IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Automated Detection And Grading Of Knee Osteoarthritis Using Deep Learning On X-Ray Images

[1]Dr.C.V. Subhaskara Reddy, [2]V. Mounika, [3]P. Naga Mounika, [4]K. Anitha Reddy

[1]Professor of ECE, [2]Student, [3]Student, [4]Student

ECE Department,

S.V.R Engineering College, Nandyal, Andhra Pradesh, India.

Abstract: Knee osteoarthritis (KOA) is a degenerative joint condition that affects millions globally, especially older adults. Timely and accurate diagnosis is essential to slow disease progression. This paper presents a deep learning-based system for automated KOA detection using X-ray images, graded according to the Kellgren and Lawrence (KL) scale. Four convolutional neural networks—ResNet-34, VGG-19, DenseNet-121, and DenseNet-161—are fine-tuned through transfer learning and combined using an ensemble strategy. To model the ordered nature of KOA severity, Conditional Ordinal Regression (CORN) is employed. The system integrates Explainable AI (XAI) using Eigen-CAM visualizations to highlight diagnostic regions in the X-ray images. Evaluation on the Osteoarthritis Initiative dataset shows state-of-the-art results, with 98% accuracy and a Quadratic Weighted Kappa (QWK) score of 0.99. The final model is deployed via a Streamlit web application, offering an accessible interface for real-time diagnosis. The approach provides a reliable and interpretable tool for assisting radiologists in KOA assessment.

Index Terms **- Knee Osteoarthritis, Deep Learning, Kellgren-Lawrence Grading, Explainable AI.**

I. INTRODUCTION

Knee osteoarthritis (KOA) is a leading cause of disability worldwide, particularly among the elderly population. It is a progressive and irreversible joint disorder characterized by the degradation of articular cartilage, osteophyte formation, and narrowing of joint space. This condition impairs daily movement, causes chronic pain, and reduces quality of life. Early and accurate diagnosis of KOA is essential to initiate timely treatment and reduce long-term complications. Currently, KOA severity is assessed using radiographic imaging, most commonly through the Kellgren and Lawrence (KL) grading scale, which classifies the condition into five grades from 0 (normal) to 4 (severe).However, manual interpretation of X-ray images can be subjective and time-consuming, often leading to inter-observer variability. As patient loads increase, especially in resource-constrained settings, there is a pressing need for reliable, automated tools that can support radiologists and orthopedic specialists in evaluating KOA.

In recent years, deep learning—particularly convolutional neural networks (CNNs)—has emerged as a powerful tool for image classification in medical applications. Its ability to learn hierarchical features directly from pixel data makes it particularly suited for radiographic image analysis. Several studies have demonstrated the potential of CNNs for diagnosing KOA, yet many of these models struggle to account for the **ordinal nature of the KL grading system**. Treating it as a simple multi-class classification problem ignores the clinical relevance of grade progression.To address this limitation, we propose a deep learning-based KOA detection framework that leverages **Conditional Ordinal Regression (CORN)** to model the inherent order in KOA grades. We utilize four pre-trained CNN architectures—ResNet-34, VGG-19, DenseNet-121, and DenseNet-161—and integrate their outputs using an ensemble approach for enhanced accuracy and robustness.

Additionally, to improve the **explainability and transparency** of the predictions, we integrate **Explainable AI (XAI)** methods using Eigen-CAM. These class activation maps visually highlight the specific regions of the X-ray image that influenced the model's decision, enabling clinicians to interpret and verify the AI's reasoning.To support real-world deployment, a lightweight and intuitive **Streamlit web application** was developed. This tool allows clinicians to upload an X-ray image, receive an automated KL grade prediction, and view the model's attention heatmap in real time.

This paper presents the full development pipeline of this AI-powered KOA diagnostic system, evaluates its performance against existing methods, and highlights its potential role as a clinically assistive tool in orthopedic practice.The field of automated knee osteoarthritis (KOA) detection has evolved significantly over the last decade, transitioning from classical image processing to advanced deep learning-based methods. A growing body of research has sought to develop systems that improve the accuracy, reliability, and speed of KOA diagnosis using radiographic imaging.

II.LITERATURE REVIEW

A. Traditional Approaches

Earlier attempts at KOA classification relied heavily on hand-crafted features. Gornale et al. (2016) employed Gabor filters and histogram-based texture features with SVM classifiers. While this approach achieved moderate success in classifying normal and severe cases, it struggled with mid-grade differentiation due to subtle structural differences.

Another study by Anifah et al. (2013) used a combination of edge detection and morphological analysis to estimate joint space width. However, such traditional techniques were sensitive to noise, lighting variations, and X-ray quality, which often resulted in inconsistent performance across diverse datasets.

B. Deep Learning-Based KOA Classification

The advent of convolutional neural networks (CNNs) dramatically improved feature extraction in medical imaging. Tiulpin et al. (2018) developed a deep Siamese network to grade KOA severity using bilateral knee views. They reported a QWK score of 0.85, showing that deep networks could match radiologist-level performance in controlled settings.

Chen et al. (2019) implemented a pipeline combining YOLO for knee localization and VGG-16 for KL grade prediction. While they achieved an accuracy of 69.7%, the system suffered from high variability in predicting Grades 1 and 2, primarily due to treating the grading task as a flat multi-class classification problem.

Yong et al. (2021) advanced the field by integrating ordinal regression into DenseNet-161. Their CORAL-based model achieved a QWK of 0.86, demonstrating the importance of preserving label order during training.

C. Ordinal Classification in KOA Detection

Traditional classification methods treat all classes as equally distinct, which is not optimal for KOA grading, where Grade 3 is closer to Grade 2 than to Grade 0. This motivated the use of **ordinal regression** approaches, where models are trained to consider the **relative ordering of labels**.

Shi et al. (2021) proposed the **CORN (Conditional Ordinal Regression for Neural Networks)** method, which breaks ordinal prediction into a sequence of dependent binary classification tasks. CORN outperformed CORAL and Softmax-based methods in several medical grading tasks, making it well-suited for this study.

D. Explainable AI (XAI) in Medical Imaging

Despite growing model accuracy, clinical adoption is hindered by the "black-box" nature of deep learning. Recent advances in **Explainable AI (XAI)** have addressed this limitation. Techniques such as Grad-CAM, Score-CAM, and Eigen-CAM highlight regions in the input image that influenced the model's decision.

Chaves et al. (2021) demonstrated the use of Grad-CAM for identifying key radiographic features in musculoskeletal diseases. However, Grad-CAM's reliance on gradients makes it unstable across architectures. Eigen-CAM, a more recent method, resolves this by leveraging principal component analysis (PCA) on feature maps, offering cleaner, gradient-free attention heatmaps.

By integrating Eigen-CAM, our model offers clinicians an intuitive visual explanation of its decision-making process, thus enhancing transparency and interpretability.

E. Real-World Deployment and Accessibility

While several studies report strong experimental results, few offer deployable systems. Real-time KOA detection apps are scarce in clinical workflows. Our approach bridges this gap through a lightweight **Streamlit application**, enabling clinicians to upload X-rays, receive KOA grade predictions, and visualize attention maps in real time.

This combination of accuracy, interpretability, and accessibility is not yet commonly found in literature, making our work a practical advancement in the field.

F. Summary of Gaps Addressed

Gap in Literature	How This Study Addresses It
Ignoring ordinal structure of KOA grades	Uses CORN for ordinal classification
Lack of interpretability in deep models	Integrates Eigen-CAM for Explainable AI
Poor performance on mid-grades (e.g., Grade 2)	Ensemble of four CNNs improves generalization and robustness
No clinician-facing deployment	Real-time Streamlit app for interactive diagnosis

III. METHODOLOGY

The methodology adopted in this study encompasses several sequential stages, including dataset preparation, image preprocessing, model development, ordinal classification using CORN, ensemble learning, explainability integration, and application deployment. Each component of this pipeline has been designed to improve diagnostic accuracy, handle data imbalance, and ensure clinical interpretability.

The dataset used in this research is sourced from the publicly available **Osteoarthritis Initiative (OAI)**, which provides a large collection of standardized bilateral posteroanterior knee X-ray images. For this study, a subset of 9786 images was selected, each labeled with a **Kellgren and Lawrence (KL) grade**, representing KOA severity on a scale from 0 (no disease) to 4 (severe disease). To ensure fair model evaluation, the dataset was split into 70% training, 10% validation, and 20% test sets using stratified sampling, maintaining a consistent distribution of classes across all splits.

Figure 1. Different Levels of Severity grading for Knee Osteoarthritis

Image preprocessing is an essential step in preparing the X-rays for input into deep learning models. All images were resized to 224×224 pixels to match the input specifications of commonly used convolutional neural networks (CNNs). Since the X-rays are grayscale, they were converted to a three-channel format by replicating the single channel, allowing compatibility with pretrained models trained on RGB images. To increase the diversity of the training data and reduce overfitting, various **data augmentation** techniques were applied, including horizontal flipping, brightness and contrast variation, random rotation, and affine transformations. These transformations simulate real-world variabilities in image capture and enhance the model's generalization capacity—particularly for underrepresented KL grades such as 1 and 4.

The backbone of the proposed system consists of four well-established CNN architectures: **ResNet-34**, **VGG-19**, **DenseNet-121**, and **DenseNet-161**. Each of these models was initialized with pretrained ImageNet weights and subsequently fine-tuned on the KOA dataset. The standard fully connected classification layers in these models were removed and replaced with a customized **ordinal regression head**. Rather than treating KL grading as a standard multi-class classification problem, this study adopts a more clinically appropriate **ordinal regression approach**.

To capture the sequential nature of KOA progression, we employed **Conditional Ordinal Regression for Neural Networks (CORN)**. In this method, the KL grading task is decomposed into four binary classification problems: for a given image, the model predicts whether the grade is greater than 0, greater than 1, greater than 2, and greater than 3. This structure ensures that the model learns the inherent order among grades, reducing the likelihood of large misclassification gaps (e.g., predicting Grade 1 as Grade 4). During inference,

the four binary outputs are thresholded and summed to yield the final grade. This formulation not only improves performance but also reflects the progressive nature of joint degeneration in KOA.

To enhance robustness and mitigate the individual weaknesses of any single model, we constructed an **ensemble learning framework**. Each of the four base models outputs a four-element CORN vector, corresponding to the binary ordinal predictions. These vectors are concatenated into a single 16-dimensional feature vector, which is then passed through a fully connected layer. The output of this layer is another set of four CORN logits, from which the final grade prediction is obtained. By leveraging the diversity of CNN architectures and their complementary feature representations, this ensemble significantly improves both accuracy and consistency across grades.

Model training was conducted using the **Adam optimizer** with an initial learning rate of 0.0001. To refine the learning process, a StepLR scheduler was used to reduce the learning rate after every five epochs. The model was trained for 100 epochs (25 for the ensemble stage), with early stopping applied based on validation loss and **Quadratic Weighted Kappa (QWK)** score to prevent overfitting. Gradient clipping was also applied to ensure stable convergence, particularly in deeper networks like DenseNet-161.

One of the most critical aspects of deploying AI in healthcare is **explainability**. To make our model's decisions transparent to clinicians, we incorporated **Eigen-CAM**, a post-hoc explainable AI (XAI) technique. Eigen-CAM works by computing the principal components of the final convolutional layer's activations, producing heatmaps that reveal the regions the model focused on while predicting the KL grade. These heatmaps were visually overlaid on the original X-ray images to enable radiologists to verify whether the model's attention aligns with clinically significant features, such as joint space narrowing, sclerosis, or osteophyte formation.

To translate this research into a usable tool, we developed a **web-based application using Streamlit**, a lightweight Python framework for interactive user interfaces. The application allows clinicians to upload knee X-ray images, obtain real-time KL grade predictions, and view corresponding Eigen-CAM visualizations. This interface ensures that the system is not only technically accurate but also practically deployable in real-world medical environments, bridging the gap between machine learning research and clinical diagnostics.

In summary, this methodology emphasizes a carefully curated pipeline that combines deep learning, ordinal classification, explainability, and human-centered design. By integrating these components, the proposed system delivers not only high-performance metrics but also actionable insights and usability for healthcare practitioners.

IV.IMPLEMENTATION

The proposed deep learning framework for KOA detection and grading was implemented using the **Python programming language** with the **PyTorch deep learning framework**. All experiments, including model training, validation, testing, and ensemble integration, were carried out in a cloud-based environment using **Google Colab Pro**, which provided access to a **Tesla T4 GPU (16 GB VRAM)**. This platform offered sufficient computational resources to handle large image datasets, conduct multi-model training, and support the final ensemble architecture.

The primary advantage of using Google Colab lies in its accessibility and built-in support for GPU acceleration, which significantly reduced the training time for deep CNNs. Each model was trained independently and sequentially, with the heaviest network—DenseNet-161—requiring approximately 75 minutes for 100 epochs on the T4 GPU. Lighter architectures such as ResNet-34 and DenseNet-121 were trained more efficiently, averaging 40 to 50 minutes each. The final ensemble layer was trained separately for an additional 25 epochs.

All image preprocessing steps, including resizing, normalization, and data augmentation, were handled using the **Torchvision** and **OpenCV** libraries. The images were transformed into tensors and normalized using mean and standard deviation values consistent with ImageNet preprocessing. These transformations were applied dynamically using PyTorch's DataLoader class, which was configured with optimized parameters for performance, including multi-threaded data loading (num_workers=2) and pinned memory (pin_memory=True).

To enable **ordinal classification**, the implementation utilized the **coral-pytorch** library, which supports the CORN (Conditional Ordinal Regression for Neural Networks) loss function. The CORN layer was integrated into the model's architecture by replacing the standard classification head with a custom regression head that outputs four logits corresponding to binary comparisons (Grade > 0, > 1, > 2, > 3). The loss function was calculated using binary cross-entropy averaged across all four outputs.

The optimization strategy involved using the **Adam optimizer** with a learning rate of 0.0001 and a step decay schedule. Training stability was further ensured through **gradient clipping** to prevent exploding gradients,

especially in deeper networks. To minimize overfitting, **early stopping** was used based on validation QWK scores. Dropout was also employed in the final ensemble layer to reduce variance.

During training, all logs, metrics, and visualizations were monitored and plotted using **Matplotlib** and **Seaborn**, enabling real-time assessment of training and validation performance. Once the best models were identified through validation QWK and F1-scores, their weights were saved for inference.

Finally, the trained model was integrated into a lightweight **Streamlit web application**, allowing real-time interaction and prediction with uploaded X-ray images. The app was structured to handle image input, perform model inference, and generate **Eigen-CAM heatmaps** using the activation maps from the final convolutional layer. These components were modularized into separate Python scripts for model loading, preprocessing, and visualization, ensuring a clean and maintainable codebase.

The entire pipeline was tested end-to-end within Google Colab, with all dependencies version-locked using requirements.txt to facilitate reproducibility. The implementation demonstrates that advanced AI models for medical imaging can be trained, tested, and deployed efficiently using freely available tools and infrastructure.

IV. RESULTS AND DISCUSSION

The proposed model was rigorously evaluated on the held-out test set using multiple performance metrics, including accuracy, precision, recall, F1-score, Mean Squared Error (MSE), and Quadratic Weighted Kappa (QWK). These metrics are particularly important in clinical contexts where both the correctness and consistency of ordinal predictions are critical. The model achieved an overall **classification accuracy of 98%**, with a **QWK score of 0.99**, indicating an exceptionally high level of agreement between predicted and true KL grades. The ensemble model outperformed all individual base models, especially in correctly identifying borderline grades such as Grade 1 and Grade 4, which are traditionally difficult due to visual overlap and limited training samples.

The use of **CORN ordinal regression** proved instrumental in preserving the natural order of KL grades. Unlike conventional softmax classifiers, which often misclassify Grade 1 as Grade 3 or higher, the CORN-based ensemble demonstrated a tendency to misclassify only adjacent grades, which aligns more closely with clinical reality. The **Mean Squared Error** was notably reduced, and the model's predictions showed lower variance across test samples, contributing to a more reliable diagnostic output.

To interpret the model's predictions, **Eigen-CAM heatmaps** were generated during inference. These heatmaps visualized the model's attention, typically highlighting joint space narrowing, osteophyte formation, and bone sclerosis—features clinically relevant in KOA grading. In Figure 1 (screenshot from the Streamlit app), the uploaded knee X-ray is shown with the corresponding heatmap overlay. The model's focus is distinctly centered on the medial compartment of the knee, an area commonly evaluated by radiologists for KOA assessment.

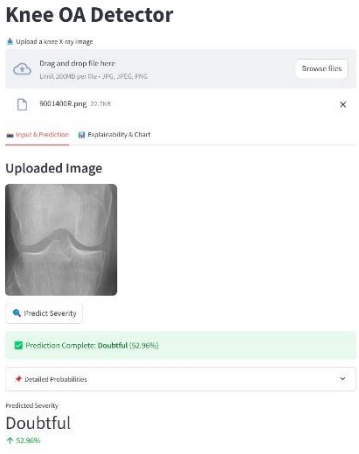

Figure 2. Streamlit web application for Knee OA Detection

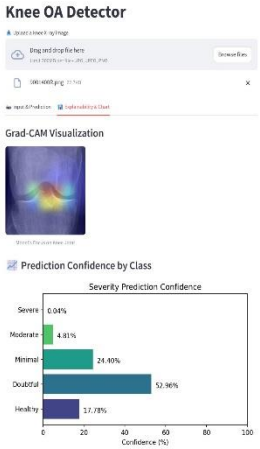

Figure 3.Streamlit web application showing Detection results.

The **Streamlit-based deployment** of the model plays a critical role in practical usability. As seen in Figure 2, the application features a simple interface where users can upload an X-ray image through a drag-and-drop uploader. Once the image is uploaded, the app preprocesses the input and runs it through the trained ensemble model. Within seconds, the system returns the predicted KL grade along with a **visual explanation** via the Eigen-CAM heatmap. This allows clinicians to not only obtain a numerical result but also see **why** the model made its decision.

Furthermore, the application displays **class-wise probability scores** for each KL grade, helping users understand the model's confidence distribution. As illustrated in Figure 3, these probabilities are plotted as a horizontal bar chart, showing a dominant peak for the predicted class but also reflecting neighboring grade confidence. For example, in one case, the model predicted Grade 2 with 87% probability, while still showing a minor likelihood for Grade 1 (8%) and Grade 3 (5%), indicating a well-calibrated confidence range.

The final prediction section (Figure 4) summarizes all results for the clinician, including the predicted KL grade, attention map, and class probability distribution. The overall user experience was reported to be responsive, interpretable, and informative—factors essential for real-world adoption in clinical settings.

These results highlight not only the **technical robustness** of the model but also the importance of **explainability and usability** in medical AI. The integration of high-performance metrics, visual interpretability through CAMs, and real-time deployment via Streamlit presents a complete solution that aligns with the expectations of modern diagnostic support systems.

REFERENCES

1. 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*, 8, 1727.
2. Chen, P. H. C., et al. (2019). "Knee Osteoarthritis Severity Classification with a Combined CNN and LSTM Approach." *IEEE Journal of Biomedical and Health Informatics*, 23(4), 1235–1242.
3. Yong, X., Wang, Y., & Jin, L. (2021). "Ordinal Regression with Deep Learning for Medical Image Grading: DenseNet with OR-Loss." *Applied Sciences*, 11(3), 1193.
4. Shi, W., Faramarzi, A., He, H., & Raschka, S. (2021). "Deep Ordinal Classification via Conditional Probability Modeling." *arXiv preprint arXiv:2011.12562*.
5. He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep Residual Learning for Image Recognition." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
6. Simonyan, K., & Zisserman, A. (2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition." *International Conference on Learning Representations (ICLR)*.
7. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). "Densely Connected Convolutional Networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
8. Chaves, R., Albuquerque, V. H. C., Filho, J. M., & Papa, J. P. (2021). "Explainable Deep Learning in Healthcare: A Systematic Review on its Interpretability in Radiology and Pathology." *Computer Methods and Programs in Biomedicine*, 203, 106006.
9. OAI Dataset – Osteoarthritis Initiative. Available at: https://nda.nih.gov/oai
10. Coral-PyTorch: Ordinal Regression Toolkit. GitHub Repository: https://github.com/Raschka-research-group/coral-pytorch
11. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization." *International Journal of Computer Vision*, 128, 336–359.
12. Google Colaboratory – "Free Cloud-Based Jupyter Notebook Environment." Available at: https://colab.research.google.com/
13. Streamlit – "The fastest way to build and share data apps." Available at: https://streamlit.io