RESEARCH PAPER

A Deep Learning Framework for Automated Detection and Grading of Knee Osteoarthritis

Eliganti Ramalakshmi[1]*, Srujana Inturi[2], Venkata Sushma Chinta[3], Navuru Sahithya[4], Vollala Pravalika[5] and Atchut Varma[6]

[1]*Department of Information Technology, Chaitanya Bharathi Institute of Technology Hyderabad, India*
[2]*Department of CSE Chaitanya Bharathi Institute of TechnologyHyderabad, India*
[3]*Department of Mechanical EngineeringChaitanya Bharathi Institute of Technology Hyderabad, India*
[4]*Department of Information Technology, Chaitanya Bharathi Institute of Technology Hyderabad, India*
[5]*Department of Information Technology Chaitanya Bharathi Institute of Technology Hyderabad, India*
[6]*Department of Information Technology Chaitanya Bharathi Institute of Technology Hyderabad, India*
[1]*eramalakshmi_it@cbit.ac.in,* [2]*isrujana_cse@cbit.ac.in,* [3]*venkatasushmachinta_mech@gmail.com,*
[4]*sahithyanavuru@gmail.com,* [5]*pravalikavollala08@gmail.com and* [6]*atchut1278@gmail.com*

Received: 24th April, 2026; Revised: 20th May 2026; Accepted: 30th June, 2026; Available Online: 10th July, 2026

ABSTRACT

Knee osteoarthritis (KOA) refers to progressive damage of the knee joint and can be characterized by pain, stiffness, and loss of movement, which negatively impacts pa-tients' quality of life. Diagnostics based on radiographic analysis and Kellgren–Lawrence (KL) classification takes considerable time and can vary from observer to observer. This paper describes an approach based on deep learning for automatic detection and grading of KOA in images acquired by X-rays. Our method uses CNNs for image features detection and classification, alongside ACO for hyperparameters tuning. It involves such stages as dataset preprocessing, augmentation, training of the classifier, and its evaluation according to metrics like accuracy, precision, recall, and F1-score. The experimental results showed that our approach demonstrated reliable efficiency, especially for moderate and severe cases; however, some problems remained in early stages because of less prominent changes in features. The developed system can provide real-time predictions while reduc-ing dependence on humans and their expertise. Further research will concentrate on interpretability, integration of different types of input data, and application in developing countries.

Index Terms— *Knee Osteoarthritis (KOA), Deep Learning, Convolutional Neural Networks (CNN), Medical Imaging, X-ray Analysis, Ant Colony Optimization (ACO), Severity Classifica-tion, Artificial Intelligence (AI), Image Processing, and Clinical Decision Support.*

How to cite this article: Ramalakshmi E, Inturi S, Chinta VS, Sahithya N, Pravalika V and Varma A., A Deep Learning Framework for Automated Detection and Grading of Knee Osteoarthritis. Int J Drug Deliv Technol. 2026;16(6): 940-947.
DOI: 10.25258/ijddt.16.6.136

Source of support: Nil.

Conflict of interest: None

I. INTRODUCTION

Knee osteoarthritis (KOA) is a common joint condition that impacts many people globally, leading to symptoms like pain, stiffness, and difficulty in moving because the cartilage and bone in the knee gradually wear down. Despite progress in medical technology, identifying KOA early is challenging since traditional methods depend on doctors manually inter-preting X-rays with the KL grading system, which can be slow and vary between different doctors. [4] To tackle these issues, scientists are using Artificial In- telligence (AI) and Deep Learning (DL) to automate how KOA is diagnosed and classified by severity. Techniques like Convolutional Neural Networks (CNNs) help by identifying important changes in X-ray images, such as bone growths, bone hardening, and the narrowing of the space in the joint. This technology improves how accurately KOA is diagnosed, minimizes human error, and speeds up the decision-making process in healthcare settings. [7] The objective of this paper is threefold. Using radiography data, we first examine current AI-based systems for KOA detection and grading. Second, we go over how explainable AI (XAI), transfer learning, and CNN architectures can enhance interpretability and clinical trust. Lastly, we point out impor- tant issues like clinical integration, model transparency, and dataset imbalance and offer future directions for intelligent, scalable healthcare solutions. [9]

II. BACKGROUND

Knee osteoarthritis (KOA) is a progressive disease associ-ated with gradual wear and tear of articular cartilage, structural changes in subchondral bone, and inflammatory changes in the adjacent soft tissue structures [7]. Gradual

**Author for Correspondence: eramalakshmi_it@cbit.ac.in*

wear and tear of knee joints can cause persistent pain, swelling, stiffness, and decreased mobility in patients, severely affecting their quality of life. Knee osteoarthritis is considered one of the most common musculoskeletal disorders and a significant cause of disability among the elderly population [3]. Some of the factors responsible for developing this condition include obesity, knee joint injury, genetic factors, and repetitive knee joint strain. With the progression of the condition, knee joint damage can impose physical constraints and create a heavy socio-economic burden due to high medical expenses, reduced efficiency at work, and prolonged medical treatments [16]. Hence, timely diagnosis and management are crucial for mitigating the progression of this debilitating condition [11].

A. *Traditional Radiographic Diagnosis*

Radiography, especially X-rays, is still the most frequently employed and economical diagnostic method for evaluating Knee Osteoarthritis [1], [3]. The degree of KOA is normally assessed through the use of the Kellgren–Lawrence (KL) clas-sification, a system that divides the condition into five classes (from Grade 0 to Grade 4) depending on visible characteristics, including joint-space narrowing, osteophytosis (formation of bone spurs), subchondral sclerosis, and bone structure de-formation [4]. Although it is widely applied clinically, the traditional approach is highly dependent on visual assessment conducted manually by radiologists and orthopedic physicians. Visual examination is both time-intensive and laborious due to the expertise required to conduct it. Furthermore, radiograph assessment is subjective and susceptible to variation among different individuals, which means that two different experts might give different grades to an identical image [3], [9]. In fact, this problem becomes even more pertinent at the initial stages of KOA since there can be a difficulty in recognizing these signs through radiography. Such circumstances can lead to improper diagnosis and corresponding treatment. In addi-tion, this method cannot be easily applied on a large scale due to high demands on manual labor [7].In order to address all of these challenges, there is growing need for automated systems capable of producing objective and accurate results. Modern developments in the realms of Artificial Intelligence (AI) and Deep Learning (DL) have helped solve these challenges by automating the process of features' extraction and making grading consistent [2], [6].

B. *AI and CNN-Based KOA Platforms*

Assessing the severity of KOA through automation has been very promising through the use of deep learning approaches, specifically CNNs. Features that are important to radiology as-sessments are automatically determined through models such as those designed by Tiulpin et al. [1], where their performance matches the expertise of experienced radiologists. Models are improved further using the methods of ensemble learning and transfer learning either through ensembling several models or employing pretrained weights. However, most models need big datasets and extensive computing resources, and therefore cannot be deployed in smaller institutions that lack such resources [17].

C. *Explainable and Multi-Modal AI*

Interpretability still remains one of the primary barriers to adoption in the clinical setting. The ability to visualize and understand the prediction made possible by the framework such as OsteoGA through the process of reconstructing healthy knees with a generative autoencoder approach to identify discrepancies in the images that correspond to pathologies [5] can help. Besides, the combination of imaging data (MRI and radiographs), clinical parameters (age, BMI, and WOMAC score), and longitudinal observations improves the early di-agnosis and prognosis for patients [12]. Though RAG-based models have yet to be adopted for KOA detection, explainable AI with multi-modal inputs holds great promise.

D. *Limitations of Current Approaches*

KOA diagnostics have improved thanks to current AI-driven techniques, but there are still a number of gaps.

Clinical utility in preventive care is limited because early-stage KOA detection is still less accurate than for advanced stages [7]. The standardized datasets used to train the majority of AI models (OAI, MOST) might not generalize well to a variety of populations or imaging conditions [3]. There is currently little integration of AI into clinical workflows, in part because of worries about interpretability and clinician trust. Multi-modal models' computational complexity and high hardware requirements make them difficult to implement in clinical settings with limited resources. Lastly, there is still room for more research into creating adaptive and scalable systems because longitudinal monitoring and continuous risk assessment—both essential for individualized treatment and intervention planning—remain understudied.

III. SYSTEM ARCHITECTURE

The designed system aims at classifying the severity of osteoarthritis in knee X-rays into KL grades (0-4). This system operates on a sequential basis whereby raw medical imaging data is converted into medically significant information using a deep learning algorithm. The system begins with an input interface where knee X-ray images are provided for analysis. These images are passed through a preprocessing module that standardizes the input format and prepares the data for model consumption. The processed images are then forwarded to a deep learning-based classification model. Ant Colony Optimization (ACO) system architecture as referred to as Fig. 2 is a decentralized bio-inspired paradigm in which multiple artificial agents, called ants, collaborate to search a graph-based solution space. Each ant builds a candidate solution by moving among nodes on a probabilistic basis according to a decision rule that takes into account pheromone trails and heuristic information. The pheromone model is a common memory that stores information about the quality of the previously found paths. Once all ants have completed their solutions in an iteration, a global update mechanism is applied. It reinforces the high-quality solutions by increasing the pheromone levels, and pheromone evaporation reduces the influence of the less optimal paths, encouraging exploration. This process is repeated iteratively until a stopping criterion is met . This iterative process enables the system to effectively converge to an optimal or near-optimal solution over time . The core component of the architecture is a Convolutional Neural Network (CNN) model trained to extract discriminative fea-tures from radiographic images. The network applies several convolutions and pooling to the data, which makes it possible for the model to detect structural abnormalities, such as joint space narrowing and bone deformities. The learned features are then fed into fully connected layers and classified using a Softmax classifier.After prediction, the output layer then produces the respective KL score along with the confidence scores.The entire model is incorporated in the application layer,allowing for the uploading of X-rays in order to generate the predictions. This architecture is scalable and ensures that inference can be done efficiently.

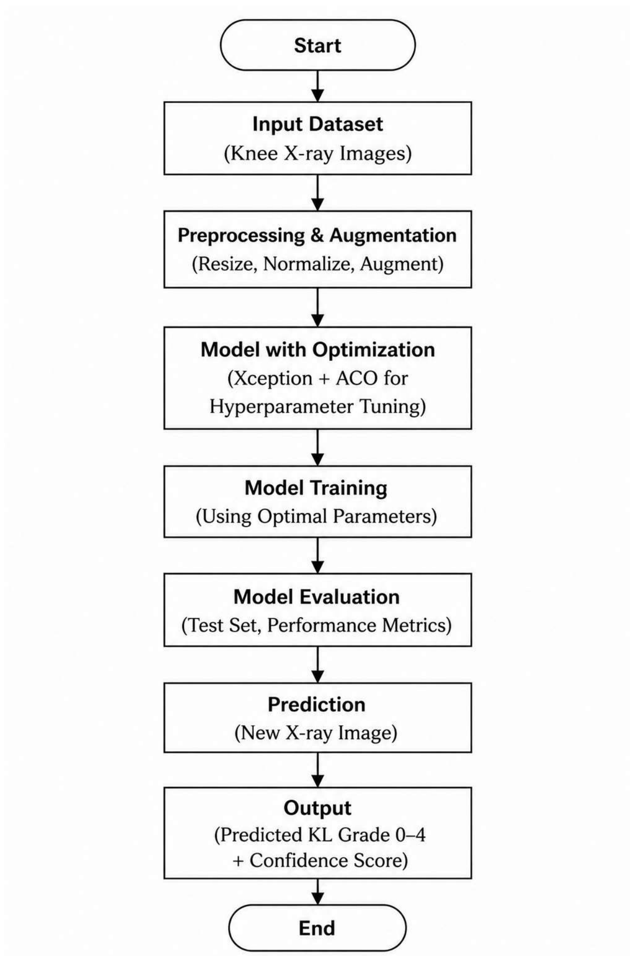

Fig. 1: System Architecture of Knee Osteoarthirites Prediction model

IV. METHODOLOGY

Methodology includes a deep learning approach towards automated detection and classification of knee osteoarthritis (KOA), which comprises the following stages: preparation and processing of the image dataset, feature extraction from these images, and classification of the severity levels based on the features, using a trained convolutional neural network combined with the optimization process. In turn, knee X-ray images will undergo augmentation and other preprocessing methods, after which the features extracted from the images will be used for classification of KOA severity.

A. ***Dataset***

The data consists of approximately 8000 knee X-ray images obtained from the open-source Osteoarthritis Initiative (OAI) dataset, which is a popular benchmark database for KOA. These images are annotated based on the Kellgren–Lawrence (KL) classification method (Grade 0–4) denoting the severity of osteoarthritis. The KL grade ranges from 0, where there is no arthritis, to 4, where there is significant deterioration of joints. In this study, the dataset consists of images that differ in terms of narrowing of joint spaces, presence of osteophytes, and alterations in bones, allowing for deep learning model training.

B. Preprocessing

The images have been resized to 224 × 224 pixels, and normalization is performed to make sure that input dimensions remain consistent and training process remains stable. Resizing helps to decrease computational load without loss of impor-tant image information. Some data augmentation techniques, including rotation, flipping, and brightness changes, are used to provide better generalization by adding artificial variability to training samples and decreasing the likelihood of overfitting. Such manipulations train the model to be more resistant to changes of image orientation, lighting conditions, and other distortions that could occur in real-life medical imaging. Dataset splitting into training, validation, and test datasets allows to train models effectively and evaluate their perfor-mance. The training dataset is used for training process, the validation one serves as a source for adjusting hyperparameters and preventing overfitting.

C. Model Development

A CNN model is implemented using TensorFlow and Keras for feature extraction and classification. The architecture in-cludes convolutional, pooling, and fully connected layers, followed by a Softmax output layer. The model is trained using the Adam optimizer and categorical cross-entropy loss function. The overall process of automated KOA detection and severity prediction using the developed model is demonstrated in Fig. 1. At first, the system receives input images of knee X-ray scans that undergo preprocessing procedures such as image resizing and normalization. These processed images are passed through a deep neural network trained on Xception architecture augmented by Ant Colony Optimization (ACO) algorithms for proper model tuning. Due to model tuning, optimal values for model parameters are selected and used in the training process. After that, the algorithm evaluates the model performance and its readiness to generate proper output based on provided input images. Finally, once the system passes the validation procedure, new images of knee X-ray scans are processed by the system and KL grades (0-4) are generated for each of them.

D. Ant Colony Optimization

The Ant Colony Optimization (ACO) algorithm, which is a type of bio-inspired meta-heuristic method used in this research, is used to optimize the hyperparameters of the pro-posed deep learning model in knee osteoarthritis classification. In this approach, many artificial ants are involved in searching for optimal solutions to the problem by choosing suitable hyperparameters like learning rate, dropout rate, number of dense layer units, and L2 regularization. This choice is done based on the concentration of pheromones, which reflects the past success rate, and some other heuristic information.

Each individual ant evaluates the suitability of its solution by performing few training epochs followed by evaluating the model using a validation criterion, such as the F1 score and accuracy metrics. Then, pheromones are updated by both evaporation and reinforcement strategies in which the ants that have achieved higher scores would leave more pheromones. These steps are iterated until the optimal hyperparameters are found, and then they are applied to train and tune the model [15].

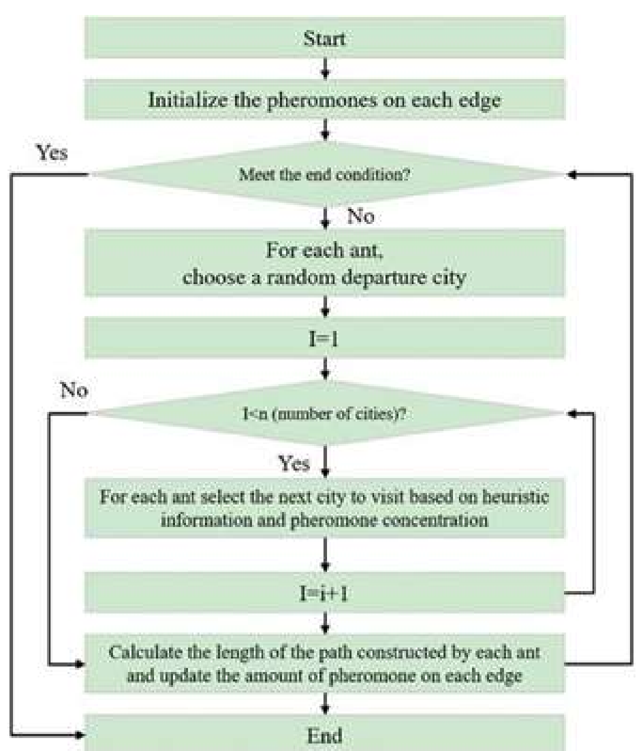

Fig. 2: Workflow of the Ant Colony Optimization (ACO) Algorithm

E. Evaluation and Prediction

The performance of the model will be evaluated using accuracy, precision, recall, and F1 score to ensure that the evaluation is reliable. Accuracy refers to the correctness of the predictions made by the model, whereas precision refers to the fraction of correctly predicted instances. Recall, on the other hand, refers to the capacity of the model to predict all positive instances, which is critical when it comes to medical diagnoses. The F1 score is used to balance precision and recall.

V. IMPLEMENTATION AND RESULTS

Python with the TensorFlow and Keras frameworks was used to implement the suggested Knee Osteoarthritis Detection and Severity Prediction System. The entire pipeline consists of preprocessing, model training, evaluation, and dataset prepara-tion. The Kellgren–Lawrence (KL) scale is used in the model to categorize knee X-ray images into various severity grades.

A. Dataset Preparation

The dataset consists of approximately 8000 knee X-ray im-ages labeled according to the Kellgren–Lawrence (KL) grading scale (0–4). The dataset is divided into training, validation, and testing subsets. Analysis of the dataset reveals an imbalance in class distribution.

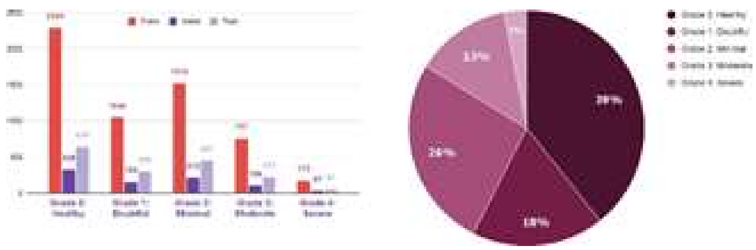

Fig. 3: Distribution of knee X-ray images across KL grades

B. Preprocessing

All images are resized to 224 × 224 pixels and normalized for consistency in the dimensions of the input and for stability during training by scaling the values of the pixels to a common scale. This also simplifies the computations performed on the data without losing important information. Techniques such as rotation, flipping, and adjusting brightness are performed on the images to enhance the generalization capabilities of the model and avoid overfitting by artificially expanding the data set.This will make the model robust against variations in orientation, lighting, and deformations that may occur in the medical images. The data set is further split into training, validation, and testing subsets in order to enable effective training and evaluation of the model. The training subset enables the training of the parameters of the model, while the validation subset is used for fine-tuning the hyperparameters and preventing overfitting.

C. Model Performance

The performance of the proposed model is measured by taking balanced accuracy as the main parameter as this mea-sure considers the class imbalance in the data set by giving equal consideration to all the classes in the process. Along with the balanced accuracy, there are other metrics for evaluating the performance of the classifier, such as precision, recall, and F1 score. The ensemble approach helps improve the classification performance because it uses the advantages of all the individual models and minimizes the variance. As per experimental observations, the proposed model exhibits good performance in identifying moderate and severe cases; however, some mistakes are made in classifying early stages because of similarities in radiographic features.

D. Confusion Matrix Analysis

Figure 5 demonstrates the distribution of images according to various KL classes in the dataset, revealing that there is class imbalance present within the dataset. It can be seen that some classes have a larger number of samples than other classes due to the fact that they fall in the category of moderate and

Classification Report:

	precision	recall	f1-score	support
Healthy (0)	0.82	0.90	0.86	328
Doubtful (1)	0.78	0.74	0.76	153
Minimal (2)	0.83	0.81	0.82	212
Moderate (3)	0.79	0.76	0.77	106
Severe (4)	0.85	0.72	0.78	27
accuracy			0.81	826
macro avg	0.81	0.79	0.80	826
weighted avg	0.81	0.81	0.81	826

Fig. 4: Confusion matrix of the proposed model on the test dataset

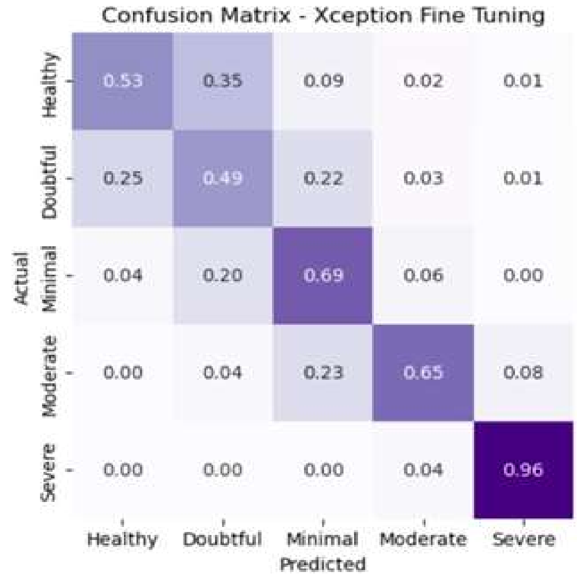

Fig. 5: Confusion matrix of the proposed model on the test dataset

minimal KL. Such an imbalance may affect the learning of the machine-learning model due to the fact that it learns according to the majority classes, hence, leading to poor performance on other classes as well.

Fig. 6 shows some prediction results obtained using the model on various X-ray images, which demonstrate the sys-tem's ability to categorize the degree of severity of KOA. The model precisely determines the severity level according to the radiographic characteristics of the knee joint, such as joint space narrowing, osteophyte presence, and irregularity in the bone structure. The confidence values are provided along with each prediction. It can be seen that the model is more efficient in identifying the moderate and severe classes due to distinct features in X-rays. On the other hand, minor errors

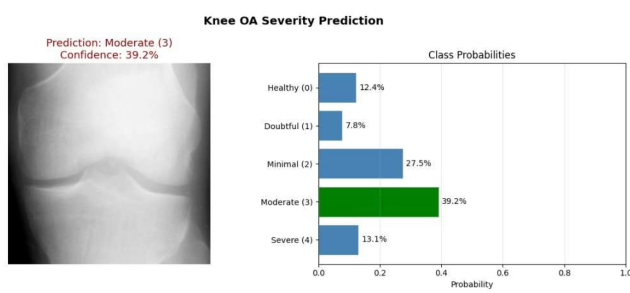

Fig. 6: Sample predictions generated by the model

in predicting the class can be made when dealing with early classes, such as healthy and doubtful cases, because there are less noticeable differences between the features.

*E. **Compartive Analysis***

A variety of AI techniques for KOA diagnosis have been suggested, each with its own pros and cons regarding clinical feasibility, model interpretability, and accuracy. The compari-son between conventional grading based on radiographs, CNN-based models, explainable AI models, and multimodal AI approaches will be discussed below.The traditional method of KOA assessment involves manual diagnosis via Kellgren-Lawrence (KL) grading scale. Although simple and inexpen-sive, this method is subjective and prone to inter-observer variation, especially in detecting early stage KOA [1], [4].

On the other hand, CNN-based models enable automation of feature extraction and classification with increased consistency and high performance compared to experts. The main problem with CNNs is the necessity for huge training datasets and substantial computing power [6], [8].Explainable AI (XAI) techniques can improve the transparency of AI models via autoencoder reconstruction, which makes these models more suitable for clinical use despite increased complexity of the models [5].Moreover, the use of multimodal AI, that combines the information from X-rays, MRI, and clinical data, has proven successful in early diagnosis and disease progression prediction. At the same time, implementation of multimodal AI is harder due to high computational demands [12]. As illus-trated in the Fig. 7, the Precision-Recall graph demonstrates the behavior of the proposed model at different threshold levels. The graph starts with a relatively high precision level when the recall level is low, implying that the proposed model predicts extremely accurate results for a few confident samples. With the increase in the recall level, the precision level falls gradually due to the trade-off between maximizing the detection rate and the precision rate of the predictions. On the side of a relatively high recall rate, a sharp fall in the precision rate can be observed, especially when the model tries to categorize almost all of the samples.

F. *Discussion*

The balance between model performance and practicality is well-defined by the results obtained. Although CNN and multi-modal models demonstrate high accuracy yet require much infrastructure, conventional methods have low accuracy yet are practical to implement. Though explainable models build confidence, they incur computational overhead. High-performing AI must be balanced with interpretability and deployability in low-resource clinical settings, according to this review. A promising way to provide scalable, clinically relevant KOA diagnosis, even in smaller hospitals or remote clinics, is to combine lightweight, optimized CNNs with explainable outputs.

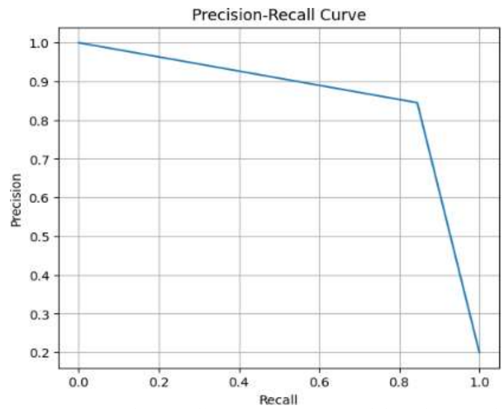

Fig. 7: Sample predictions generated by the model

TABLE I: Critical comparison of KOA diagnostic methods

Method	**Advantages**	**Primary Bottleneck**
Manual X-ray Analysis	Well-established, interpretable by radiologists	Subjective, time- consuming, prone to inter-observer variability
CNN-based AI	Fast, automated grading, can detect subtle radiographic features	Requires high- quality datasets and computational resources
Multi-modal AI (X-ray + MRI + Clinical)	Improved early detection and progression forecasting	High computa- tional cost, limited deployment in small clinics
Explainable AI	Highlights pathological regions, improves interpretability	Still limited for real-time clinical decision support, requires model integration

A comparative analysis of the discussed methodologies is provided in Table I . It presents the most important pros and cons for each methodology, including traditional diagnostic tools, CNN-based algorithms, explainable AI, and multimodal approaches.

VI. CONCLUSION

This paper provides a comprehensive approach to automate the process of detecting and grading the severity of KOA from the radiographic images of knees through convolutional neural networks. The proposed system not only minimizes the need for expert intervention but also ensures consistency in pre-dicting the severity level according to the Kellgren–Lawrence scale.

The designed machine learning model, which has been created using TensorFlow and Keras, is a good example of applying deep learning algorithms for extracting significant radiographic characteristics such as joint space narrowing and bone deformation. Preprocessing and data augmentation also contribute to the improvement of the learning algorithm and ensure better generalization. The experimental findings indicate that the trained model attains a high level of accuracy of about 81%. Besides, the accuracy level for moderate and severe classes is higher than the other two categories. However, there may be some errors for early stages of the disease because of the slight changes in the features.

Overall, the proposed method serves as a valuable tool for diagnosing KOA in a cost-effective manner. Moreover, the system helps to decrease the consumption of time and resources. Thus, healthcare practitioners can make faster and more reliable decisions regarding patients' health status.

REFERENCES

[1] A. Tiulpin *et al.*, "Automatic Grading of Individual Knee Osteoarthritis Features and Severity Using Deep Learning," *Medical Image Analysis*, vol. 65, p. 101789, 2020.

[2] P. Chen *et al.*, "Fully Automatic Knee Osteoarthritis Severity Grading from Radiographs using Deep Convolutional Neural Networks," *Com-puter Methods and Programs in Biomedicine*, 2019.

[3] Eliganti Ramalakshmi, Loshma Gunisetti, Sumalatha Lingamgunta, "Comparative Analysis of Different U-Net Variants for Breast Cancer Detection in Histological Images," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 17, pp. 350–361, 2025.

[4] Eliganti Ramalakshmi, Loshma Gunisetti, Sumalatha Lingamgunta, "Breast Cancer Detection Using Hybrid Optimised Deep Learning Method," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 48, no. 3, pp. 149–163, 2025.

[5] "OsteoGA: An Explainable AI Framework for Knee Osteoarthritis Severity Assessment," *Proceedings of ACM Conference on Health Informatics*, 2023.

[6] S. W. Pi *et al.*, "Ensemble Deep-Learning Networks for Automated Kellgren–Lawrence Grading of Knee Osteoarthritis," *Scientific Reports*, 2023.

[7] S. Rani *et al.*, "Deep Learning to Combat Knee Osteoarthritis and Severity Assessment: A Review," 2024.

[8] T. Momenpour *et al.*, "Optimizing CNN-Based Diagnosis of Knee Osteoarthritis," *Journal of Medical Imaging*, 2025.

[9] D. Nasef *et al.*, "Deep Learning for Automated Kellgren–Lawrence Grading: Comparative Study Across Datasets," *Journal of Imaging (MDPI)*, 2024.

[10] N. Belton, "An AI System for Continuous Knee Osteoarthritis Severity Grading," *arXiv preprint*, 2024.

[11] T. Wang *et al.*, "Predicting Knee Osteoarthritis Progression Using Neural Networks and Longitudinal MRI," 2025 (preprint/accepted).

[12] W. Li *et al.*, "The Value of Multiview X-Ray Images and Prior Knowledge for KOA Grading," *Journal of Imaging*, 2023.

[13] M. Soufi *et al.*, "Validation of a Musculoskeletal Segmentation Model with Uncertainty Estimation," *Medical Image Analysis*, 2025.

[14] S. Olsson *et al.*, "Automating KL Classification in an Unfiltered Adult Population Using Deep Learning," *npj Digital Medicine*, 2021.

[15] N. Belton, "Anomaly Detection Inspired Continuous KOA Grading with Few Labels (SS-FewSOME)," *arXiv:2407.11500*, 2024.

[16] S. Seoni *et al.*, "Application of Uncertainty Quantification to AI for Healthcare," *Computers in Biology and Medicine*, 2023.

[17] O. Esmez *et al.*, "TurkerNeXtV2: A Lightweight CNN for Knee Os-teoarthritis Detection," *Diagnostics (MDPI)*, 2025.

[18] J. Amjad *et al.*, "Knee Osteoarthritis Detection and Classification Using IoT-enabled Imaging and CNNs," *MDPI*, 2025.

[19] S. Olsson *et al.*, "Automating Classification of Osteoarthritis According to Kellgren–Lawrence in the Knee Using Deep Learning in an Unfiltered Adult Population," *Scientific Reports*, 2021.