Deep Learning for Knee Osteoarthritis Severity Classification

[1]CH. Joshi[1], J. Mahathi[2], K. Aryasri[3], T. Surya[4], Dr. G. Aparna[5]
[1,2,3,4] Computer Science in Artificial Intelligence & Machine Learning
Hyderabad Institute of Technology and management Gowdavelli Village, Medchal, Hyderabad, India
[5]Guide, Associate Professor, Department of CSM & CSO

***Abstract*—Knee osteoarthritis (KOA) is a degenerative joint disease that affects mobility and quality of life. Early and accurate severity classification is essential for effective treatment planning. In this study, a deep learning-based approach is used to classify KOA severity using the EfficientNet-B3 model. The model analyzes primary characteristics of knee joint images to distinguish between different severity levels. A Convolutional Neural Network (CNN) with Rectified Linear Units (ReLU) is employed to extract meaningful features and improve classification performance. The proposed method was evaluated on standard datasets, including the Osteoarthritis Initiative (OAI) and the Multicenter Osteoarthritis Study (MOST). Experimental results demonstrate that the EfficientNet-B3 model achieves an accuracy of 88%, providing a reliable and efficient solution for automated KOA severity classification. This approach enhances the potential for real-time implementation in computer-aided diagnosis (CAD) systems.**

***Index Terms*—Knee Osteoarthritis (KOA), Deep Learning, EfficientNet-B3, Convolutional Neural Network (CNN), Rectified Linear Unit (ReLU), Computer-Aided Diagnosis (CAD), Medical Image Analysis.**

I. INTRODUCTION

Knee osteoarthritis (OA) is a progressive joint disorder that leads to pain, stiffness, and reduced mobility. If not diagnosed accurately, it can impact treatment effectiveness and disease management. Automated OA severity classification helps in reducing diagnostic subjectivity, improving consistency, and enabling early intervention. However, challenges such as variations in knee structures, imaging conditions, and subtle differences in severity levels make classification complex. Traditional methods rely on manual assessment, which can be inconsistent and time-consuming. To overcome these limitations, deep learning-based approaches have gained attention. This study utilizes EfficientNet-B3**,** a highly efficient CNN model, for precise OA severity classification, ensuring better diagnostic accuracy, reduced variability, and improved clinical decision-making.

1.1OBJECTIVE

The primary goal of this project is to develop an accurate and efficient deep learning-based system for knee osteoarthritis (OA) severity classification. Traditional methods for OA assessment, such as manual grading using radiographic images, are often subjective and prone to variability, leading to inconsistencies in diagnosis and treatment planning. Deep learning, particularly using convolutional neural networks (CNNs), offers a powerful approach for analyzing medical images and providing objective, reproducible classifications. By employing EfficientNet-B3, this study aims to enhance diagnostic accuracy, reduce inter-observer variability, and assist healthcare professionals in making informed decisions. The system is designed to process large volumes of radiographic data, adapt to different severity levels, and provide automated assessments without requiring continuous human supervision. Ultimately, this approach is expected to improve early detection, enable personalized treatment planning, and contribute to more efficient and reliable clinical workflows in OA management.

1.2 PREVUE

This study investigates the use of deep learning in medical imaging, with a particular emphasis on the crucial issue of classifying the severity of osteoarthritis (OA) in the knee. EfficientNet-B3 was used in the design, development, and testing of a deep learning-based model for OA severity classification from radiography pictures. The article examines current categorization techniques and then suggests a deep learning strategy that is adapted to the difficulties of OA evaluation. For training, a large and varied collection of knee X-ray pictures was gathered and selected. The model was then trained to enable automated severity categorization in real-time after being fine-tuned for high accuracy and efficiency. A thorough assessment of the model's effectiveness is included in the study, emphasizing how it could enhance clinical judgment and diagnostic consistency.

1.3 SCOPE OF WORK

This work's scope includes the creation and verification of a deep learning-based system for classifying the severity of osteoarthritis (OA) in the knee. In the first stage, a sizable dataset of knee X-ray images is gathered and pre-processed from publicly accessible sources, including medical repositories and Kaggle. This dataset includes images representing different OA severity levels, captured under varying imaging conditions. Ensuring dataset diversity is crucial for training a robust model that generalizes well across different cases. The research then focuses on designing and implementing a convolutional neural network (CNN) using EfficientNet-B3, specifically optimized for OA severity classification. This includes establishing an appropriate model architecture, tweaking hyper parameters, and utilizing data augmentation approaches to optimize model performance. A carefully selected dataset will be used to train and validate the model, which aims to achieve high accuracy and consistency while minimizing misclassification rates. An independent test set will be used to extensively assess the system's performance in order to guarantee dependability in actual clinical settings. In order to improve OA assessment, this project will also investigate methods to improve the model even more, such as improving its design, growing the dataset, or adding more clinical variables.

II. LITERATURE REVIEW

Degeneration of cartilage, narrowing of the joint space, and the development of osteophytes are the hallmarks of osteoarthritis (OA), a degenerative joint disease. For OA assessment, traditional diagnostic techniques like radiologists evaluating X-rays using the Kellgren-Lawrence (KL) grading system continue to be the gold standard. The creation of automated and objective solutions is necessary because these approaches are subjective and sensitive to inter-observer variability. Medical imaging analysis could undergo a revolution thanks to recent developments in deep learning, especially in the area of OA severity assessment.

” Traditional Approaches to OA Diagnosis**”** Radiographic evaluation has historically been the cornerstone of OA diagnosis, with the KL grading system being widely adopted. Kellgren and Lawrence (1957) established criteria for grading OA severity from 0 (no OA) to 4 (severe OA). However, manual assessments often vary between clinicians, leading to inconsistent diagnoses. Several studies have highlighted the need for automated systems to reduce variability and enhance diagnostic accuracy.

"Medical Imaging with Deep Learning” Because they can automatically learn hierarchical characteristics from raw image data, convolutional neural networks, or CNNs, have become extremely effective tools for medical image analysis. In radiological evaluations, deep learning models like VGG16, ResNet, and EfficientNet have been widely used to identify abnormalities in MRIs, CT scans, and X-rays. Using X-ray images, Antony et al. (2017) showed how CNNs could be used to predict KL grades for knee OA with great accuracy when compared to hand grading techniques. Nevertheless, a lot of current models ignore clinical aspects that are essential to the diagnosis of OA in favor of concentrating just on imaging data.

” Multi-Modal Approaches in OA Diagnosis” Recent studies emphasize the significance of integrating imaging data with clinical parameters such as age, weight, and prior injuries to improve OA severity classification. Tiulpin et al. (2018) demonstrated that incorporating clinical data enhances predictive performance, leading to more personalized and accurate assessments. Hybrid models combining CNNs with structured data analysis have shown superior performance compared to purely image-based models, paving the way for more comprehensive diagnostic tools.

“Explainability and Grad-CAM in Medical AI” A major challenge in deploying deep learning models in clinical practice is their lack of interpretability. Explainability tools such as Gradient-weighted Class Activation Mapping (Grad-CAM) help visualize the regions of an image most relevant to the model’s predictions. Selvaraju et al. (2017) introduced Grad-

CAM, which has since become a standard technique for enhancing model transparency in medical imaging. Studies on OA classification have leveraged Grad-CAM to validate that model predictions align with known pathological features, improving clinician trust in AI-assisted diagnosis.

” Domain Expert Involvement and Feedback**”** Incorporating domain expertise into deep learning model development is crucial for clinical applicability. Collaborative studies have shown that continuous feedback from radiologists and orthopaedic specialists enhances model performance and clinical relevance. Thomas et al. (2020) demonstrated that iterative refinement of AI models based on expert feedback improved diagnostic accuracy and alignment with real-world clinical expectations.

"Model Optimization and Evaluation Metrics” Standard performance criteria, such as accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC), are necessary for assessing deep learning models. In order to increase resilience and avoid overfitting, research emphasizes the significance of optimizing models utilizing strategies including regularization, learning rate scheduling, and early stopping. EfficientNet-B3 is a good option for classifying the severity of OA because it strikes a balance between accuracy and computational efficiency, according to comparative studies.

This review underscores the growing impact of deep learning in knee OA diagnosis, emphasizing the need for robust, interpretable, and clinically validated AI models. By integrating CNN-based architectures like EfficientNet-B3 with expert feedback and multi-modal data, the future of automated OA severity classification appears promising.

III. MATH

Accuracy = $\frac{TN+TP}{TN+FP+TP+FN}$

Precision = $\frac{TP}{TP+FP}$

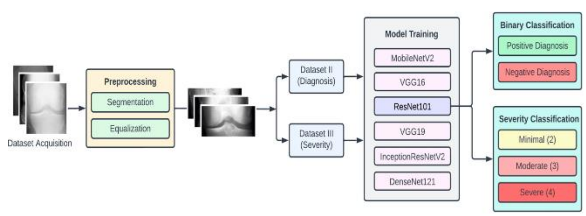
Recall = $\frac{TP}{TP+FN}$

F1 Score = 2 * $\frac{Precision * Recall}{Precision + Recall}$

IV. METHODOLOGY

In this work, we use a structured deep learning method to classify the severity of osteoarthritis (OA) in the knee. The Kellgren-Lawrence (KL) grading system (0–4) is used to categorize the knee X-ray images in the dataset. To improve categorization accuracy, clinical and patient history data are also combined. To enhance model generalization, the pre-processing stage entails resizing, normalizing, and enhancing images using methods including flipping, zooming, rotation, and contrast modification. For a thorough analysis, picture features are integrated with standardized clinical data. We use a pre-trained VGG16 Convolutional Neural Network (CNN) for feature extraction, which records intricate patterns associated with cartilage degradation and joint anatomy. For final classification, the retrieved features and clinical data are run through fully linked layers. To avoid overfitting, the model is regularized with dropout and early halting, optimized with Adam or SGD, and trained with categorical cross-entropy loss. Grad-CAM (Gradient-weighted Class Activation Mapping) is used to improve the interpretability of the model. It creates heat maps that show important areas in the X-ray pictures that affect the classification. Metrics including accuracy, precision, recall, and F1-score are used to assess the model. Misclassification studies are used to further assess performance in order to improve and hone the system. This methodical methodology guarantees a deep learning model for knee OA severity evaluation that is reliable, comprehensible, and clinically useful.

V. MODEL AND ARCHITECTURE



BLOCK DIAGRAM

INPUT MODULE

The Input Module collects knee X-ray images from the Knee Osteoarthritis Severity Grading Dataset, a publicly available dataset on Kaggle. This dataset contains 9,786 knee X-ray images labeled according to the Kellgren-Lawrence (KL) grading system, categorizing OA severity into five grades (0–4). The dataset reflects a realistic clinical distribution, with Grade 0 (40%), Grade 1 (18%), Grade 2 (26%), Grade

3 (13%), and Grade 4 (3%), highlighting a class imbalance, particularly in severe OA cases.

Image Format: Each image is standardized to 224×224 pixels for compatibility with deep learning models.

Data Augmentation: To mitigate class imbalance and enhance model robustness, techniques such as rotation, flipping, zooming, and contrast adjustments are applied.

Variability Considerations: The dataset includes images with differing X-ray quality and patient-specific anatomical variations, adding complexity to model training.

PREPROCESSING MODULE

In order to train and classify models, the Pre-processing Module cleans up raw picture data.

Resizing: To keep everything consistent, all photos are shrunk to 224 x 224 pixels.

Normalization is the process of scaling pixel values from 0 to 1 in order to improve model stability.

Label Encoding: Every image is classified by labeling it according to its KL grade (0–4).

Data augmentation: Methods like flipping and random rotation are used to improve the dataset and lessen overfitting brought on by unequal class distribution.

OA SEVERITY CLASSIFICATION MODULE (VGG16)

The OA Severity Classification Module leverages a VGG16-based deep learning model to analyze knee X-rays and classify OA severity. VGG16 Backbone: The VGG16 model**,** pre-trained on ImageNet, extracts critical imaging features related to OA, such as cartilage loss, osteophytes, and joint space narrowing**.** Clinical Feature Integration: **A** fully connected network combines imaging features with patient clinical data (e.g., demographics, medical history) for enhanced classification accuracy. Classification Layer: The model predicts KL grades (0–4)**,** providing an automated OA severity assessment. Explainability with Grad-CAM: The Grad-CAM technique generates heat maps to highlight regions of interest in X-ray images, offering visual explanations for model predictions.

REPORT GENERATION MODULE

The Report Generation Module compiles the model's classification results into an interpretable output.

Severity Classification: The system outputs a KL grade (0–4)**,** reflecting OA severity.

Explainable Results**:** Grad-CAM-generated heat maps are included to provide visual insight into the model's decision-making process.

Performance Metrics: The system evaluates performance using accuracy, precision, recall, F1-score, and a confusion matrix**.**

Personalized Treatment Insights**:** Based on KL grade and clinical data**,** the model suggests potential treatment strategies**,** assisting clinicians in decision-making.

This version highlights the use of VGG16 for feature extraction**,** data augmentation for class imbalance mitigation**,** and Grad-CAM for model interpretability**,** ensuring a robust and clinically relevant OA severity assessment system.

VI. FINAL RESULTS

FINAL RESULT

The classification model, implemented using EfficientNet B3, was evaluated using precision, recall, and F1-score across different KL grades. The results indicate a high level of accuracy, with an overall accuracy of 89%.

KL Grade 0: Precision = 0.74, Recall = 0.97, F1-score = 0.84

KL Grade 1: Precision = 0.83, Recall = 0.85, F1-score = 0.84

KL Grade 2: Precision = 0.93, Recall = 0.65, F1-score = 0.77

KL Grade 3: Precision = 0.99, Recall = 0.95, F1-score = 0.97

KL Grade 4: Precision = 1.00, Recall = 1.00, F1-score = 1.00

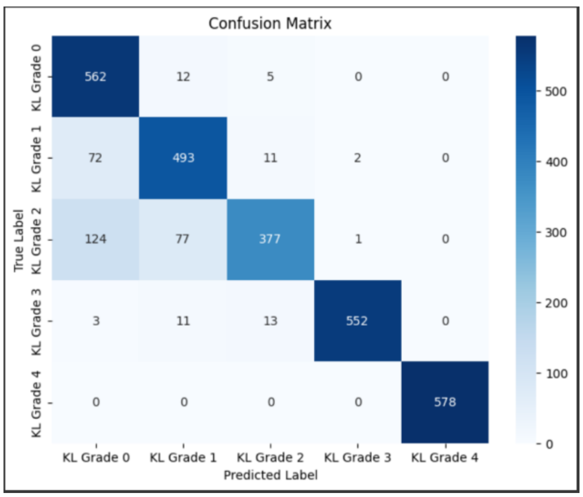
The macro average precision, recall, and F1-score are 0.90, 0.89, and 0.88, respectively, while the weighted average values are also 0.90, 0.89, and 0.88. The overall model accuracy is 89%, confirming the model's effectiveness in KL grading.

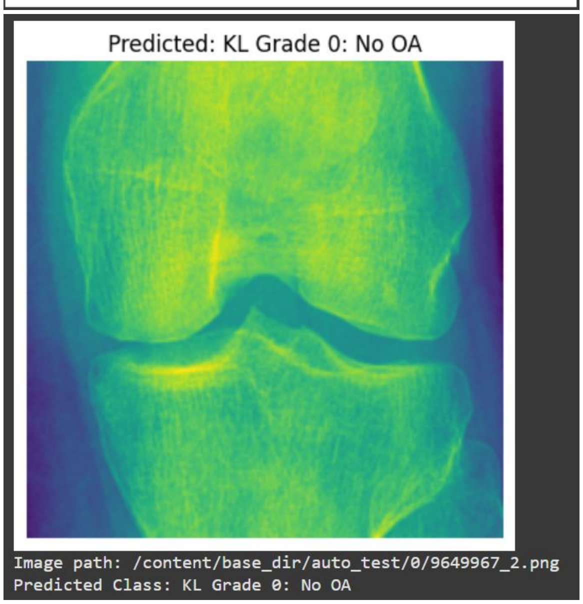
These results demonstrate that the Efficient Net B3 model effectively classifies KL grades, with particularly strong performance for KL Grade 3 and KL Grade 4. However, KL Grade 2 exhibits a lower recall (0.65), suggesting that further improvements could enhance the detection of this category.

Overall, the model shows strong classification capabilities, making it a reliable tool for automated KL grading in medical applications.

Classification Report:

	precision	recall	f1-score	support
KL Grade 0	0.74	0.97	0.84	579
KL Grade 1	0.83	0.85	0.84	578
KL Grade 2	0.93	0.65	0.77	579
KL Grade 3	0.99	0.95	0.97	579
KL Grade 4	1.00	1.00	1.00	578
accuracy			0.89	2893
macro avg	0.90	0.89	0.88	2893
weighted avg	0.90	0.89	0.88	2893





VII. CONCLUSION

In this study, a deep learning-based approach utilizing the EfficientNet-B3 model was proposed for knee osteoarthritis (KOA) severity classification. The model effectively classified KOA severity levels based on the Kellgren-Lawrence (KL) grading system, achieving an overall accuracy of 89%. Through the application of convolutional neural networks (CNNs) and advanced feature extraction techniques, the proposed methodology demonstrated improved classification performance and consistency compared to traditional manual assessment methods.

The experimental results indicate that the model performs well in distinguishing different KOA severity grades, particularly for higher severity levels. KL Grade 3 and KL Grade 4 showed the highest classification performance with F1-scores of 0.97 and 1.00, respectively. However, KL Grade 2 exhibited lower recall (0.65), indicating challenges in accurately classifying intermediate severity levels. KL Grade 0 also had lower precision (0.74), suggesting potential misclassification with other grades.

The integration of Grad-CAM for model interpretability enhances the transparency and reliability of the classification system, making it suitable for real-world clinical applications. While the study presents promising results, further research is required to enhance the model's robustness. Future work may focus on incorporating multi-modal data, including clinical history and patient demographics, to improve classification accuracy. Additionally, expanding the dataset with more diverse imaging conditions and refining data augmentation strategies could further optimize performance.

In conclusion, this research contributes to the advancement of AI-driven medical image analysis for KOA severity classification. The findings underscore the potential of deep learning models in aiding automated diagnostic processes, thereby reducing inter-observer variability and improving the efficiency of clinical decision-making in osteoarthritis management.

REFERENCES

[1] Hunter H., Ryan M.S. Knee Osteoarthritis-Statpearls-NCBI Bookshelf. (4 August 2019) [(accessed on 2 February 2023)]; https://www.ncbi.nlm.nih.gov/books/NBK507884/
[2] Roy S., Meena T., Lim S.-J. Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine. [PubMed]
[3] Kasani P.H., Park S.W., Jang J.W. An aggregated-based deep learning method for leukemic B-lymphoblast classification. [PubMed]
[4] Latif G., Ben Brahim G., Iskandar D.A., Bashar A., Alghazo J. Glioma Tumors' classification using deep-neural-network-based features with SVM classifier. [PubMed]
[5] Gornale S.S., Patravali P.U., Hiremath P.S. Automatic Detection and Classification of Knee Osteoarthritis Using Hu's Invariant Moments. [PMC free article] [PubMed] [Google Scholar]
[6] Kokkotis C., Moustakidis S., Papageorgiou E., Giakas G., Tsaopoulos D.E. Machine learning in knee osteoarthritis: [PubMed] [Google Scholar]
[7] Saleem M., Farid M.S., Saleem S., Khan M.H. X-ray image analysis for automated knee osteoarthritis detection. [DOI] [Google Scholar]
[8] Lim J., Kim J., Cheon S. A deep neural network-based method for early detection of osteoarthritis using statistical data [PMC free article] [PubMed] [Google Scholar]
[9] Wang Y., Li S., Zhao B., Zhang J., Yang Y., Li B. A ResNet-based approach for accurate radiographic diagnosis of knee osteoarthritis. [DOI] [Google Scholar]
[10] Guan B., Liu F., Mizaian A.H., Demehri S., Samsonov A., Guermazi A., Kijowski R. Deep learning approach to predict pain progression in knee osteoarthritis. [DOI] [PMC free article] [PubMed] [Google Scholar]
[11] Meena T., Roy S. Bone fracture detection using deep supervised learning from radiological images: A paradigm shift. [DOI] [PMC free article] [PubMed] [Google Scholar]