

Automated Knee Osteoarthritis Classification from X-ray Images Using the VGG-16 Model

Suman Rani^{1,2*}, Minakshi Memoria³, Kapil Joshi⁴ and Anil Kumar Lamba⁵

¹Department of Computer Science, Uttarakhand University, Dehradun, India.

²Department of Information Technology, IIMT College of Engineering, Greater Noida, India.

³Department of Computer Science, King Khalid University, Saudi Arabia.

⁴Department of Computer Science, Uttarakhand University, Dehradun, India.

⁵ Department of Computer Science, MMEC, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, India.

*Corresponding Author E-mail: sumanaggrawal@gmail.com

<https://dx.doi.org/10.13005/bpj/3083>

(Received: 26 November 2024; accepted: 13 January 2025)

Knee arthritis is the most frequent ailment among the senior population. This illness affects a large number of people worldwide. The biggest issue is with the joints. The higher joints are the femur, the lower joints are the tibia, and the patella is the kneecap. There is cartilage loss, which causes a difficulty with mobility. To diagnose this condition manually, knee scans are analysed and divided into five groups using the Kellgren-Lawrence (KL) approach. This process requires extensive healthcare expertise, significant experience, and considerable time, yet it remains susceptible to errors. Consequently, the era of artificial intelligence has arrived. AI is making a profound impact on the healthcare sector. This work classified the KOA using the publicly accessible OAI (Osteoarthritis Initiative) dataset. This work primarily uses deep learning, a specialisation of AI, for the categorisation and severity detection of the Knee Osteoarthritis. This study primarily uses the VGG-16 DNN model for binary classification as well as multiclassification. Using this model results in optimised efficiency and higher accuracy than previous models. In the future, we will work with genuine data collected from numerous hospitals.

Keywords: OAI, CNN, Deep Neural Networks, KOA, VGG-16, X-Ray.

The cartilage that lies between the knee joints deteriorates with time, leading to knee osteoarthritis (KOA), A health issue that mainly impacts older individuals. ¹In the United States, it is the most prevalent joint condition, affecting 10% of male and 13% of female over 60. Knee osteoarthritis (KOA) is caused by cartilage loss in the knees and can cause years of disability. ²An arthroplasty is the final choice for healing from

this condition, but it is prohibitively costly and not accessible to all. The purpose of this study is to automatically detect early stages arthritis in the knee. Using the wealth of available data, this study attempts to examine hitherto unknown aspects of healthcare decision-making and precise diagnosis in the field of KOA. ³ The Kellgren and Lawrence (KL) grading scheme is a widely used method in knee osteoarthritis (KOA) research. ⁴It assigns

ordinal grades to radiographs according to the degree of osteoarthritis, with Grade 0 showing no evidence of KOA and Grade 4 indicating substantial joint degradation.⁵ This semiquantitative technique enables researchers and doctors to assess the severity of knee osteoarthritis (KOA) using particular radiographic signs such as subchondral bone alterations, osteophyte formation, and joint space narrowing.⁶ Furthermore, X-rays, which are readily available and reasonably cheap, can be used to detect bone spur growth, articular cartilage degeneration, and an increase in the space between bones. The KL scale is a popular radiographic classification approach that was created to measure the amount of osteoarthritis in the KOA knee using X-rays. The severity of osteophyte and joint space narrowing (JSN) symptoms divides KOA into five categories. Figure 1 depicts the knee with and without KOA.

As it demonstrates⁷, the creation of osteophytes and the loss of cartilage between the joint of the knee and the bony spur are among the many symptoms of KOA. These symptoms are primarily brought on by ageing or any specific illness.⁸ These symptoms are primarily present in females. The main causes of KOA are shown in fig. 2.

KL grade system is shown in table 1. There are four categories of KL Grade.

⁹This study primarily utilizes a deep learning approach. we have studied various papers based on healthcare solution using the DL. This study applies the VGG-16 model. The main contribution of the study is:

- Understanding the symptoms and causes of osteoarthritis in the knee is the primary contribution of this study. This study primarily uses X-ray pictures from the OAI collection.

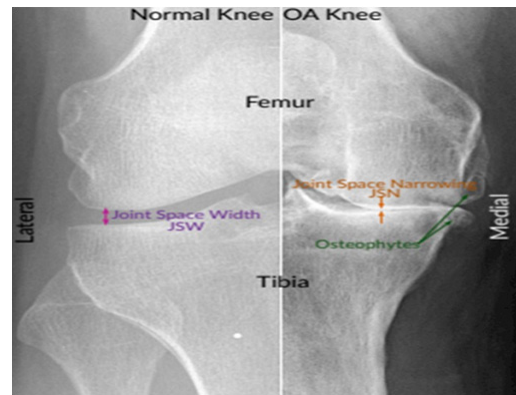


Fig. 1. Normal Knee and with with KOA

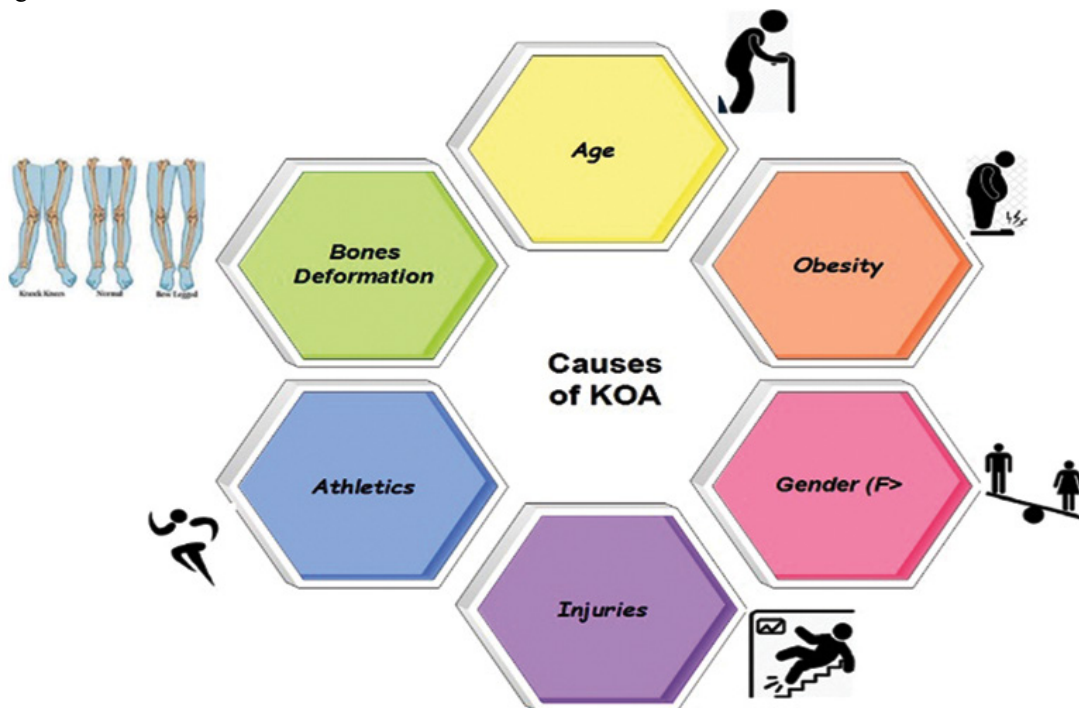


Fig. 2. Causes of KOA [8]

- In this study, the VGG-16 model, a deep learning method, is employed for binary classification to assess whether an individual has the condition or not.
- Using the VGG-16 model, we performed multiclassification at the next level.

The paper’s format is as follows: Part 1 presents the introduction, Section 2 consists of a literature review, and Section 3 displays the methodology section. Section 4 presents the experimental results, Section 5 examines and analyses them, and Section 6 concludes.

Review of literature

Several studies have examined sophisticated machine learning and deep neural network models for the diagnosing and classifying knee osteoarthritis (KOA).⁹According to the Kellgren and Lawrence (KL) scale, one study’s 95.1% accuracy rate in detecting and rating KOA using an ensemble model and InceptionResNetV2, which has the potential to be a trustworthy tool for radiologists. For binary and severe classification, a 12-layer Convolutional Neural Network (CNN) was constructed using Osteoarthritis Initiative

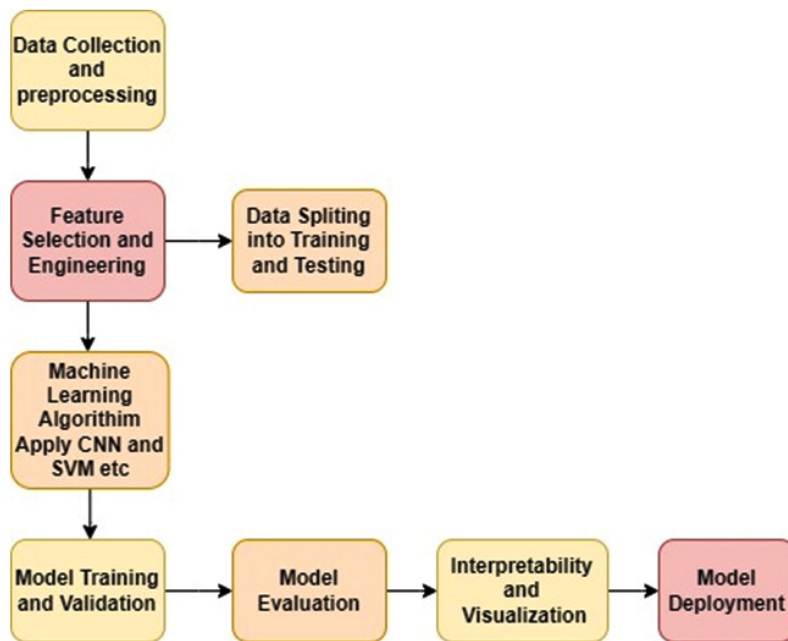


Fig. 3. General Methodology for KOA detection and classification

Table 1. KL Grading Levels

Phase 0: Healthy	No signs of osteoarthritis (OA) are visible
Phase 1: Doubtful	X-rays may not show clear damage, but small bone growths might indicate early OA. Pain or discomfort is generally absent, and the joint appears normal on X-ray.
Phase 2: Mild	X-rays may show bone spurs and joint space narrowing. Mild joint stiffness and pain might begin. Bone and tissue junctions become denser, and bones thicken. A thin bone layer grows beneath cartilage, and interstitial fluid helps reduce resistance, improving mobility
Phase 3: Moderate	X-rays reveal more joint space narrowing, bone spur growth, and potential bone deformities. Activities like walking or bending may cause soreness. Tissues thin, and bones start developing outgrowths as OA progresses
Phase 3: Severe	X-rays show significant joint space reduction, causing bones to grind against each other. Tissue damage is extensive, with cartilage erosion and visible bone deformities.

Table 2. Literature Review of various studies

Schemens	Algorithms/Methods	Discovery	Cobnclusion
[9]	Ensemble Model (Xception + InceptionResNetV2)	<ul style="list-style-type: none"> Combined Xception and InceptionResNetV2 architectures for enhanced predictive accuracy, • 95.1% accuracy 	<ul style="list-style-type: none"> • The ensemble model demonstrated strong potential for accurate early-stage knee OA detection • offering a reliable tool for radiologists. • contributing significantly to medical image analysis in osteoarthritis diagnosis.
[10]	12-layer Convolutional Neural Network (CNN)	<ul style="list-style-type: none"> • Utilized CNN on Osteoarthritis Initiative (OAI) data for binary classification of KOA and severity classification using the Kellgren-Lawrence (KL) grading scale. • Achieved 92.3% accuracy in binary classification and 78.4% in multiclassification 	<ul style="list-style-type: none"> • The model demonstrated high accuracy for KOA detection and severity. • Grading suggests the potential for wider application and reduced need for radiologist intervention. • Future work aims to improve performance through diverse datasets.
[11]	Random Forests (RF), Support Vector Machines (SVM), Logistic Regression (LR), Decision Tree (DT), Bayesian (Bayes), Convolutional Neural Network (CNN) with Class Activation Mapping (CAM)	<ul style="list-style-type: none"> • Utilized semi-quantitative knee radiograph assessments, MRI assessments. • Tested various machine learning models, finding no significant performance difference across imaging data. • CAM identified cartilage loss (30.6%) and synovitis/effusion (30.9%) as key factors related to pain severity. 	<ul style="list-style-type: none"> • Machine learning models, particularly CNN with CAM, effectively identify structural factors associated with pain severity in knee osteoarthritis. • Cartilage loss and synovitis/effusion were highlighted as key factors, suggesting machine learning's potential in pain assessment and severity prediction from radiographic data.
[12]	Two-stage Balanced Random Forest model	<ul style="list-style-type: none"> • Classified future KOA severity into three grades (KL01, KL2, KL34) using two binary classifications. • The model, trained with demographic and quantitative knee morphology data from 8-year longitudinal data of 1213 knees, identified baseline KL grade and weight as key features. • Achieved a weighted F1 score of 79.0%, balanced accuracy of 65.9%, and AUC of 83.0% and 86.6% for the two stages. 	<ul style="list-style-type: none"> • The two-stage classification model significantly outperformed traditional single-stage models in KOA severity classification. • showing potential for future application in clinical settings for more accurate severity prediction and early intervention.
[13]	High-Resolution Network (HRNet) with Attention Mechanism, Gradient-based Class Activation Maps (Grad-CAMs)	<ul style="list-style-type: none"> • Developed OsteoHRNet, a deep learning model based on HRNet to classify Knee OA severity according to the Kellgren and Lawrence (KL) grading system from X-ray images. • The model achieved a multi-class accuracy of 71.74% and Mean Absolute Error (MAE) of 0.311 on the OAI dataset, surpassing prior methods. 	<ul style="list-style-type: none"> • OsteoHRNet's multi-scale feature extraction and attention mechanism enhanced KL grade classification accuracy. • Demonstrating its potential as an effective tool for automated Knee OA severity assessment. Grad-CAMs provided interpretability. • helping to validate the network's learning process.
[14]	Five fine-tuned state-of-the-art (SOTA) DL models with Gradient-weighted Class Activation Mapping (GradCAM)	<ul style="list-style-type: none"> • Applied divide-and-conquer approach to classify knee OA using both multi-class and binary-class methods. • enhancing interpretability through GradCAM. EfficientNetb7 model achieved 99.13% accuracy for distinguishing normal and severe cases, but only 67% for intermediate cases. 	<ul style="list-style-type: none"> • The study highlights the potential of XAI in knee OA classification. EfficientNetb7 showed high accuracy for extreme cases but struggled with intermediate classifications. • suggesting a need for further model refinement to match clinician accuracy.

[15]	Transfer learning with Inception V3, compared with Xception, VGG16, VGG19, ResNet (various versions), DenseNet (various versions)	<ul style="list-style-type: none"> • Inception V3 achieved the highest performance, with 91% training accuracy and 67% testing accuracy. • effectively identifying OA severity based on the Kellgren–Lawrence scale and predicting knee replacement needs. 	<ul style="list-style-type: none"> • The Inception V3 model demonstrates strong potential for supporting radiologists in knee OA diagnosis. • Treatment planning, facilitating accurate, timely healthcare decisions for OA patients.
------	---	--	---

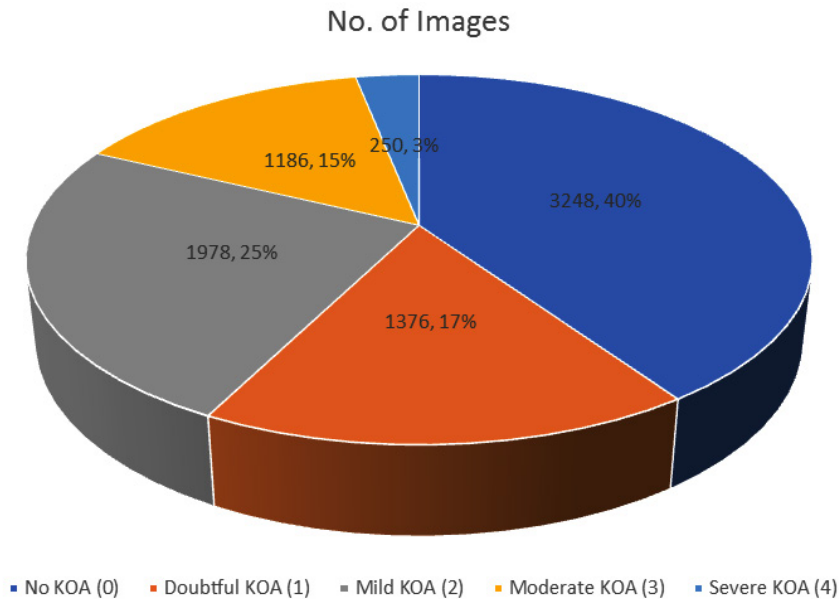


Fig. 4. OAI Dataset

(OAI) data.¹⁰ It proved highly effective in detecting and grading KOA, achieving an accuracy of 78.4% in multiclassification and 92.3% in binary classification. In another work, many machine learning models—including Random Forest, Support Vector Machine, and Bayesian classifiers—has been contrasted using several imaging modalities. The model showed promise in predicting the pain level by identifying cartilage loss and synovitis/effusion as significant pain-related indicators, despite the fact that there were no appreciable differences in model performance.¹¹ An extra study created a two-stage Balanced Random Forest model with a balanced accuracy of 65.9% and an F1 score of 79.0%, which classified KOA severity into three stages. This strategy performed better than previous single-stage methods, indicating that it might be used to predict the severity of KOA accurately.¹² With a multi-class accuracy of 71.74% and a Mean Absolute Error (MAE) of 0.311, the OsteoHRNet model—which

is based on the Excellent quality Network (HRNet) architecture and incorporates a focus mechanism and Gradient-based Class Activation Maps (Grad-CAMs)—proved its efficacy in automatic KOA inspection. For interpretable KOA classification, another study used GradCAM in conjunction with five state-of-the-art deep learning models.¹³ It achieved 99.13% accuracy in differentiating between normal and severe cases but only 67% for intermediate cases, highlighting the need for further model improvement. Finally, research using transfer learning with the Inception V3 model compared it to other deep learning architectures, reaching 91% training accuracy and 67% testing accuracy.¹⁴ This model showed potential in predicting KOA severity and assessing knee replacement demands, highlighting its utility in assisting radiologists with KOA diagnosis and treatment.¹⁵ Table -2 Provide a detailed description of the literature review.

MATERIALS AND METHODS

The four primary phases of this methodology are data collection, dataset preparation, model training, and classification. First, we obtained the KOA X-ray imaging dataset from the Osteoarthritis Initiative (OAI), which is available on Kaggle.¹⁵ This dataset has five different picture classifications labelled as 0 (healthy), 1 (doubtful), 2 (mild), 3 (moderate), and

4 (severe). The dataset’s knee joint X-ray pictures lack the quality and localization required to be directly fed into deep learning models. To improve the images a data preprocessing stage is required, with an emphasis on the joint area where useful KOA information is most likely to be detected. Initially, image segmentation included cropping each image to isolate the desired knee region and efficiently deleting any superfluous elements.

Fig. 3 depicts a methodical approach to designing and implementing a machine learning model in healthcare, specifically for medical imaging activities. The procedure begins with data collection and preprocessing, which involves

Table 3. Binary Classification Results

F1 Score	Recall	Precision	Accuracy	Loss
89.8	95.6	84.6	88.8	24.9

Table 4. Classification report of Binary classification using VGG-16

Class	Precision	Recall	F1-Score	Support
0	0.95	0.81	0.88	151
1	0.85	0.96	0.90	161
Macro Avg.	0.90	0.89	0.89	312
Weighted Avg.	0.89	0.89	0.89	312

Table 5. VGG-16 model config

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 7, 7, 512)	14714688
flatten (Flatten)	(None, 25088)	0
dropout (Dropout)	(None, 25088)	0
dense (Dense)	(None, 64)	1605696
dense_1 (Dense)	(None, 3)	195

Total params: 16320579 (62.26 MB)
 Training parameters: 1605891 (6.13 MB)
 Non-training parameters: 14714688 (56.13 MB)



Fig. 5. Dataset Images

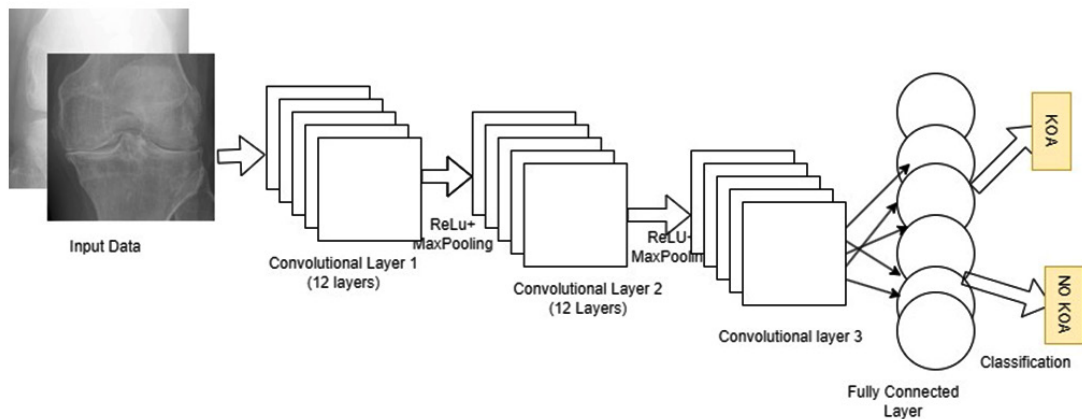


Fig. 6. CNN Architecture for KOA Classification

gathering and cleaning medical pictures. The next step is feature selection and engineering, which identifies critical data features for improved accuracy. The dataset is then divided into training and test sets.¹⁶ In the Machine Learning Algorithm stage, algorithms such as CNN and SVM are employed to extract patterns from data. Training and validation improve the model's parameters. Model evaluation evaluates accuracy using test data. Once the model has performed successfully, Interpretability .

Dataset Description

The knee X-ray images used to train the model in this work came from the Knee Osteoarthritis Severity Grading dataset, which is accessible on Kaggle and organised by the Osteoarthritis Initiative (OAI). The dataset includes 9,786 knee pictures classified into five severity levels using the Kellgren-Lawrence (KL) grading system: 0 (healthy), 1 (doubtful), 2 (minimum), 3 (moderate), and 4 (severe). Each picture has a resolution of 224 × 224 pixels. The

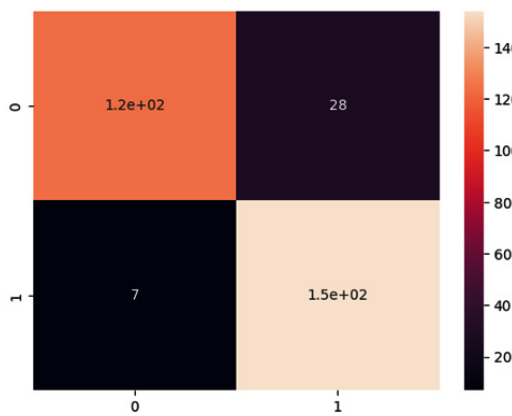


Fig. 7. Confusion Matrix for binary classification

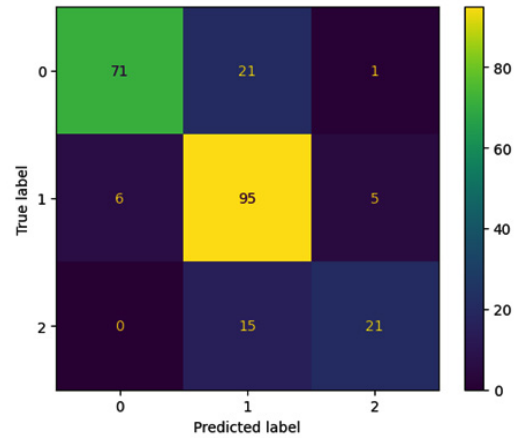


Fig. 8. Confusion matrix for Multiclassification using VGG-16

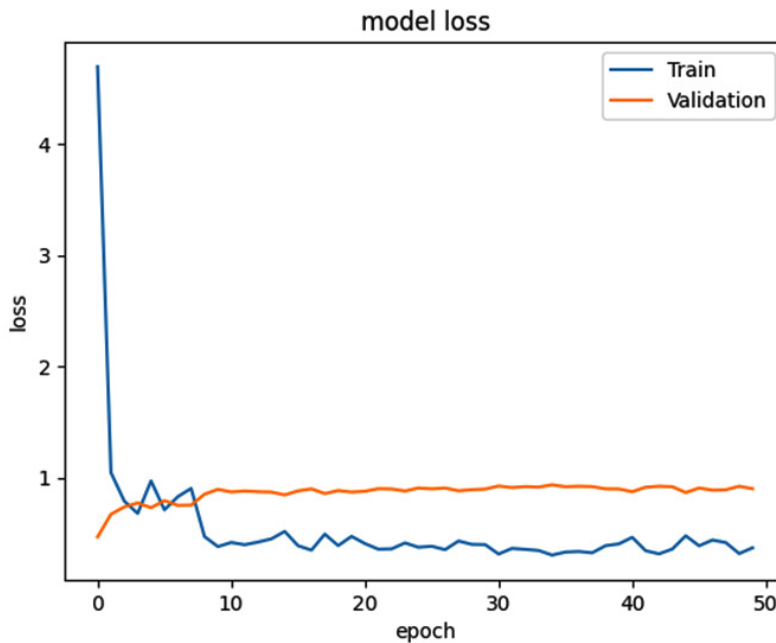


Fig. 9. Accuracy Loss Curve

Table 6. Multi-Classification report

Class	Precision	Recall	F1-Score	Support
0	0.92	0.76	0.84	93
1	0.73	0.90	0.80	106
2	0.85	0.96	0.90	36
Macro Avg.	0.81	0.75	0.77	235
Weighted Avg.	0.81	0.80	0.79	235

distribution of images across categories is as follows: approximately 40% are categorized as healthy, 18% as dubious, 26% as minimal, 13% as moderate, and just over 3% as severe. Fig. 4 describe the OAI dataset.

Dataset images are given in Fig.5. These images are based on the classes of KL grade system.

The distribution is significantly skewed, with the majority of photos in the “No KOA” and “Mild KOA” categories and only a handful in the “Severe KOA” category. This imbalance may have an influence on the model’s performance, particularly in properly recognising severe KOA instances.

Convolution Neural Network Architecture

Artificial intelligence (AI) has advanced rapidly in recent years, with applications in a variety of industries, including computer vision. The primary aim of computer vision is to make it possible for computers to understand visual information much like people do. The goal is to digitally capture, collect, and process pertinent environmental data. Many techniques have been developed to do this, with convolutional neural networks (CNNs) showing the most promise. In the context of image processing, a CNN is a deep learning algorithm that analyses a picture and assigns weights to various attributes to differentiate it from previous images processed by the same method.¹⁷ CNN architecture uses the VGG-16 model. We can say that the VGG-16 model is subpart of CNN as there is a difference in the number of layers. The VGG-16 model is a deep convolutional neural network (CNN) architecture developed by Oxford University’s Visual Geometry Group, and it is well-known for its simplicity and efficacy in image categorisation.¹⁸ “VGG-16” refers to the network’s depth, which consists of 16 learnable weight layers (13 convolutional layers and 3 fully connected layers). Architecture of CNN is described in Fig. 6.

The convolution operation, which is a critical component of the technique, may be stated as shown in Eq. [1] below :

$$(I * K)(i,j) = \sum_m I(m,n).K(i - m, j - n) \dots(1)$$

Equation 1: CNN’s convolution layer

Where S(i,j) represents the value at location (i,j) in the output feature map, I is the input image, and K is the convolution kernel.

Performance Metrics

This section briefly describes the performance indicators we utilised to evaluate our classifiers. Before deploying machine learning models, it is critical to assess their efficacy. The F1 score and classification accuracy are two common measures for classifier assessment.

Classification Accuracy is calculated as the number of accurately predicted samples divided by the total number of samples in the dataset.¹⁷ For accuracy to be useful, the dataset should be balanced; otherwise, high accuracy might be deceptive. In imbalanced datasets, models can show high accuracy by focusing on the majority class, yet they tend to underperform on minority classes, which contribute less to the overall accuracy.

The F1 Score, another important statistic, finds a balance between Precision and Recall by computing their harmonic mean. Precision measures the model’s ability to classify a sample as positive only when it actually belongs to the positive class (the ratio of true positives to projected positives). High precision means that favourable forecasts are likely to be accurate. Recall measures the ability of a model to identify all accurate positive samples. High recall indicates the model’s ability to correctly identify samples with positive results.¹⁸

Equations (2)-(5) include the equations for these metrics, with TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) reflecting different types of categorisation outcomes. CA refers to the accuracy of classification in this case.

$$CA = \frac{TP+TN}{TP+FN+TN+FP} \dots(2)$$

$$\text{PRECISION} = \frac{\text{TP}}{\text{TP}+\text{FP}} \quad \dots(3)$$

$$\text{RECALL} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad \dots(4)$$

$$\text{F1 SCORE} = 2X \frac{\text{PRECISION}*\text{RECALL}}{\text{PRECISION}+\text{RECALL}} \quad \dots(5)$$

RESULTS

This section discusses the use of the VGG-16 model to solve binary and multi-class classification problems. The dataset is divided into three subgroups for this study using a 7:2:1 split ratio: 70% for training, 20% for testing, and 10% for validation. This method ensures that the model has sufficient data for training while retaining enough data for hyperparameter tuning and generalization testing. The depth and efficacy of the VGG-16 model, a prominent convolutional neural network architecture, in recognising characteristics in pictures influenced its adoption. In order to meet the computational needs of training a deep network like VGG-16, we placed the model on Google Colab and used its powerful GPUs. In addition, we created, trained, and evaluated the model on Tensor Flow, a widely used deep learning platform. Tensor Flow offers the required characteristics to interact with VGG-16, such as previously trained weights for transfer learning, which can increase performance and speed up training, particularly when there is limited input. The model learns to differentiate between many classes for multi-class classification, although it was taught to discriminate between two distinct groups for binary categorisation. The model’s architecture was modified for each categorisation type by changing the output layer to correspond to the number of categories in each scenario. This implementation strategy guaranteed that VGG-16 was optimised for both binary and multi-class classification tasks, giving it versatility in its application to a variety of image classification issues.

Classification Report: Table 4 represent the classification report of the binary classification.

confusion matrix:

```
[[123 28]
 [ 7 154]]
```

Multiclassification Results

In multiclassification and severity detection using KL grade system. In this study we used three classes.

0 ,1 and 2 i.e 0 is healthy knee,1 is mild and 2 is severe.

accuracy: 0.7957 that means 79.6 and loss is 0.772 i.e 77.2. It indicated that as the accuracy Increase then the loss decreases.

Confusion matrix is shown in fig.7 for multiclassification.

```
array([[71, 21, 1],
       [ 6, 95, 5],
       [ 0, 15, 21]])
```

The model’s validation and training losses across 50 epochs are shown in the figure.

The orange line indicates the validation loss, while the blue line corresponds to the training loss.

Initially, both losses are significant, with the training loss rapidly decreasing during the first few epochs, showing efficient learning from the training data. Around the tenth epoch, the training loss stabilises and stays low, indicating that the model is successfully learning the patterns in the training data.

On the other hand, the validation loss initially decreases but then levels off, with minor volatility as training progresses. The considerable difference in training and validation loss suggests probable overfitting, since the model performs better on the training data than on the validation set. Despite this, the rather steady validation loss indicates that overfitting is insignificant. Further reducing this gap might assist improve the model’s generalisability.

DISCUSSION

The Binary classification using the VGG-16 model F1-score is 89.8%, Recall is 95.7%, Precision is 84.6 % and the accuracy is 89%. The high recall of 95.7% indicates that the model

successfully identifies most positive samples, which is valuable when false negatives are costly. However, the precision is slightly lower, at 84.6%, meaning there are some false positives where the model incorrectly classifies negative samples as positive. The F1-score of 89.8% reflects a balanced performance, capturing both precision and recall. The confusion matrix shows that:

- Class 0 (negative class) had a recall of 81%, suggesting the model missed some negative cases.
- Class 1 (positive class) had a recall of 96%, reflecting strong performance in correctly identifying positive cases.

In case of Multiclassification the accuracy of 80% is reasonable; nevertheless, performance varies by class. Class 0 has the highest accuracy (92%), indicating that the model can reliably identify instances of Class 0. However, the reduced recall (76%) indicates that some real Class 0 samples were misclassified. Class 1 has a high recall (90%) but a lower accuracy (73%), indicating that the model is more likely to categorise samples as Class 1, which might lead to more false positives. Class 2 has the lowest performance, with a recall of 58%, suggesting that it is the most challenging class for the model to recognise correctly.

The approach works well in both binary and multiclass scenarios, although there is potential for improvement, particularly in managing minority classes in multiclass classification. The model's efficacy may be enhanced by addressing class imbalance, improving recall for difficult courses, and balancing accuracy and recall levels. By putting these ideas into reality, the model may become more resilient and efficient in real-world applications by enhancing its consistency and reliability in binary and multiclass categories.

CONCLUSION

The illness known as KOA has propagated throughout the world. AI is being used effectively in the healthcare business. KOA is a joint condition, as is generally understood. Our bodies depend on joints for movement. The VGG-16 algorithm is being used in deep learning to identify and categorise KOA. Both the binary and multiclass classification tests performed well with this model. The model obtained a strong F1-score and a high accuracy of 89% for binary classification,

indicating a reasonable trade-off between recall and precision. This reveals that the model is good at discriminating between positive and negative instances of knee osteoarthritis (KOA), which is especially useful in situations where positive cases must be precisely identified. In multiclass classification, the model performed well with Class 0 (No KOA) and Class 1 (Mild KOA), achieving 80% overall precision. Fortunately, Class 2 (Moderate KOA) had a lower recall. This divergence indicates that the model struggles to discriminate between certain KOA stages, most likely because to disparity in class or visually trait overlap. Class 2 is very difficult to recognise accurately due to the small sample size, and accuracy and memory levels vary throughout classes. Despite these challenges, the VGG-16 model's effectiveness as a diagnostic tool for KOA categorisation in medical imaging is promising, particularly in binary classification tasks where it excels. Future studies will use huge, real-life data sets for validation as well as training to increase the model's capacity to generalise across different patient demographics and clinical conditions. Furthermore, using interpretability tools such as Grad-CAM or saliency maps would clarify the model's decision-making process, boosting its credibility and usefulness in healthcare contexts.

ACKNOWLEDGEMENT

I want to express my sincere thanks to everyone who helped with our study on the automatic classification of osteoarthritis in the knee using X-ray images. The VGG-16 Model is used. The study team deserves special recognition for their hard work and insightful contributions. Our work relied heavily on the Osteoarthritis Initiative (OAI) dataset, which we are grateful for using. This effort would not have been feasible without the constant direction and assistance of our mentor, Supervisor Dr. Minakshi Memoria, Co-Supervisor Dr. Kapil Joshi, and our Uttaranchal University.

Funding Sources

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

The author(s) do not have any conflict of interest.

Data Availability Statement

This statement does not apply to this article.

Ethics Statement

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

Informed Consent Statement

This study did not involve human participants, and therefore, informed consent was not required.

Clinical Trial Registration

This research does not involve any clinical trials

Authors Contribution

Ms. Suman Rani drafted the original manuscript; Dr. Meenakshi Memoria Supervisor help me in writing, review and editing; Dr. Kapil Joshi provide contious support in data collection and methodology implementation; Dr. Anil Lamba help me in cleaning and processing of data.

REFERENCES

- Mohammed AS, Hasanaath AA, Latif G, Bashar A. Knee osteoarthritis detection and severity classification using residual neural networks on preprocessed x-ray images. *Diagnostics*. 2023;13(8):1380.
- Rani S, Memoria M. Deep learning to combat knee osteoarthritis and severity assessment by using CNN-based classification. *BMC Musculoskelet Disord*. 2024;25(1):817.
- Rani S, Memoria M, Choudhury T, Sar A. A comprehensive review of machine learning's role within KOA. *EAI Endorsed Trans Internet Things*. 2024;10.
- Roy S, Meena T, Lim SJ. Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine. *Diagnostics*. 2022;12(10):2549.
- Mohammed AS, Hasanaath AA, Latif G, Bashar A. Knee osteoarthritis detection and severity classification using residual neural networks on preprocessed x-ray images. *Diagnostics*. 2023;13(8):1380.
- Wahyuningrum RT, Anifah L, Purnama IKE, Purnomo MH. A new approach to classify knee osteoarthritis severity from radiographic images based on CNN-LSTM method. In: *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)*. IEEE; 2019:1-6.
- Tri Wahyuningrum R, Yasid A, Jacob Verkerke G. Deep neural networks for automatic classification of knee osteoarthritis severity based on X-ray images. In: *Proceedings of the 2020 8th International Conference on Information Technology: IoT and Smart City*. 2020:110-114.
- Qadir A, Mahum R, Aladhadh S. A robust approach for detection and classification of KOA based on BILSTM network. *Comput Syst Sci Eng*. 2023;47(2).
- Reddy TR, Lakshmi KV, Nikhila L, Koneru S. Knee osteoarthritis detection and classification using ensemble model of Xception and Inception ResNetV2. In: *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*. IEEE; 2024:607-612.
- Yunus U, Amin J, Sharif M, Yasmin M, Kadry S, Krishnamoorthy S. Recognition of knee osteoarthritis (KOA) using YOLOv2 and classification based on convolutional neural network. *Life*. 2022;12(8):1126.
- Zhao Z, Zhao M, Yang T. Identifying significant structural factors associated with knee pain severity in patients with osteoarthritis using machine learning. *Sci Rep*. 2024;14(1):14705.
- Nurmrinta TA, Turunen MJ, Korhonen RK. Two-stage classification of future knee osteoarthritis severity after 8 years using MRI: data from the osteoarthritis initiative. *Ann Biomed Eng*. 2024:1-12.
- Jain RK, Sharma PK, Gaj S, Sur A, Ghosh P. Knee osteoarthritis severity prediction using an attentive multi-scale deep convolutional neural network. *Multimed Tools Appl*. 2024;83(3):6925-6942.
- Ahmed R, Imran AS. Knee osteoarthritis analysis using deep learning and XAI on X-rays. *IEEE Access*. 2024.
- Kinger S. Deep learning for automatic knee osteoarthritis severity grading and classification. *Indian J Orthop*. 2024:1-16.
- Goswami AD. Automatic classification of the severity of knee osteoarthritis using enhanced image sharpening and CNN. *Appl Sci*. 2023;13(3):1658.
- Rani S, Memoria M, Rani M, Kumar R. Breast cancer detection using mammographic images over convolutional neural network. In: *2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE; 2023:73-78.
- Lee DW, Song DS, Han HS, Ro DH. Accurate, automated classification of radiographic knee osteoarthritis severity using a novel method of deep learning: Plug-in modules. *Knee Surg Relat Res*. 2024;36(1):24.