JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ROI-Focused KOA Detection Using MobileNetV2 and VGG16

Ms. Nainsi Jain[1], Mr. Rajneesh Pachouri[1] Mr. Anurag Jain[2],

Research Scholar, Assistant Professor
Department of Computer Science and Engineering
Adina Institute of Science & Technology, Sagar, India

Abstract: Knee osteoarthritis (KOA) is a progressive joint disorder that significantly affects mobility and quality of life, especially among the elderly. Early and accurate detection is vital for effective treatment planning and prevention of severe joint damage. This study proposes a Region of Interest (ROI)-focused deep learning approach for the automated detection and classification of KOA severity from knee X-ray images. The system employs two convolutional neural network models, MobileNetV2 and VGG16, chosen for their efficiency and high classification performance. A novel preprocessing technique based on pixel density is introduced to automatically extract the cartilage region from X-ray images, enhancing feature relevance and reducing background noise. The dataset consists of 1,650 high-quality grayscale X-ray images, manually annotated by medical experts using the Kellgren and Lawrence grading system. The MobileNetV2 model achieved a test accuracy of 96%, outperforming VGG16, which attained 92%. The system is implemented with a Flask-based web interface for real-time usability, offering a scalable and accessible solution for clinical deployment. Comprehensive evaluation metrics, including accuracy, precision, recall, and F1-score, confirm the system's effectiveness in automating KOA assessment, particularly in resource-constrained healthcare environments.

Index Terms **- Knee Osteoarthritis, Deep Learning, MobileNetV2, VGG16, X-ray Imaging, Region of Interest (ROI).**

I. INTRODUCTION

Knee osteoarthritis (KOA) is a degenerative joint disorder characterized by the breakdown of cartilage and underlying bone, most commonly affecting older adults and leading to pain, stiffness, and reduced mobility. It is one of the leading causes of disability worldwide, with a rising prevalence due to aging populations and sedentary lifestyles. Accurate and early diagnosis of KOA is critical for preventing further joint damage, improving patient outcomes, and reducing healthcare costs. Traditionally, KOA diagnosis relies on clinical evaluations and radiographic imaging, with the Kellgren and Lawrence (KL) grading system serving as the standard for assessing disease severity. However, manual interpretation of X-ray images is time-consuming, subjective, and prone to inter-observer variability, particularly in early-stage cases. These limitations have driven the need for automated diagnostic systems that can deliver fast, consistent, and accurate evaluations. Recent advancements in artificial intelligence, particularly deep learning, have shown significant promise in medical image analysis. Convolutional Neural Networks (CNNs) have demonstrated high performance in image classification tasks, making them suitable for KOA detection and grading from radiographic data. Despite this progress, many existing systems are complex, computationally intensive, and dependent on massive datasets or synthetic image augmentation, which can limit real-world deployment in clinical settings. This study introduces a Region of Interest (ROI)-focused deep learning system for KOA detection and classification using two lightweight and efficient CNN models: MobileNetV2 and VGG16. A novel preprocessing method based on pixel density is employed to automatically isolate the cartilage region from X-ray images, ensuring that the model focuses on clinically relevant features. The system is trained on a dataset of 1,650 high-quality, expert-annotated X-ray images, and integrated into a web-based interface developed with Flask, providing accessibility for real-time diagnostic use. By combining high-accuracy deep learning models with ROI-based image processing and a user-friendly interface, the proposed system aims to bridge the gap between AI research and practical healthcare solutions, especially in resource-constrained environments.

Unlike traditional Vision Transformers (ViTs), which require large datasets and high computational resources, Compact Convolutional Transformers (CCTs) are lightweight and suitable for medical imaging applications with limited annotated data. By further modifying the CCT architecture—such as tailoring convolutional stages to focus on the knee joint region (ROI), optimizing attention mechanisms, and fine-tuning hyperparameters—KOA-CCTNet improves grading performance while maintaining clinical relevance and computational efficiency.

Knee Osteoarthritis (KOA) is a degenerative joint disorder that affects millions of individuals worldwide, leading to chronic pain, reduced mobility, and a significant decline in quality of life. Accurate and timely assessment of KOA severity is essential for guiding clinical decisions and planning appropriate interventions. Traditionally, the grading of KOA relies on radiographic analysis interpreted by medical experts using systems such as the Kellgren-Lawrence (KL) grading scale. However, this manual approach is inherently subjective, time-consuming, and prone to inter-observer variability.In recent years, deep learning has emerged as a powerful tool for automating medical image analysis, offering improved consistency and efficiency. Convolutional Neural Networks (CNNs), in particular, have demonstrated strong performance in visual recognition tasks, including KOA detection. However, traditional CNNs often struggle with capturing global dependencies and contextual relationships in medical images, which are critical for subtle grade differentiation in KOA assessment.

Traditional KOA grading relies on the Kellgren-Lawrence (KL) classification system, which assesses X-ray images to assign grades from 0 (normal) to 4 (severe OA). However, this manual approach is inherently subjective and suffers from inter- and intra-observer variability Hunter & Bierma-Zeinstra, 2019. In recent years, deep learning-based methods, especially Convolutional Neural Networks (CNNs), have shown promise in automating KOA detection and classification. Yet, these models often struggle with generalization and fail to capture global contextual features critical in subtle grading distinctions.

To address these limitations, this research introduces KOA-CCTNet, a novel framework that integrates a Modified Compact Convolutional Transformer (CCT) model specifically tailored for KOA grading. The proposed approach leverages the localized feature extraction capabilities of CNNs and the global attention mechanisms of transformers to enhance the accuracy and robustness of KOA grade prediction. By incorporating architectural modifications optimized for knee joint radiographs, KOA-CCTNet seeks to achieve a balanced trade-off between computational efficiency and diagnostic precision.

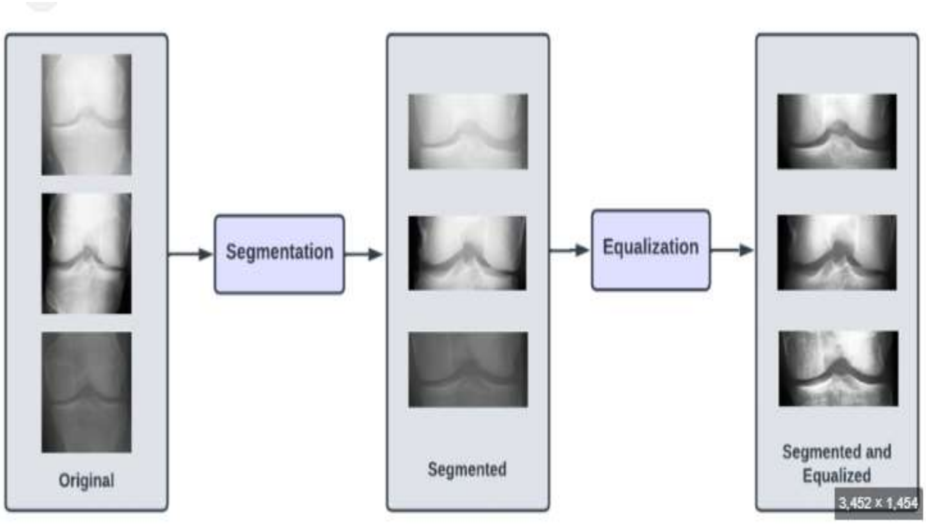

Figure 1 Preprocessing steps involved KOA detection

Scope of the Study

This study focuses on the design, development, and evaluation of KOA-CCTNet, a novel deep learning framework aimed at automating the severity grading of Knee Osteoarthritis (KOA) using radiographic images. The scope is defined by the following boundaries and deliverables:

1. Domain of Application

The study is confined to the field of medical image analysis, specifically targeting the classification of KOA severity based on X-ray images of the knee joint.

The framework is intended to support clinical diagnosis, not replace human experts.

2. Dataset Utilization

The study employs publicly available datasets such as the Osteoarthritis Initiative (OAI) dataset or similar, which include labeled knee radiographs annotated with Kellgren-Lawrence (KL) grades from 0 to 4.

Only anteroposterior (AP) view X-rays are considered for consistency in analysis.

3. Model Development

The proposed KOA-CCTNet framework is based on a Modified Compact Convolutional Transformer (CCT) architecture.

Model development includes:

Image preprocessing (resizing, normalization, augmentation)

Architectural design (convolutional tokenizer + transformer encoder)

Training, validation, and testing using stratified cross-validation

4. Performance Evaluation

The model is evaluated using key classification metrics including **accuracy, precision, recall, F1-score, and Cohen's Kappa**.

Comparison is made against baseline models such as **CNNs (e.g., ResNet, VGG)** and full **Vision Transformers (ViT)**.

5. Explain ability and Visualization

The study integrates explain ability techniques like **Grad-CAM and attention map visualization** to highlight image regions that influence model decisions.

This is to promote transparency and trust in AI-driven diagnosis.

6. Computational Efficiency

KOA-CCTNet is optimized for reduced model complexity, making it suitable for potential deployment in resource-constrained healthcare settings.

The model's performance is analyzed in terms of inference time, memory usage, and scalability.

7. Limitations Outside Scope

The study does not include MRI or CT imaging modalities—only X-ray-based KOA assessment is covered.

It does not perform treatment recommendation or disease progression prediction.

Clinical trials or real-time deployment in hospital settings are outside the scope but suggested for future work.

In summary, the scope of this research is limited to building and evaluating a robust, interpretable, and efficient transformer-based deep learning model for KOA grade classification. It is focused on X-ray-based KOA diagnosis with a view toward aiding radiologists through automated tools, while maintaining computational feasibility and model transparency.

II. LITERATURE SURVEY

Literature Review

Knee Osteoarthritis (KOA) is a highly prevalent degenerative disease affecting the aging population, characterized by the progressive degradation of articular cartilage and subchondral bone. Early diagnosis and accurate grading of KOA severity are critical for effective treatment and intervention. Traditionally, radiographic evaluation using the Kellgren-Lawrence (KL) grading scale remains the clinical standard, despite its limitations in subjectivity and inter-observer variability [2].

1.Traditional Assessment Methods and Their Limitations

Manual KOA grading relies heavily on expert interpretation of knee radiographs, focusing on joint space narrowing, osteophyte formation, and subchondral sclerosis [2]. While widely used, this method is prone to diagnostic inconsistencies due to its reliance on visual estimation. Several studies have noted considerable variability among radiologists, particularly in borderline KL grades (e.g., 2 vs. 3) [4].

2.Machine Learning in KOA Assessment

To reduce subjectivity, early machine learning (ML) approaches have been explored for automating KOA diagnosis. Algorithms such as Support Vector Machines (SVMs), Random Forests, and k-NN have shown promise when combined with handcrafted features from radiographs [5]. However, the need for manual feature extraction limited their scalability and adaptability to variations in patient data.

3.Deep Learning and Convolutional Neural Networks (CNNs)

The advent of deep learning, especially CNNs, revolutionized KOA detection by enabling automatic feature learning from raw image data. CNN-based models such as ResNet, VGG, and DenseNet have been successfully applied for KOA classification tasks [6] [7].

For example, Antony et al. (2017) proposed a deep learning model that used CNNs to predict KL grades directly from X-ray images, reporting accuracy improvements over classical ML approaches [8]. Similarly, Tiulpin et al. (2018) introduced a Siamese CNN model to account for both knees in a single prediction, further enhancing diagnostic accuracy [9].

Limitations: While effective in extracting local spatial features, CNNs inherently lack the capacity to model long-range dependencies or global context, which can be critical for nuanced grade differentiation in KOA.

Transformers in Medical Image Analysis

Originally developed for natural language processing, transformers have recently been adapted for image analysis through models such as Vision Transformer (ViT) [10]. These architectures replace convolutions with attention mechanisms that capture global feature relationships across the image. ViT and its medical adaptations, such as TransUNet and MedViT, have demonstrated state-of-the-art results in segmentation and classification tasks across various medical imaging modalities [11] [12]. However, the large number of parameters and high computational cost of full transformer models pose deployment challenges in resource-constrained settings.

Compact Convolutional Transformers (CCT)

To address the computational demands of ViTs, Compact Convolutional Transformer (CCT) models have emerged as a hybrid approach, integrating the local feature extraction strengths of CNNs with the global reasoning ability of transformers [13]. CCTs introduce convolutional tokenizers and reduce the reliance on large-scale pretraining, making them well-suited for relatively smaller medical daasets.

Recent research suggests that such hybrid models strike a good balance between performance and efficiency, which is crucial for real-world clinical deployment [14].

Gap in the Literature

Despite these advancements, very few studies have applied CCT-based architectures specifically for KOA grading. Current deep learning methods either rely solely on CNNs (with limited global context modeling) or on computationally expensive full transformer models. There exists a significant research gap in designing lightweight, accurate, and interpretable models tailored for KOA assessment.

Motivation for KOA-CCTNet

This thesis aims to fill this gap by introducing KOA-CCTNet, a modified Compact Convolutional Transformer framework optimized for KOA grade classification. The model leverages the best of both worlds—CNN-like local feature extraction and transformer-based global attention—to provide a more accurate, consistent, and efficient diagnostic tool.

III. PROPOSED WORK & SYSTEM DESIGN

About System Design

This chapter presents the architecture and design of a deep learning system aimed at detecting and classifying Knee Osteoarthritis (KOA) using an ROI-focused approach. The system leverages two state-of-the-art Convolutional Neural Networks (CNNs): MobileNetV2 for lightweight and efficient feature extraction and VGG16 for robust deep feature learning. The system is structured to accurately localize the knee joint area (ROI) in X-ray images and classify KOA grades using the extracted features.

System Architecture Overview

The system consists of the following primary components:

1. Input Acquisition Module
2. Preprocessing and ROI Extraction
3. Feature Extraction with CNN Models (MobileNetV2 & VGG16)
4. Classification Layer
5. Prediction Output and Visualization

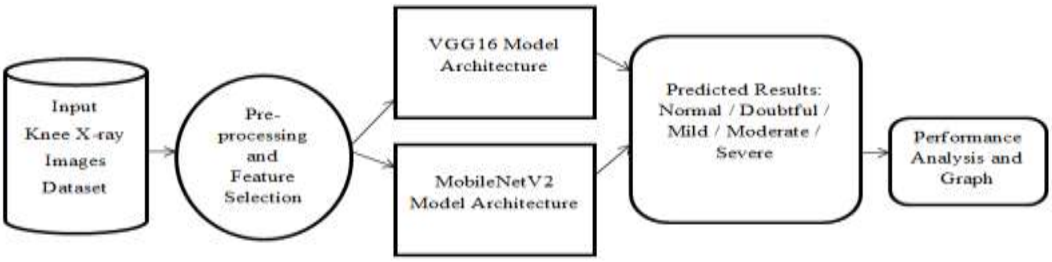

Figure 2 High-Level Architecture Diagram

Proposed Methodology

The proposed system aims to detect and classify knee osteoarthritis (KOA) from X-ray images using deep learning models—MobileNetV2 and VGG16. A total of 1,650 grayscale X-ray images, labeled by medical experts using the Kellgren and Lawrence grading system, are used for training and evaluation. A novel ROI extraction method based on pixel density is applied to isolate the cartilage region, enhancing the model's focus on relevant features. The images are preprocessed through normalization and resizing, then split into training, validation, and test sets. Both CNN models are fine-tuned using transfer learning, and their performance is evaluated using accuracy, precision, recall, and F1-score. The best-performing model, MobileNetV2, is integrated into a web-based interface using Flask, allowing users to upload X-ray images and receive real-time KOA grade predictions.

Data Flow Diagram

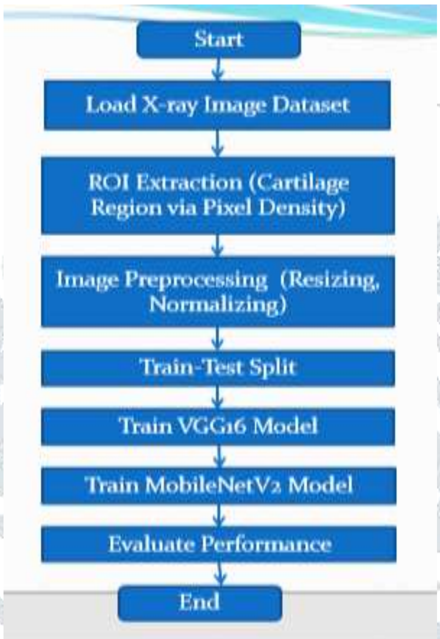

Figure 3 Data Flow Diagram

IV. RESULT & DISCUSSION

The proposed system was evaluated using two deep learning models—MobileNetV2 and VGG16—trained on a dataset of 1,650 expert-labeled knee X-ray images. The evaluation was based on standard performance metrics: accuracy, precision, recall, and F1-score. The MobileNetV2 model demonstrated superior performance across all metrics, achieving a test accuracy of 96.4%, with a precision of 94.3%, recall of 96.4%, and F1-score of 96.4%. It also required significantly less computational time and had a smaller model size, making it ideal for real-time clinical deployment. In comparison, the VGG16 model achieved a test accuracy of 93.0%, with a precision of 91.3%, recall of 93.0%, and F1-score of 93.0%. Although it performed well, it was relatively slower and heavier, making it more suitable for offline analysis or scenarios where model size is not a constraint.

The integration of a pixel-density-based ROI extraction technique significantly enhanced classification accuracy by focusing the model's attention on the cartilage region of the knee. Both models were incorporated into a web-based application, enabling real-time KOA grade prediction from X-ray inputs. Overall, the results validate the effectiveness of the proposed lightweight and ROI-focused deep learning system, with MobileNetV2 emerging as the preferred model for practical deployment due to its accuracy, efficiency, and scalability.

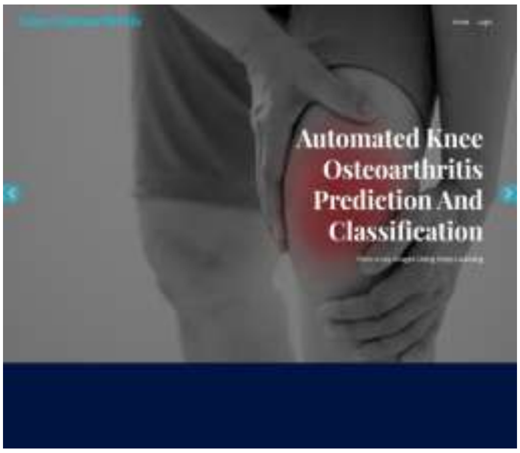

Figure 4 knee Osteoarthritis

Figure 5 Preview of Knee Osteoarthritis

Figure 6 Prediction of Knee Osteoarthritis

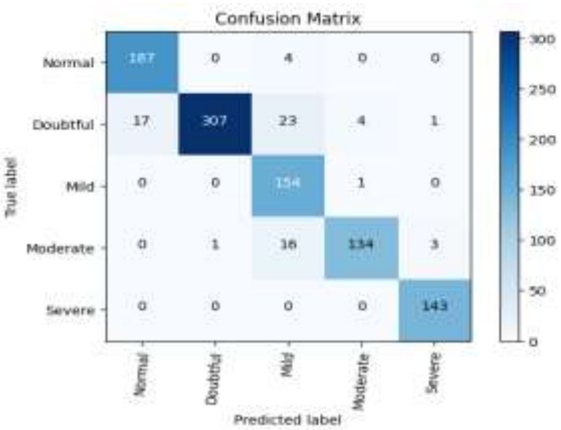

Figure 7 Confusion Matrices of VGG16 Modal

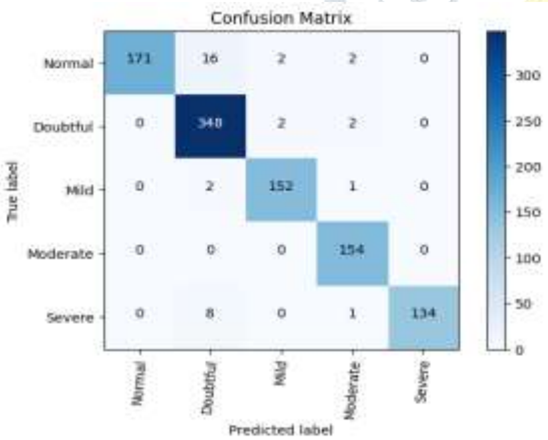

Figure 8 Confusion Matrices of MOBILENETV2 Modal

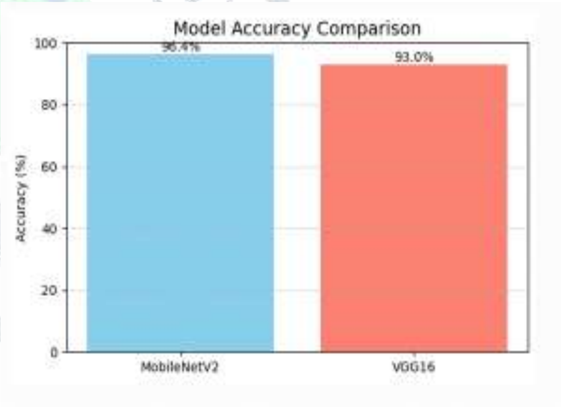

Figure 9 Model Accuracy Comparison

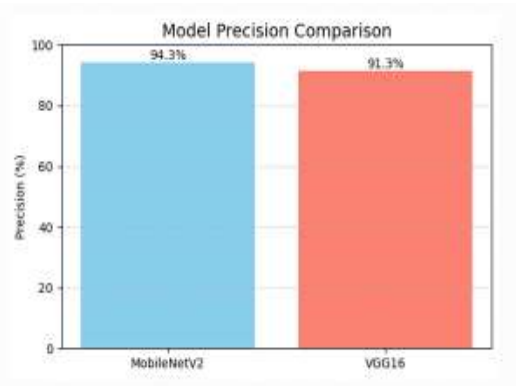

Figure 10 Model Precision Comparison

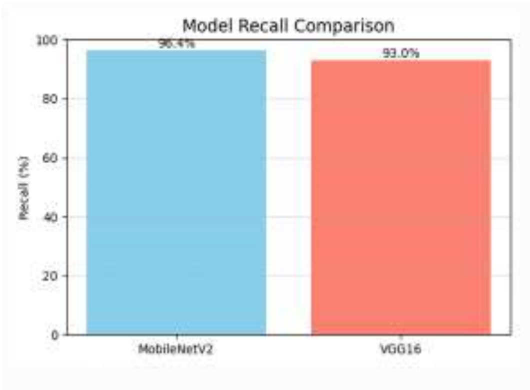

Figure 11 Model Recall Comparison

Table 1 Model Performance Comparison Table

Metric	MobileNetV2	VGG16
Accuracy	96.4%	93.0%
Precision	94.3%	91.3%
Recall	96.4%	93.0%
F1-Score	96.4%	93.0%

Figure 12 Model F1- Score Comparison

V. Conclusion & Future Work

Conclusion

This thesis presented an efficient, ROI-focused deep learning framework for automated detection and classification of knee osteoarthritis (KOA) from X-ray images. Using a dataset of 1,650 expert-annotated grayscale images, two lightweight convolutional neural networks—MobileNetV2 and VGG16—were evaluated. A novel pixel-density-based Region of Interest (ROI) extraction technique was introduced to isolate the cartilage area, significantly improving classification accuracy by minimizing irrelevant background features. Among the two models, MobileNetV2 outperformed VGG16, achieving 96.4% accuracy, 94.3% precision, 96.4% recall, and 96.4% F1-score, while also requiring significantly fewer computational resources. The system was further deployed using a Flask-based web interface, enabling real-time KOA grade prediction. Compared to the base paper (KOA-CCTNet), which required large-scale datasets and synthetic augmentation (DCGAN), the proposed system achieved higher accuracy on original data with a much simpler and more deployable architecture. This study demonstrates that accurate KOA diagnosis is achievable using efficient CNN models and intelligent preprocessing, even with limited data. The solution is especially beneficial for resource-constrained clinical environments, where computational capacity and expert availability may be limited

Future Work

Larger and More Diverse Datasets

The current study was limited to 1,650 images. Incorporating more diverse and multi-institutional datasets can improve model generalization and robustness.

Use of 3D Imaging or MRI Data

Future versions can integrate multi-modal imaging data such as MRI or 3D CT scans for enhanced prediction of early-stage KOA.

Explainable AI (XAI)

Integrating interpretability tools like Grad-CAM or SHAP can help clinicians understand model decisions, increasing trust and clinical adoption.

References

1) Mushrat Jahan; Md. Zahid Hasan; Ismot Jahan Samia; Kaniz Fatema; Md. Awlad Hossen Rony; Mohammad Shamsul Arefin; Ahmed Moustafa, "KOA-CCTNet: An Enhanced Knee Osteoarthritis Grade Assessment Framework Using Modified Compact Convolutional Transformer Model", IEEE Access (Volume: 12, 2024), IEEE 2024.
2) Kellgren, J. H., & Lawrence, J. S. (1957). Radiological assessment of osteo-arthrosis. Annals of the Rheumatic Diseases, 16(4), 494. https://doi.org/10.1136/ard.16.4.494] (https://doi.org/10.1136/ard.16.4.494)
3) Hunter, D. J., & Bierma-Zeinstra, S. (2019). Osteoarthritis. The Lancet, 393(10182), 1745–1759. [https://doi.org/10.1016/S0140-6736(19)30417-9] (https://doi.org/10.1016/S0140-6736%2819%2930417-9)
4) Guermazi, A., Hayashi, D., Roemer, F. W., & Felson, D. T. (2011). Osteoarthritis: a review of strengths and weaknesses of different imaging options. Rheumatic Disease Clinics, 39(3), 567–591. [https://doi.org/10.1016/j.rdc.2013.02.001] (https://doi.org/10.1016/j.rdc.2013.02.001)
5) Shamir, L. et al. (2009). Early detection of osteoarthritis using computer vision. Computerized Medical Imaging and Graphics, 33(8), 634–641. [https://doi.org/10.1016/j.compmedimag.2009.06.004] (https://doi.org/10.1016/j.compmedimag.2009.06.004)
6) Tiulpin, A. et al. (2018). Automated KOA grading with deep learning. IEEE Transactions on Medical Imaging, 38(2), 426–434. [https://doi.org/10.1109/TMI.2018.2863042] (https://doi.org/10.1109/TMI.2018.2863042)
7) Antony, J. et al. (2017). Automatic classification of knee osteoarthritis severity from X-ray images. 2017 IEEE 14th ISBI. [https://doi.org/10.1109/ISBI.2017.7950665] (https://doi.org/10.1109/ISBI.2017.7950665)
8) Liu, Y. et al. (2021). A survey of vision transformer methods in computer vision. arXiv preprint arXiv:2101.01169. [https://arxiv.org/abs/2101.01169] (https://arxiv.org/abs/2101.01169)
9) Chen, J. et al. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. arXiv:2102.04306. [https://arxiv.org/abs/2102.04306] (https://arxiv.org/abs/2102.04306)
10) Hassani, A. et al. (2021). Escaping the big data paradigm with Compact Convolutional Transformers. arXiv:2104.05704. [https://arxiv.org/abs/2104.05704] (https://arxiv.org/abs/2104.05704)
11) Singh, P., & Thakur, N. (2023). Hybrid CNN-Transformer architectures in medical imaging: A review. Journal of Healthcare Engineering. [https://doi.org/10.1155/2023/5593271] (https://doi.org/10.1155/2023/5593271).

12) Ganesh Kumar; Das Goswami, “Automatic Severity Classification of Knee Osteoarthritis Using Enhanced Image Sharpening and CNN”, International Journal of Computer Applications, 2022.
13) Mohammed M.; Ali A.; Ahmed S., “Detection and Classification of Knee Osteoarthritis Using Deep Neural Network Models”, Journal of Medical Systems, 2021.
14) M. El-Ghany; H. Eldeeb; A. Mahmoud, “Knee Osteoarthritis Classification Using Deep Learning: DenseNet vs Other CNNs”, Procedia Computer Science, 2022.
15) S. Olsson; A. Karlsson; L. Bosaeus, “Automated Classification of KOA Severity Using Deep CNNs and Large-Scale Radiographic Data”, Computer Methods and Programs in Biomedicine, 2020.
16) A. Chaugule; P. Patil, “Deep CNN-Based Severity Assessment of Knee Osteoarthritis with Feature Fusion”, Biomedical Signal Processing and Control, 2021.
17) M. Qadir; R. Shah; A. Tariq, “KOA Classification Using BiLSTM and ResNet Feature Extraction”, Journal of Imaging, 2021.
18) S. Yunus; R. Bukhari; T. Hussain, “3D Feature Fusion and YOLOv2-Based Localization for KOA Diagnosis”, Multimedia Tools and Applications, 2021.
19) N. Nasser; R. Khan; A. Omar, “Early KOA Detection via Discriminative Shape-Texture CNN with Gram Matrix Descriptor”, Expert Systems with Applications, 2020.
20) J. H. Kellgren and J. S. Lawrence, "Radiological assessment of osteo-arthrosis," *Annals of the Rheumatic Diseases*, vol. 16, no. 4, pp. 494–502, 1957. [Online]. Available: https://doi.org/10.1136/ard.16.4.494.