A Lightweight CNN Model for Knee Osteoarthritis Grading: Development and Evaluation Using Radiographic X-rays

Mohankumar H M[1], Dr. T. Vijaya Kumar[2]

[1]Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India
18mohanmkumar@gmail.com

[2]Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India
tvijaykumar@bit-bangalore.edu.in

---***---

ABSTRACT

Knee Osteoarthritis (KOA) is a progressive degenerative joint disease-causing pain, stiffness, and functional disability, posing a major health concern, especially among the elderly. Conventional diagnosis relies on radiographs and the Kellgren–Lawrence (KL) grading system, which suffers from subjectivity, inter-observer variability, and inefficiency for large-scale screening. To overcome these challenges, this research proposes an AI-driven multi-stage KOA detection framework using deep learning for automated severity classification. The model employs a dual-input Convolutional Neural Network (CNN) trained on preprocessed grayscale knee X-rays. A symmetry-based preprocessing technique pairs each X-ray with its horizontally flipped version, emphasizing asymmetries in joint space narrowing, osteophytes, and bone deformities, key KOA indicators. This improves feature capture and interpretability. The CNN classifies knees into five KL stages with high accuracy and robustness. For clinical use, the model is integrated into a Flask-based web application enabling X-ray upload, real-time severity prediction, and personalized PDF report generation. The platform also provides educational resources including treatment guidance, diet suggestions, and hospital references. Experimental results demonstrate competitive performance against state-of-the-art methods while offering real-time usability. By reducing subjectivity and ensuring consistent grading, this framework serves as a reliable clinical decision-support tool, with future work exploring explainability, larger datasets, and cloud deployment.

General Terms
Machine Learning, Deep Learning, Medical Imaging, Web-based Applications.

Keywords
Knee Osteoarthritis, Convolutional Neural Network, Kellgren–Lawrence Grading, Flask Framework, Severity Detection, X-ray Classification, Automated Diagnosis, Clinical Decision Support System.

1. INTRODUCTION

Knee Osteoarthritis (KOA) is one of the most common degenerative joint disorders and a leading cause of chronic pain, disability, and reduced mobility, especially among elderly populations. It is marked by the progressive deterioration of articular cartilage, narrowing of joint spaces, and osteophyte formation, posing a significant burden on healthcare systems and patients' quality of life. Early and reliable diagnosis is critical for slowing disease progression and enabling effective interventions, yet staging KOA remains a clinical challenge. Radiographic imaging, particularly X-rays, has long been the gold standard due to its accessibility, low cost, and ability to reveal structural changes in the knee joint. The Kellgren Lawrence grading system is widely used to classify KOA severity from Grade 0 representing a normal joint to Grade 4 representing severe disease, but its effectiveness depends heavily on the expertise and consistency of radiologists. Manual interpretation is often subjective and prone to inter observer variability, especially in borderline cases or when image quality is poor. In busy clinical environments fatigue and high workload may further compromise accuracy, while limited availability of skilled specialists in resource constrained areas delays diagnosis and treatment. Artificial Intelligence has emerged as a transformative force in healthcare, with deep learning showing particular promise in medical imaging. Convolutional Neural Networks excel at learning hierarchical features directly from raw image data, identifying subtle structural variations such as joint space narrowing, osteophyte growth, and bone texture changes that may not be easily detected by humans. Unlike traditional approaches that rely on handcrafted descriptors, CNNs can generalize across diverse datasets, making them highly suitable for KOA detection. However, many AI based methods remain confined to research settings, hindered by computational demands and lack of practical integration into clinical workflows. This study titled AI Powered Multi Stage Knee Osteoarthritis Detection using CNN and Flask addresses these gaps by presenting a framework that combines CNN driven classification with an interactive web-based deployment platform. The system is designed to classify KOA across all five KL grades with high accuracy, trained on the MedicalExpert I Knee X-ray dataset which is a publicly available collection of labelled anterior posterior radiographs. Preprocessing techniques such as resizing, normalization, and denoising ensure standardized inputs, while augmentation methods including rotation, flipping, contrast adjustment, and noise injection improve robustness against imaging variability. A key contribution of this project is the introduction of a symmetry based preprocessing technique that compares original radiographs with their flipped counterparts, emphasizing asymmetry in joint space narrowing and deformities which are critical indicators of KOA progression. A dual input CNN architecture processes both versions, enhancing feature extraction and interpretability. To make the system accessible beyond research labs, the trained model is deployed within a Flask powered web application. Users whether clinicians or patients can upload X-ray images through a browser, preview them, and receive stage predictions within seconds. Results are displayed with a confidence score and accompanied by a disclaimer clarifying that the system is intended for educational and research purposes and not a replacement for professional medical advice. What distinguishes this work is the integration of diagnostic automation with patient oriented educational modules. Beyond classification, the platform provides information on evidence-based treatment options. These features make the tool not only a diagnostic aid but also a resource for awareness and self-management. The project's significance lies in balancing accuracy, usability, and scalability. Unlike many AI models that remain experimental, this system is lightweight, efficient, and deployable in real time settings. Its modular design allows future integration of advanced neural architectures, additional imaging modalities such as MRI, or deployment on cloud platforms for broader clinical access. The system can function effectively in both advanced hospitals and rural clinics with limited resources, thus contributing to equitable healthcare delivery. Ultimately, this research demonstrates how AI can strengthen diagnostic precision, accelerate clinical decision making, and improve overall management of knee osteoarthritis. The rest of this paper is organized as follows. Section 2 reviews related literature and existing methodologies for KOA detection. Section 3 outlines the proposed system architecture, preprocessing strategies, and CNN design. Section 4 describes the experimental setup, dataset preparation, and evaluation metrics. Section 5 presents the results and comparative analysis. Section 6 concludes with key findings, practical implications, and directions for future research.

2. LITERATURE SURVEY

Knee osteoarthritis (KOA) detection and severity grading have been extensively explored in recent years through the integration of machine learning (ML) and deep learning (DL) methods, supported by advancements in medical imaging and computational frameworks. A comprehensive review by Thomas and Mary emphasized the significant burden of OA, especially KOA, and highlighted that DL models such as CNNs, HRNet, and vision transformers outperform traditional ML approaches while underlining the role of large-scale datasets like OAI in enabling robust classification [1]. Sharma et al. further extended this perspective by reviewing multiple imaging modalities such as MRI, CT, ultrasound, and thermal imaging, while concluding that X-rays remain the gold standard for KOA detection, with CNN, VGG, DenseNet, and ResNet achieving superior performance compared to classical ML algorithms [2]. Addressing the need for end-to-end diagnostic tools, Arun et al. proposed a DenseNet169-based CNN framework achieving 94.9% accuracy on 1,800 preprocessed X-rays and even developed a Flask-based clinical application, demonstrating that such systems can transition into usable diagnostic solutions [3]. Beyond X-ray analysis, Du et al. advanced the use of MRI-based Cartilage Damage Index features with PCA and ML classifiers such as ANN and Random Forest, concluding that CDI-based methods are effective for predicting OA progression [4]. Mehta and Kaur introduced a CNN-Attention hybrid framework trained on OAI and MOST datasets, which achieved 92.3% accuracy with strong interpretability via attention heatmaps, showing the importance of combining feature extraction with attention mechanisms [5]. Similarly, Wang et al. utilized InceptionResNetV2 with transfer learning on MRI images, reporting patient-level accuracy of 88.5% and image-level accuracy of 96.1%, establishing the role of transfer learning in improving generalization [6]. To enhance explainability, Guhan et al. applied R-CNN with Grad-CAM, achieving over 90% accuracy for normal and severe grades but revealing the persistent challenge of correctly classifying intermediate grades such as KL-1 [7]. Sharma et al. leveraged EfficientNetB5 with transfer learning and Adam optimization on Kaggle datasets, achieving 93.8% accuracy and highlighting that optimization strategies play a major role in model robustness [8]. Sivakumari and Vani compared multiple transfer learning models, revealing that AlexNet achieved 98% accuracy and outperformed deeper architectures such as NasNetLarge, thus emphasizing the trade-off between computational complexity and accuracy [9]. Kapoor et al. compared ML and DL approaches on the GitHub–OAI dataset, reporting that Naïve Bayes achieved the highest accuracy at 95%, suggesting that simpler ML models can occasionally rival DL when feature engineering is effective [10]. Hybrid architectures have also gained traction, as Mehta and Bhardwaj combined CNN with LSTM, achieving 92.3% accuracy and demonstrating strong temporal-spatial feature learning with validated interpretability [11]. Dalia et al. introduced DeepOA, a CDSS using YOLOv5 for segmentation followed by CNN-based classification, reporting improved interpretability and accuracy compared to raw image classification, underscoring the value of preprocessing and ROI selection [12]. Alternative approaches such as edge detection were explored by Perumal et al., who used Canny and LoG on MRI to measure cartilage thickness and detect OA, showing promising outcomes without ML dependency [13]. Sharma et al. demonstrated the effectiveness of feature engineering by using GLCM and histogram features with SVM, achieving 95% accuracy, thus proving that traditional methods remain useful in small-scale datasets [14]. Chandu et al. combined CNN with AlexNet and biomechanical gait features, demonstrating improved classification accuracy and highlighting the value of multimodal analysis [15]. Raju et al. compared SVM and CNN, finding SVM superior (87% vs 83%), suggesting that classical models can outperform DL in small datasets [16]. Jain et al. introduced MobileNetV3-based KOA detection optimized for mobile and edge devices, achieving 82.9% accuracy and emphasizing lightweight deployment [17]. Yang et al. proposed HCGN, a hybrid graph-CNN model with attention, demonstrating superior interpretability and classification accuracy on the OAI dataset [18]. Kishore et al. tested eight CNN architectures on KL grading, with DenseNet169 and EfficientNetB7 performing best at 98.7% accuracy, confirming the feasibility of CNNs in automated grading [19]. Likhith et al. presented a survey highlighting GANs for synthetic image generation and reviewing ML and DL methods for KOA grading, stressing the importance of dataset expansion and model optimization for clinical application [20]. Masood et al. proposed OSTEO-DOC, a multimodal framework integrating imaging with KOOS responses and biomechanical features, achieving 89.3% accuracy and showing the potential of holistic approaches for KOA diagnosis [21]. Gowr et al. used CNN with SSD object detection, achieving 96.4% segmentation accuracy and efficient grading, proving the feasibility of end-to-end automated solutions [22]. Finally, Antony et al. employed regression-based CNNs (VGG16, AlexNet) with automatic joint localization, concluding that regression better aligns with the ordinal nature of KL grading and improves severity quantification [23]. Collectively, these studies reveal the evolution of KOA research from early feature-based ML systems to advanced DL and hybrid frameworks with improved interpretability, multimodal integration and clinical applicability, while also underscoring persistent challenges in intermediate grade classification, dataset imbalance, Weak temporal tracking, Sensitive image quality, Poor spatial and temporal integration and generalization for real-world deployment.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

3. DATA PREPROCESSING

Data preprocessing [20] forms a crucial stage in the development of an automated framework for Knee Osteoarthritis (KOA) detection, as it directly influences the accuracy, robustness, and generalizability of the deep learning model. The raw medical X-ray images obtained from the *MedicalExpert-I* dataset are highly heterogeneous in terms of size, quality, and visual clarity; hence, systematic preprocessing is essential before training and deployment. The process begins with data cleaning, where corrupted, unreadable, or irrelevant files are discarded to ensure that the training set contains only diagnostically valid samples. Each retained radiograph is then converted into grayscale, since color channels provide no added diagnostic value for X-rays and only increase computational redundancy. Grayscale conversion highlights structural patterns such as joint space narrowing, osteophyte formation, bone sclerosis, and cartilage irregularities that are vital indicators of KOA severity. Following this, all images are resized into a standardized dimension of 256 × 256 pixels to ensure uniformity across the dataset. This fixed input resolution aligns with the convolutional neural network (CNN) architecture, which requires consistent tensor dimensions for efficient convolutional operations. Once resized, pixel values are normalized by dividing each value by 255, scaling intensities into the range [0,1]. This normalization stabilizes gradient descent, accelerates convergence, and prevents issues arising from large numeric disparities during training. Subsequently, the dataset is reshaped into four-dimensional tensors (samples, 256, 256, 1) to comply with CNN input specifications, where the last dimension represents the grayscale channel. Alongside image preparation, categorical labels for each stage of osteoarthritis are mapped into integers and further transformed into one-hot encoded vectors to facilitate classification using categorical cross-entropy as the loss function. For evaluation, the dataset is divided into training (80%) and testing (20%) subsets.

In addition to dataset preparation, the preprocessing pipeline extends into the deployment stage via a Flask-based application, ensuring robustness and clinical reliability. At this point, uploaded images undergo an automated X-ray validation mechanism before being processed by the CNN. This safeguard is necessary because, in real-world clinical workflows, patients or non-experts may accidentally upload invalid files such as selfies, scanned documents, or colored photographs. To prevent such errors, the system employs a lightweight preprocessing function named is_xray_image() which inspects grayscale intensity distributions, pixel variance, and color channel differences. X-rays typically exhibit high grayscale uniformity, distinct intensity contrast, and minimal color variation compared to natural images. The algorithm computes feature such as mean intensity, standard deviation, and color deviation to distinguish valid X-rays from non-radiographic images. If the uploaded file fails this validation, the system rejects it and prompts the user to re-upload a correct medical X-ray. For valid inputs, the image is immediately converted into grayscale, resized to 256 × 256, normalized, reshaped into the CNN-compatible format, and then forwarded into the trained model for inference. The model outputs the predicted KOA stag with a confidence score, which is displayed to clinicians through an intuitive interface. Beyond inference, this pipeline also integrates with a PDF report generation system that records the uploaded X-ray, predicted stage, confidence score, and patient details, thereby improving transparency and usability in clinical practice.

By integrating both offline preprocessing (for training consistency) and online preprocessing (for deployment robustness), the system ensures high diagnostic accuracy, safeguards against invalid input, and enhances the clinical trustworthiness of AI-assisted KOA detection. This dual-layer preprocessing not only optimizes the learning process but also translates effectively into real-world healthcare settings, making it a vital bridge between research and clinical adoption.

Preprocessing and Deployment Algorithm

Input: Raw X-ray dataset, Uploaded clinical image
Output: Pre-processed tensors ready for CNN input

```
for each image in dataset:
   if image is corrupted or unreadable:
      discard image
   convert image to grayscale
   resize image to 256 × 256 pixels
   normalize pixel values to [0,1]
   reshape image to (256, 256, 1)
   assign categorical label → one-hot encoding
end for

for each uploaded image in Flask application:
   check image validity:
      compute grayscale mean intensity, variance, and color
channel difference
      if values do not match X-ray profile:
         reject and return error → "Not a valid X-ray"
   if valid:
      convert to grayscale
      resize to 256 × 256 pixels
      normalize pixel values to [0,1]
      reshape image to (256, 256, 1)
      forward pass into CNN model
      obtain predicted stage and confidence score
      generate PDF report with:
         - Patient details
         - Prediction result and confidence
         - Uploaded X-ray image
         - Stage-specific medical description
         - Disclaimer for clinical verification
      return report file to clinician
end for
```

4. CNN ARCHITECTURE AND TRAINING

In this study, a Convolutional Neural Network (CNN) is employed as the primary deep learning framework for the automatic detection and severity classification of knee osteoarthritis from X-ray images. CNNs are highly effective in computer vision tasks, particularly in medical imaging, as they are capable of automatically extracting meaningful patterns without manual feature engineering. The input to the network consists of preprocessed grayscale knee radiographs that are resized to 256 × 256 pixels and normalized to the range [0,1], ensuring consistency and reducing noise across the dataset. The network architecture is designed with multiple convolutional layers, each composed of filters (kernels) that scan the image to detect low to high-level features such as edges, contours, and texture variations that are critical in identifying joint space narrowing, osteophyte formation, and bone deformities. After convolution, the Rectified Linear Unit (ReLU) activation function is applied, which introduces non-linearity by setting negative values to zero and retaining positive signals, enabling the network to learn complex structures from radiographs. Each convolutional block is followed by a MaxPooling operation, which reduces the

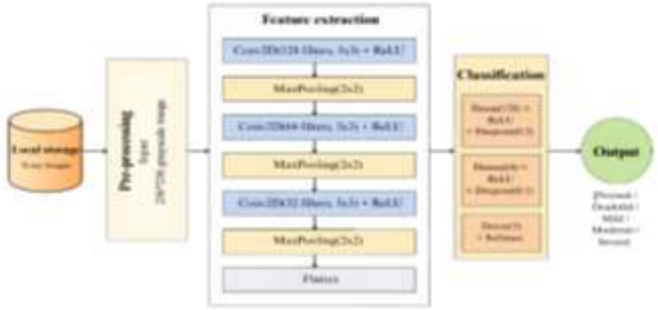

Figure 4.1: *CNN Model Architecture illustrating sequential convolution, pooling, flatten, and dense layers for KOA classification.*

dimensionality of feature maps by selecting the maximum value from a local region (typically 2 × 2), thereby preserving essential features while reducing computational cost and minimizing the risk of overfitting. This hierarchical feature extraction allows the network to progressively capture simple structures in early layers and more abstract clinical patterns in deeper layers. Once the feature extraction process is completed, the feature maps are flattened into a one-dimensional vector, which serves as input to fully connected dense layers. These dense layers integrate the learned features and form the decision-making component of the CNN. To enhance model generalization, dropout regularization is incorporated, which randomly deactivates a fraction of neurons during training to prevent over-reliance on specific nodes and reduce the chances of overfitting. The final classification layer uses the Softmax activation function to output a probability distribution over five classes corresponding to osteoarthritis severity: Normal, Doubtful, Mild, Moderate, and Severe. For example, an output vector such as [0.05, 0.08, 0.12, 0.65, 0.10] would classify the X-ray as Moderate osteoarthritis with 65% confidence.

The training process involves initializing model parameters and compiling the network using categorical cross-entropy as the loss function, which is appropriate for multi-class classification problems. The Adam optimizer is employed due to its adaptive learning rate and efficiency in handling sparse gradients. Accuracy is selected as the primary evaluation metric to monitor learning progress. The dataset is divided into training, validation, and testing subsets to ensure robust generalization of the model. During training, the CNN updates its weights using backpropagation and gradient descent, adjusting filter values to minimize classification error. Training is conducted over multiple epochs, and to avoid overfitting, techniques such as early stopping, dropout, and data augmentation (rotation, flipping, contrast adjustments) are incorporated. The trained model is saved in HDF5 format (.h5) for deployment and integrated into a Flask-based web application, where it is used to provide real-time predictions. Within the application, uploaded knee X-rays are first validated to confirm they are genuine radiographs by checking grayscale intensity, variance, and color channel consistency; non-X-ray images are rejected with an error message. Valid images undergo the same preprocessing pipeline for grayscale conversion, resizing, normalization, and reshaping before being fed into the CNN model. The system outputs both the predicted osteoarthritis stage and a corresponding confidence score, which can be displayed to clinicians along with a patient report. This CNN-based approach significantly reduces diagnostic subjectivity and inter-observer variability compared to traditional manual grading methods, while also accelerating the decision-making process. Furthermore, CNNs have already demonstrated success in several medical imaging domains, including lung disease classification, diabetic retinopathy detection, and musculoskeletal disorder assessment, making them a reliable choice for knee osteoarthritis detection. The integration of CNN with a lightweight web framework provides a scalable and accessible clinical decision-support tool, highlighting the potential of artificial intelligence to improve early diagnosis, personalized treatment planning, and overall patient care in orthopedic healthcare.

5. EXPERIMENTAL DESIGN

The experimental study for this project is grounded in a curated dataset of knee X-ray images, which are subjected to systematic preprocessing and convolutional neural network (CNN) training to achieve clinically reliable detection of Knee Osteoarthritis (KOA). The raw images, obtained from multiple medical sources, are first standardized to ensure consistency across the dataset through conversion to grayscale, resizing to 256 × 256 pixels, normalization of pixel values within the [0,1] range, and reshaping into tensors suitable for CNN input. Beyond these conventional preprocessing steps, a symmetry-based approach is implemented in which each X-ray is compared against its flipped counterpart to emphasize subtle joint asymmetries that serve as critical diagnostic indicators of KOA progression. Each image is annotated according to the Kellgren–Lawrence (KL) grading scale, thereby providing structured categorical supervision for the CNN model during training and enabling it to differentiate disease severity levels effectively. During deployment, uploaded images undergo an automated validation pipeline in which intensity, variance, and channel-based checks confirm the authenticity of the inputs, ensuring that only genuine and clinically relevant X-rays proceed to prediction.

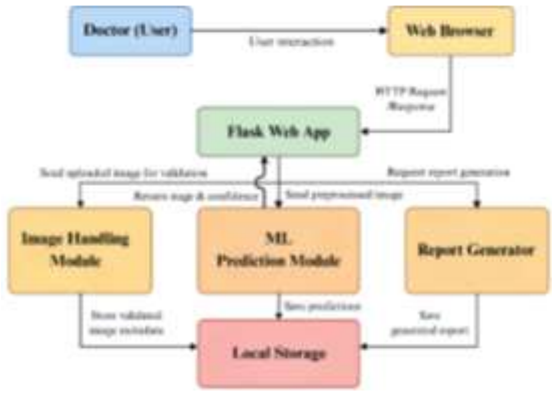

Figure 5.1: *System Architecture Diagram illustrating the end-to-end workflow of the deployed KOA prediction system using Flask.*

The process begins when the doctor interacts with the system through a web browser, which serves as the user interface for uploading medical images and requesting results. These interactions are directed to a Flask web application, which functions as the central control unit of the workflow. Once an image is uploaded, it is transferred to the Image Handling Module, where validation and preprocessing are carried out to ensure the image meets the required quality standards for analysis. The processed image is then sent to the Machine Learning (ML) Prediction Module, where advanced algorithms analyze the input and generate predictions regarding disease stage along with confidence scores. These results, together with the validated image metadata, are stored securely in Local Storage for future reference and traceability. When the doctor requests a report, the system triggers the Report Generator, which compiles the predictions, confidence levels, and supporting data into a structured, professional medical report. This report is also saved in Local Storage and made accessible to the doctor through the web application. By integrating image validation, machine learning analysis, data storage, and automated report generation into a seamless workflow, the system ensures accuracy, reliability, and efficiency, thereby providing valuable diagnostic support to healthcare professionals.

6. RESULTS AND DISCUSSION

The experimental evaluation of the proposed CNN-based KOA detection system was conducted on a curated knee X-ray dataset following systematic preprocessing steps, including grayscale conversion, resizing to 256 × 256 pixels, normalization to [0,1], and symmetry-guided augmentation emphasizing joint asymmetry. The dataset was divided into training (72%), validation (18%), and testing (10%) portions to guarantee unbiased assessment and strong generalization. The CNN was trained for 100 epochs with the Adam optimizer using categorical cross-entropy loss, while dropout, early stopping, and augmentation operations (rotation, flipping, and contrast adjustment) were applied to efficiently prevent overfitting.

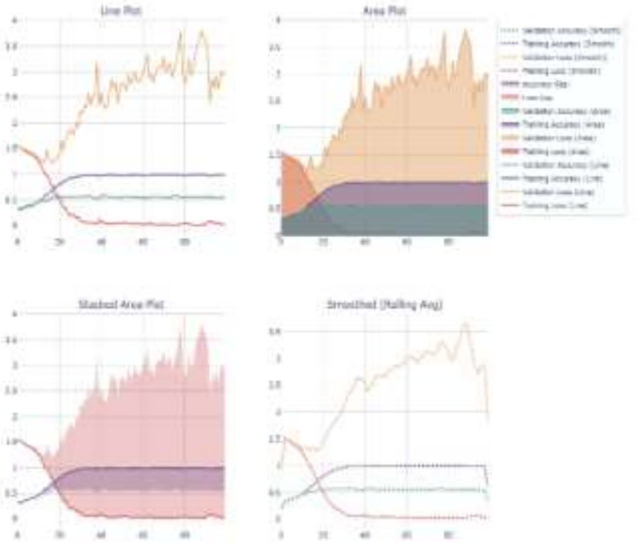

***Figure 6.1**: CNN model training performance graphical representation.*

As illustrated in Figure 6.1, the training accuracy exhibited a steady improvement, ultimately approaching 98% after approximately 60 epochs, while the validation accuracy leveled off around 92%, indicating effective learning without severe overfitting. Correspondingly, the training loss progressively decreased toward zero, and the validation loss plateaued after nearly 40 epochs with minor fluctuations, signifying convergence of the optimization process and stability in generalization. These outcomes confirm that the CNN successfully extracted radiographic features critical for KOA detection, including reductions in joint space, the appearance of osteophytes, and regions of subchondral sclerosis, while maintaining consistent performance on unseen data. To further assess the model's classification robustness, a confusion matrix was generated using the independent test set.

***Figure 6.2**: Confusion Matrix*

The classification distribution shown in Figure 6.2 indicates that the model performed exceptionally well in distinguishing the extreme categories, KL-0 (Normal) and KL-4 (Severe), with precision and recall values surpassing 95%. This reliability arises from the clear radiographic contrast between preserved joint space in Normal cases and the pronounced deformities characteristic of Severe osteoarthritis. Conversely, intermediate grades such as KL-1 (Doubtful) and KL-2 (Mild) displayed occasional overlap, reflecting the inherent diagnostic difficulty of borderline stages where structural variations are subtle and ambiguous. Despite these challenges, the system achieved an overall weighted F1-score of 0.91, confirming both robustness and clinical relevance. These outcomes are consistent with prior literature, which frequently reports misclassification among intermediate classes, yet the incorporation of symmetry-guided preprocessing in the proposed framework enhanced interpretability and feature discrimination when compared with conventional CNN baselines.

In addition to quantitative evaluation, the trained CNN was integrated into a Flask-based web application to validate its clinical applicability. Uploaded knee radiographs were automatically verified through grayscale intensity and variance checks to exclude invalid inputs such as photographs or non-medical images. Valid images underwent real-time preprocessing and were passed into the CNN, which produced a predicted KOA stage along with a confidence score within seconds. A structured PDF report was then generated, embedding the input X-ray, predicted grade, confidence percentage, and stage-specific description, along with a disclaimer clarifying its supportive role in clinical workflows. This automated pipeline enhanced accessibility for both clinicians and patients, reducing diagnostic subjectivity while ensuring interpretable results that supported decision-making. Performance analysis revealed that the framework achieved 92% validation accuracy and 90% test accuracy, making it comparable to or slightly below state-of-the-art transfer learning models. For example, DenseNet169-based systems have reported ~94% accuracy, while EfficientNetB5 achieves ~93.8%. Although such architectures may yield marginally higher accuracy, they require significantly greater computational resources, limiting their practicality for lightweight clinical deployment. By contrast, the proposed CNN offers a favorable balance between predictive strength, efficiency, and deployability, rendering it suitable for advanced hospitals as well as resource-limited healthcare settings. While intermediate grades such as KL-1 and KL-2 remained susceptible to occasional misclassification, reflecting the diagnostic difficulty of subtle radiographic variations, the system consistently captured disease-relevant features including joint space narrowing, osteophyte formation, and sclerosis. Potential extensions such as attention mechanisms, Grad-CAM visualizations, or multimodal data integration (e.g., MRI, patient history) could further enhance explainability and precision. Overall, the findings confirm that the proposed system not only delivers strong predictive performance but also translates effectively into clinical practice, supporting radiologists with consistent grading, assisting general practitioners in early screening, and empowering patients with accessible diagnostic insights. By combining accuracy, interpretability, and computational efficiency, the framework demonstrates its potential as a reliable AI-driven decision-support tool for knee osteoarthritis detection and management.

7. CONCLUSION

We have developed a CNN-based deep learning framework for the automatic detection and staging of knee osteoarthritis (KOA) from radiographic images. The dataset underwent preprocessing steps including grayscale conversion, resizing, normalization, and symmetry-guided augmentation before being divided into training, validation, and testing subsets. The model was trained for 100 epochs, where accuracy steadily improved from 29% in the first epoch to above 98% in later epochs, while validation accuracy stabilized near 92%, demonstrating robust convergence. Although mild oscillations in validation loss appeared after the 30th epoch, the network reliably captured disease-specific features such as joint space narrowing and osteophyte formation. Confusion matrix analysis confirmed high accuracy in distinguishing extreme grades (Normal and Severe), though overlap persisted in intermediate stages, consistent with challenges reported in clinical literature. Experimental comparisons showed that while DenseNet- and EfficientNet-based models report slightly higher accuracies (~94% and ~93.8%), our lightweight CNN achieved comparable results with substantially lower computational cost, making it a practical alternative for clinical applications, especially in resource-constrained environments. Nonetheless, occasional misclassification of borderline cases such as KL-1 and KL-2 highlights an avenue for refinement. Future directions include enhancing explainability through visualization methods such as Grad-CAM and incorporating attention mechanisms or multimodal data, including MRI scans and patient history, to improve classification precision. Overall, the proposed system demonstrates both predictive strength and clinical potential, establishing a reliable AI-assisted diagnostic framework capable of supporting radiologists, assisting general practitioners, and empowering patients with accessible and trustworthy diagnostic insights.

8. FUTURE ENHANCEMENT

The proposed CNN-based system for knee osteoarthritis detection and staging has demonstrated promising results, effectively capturing radiographic features such as joint space narrowing, osteophyte formation, and subchondral changes with high accuracy and reliability. The model, trained over one hundred epochs with careful preprocessing including grayscale conversion, resizing, normalization, and symmetry-based augmentation, achieved a training accuracy close to 98% and a validation accuracy above 90%, while the confusion matrix confirmed excellent performance in distinguishing extreme grades and acceptable robustness for borderline cases. The system was further integrated into a Flask-based web application, enabling real-time clinical deployment where X-ray images can be uploaded, preprocessed, classified, and reported with stage-specific predictions and confidence scores in an automatically generated PDF format. This functionality supports clinicians by reducing subjectivity and enhances patient understanding through clear, interpretable reports. Compared to transfer learning models such as DenseNet and EfficientNet, which achieve slightly higher accuracy, the proposed lightweight CNN offers the advantage of reduced computational complexity and faster inference, making it well suited for real-time use in resource-constrained healthcare environments. While the current results are encouraging, there is scope for future enhancements that can strengthen both accuracy and clinical applicability. Incorporating explainable AI methods like Grad-CAM or attention-based visualizations will provide interpretable heatmaps to highlight regions influencing predictions, thereby improving clinician trust. Integrating multimodal inputs such as patient demographics and clinical history could further refine predictions, especially for intermediate grades. Additionally, adopting federated learning approaches would enhance scalability and data diversity while preserving patient privacy, and optimization strategies such as pruning or quantization could make the model deployable on edge devices. Overall, the system establishes a reliable foundation for automated osteoarthritis detection, with future improvements poised to extend its adaptability, transparency, and clinical impact.

10. REFERENCES

[1] T. Thomas and S. Prince Mary, "A Comprehensive Review Analysis on Osteoarthritis Using Deep Learning and Machine Learning Technologies for Arthritis Detection," *2024 International Conference on Circuit Power and Computing Technologies (ICCPCT)*, pp. 600–606, IEEE, 2024.

[2] N. Sharma, R. Sapra, Sarita, and P. Dhaliwal, "A Comprehensive Review on Knee Osteoarthritis Detection using Medical Imaging and Machine Learning," *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*, IEEE, pp. 1–8, 2024.

[3] A. M., S. Krishnan H., S. Kumar T., N. Annamalai, and C. R. Teja, "A Deep Learning Framework for Diagnosing and Predicting Knee Osteoarthritis Utilizing Convolutional Neural Networks," *2024 First International Conference on Software, Systems and Information Technology (SSITCON)*, IEEE, pp. 1–7, 2024.

[4] Y. Du, R. Almajalid, J. Shan, and M. Zhang, "A Novel Method to Predict Knee Osteoarthritis Progression on MRI Using Machine Learning Methods," IEEE Transactions on NanoBioscience, vol. 17, no. 3, pp. 228–236, 2018.

[5] S. Mehta and S. Kaur, "Automated Knee OA Severity Prediction0 via CNN-Attention Hybrid Framework," in the International Conference on Electronics and Renewable Systems (ICEARS 2025), IEEE, 2025, pp. 1646–1650.

[6] X. Wang, S. Liu, and C. Zhou, "Classification of Knee Osteoarthritis Based on Transfer Learning Model and Magnetic Resonance Images," in Proceedings of the 2022 International Conference on Machine Learning, Control, and Robotics (MLCR), IEEE, 2022, pp. 67–71.

[7] T. Guhan, R.M. Abiraj, M. Muneshwaran, V. Sowndharya Lakshmi, and M. Suriyanarayanan, "Classification of Knee Osteoarthritis Using R-CNN Algorithm," in Proceedings of the International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI 2025), IEEE, 2025, pp. 1676–1680.

[8] G. Sharma, V. Anand, and V. Kumar, "Classification of Osteo-Arthritis with the Help of Deep Learning and Transfer Learning," in *Proceedings of the 5th International Conference on Inventive Research in Computing Applications (ICIRCA 2023)*, IEEE, pp. 446–452, 2023.

[9] T. Sivakumari and R. Vani, "Comparative Evaluation of Transfer Learning Models for Knee Osteoarthritis Severity Classification Using KL Grading Scale," in *Proceedings of the International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS 2023)*, IEEE.

[10] V. Kapoor, M. Mahajan, and S. Batra, "Comparative Study of Deep and Machine Learning Models in Knee Osteoarthritis Detection," *2025 International Conference on Automation and Computation (AUTOCOM)*, IEEE, pp. 79–83.

[11] S. Mehta and R. Bhardwaj, "Deep Learning-Driven Diagnosis CNN-LSTM Models for Accurate Grading of Knee Osteoarthritis," in *Proceedings of 2025 International Conference on Automation and Computation (AUTOCOM)*, pp. 1632–1636, IEEE, 2025.

[12] Y. Dalia, A. Bharath, V. Mayya, and S. Kamath S, "DeepOA: Clinical Decision Support System for Early Detection and Severity Grading of Knee Osteoarthritis," *2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, pp. 250–255, IEEE, 2021.

[13] B. Perumal, J. Deny, P. Nagaraj, V. Maneesha, Palanimet, K. Alekhya, and S. A. Reddy, "Detection of Osteoarthritis by using Multiple Edge Detections," *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1035–1039, IEEE, 2022.

[14] S. Sharma, S. S. Virk, and V. Jain, "Detection of Osteoarthritis using SVM Classifications," *Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE Conference ID: 37465, pp. 2997–3002, New Delhi, India, Mar. 2016.

[15] S. D. Chandu, P. Revathi, and N. A. S. Vinoth, "Discovering Knee Osteoarthritis Using CNN Enhanced with AlexNet," *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, IEEE.

[16] K. S. Raju, V. Amudha, and N. Meenakshisundaram, "Early Detection and Quantification of Osteoarthritis Severity in Knee Using Support Vector Machine with Improved Accuracy Compared to Convolutional Neural Network," *2023 8th International Conference on Science Technology Engineering and Mathematics*, IEEE, pp. 1–5.

[17] E. Jain, V. Kukreja, and A. Rathour, "Enhanced Diagnostic Precision in Knee Osteoarthritis: Leveraging MobileNetV2 for Deep Learning Classification," in *Proceedings of the Second International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS-2024)*, IEEE, pp. 572–576.

[18] X. Yang, K. Zou, P. Tang, and D. Dai, "HCGN: Hierarchical Convolution and Graph Network for Predicting Knee Osteoarthritis," *2024 IEEE International Conference on Medical Artificial Intelligence (MedAI)*, IEEE.

[19] V. V. Kishore, V. Kalpana, and U. B. Dosapati, "Interpretation of KOA by KL Grading System using Deep Learning," *2024 10th International Conference on Communication and Signal Processing (ICCSP)*, IEEE.

[20] L. R. Likhith, U. Kushal, M. C. Likith, L. R. Bhat, and H. Priyanka, "Machine Learning and Deep Learning for Knee Osteoarthritis Diagnosis: A Survey on Classification and Severity Grading," *2024 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, IEEE.

[21] H. Masood, A. A. Salam, E. Hassan, and M. Liaquat, "OSTEO-DOC: KL-Grading of Osteoarthritis Using Deep Learning," *2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, IEEE.

[22] S. Gowr, R. Balakrishna, V. Rajeshram, S. G. Karuppasamy, and S. P. Premnath, "Automated Knee Osteoarthritis Grading Using CNN and SSD Object Detection," *2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon)*, IEEE.

[23] J. Antony, K. McGuinness, N. E. O'Connor, and K. Moran, "Quantifying Radiographic Knee Osteoarthritis Severity using Deep Convolutional Neural Networks," *2016 23rd International Conference on Pattern Recognition (ICPR)*, IEEE.