinal loss function to reflect the ordered nature of Kellgren–Lawrence grades. When evaluated on the Osteoarthritis Initiative dataset, the ensemble outperformed earlier methods, showing improved alignment with expert radiologist assessments and more reliable calibration across severity levels. The integration of precise localization with an ordinal learning strategy was a major strength, contributing to more clinically meaningful predictions. However, the approach remains limited by the inherent constraints of radiographic imaging, which cannot capture biochemical cartilage changes, and by the continued reliance on KL grading as an imperfect reference standard.

[17] Jiang et al. (2024) presented an extensive review of radiomics approaches applied to osteoarthritis, summarizing research that extracts high-dimensional quantitative imaging features—such as texture, shape, and intensity—from modalities including X-ray, CT, and MRI. These radiomic signatures have been used in machine learning models to support early disease detection, severity grading, and prediction of future disease progression. The authors report that combining radiomics with clinical information often leads to notable improvements in AUC values. However, they also highlight persistent challenges, including small and heterogeneous study cohorts, limited harmonization of imaging protocols across centers, and inconsistent feature-extraction pipelines. The strengths of this work lie in its broad and practical overview of the field and its clear recommendations for improving reproducibility through harmonization and multi-center research. Nonetheless, methodological variability across studies limits the ability to draw unified quantitative conclusions.

[18] Zhao et al. (2024) conducted a systematic review and meta-analysis evaluating the effectiveness of deep learning models for detecting and classifying knee osteoarthritis severity from X-ray images. Their analysis incorporated 19 studies comprising 62,158 radiographs. The pooled results showed strong sensitivity for identifying more advanced disease stages—84.8% for KL-3 and 90.3% for KL-4—while performance dropped notably for earlier grades, with sensitivities of 64.0% for KL-1 and 75.0% for KL-2. These findings indicate that deep learning methods are generally reliable for moderate to severe osteoarthritis but face challenges in early detection due to subtle radiographic features and inconsistencies in KL grading across studies. Strengths of the review include its large aggregated sample and grade-specific performance assessment. However, limitations such as publication bias, variability in preprocessing pipelines, and inconsistent labeling practices across the included studies affect the robustness of the conclusions.

[19] Cueva et al. (2022) developed a convolutional neural network approach for grading knee osteoarthritis

severity, emphasizing the importance of preprocessing and addressing class imbalance. Using an institutional dataset supplemented with augmentation techniques, their model achieved competitive performance across multiple KL categories. Grad-CAM visualizations were employed to improve interpretability by highlighting the regions influencing the network's decisions. The study demonstrated that accurate region-of-interest extraction and balanced sampling are essential for reliable classification. However, the work was limited by minimal external validation and reduced sensitivity when identifying KL-1 cases.

[20] Bayramoglu, Nieminen, and Saarakkala (2022) investigated whether texture patterns extracted from the patella on knee radiographs can serve as useful biomarkers for identifying patellofemoral osteoarthritis. Using trabecular texture descriptors from the Osteoarthritis Initiative dataset along with machine learning classifiers, the study demonstrated that patellar texture provides information complementary to traditional tibiofemoral KL grading and improves the detection of patellofemoral disease. A key strength of this work is its focus on an often overlooked compartment of the knee and the integration of biomarkers that capture different structural characteristics. However, variations in patellofemoral anatomy and uncertainty about how well these findings translate to routine clinical radiographs remain important limitations.

III. METHOD OVERVIEW

To identify early osteoarthritis (OA) in knee X-ray pictures, four stages are necessary: preprocessing, segmentation, feature extraction, and classification. Each of the four processes is critical in transforming raw radiographs to clinically meaningful imaging biomarkers, establishing reliability and consistency in identification at mildly early stages (KL 0 vs KL 1).

A. Preprocessing:

Preprocessing is used to increase the quality of X-ray pictures and make them uniform across data sets, *allowing joint structures and minor OA-related changes to be more easily identified.* Normalization, contrast enhancement, and noise reduction are common preprocessing techniques. Noise can be decreased using Gaussian smoothing, which uses average pixel values within local areas, but histogram equalization improves contrast, which is useful for detecting fine structures such as joint space constriction. *Normalization allows for standardized intensity ranges across several sources,* which is required for your AI/deep learning model to learn well from normalized data.

B. Segmentation:

The area of interest (ROI), which is usually the tibiofemoral joint space in the knee, can be isolated using segmentation after preprocessing is finished. Deep learning-based systems like U-Net and its variations have replaced more conventional methods like thresholding and edge detection in segmentation. U-Net's encoder-decoder design provides powerful localization and relatively efficient feature learning. The encoder compresses the image and identifies key features. The decoder reconstructs the finer spatial details. *Skip connections do an excellent job of keeping essential contextual boundary information,* ensuring that delicate structures such as early osteophytes or modest joint space narrowing do not disappear during preprocessing and segmentation.

C. Feature Extraction:

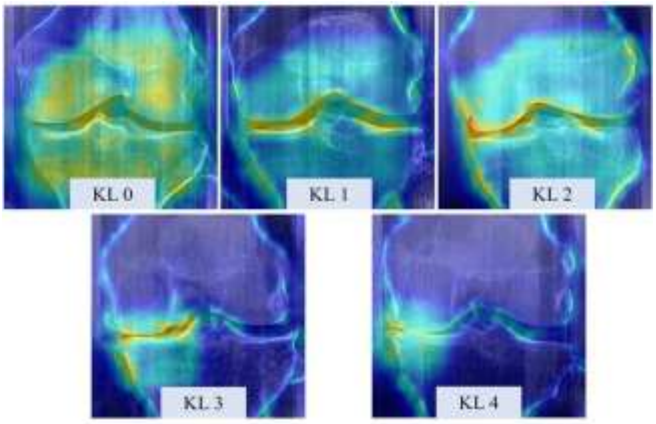
Once the Region of Interest has been defined, feature extraction investigates certain patterns that indicate the presence of early OA, such as texture, shape, and intensity features. *For eg., texture features may detect anomalies in cartilage appearance, shape features may capture crookedness in the joint space, and* *intensity patterns may detect the first density changes in subchondral bone.* Convolutional layers of convolutional neural networks are used in deep learning pipelines to learn these characteristics, which are then transformed into feature vectors that characterize the specific characteristics of the knee X-ray.

D. Classification:

The last step is the classification stage, where labels associated with the severity of OA are predicted by machine or deep learning models using the collected features. The ability of CNNs like ResNet-50, VGG-16, and EfficientNet to automatically learn distinguishing features between KL 0 (normal) and KL 1 (early OA) has demonstrated their great efficacy. To ascertain if the knee is healthy or has changed due to early OA, each of these models produces probability scores. Accuracy, precision, recall, and other metrics are used to measure categorization performance.

F1-score, and AUC.

Furthermore, *the interpretability of these models can be improved by employing Grad-CAM heatmaps,* which can identify and show the regions (e.g., osteophyte edges or shrinking joint spaces) that drive model prediction.



IV. RESEARCH GAPS

There is still a gap in the accurate classification of early illness (Kellgren-Lawrence grades 0–1), despite numerous deep learning and machine learning research initiatives aiming to automate knee osteoarthritis (OA) diagnosis and grading. There is a great deal of work that demonstrates adequate performance on moderate to severe OA classification (KL grades 2–4), but there are also a number of studies that consistently show lower performance on early OA due to subtle radiographic changes, class imbalance, and overall lower sensitivity to small structure differences.

Additional constraints to generalization include domain shift variability between datasets, reliance on cohort homogeneity (e.g., OAI), and the lack of multi-center external validation. Thus, *there is a need to develop frameworks specifically designed for early OA diagnosis using X-rays,* such as enhanced feature extraction, domain adaptation approaches, and explainable AI.

V. CONCLUSION

The early identification of knee osteoarthritis (OA) is critical for limiting disease progression, improving patient outcomes, and lowering long-term healthcare expenditures. Radiographic imaging, particularly X-rays, remains the most accessible diagnostic modality, but early-stage OA has modest alterations that doctors find difficult to detect reliably. *The combination of artificial intelligence (AI) and deep learning* *approaches, such as CNNs, transfer learning, and hybrid architectures, has shown great potential in automating the identification and assessment of early OA.*

Despite development, a number of obstacles remain: *Differentiating between Kellgren-Lawrence grades 0 and 1 is extremely difficult because of modest radiographic changes, resulting in inconsistent ground truth labels and decreased model sensitivity.* Models trained on homogeneous datasets frequently fail to generalize to other demographics, imaging centers, or ethnic groupings. *Class imbalance and computational needs further impede early OA identification and real-time implementation,* while *the interpretability of CNN-based models remains a hurdle to clinical trust.*

To address these problems, future efforts will include multi-center and cross-population validation, architectures customized for subtle biomarkers of early OA, and the integration of radiomics characteristics with deep learning. Explainable AI techniques beyond Grad-CAM, lightweight architectures, multimodal integration (clinical, demographic, and genetic data), and longitudinal imaging investigations can all improve performance and dependability.

Addressing these challenges would allow AI-powered tools to assist physicians in providing consistent, accurate, and timely assessments of early OA, thereby enhancing patient care and enabling proactive intervention techniques.

REFERENCES

1. Kundu, S., Ashinsky, B. G., Bouhrara, M., Dam, E. B., Demehri, S., Shifat-E-Rabbi, M., ... & Rohde, G. K. (2020). Enabling early detection of osteoarthritis from presymptomatic cartilage texture maps via transport-based learning. Proceedings of the National Academy of Sciences, 117(40), 24709-24719.
2. Thomas, K. A., Kidziński, Ł., Halilaj, E., Fleming, S. L., Venkataraman, G. R., Oei, E. H., ... & Delp, S. L. (2020). Automated classification of radiographic knee osteoarthritis severity using deep neural networks. Radiology: Artificial Intelligence, 2(2), e190065.
3. Lee, D. W., Song, D. S., Han, H. S., & Ro, D. H. (2024). Accurate, automated classification of radiographic knee osteoarthritis severity using a novel method of deep learning: Plug-in modules. Knee Surgery & Related Research, 36(1), 24.
4. Pi, S. W., Lee, B. D., Lee, M. S., & Lee, H. J. (2023). Ensemble deep-learning networks for automated osteoarthritis grading in knee X-ray images. Scientific Reports, 13(1), 22887.
5. Ren, X., Hou, L., Liu, S., Wu, P., Liang, S., Fu, H., ... & Cheng, Y. (2025). OA-MEN: a fusion deep learning approach for enhanced accuracy in knee osteoarthritis detection and classification using X-Ray imaging. Frontiers in Bioengineering and Biotechnology, 12, 1437188.
6. Tiulpin, A., & Saarakkala, S. (2020). Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks. Diagnostics, 10(11), 932. Tiulpin, A., Thevenot, J., Rahtu, E., Lehenkari, P., & Saarakkala, S. (2018). Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. Scientific reports.
7. Alshamrani, H. A., Rashid, M., Alshamrani, S. S., & Alshehri, A. H. (2023, April). Osteo-net: An automated system for predicting knee osteoarthritis from x-ray images using transfer-learning-based neural networks approach. In Healthcare (Vol. 11, No. 9, p. 1206). MDPI.
8. Podsiadlo, P., Cicuttini, F. M., Wolski, M., Stachowiak, G. W., & Wluka, A. E. (2014). 6-year prospective follow-up research found that individuals with osteoarthritis who had trabecular bone texture on plain radiography had a higher chance of needing a knee replacement. Osteoarthritis and

cartilage, 22(1), 71-75.

9. Almhdie-Imjabbar, A., Lespessailles, E., & Toumi, H. (2021). Using data from the Osteoarthritis Initiative cohort, trabecular bone texture analysis of conventional radiographs is used to predict the likelihood of total knee replacement. Osteoarthritis and Cartilage, 29, S65.
10. Kondal, S., Kulkarni, V., Gaikwad, A., Kharat, A., & Pant, A. (2022, January). Automatic grading of knee osteoarthritis on the Kellgren-Lawrence scale from radiographs using convolutional neural networks. In Advances in Deep Learning, Artificial Intelligence and Robotics: Proceedings of the 2nd International Conference on Deep Learning, Artificial Intelligence and Robotics,(ICDLAIR) 2020 (pp. 163-173). Cham: Springer International Publishing.
11. Touahema, S., Zaimi, I., Zrira, N., Ngote, M. N., Doulhousne, H., & Aouial, M. (2024). MedKnee: a new deep learning-based software for automated prediction of radiographic knee osteoarthritis. Diagnostics, 14(10), 993.
12. Sekiya, I., Katano, H., Guermazi, A., Miura, Y., Okanouchi, N., Tomita, M., ... & Ozeki, N. (2023). Association of AI-determined Kellgren–Lawrence grade with medial meniscus extrusion and cartilage thickness by AI-based 3D MRI analysis in early knee osteoarthritis. Scientific Reports, 13(1), 20093.
13. Tiulpin, A., *et al.* "Automatic severity assessment of knee osteoarthritis from X-rays using an ensemble of deep convolutional neural networks." *Scientific Reports*, 2020.
14. Jiang, T., Lau, S. H., Zhang, J., Chan, L. C., Wang, W., Chan, P. K., ... & Wen, C. (2024). Radiomics signature of osteoarthritis: Current status and perspective. *Journal of orthopaedic translation*, *45*, 100-106.
15. Zhao, H., Ou, L., Zhang, Z., Zhang, L., Liu, K., & Kuang, J. (2025). A comprehensive evaluation and meta-analysis of the usefulness of deep learning-based X-ray methods for identifying and categorizing KL grades of knee osteoarthritis. European Radiology, 35(1), 327-340.
16. Ahmed, S. M., & Mstafa, R. J. (2022). Identifying severity grading of knee osteoarthritis from x-ray images using an efficient mixture of deep learning and machine learning models. Diagnostics, 12(12), 2939.
17. Tiulpin, A., *et al.* "Automatic Tibiofemoral Osteoarthritis Severity Grading from Radiographs Using a Deep Learning-Based Framework." *Diagnostics*, 2021.
18. Kokkotis, C., Ntakolia, C., Moustakidis, S., Giakas, G., & Tsaopoulos, D. (2022). Explainable machine learning for knee osteoarthritis diagnosis based on a novel fuzzy feature selection methodology. Physical and Engineering Sciences in Medicine, 45(1), 219-229.
19. Goswami, A. D. (2023). Automatic classification of the severity of knee osteoarthritis using enhanced image sharpening and CNN. Applied Sciences, 13(3), 1658.