SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

# Federated Attention-Based Deep Learning Framework For Privacy-Preserving Multi-Center Knee Osteoarthritis Diagnosis From Heterogeneous Radiographic Images

Ms. Swathi Kumari R[1], Dr. S. Geetha[2], Dr. P. S. Rajakumar[3]

[1]Research Scholar Department of Computer Science and Engineering, Dr. M.G.R.Educational and Research institute-Chennai - 600095, Tamil Nadu, India. Email: swathi.mca@drmgrdu.ac.in
[2]Professor and Head, Department of Computer Science and Engineering, Dr.M.G.R.Educational and Research institute-Chennai -95, Tamil Nadu, India. Email: geethasoman@drmgrdu.ac.in
[3]Professor Department of Computer Science and Engineering Dr.M.G.R.Educational and Research institute-Chennai -95, Tamil. Email: rajakumar.subramanian@drmgrdu.ac.in

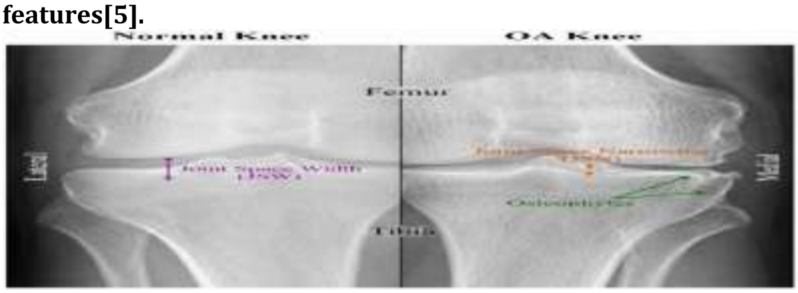
**Abstract**
The diagnosis of Knee Osteoarthritis (KOA) using radiography is faced with issues such as different imaging conditions, unbalance in classes, low inter-center cooperation, and privacy of patients. The research presents a Distributed Attention-Based Deep Learning System that enables KOA severity classification while preserving privacy across medical facilities. In the system, image preprocessing and DenseNet121-based extraction of features are combined with the method of attention, as well as Federated Averaging (FedAvg) and Kellgren-Lawrence (KL) classification. The preprocessing of images implies activities such as resizing (224 x 224), intensity normalization, contrast enhancement, noise reduction, region of interest extraction, augmentation, and client-based systematization is applied. The system development is based on 9,547 x-ray images, including: 5,610 images used for training, 810 images for validation, 1,630 images for internal testing, and 1,497 images for external testing. The obtained images have 3,740 grade 0 images (39.17%) and 287 grade 4 images (3.01%), which shows the significant difference across classes. The knowledge gained in federated training comes from model parameters update, as the original x-rays remain in the owners' location.

**Keywords:** Knee Osteoarthritis (KOA), Federated Learning, Deep Learning; DenseNet121, Attention Mechanism, Privacy-Preserving Healthcare.

## 1. Introduction

Knee Osteoarthritis (KOA) is one of the most common chronic musculoskeletal conditions and a major cause of pain, disability and impaired quality of life in the elderly population worldwide [1]. The disease is essentially a progressive degeneration of articular cartilage, subchondral bone remodelling, formation of osteophytes and synovial inflammation that leads to loss of mobility and joint stiffness [2]. The burden of KOA is continually growing worldwide as obesity, sedentary lifestyle and a growing elderly population becomes more common. The early diagnosis and early intervention are crucial to control the progression of the disease and optimize treatment strategies to enhance long-term clinical outcomes [3]. Radiographic imaging, especially X-ray examination, is still the main diagnostic tool as it is accessible, is a low cost imaging technique, and it shows the structural abnormalities, which are associated with osteoarthritis [4]. Figure 1 compares the X-rays of a normal knee to an osteoarthritic knee and shows joint space narrowing and development of osteophytes.

**Figure 1. Comparison of normal and osteoarthritic knee radiographs highlighting key pathological features[5].**



The automated interpretation of medical images is significantly improved with recent progress in Artificial Intelligence (AI), particularly Deep Learning (DL) [6]. DL models like Convolutional Neural Networks (CNNs) and attention-based models have shown impressive ability in detecting fine-grained radiographic features linked to KOA, which allows for precise diagnosis and disease severity estimation. In addition, attention mechanisms enhance the diagnostic accuracy by selectively attending to clinically relevant areas of the knee joint, such as cartilage space narrowing, osteophytes, and subchondral sclerosis [8]. Although these developments are promising, the vast majority of current DL models are still centralized, meaning patient images must be sent to a shared server for model development by medical institutions. These centralized strategies raise significant privacy, regulatory, ownership, and cybersecurity concerns [9].
Medical imaging data is highly sensitive and is subject to strict privacy regulations like the General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA) [10]. Patient data is not readily shared across institutional boundaries, especially by hospitals or healthcare organizations, which are often legally prohibited or unwilling. Therefore, DL models learned from a single data set often have restricted generality because of data scarcity. In addition, there are significant differences in the scans obtained from various medical centres due to differences in imaging devices, imaging protocols, patient population, and annotations [11]. The differences result in domain shifts that diminish the robustness and reliability of the conventional diagnostic models used in a real-world multi-center healthcare environment [12].

In healthcare, Federated Learning (FL) is a novel paradigm that allows multiple healthcare institutions to collaborate on training a global model without sharing patient data directly [13]. The participating centers locally update the model parameters and upload only the aggregated or encrypted weights to an aggregation server [14]. This distributed learning approach keeps patient information confidential while also taking advantage of the expertise spread across widely-separated data sets. One of the key advantages of federated learning is its ability to overcome the regulatory hurdles and boost model generalization by utilizing a wide range of clinical data from various institutions. But traditional federated learning systems still suffer from performance degradation due to data distribution heterogeneity, communication inefficiency, and inconsistent feature representation among these different centers [15,16].
To address these challenges, incorporating attention mechanisms into federated DL systems is a potential approach. The attention-related architectures improve feature extraction by focusing on diagnostic anatomical parts and ignore irrelevant background information. In federated learning, attention modules can be used to adapt the global model to learn more discriminative and transferable representations of data even in the presence of imaging differences across institutions [17]. These frameworks can enhance the accuracy, robustness and adaptability of diagnosis, while preserving privacy throughout the collaborative learning process. Furthermore, in multi center medical imaging applications, where data is typically non independent and identically distributed (non-IID), advanced aggregation methods can help reduce the impact of non-IID data [18].

In this study, a Federated Attention-Based DL Framework for Privacy-Preserving Multi-Center KOA Diagnosis from Heterogeneous Radiographic Images is proposed. The proposed framework involves the use of privacy-preserving federated learning and attention mechanism-based deep neural networks for collaborative KOA diagnosis while avoiding the sharing of protected patient information. The framework is designed to learn from diverse radiographic datasets across different healthcare centers, enhancing diagnostic accuracy, robustness and generalization, while respecting medical data privacy laws and regulations. The proposed approach will help to develop secure, scalable and clinically deployable AI systems that can aid to reliable diagnosis of KOD in distributed healthcare settings. Here are the research objectives as shown below:

- To develop a Federated Attention-Based DL Framework for privacy-preserving multi-center KOA diagnosis using heterogeneous radiographic images.
- To preprocess heterogeneous knee radiographs for improving image quality, reducing domain variability, and enhancing feature consistency across distributed healthcare institutions.
- To extract discriminative radiographic features using DenseNet121 integrated with an attention mechanism for accurate KOA severity classification.
- To implement a federated learning strategy for collaborative model training and global model aggregation without sharing sensitive patient data.
- To evaluate the proposed framework using standard performance metrics for reliable and robust Kellgren–Lawrence (KL) grade classification across heterogeneous clinical datasets.

## 2. Review of Literature

In recent years, DL, attention mechanism, transfer learning and hybrid architectures have been investigated in detail to enhance the automated diagnosis and severity grading of KOA from radiographic images. An attention-assisted convolutional autoencoder combined with machine learning classifiers and dimensionality reduction methods was proposed by Wang et al. (2025) [19] that used attention mechanisms to greatly improve the feature representation, yielding an AUC of 96.5% and an F1-score of 93.5%. Zhang et al. (2025) [20] proposed a novel large separable kernel attention-based multidimensional MRI classification framework that seamlessly combined shallow and deep features and achieved 99.7% classification accuracy. Pugazharasi et al. (2026) [21] proposed a model named Radiomic Feature Fusion and Joint Space Segmentation (RFF-JSS), which integrated the U-Net segmentation model, PCA assisted radiomic fusion with Random Forest classification, and achieved an accuracy of 94.87% and a ROC-AUC value of 0.972. Likewise, Pi et al. (2023) [22] proposed an ensemble model that combined multiple image resolutions and Grad-CAM visualization, with a success rate of 76.93% in predicting KL grade. Yasin et al. (2025) [23] presented a system based on YOLOv11 for the detection of both soft-tissue and osteoarthritis abnormalities directly from X-ray images, whereas Sudha et al. (2026) [24] concluded that the Swin-O-NETS architecture with Modified Swin Transformer and Fast Extreme Learning Network yielded the highest accuracy of 99.4%. Additionally, Tiwari et al. (2022) [25] tested 8 transfer learning models and found that DenseNet achieved the highest classification accuracy of 93% while Nguyen et al. (2026) [26] fused DL and a Large Language Model (LLM) to automatically produce structured clinical reports, along with KOA grading. Furthermore, Ahmad et al. (2026) [27] introduced a multimodal framework called RAD-OAEnsemble which combined radiographs and EHRs through graph fusion and attention modules, resulting in an overall accuracy of 99.275%, underscoring the significance of multimodal learning and Explainable Artificial Intelligence (XAI) in the diagnosis of musculoskeletal diseases. Table 1 is a summary of the recent studies related to DL-based KOA diagnosis, detailing the methodology, datasets, performance, and research limitations.

**Table 1: Summary of Recent Literature on KOA Diagnosis**

| Author(s) | Proposed Method | Dataset | Key Findings | Limitation |
|---|---|---|---|---|
| **Wang et al. (2025)** | Attention-assisted Convolutional Autoencoder with PCA/RFE and ensemble classifiers | 5,987 knee X-ray images | Achieved AUC of 96.5% and F1-score of 93.5%, demonstrating improved feature representation and interpretability. | Centralized learning; no privacy preservation or multi-center collaboration. |
| **Rifat et al. (2023)** | Federated Learning using DenseNet-169, Inception-v2, and MobileNet-v2 | Distributed KOA datasets | DenseNet-169 achieved 82% accuracy and 81% F1-score while preserving patient privacy through federated learning. | Limited number of clients and moderate classification performance. |
| **Zhang et al. (2025)** | Large Separable Kernel Attention with Cross-Level Feature Fusion | Hospital-based MRI dataset | Achieved 99.7% accuracy, demonstrating highly effective hierarchical feature learning. | Relies on MRI imaging instead of widely available radiographs. |
| **Pugazharasi et al. (2026)** | Radiomic Feature Fusion and Joint Space Segmentation (RFF-JSS) | OAI and MOST datasets | Achieved 94.87% accuracy and 0.972 ROC-AUC using radiomics and anatomical segmentation. | Computationally intensive and lacks distributed learning capability. |
| **Pi et al. (2023)** | Multi-resolution Ensemble DL with Grad-CAM | OAI dataset | Improved KL grading with 76.93% accuracy and | Lower classification accuracy compared with recent hybrid models. |

| | | | | |
|---|---|---|---|---|
| | | | enhanced model explainability. | |
| **Yasin et al. (2025)** | YOLOv11 with Multiple CNN Models and Explainable AI | 4,215 knee X-rays from two medical centers | Automated knee localization and multi-label classification with 90.1% mAP. | Focused mainly on soft-tissue abnormalities rather than privacy-preserving collaboration. |
| **Alkhatatbeh et al. (2026)** | Multi-task ConvNeXt Framework | OAI dataset | Simultaneously predicted MRI-defined pathology and KL grades with QWK = 0.8284. | Performance decreases for early-stage pathology and centralized model training. |
| **Sidha et al. (2026)** | Swin-O-NETS (Modified Swin Transformer + FELN) | OAI dataset | Achieved 99.4% accuracy with reduced computational complexity. | Does not address patient privacy or heterogeneous institutional data. |
| **Prezja et al. (2022)** | GAN-based Synthetic Knee X-ray Generation | 5,556 real and 320,000 synthetic images | Synthetic images significantly improved KOA classification under limited training data. | Focuses on data augmentation rather than diagnosis across multiple centers. |
| **Tiwari et al. (2022)** | Comparative Transfer Learning Models | 2,068 clinical radiographs | DenseNet achieved the highest accuracy (93%) among eight transfer learning models. | Limited dataset and absence of explainability or federated learning. |
| **Nguyen et al. (2026)** | Hybrid DL + LLM | Vietnam Osteoporosis Study | Achieved 79.1% KL grading accuracy while automatically generating structured clinical reports. | Moderate grading accuracy and centralized architecture. |
| **Ahmad et al. (2026)** | RAD-OAEnsemble (Radiographs + EHR Fusion with Attention) | Multimodal KOA dataset | Achieved 99.275% accuracy through multimodal feature fusion and explainable AI. | Requires multimodal clinical data, limiting applicability in routine screening. |

Additionally, privacy-preserving learning, multi-center collaboration, synthetic data generation, and strong clinical generalization are key research areas to be explored. Rifat et al. (2023) [28] proposed a federated learning framework based on Deep neural networks (DenseNet-169, Inception-v2, and MobileNet-v2) on distributed clients showing that federated learning can achieve 81% F1 score and 82% accuracy for diagnostics tasks, which is comparable to central learning, without compromising patient privacy. Alkhatatbeh et al. (2026) [29] presented a multi-task ConvNeXt network that was capable of predicting MRI-defined structural abnormalities and the KL grade in radiographs, achieving excellent generalization within early-stage osteoarthritis. Prezja et al. (2022) [30] explored the Generative Adversarial Networks (GANs) to create realistic knee radiographs, producing more than 320,000 synthetic radiographs that improved classification accuracy in small training sets.

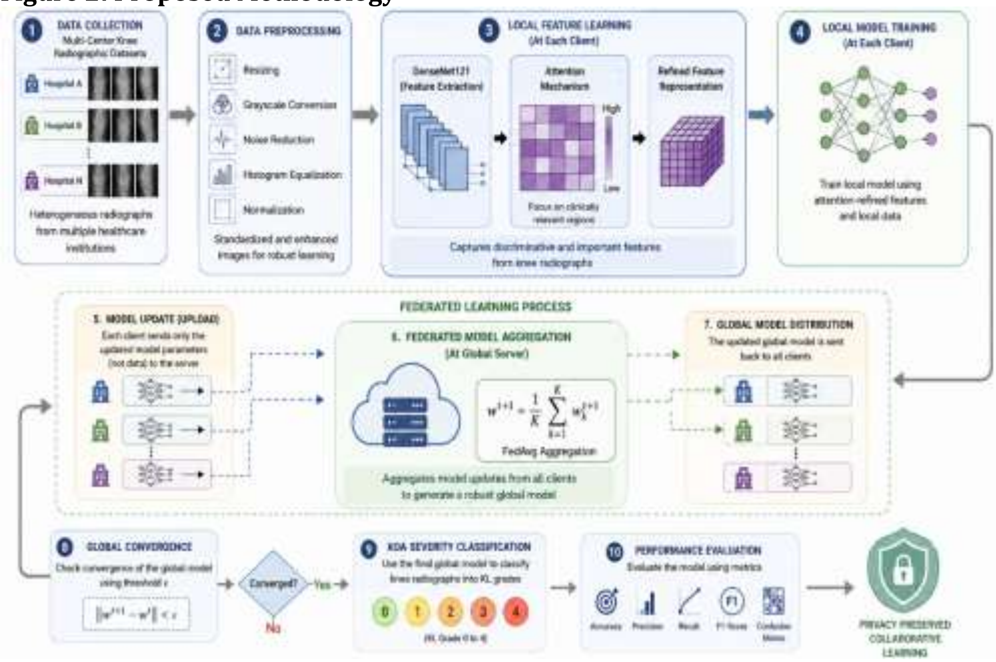
Rawat et al. (2026) [31] reviewed 18 studies on AI-based KOA grading and reported accuracy in the range of 75–98%, AUC up to 0.98, and Cohen's κ of 0.67–0.86. The study highlighted the issues of labelling subjectivity, class imbalance, and limited external validation. Sundaramurthy et al. (2026) [32] investigated Quantum Convolutional Neural Networks (QCNNs) and demonstrated improved robustness, computational efficiency, and KOA severity classification compared to traditional CNNs. Alkhatatbeh et al. (2026) [33] proposed KL-FuseNet, 70.3% internal accuracy, QWK 0.881. Selective fine-tuning improved external accuracy to 80.0%, QWK 0.950, and AUC 0.984. Owida et al. (2026) [34] 99.50% accuracy using ANN-SGD. Abd Al Nabi et al . (2025) [35] used fusion models with 70% and 90.61% accuracy. EfficientNet-B0 with ECA has shown better early-stage OA detection as per Abdusalomov et al. (2025) [36].

All of these studies have shown that there has been substantial advancement in the field of attention-guided learning, transformer architecture, multimodal fusion, federated learning, and explainable AI in the field of KOA diagnosis. However, most current methods require centralised training, rely on data from a single institution or fails to adequately account for the heterogeneity of multi-center radiographic variability and strict patient privacy requirements. These limitations prompt the creation of a Federated Attention-Based DL Framework that can learn together from multiple distributed heterogeneous radiography datasets while respecting patient privacy and enhancing the diagnostic performance in multiple healthcare units.

## 3. Research Methodology

Figure 2 illustrates the complete research methodology of the proposed Federated Attention-Based Deep Learning Framework, including multi-center dataset collection, image preprocessing, local feature learning using DenseNet121 and attention mechanism, federated model aggregation, global model distribution, KOA severity classification, and performance evaluation while preserving patient privacy.

**Figure 2: Proposed Methodology**




### 3.1Dataset Description

In this study, the OAI dataset and the Mendeley Knee Osteoarthritis Dataset were used to train, validate, and evaluate the proposed framework.

### 3.1.1 Osteoarthritis Initiative (OAI) Dataset

The dataset of the Osteoarthritis Initiative (OAI) [37] is a large-scale and publicly available dataset that is widely used for KOA research and computer-aided diagnosis. It consists of thousands of standard knee X-rays from people with different severities of OA and healthy individuals. Each image is labelled using the Kellgren-Lawrence (KL) grading system which allows the disease severity of the image to be accurately rated from KL Grade 0 to KL Grade 4. Radiographic images are not the only valuable information provided in the dataset; it also contains useful clinical data, including age, gender, Body Mass Index (BMI), and disease progression. It is a diverse and high-quality imaging dataset, making it an ideal benchmark for DL models and federated learning models for automated KOA diagnosis.

### 3.1.2 Mendeley Knee Osteoarthritis Dataset

The Knee Osteoarthritis Dataset [38] with Severity Grading is a public radiographic dataset for automated diagnosis and severity grading of KOA. The dataset consists of anterior-posterior knee X-ray images with the severity of the disease annotated by KL grading system, which assigns 5 classes to the severity of the disease: Grade 0 (Healthy), Grade 1 (Doubtful), Grade 2 (Minimal), Grade 3 (Moderate), and Grade 4 (Severe). The images feature a range of grades of joint space narrowing, osteophyte formation, and subchondral sclerosis, for a thorough assessment of the DL models. The dataset is well annotated and uniformly graded and hence can be used for external validation, testing the generalization capability of the proposed federated framework under diverse clinical imaging conditions and heterogeneous radiographic images.

**Table 2: Class-wise Distribution of KOA Radiographic Images Used for Federated Learning**

| KL Grade | Clinical Severity | Training | Validation | Internal Test | External Test |
|---|---|---|---|---|---|
| 0 | Healthy | 2200 | 320 | 630 | 590 |
| 1 | Doubtful | 1020 | 150 | 290 | 270 |
| 2 | Minimal | 1480 | 210 | 440 | 400 |
| 3 | Moderate | 740 | 105 | 220 | 195 |
| 4 | Severe | 170 | 25 | 50 | 42 |
| Total | — | 5610 | 810 | 1630 | 1497 |

### 3.2Image Preprocessing

The proposed Federated Attention-Based DL Framework requires image preprocessing to enhance the quality and uniformity of the heterogeneous knee radiographic images prior to feature extraction and federated model training. X-ray images are captured at many health facilities, on different imaging equipment and with different acquisition techniques, making it impossible to avoid differences in the resolution, brightness, contrast, noise and orientation of the images. To enhance the robustness, convergence and generalization capability of the proposed federated learning framework, thus, a comprehensive preprocessing pipeline is employed to standardize image characteristics while preserving clinically important anatomical structures.

#### 3.2.1 Image Resizing and Intensity Normalization

The spatial resolution and image size of radiographs taken in different medical centres differ. All images are scaled to the same input size as the model requires for training, which for DenseNet121 is 224 × 224 pixels. Resizing makes spatial representation consistent across the different data types and simplifies the computing. Following this, pixel intensities are normalized to the range [0,1] to reduce the variation in intensities arising from different X-ray acquisition systems and accelerate network convergence.
The normalization process is expressed as:

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}} \quad (1)$$

where:
$I$denotes the original pixel intensity,
$I_{min}$represents the minimum pixel value,
$I_{max}$represents the maximum pixel value,
$I_{norm}$is the normalized pixel intensity.

#### 3.2.2 Contrast Enhancement and Noise Reduction

Medical radiographs are frequently poorly illuminated and contrasted, and have imaging noise that can conceal clinically relevant features including joint-space narrowing and osteophytes. Contrast enhancement is used prior to feature extraction to enhance the anatomical visibility. Histogram Equalization (HE) or Contrast Limited Adaptive Histogram Equalization (CLAHE) improves the contrast of the local image without over emphasizing the noise. Furthermore, Gaussian filtering is used to remove high frequency noise and retain edges.
The Gaussian filter is defined as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (2)$$

where:
$\sigma$denotes the standard deviation controlling the smoothing effect.
This operation improves image clarity and facilitates more reliable extraction of disease-related features.

#### 3.2.3 Region of Interest (ROI) Extraction

Only the knee joint area is chosen for analysis, leaving out the entire radiograph. ROI extraction helps remove irrelevant anatomical background, lower the computational cost, and allow the DL model to pay attention only to the disease-related structures including cartilage boundaries, osteophytes and tibiofemoral joint space.
The extracted ROI is represented as

$$ROI = I(x_1, : x_2 \; y_1 : y_2) \quad (3)$$

where:
$x_1, x_2$ represent horizontal image boundaries,
$y_1, y_2$ denote vertical boundaries of the knee joint.
This process increases feature discrimination while reducing redundant information.

### 3.2.4 Data Augmentation

Data augmentation methods are used during local training on each federated client to increase model generalization and reduce overfitting due to small class sizes. These are random rotations, horizontal flipping, scaling, translation, zooming and brightness adjustments. Data augmentation introduces different imaging variations to the model without compromising the diagnostic features of KOA.
Mathematically, image transformation is expressed as

$$I' = T(I) \tag{4}$$

where:
$I$ is the original image,
$T(\cdot)$ denotes the augmentation transformation,
$I'$ represents the augmented image.
The image-augmented samples are more robust to imaging variability between different healthcare institutions.

### 3.2.5 Intensity Standardization Across Federated Clients

The proposed framework is based on radiograms from several hospitals, and the intensity distributions are much different because of variations in imaging devices, exposure parameters, and acquisition protocols. In order to reduce discrepancies between different clients' domains, Z-score standardization is conducted separately at each of the medical centers involved in federated training.
The standardized intensity is computed as

$$Z = \frac{I-\mu}{\sigma} \tag{5}$$

where:
$\mu$ denotes the mean pixel intensity, and $\sigma$ denotes the standard deviation.
This normalization reduces domain shift among distributed datasets and improves the stability of federated optimization.

### 3.2.6 Image Quality Verification

All preprocessed images are checked for quality before they can be used for local model training, discarding corrupted, blurred, duplicated and incomplete radiographs. Images with quality metrics that do not meet the set thresholds are ignored, meaning that only images with diagnostic value are used in the federated learning process. This helps to improve robustness of the model as it improves the ability to prepare the global parameters without being affected by noisy or low-quality data.

## 3.3 Local Feature Learning

Local feature learning is an important phase in the proposed Federated Attention-Based DL Framework, during which each participating healthcare institution is allowed to train discriminative representations from its locally stored knee X-rays images. To protect patient privacy, each federated client trains the model with their own patient information, while aligning the model with their local patient population to learn disease-specific features linked to KOA. The proposed framework uses the DenseNet121 as the backbone network to extract hierarchical features and an attention mechanism to highlight clinically relevant regions and suppress irrelevant background details. These locally learnt feature representations are then optimised for the model and sent to a federated server for global aggregation. The detailed working principle and architecture of DenseNet121 and the attention mechanism are explained in following subsections.

### 3.3.1 DenseNet121

DenseNet121 is a popular CNN design with dense connections and efficient feature propagation[39]. The layers are different from conventional CNNs, where each layer only has its own data. The difference is that each layer has data from all previous layers, allowing for the reuse of features and better flow of information across the network[40]. This architecture alleviates vanishing gradient problem, drastically reduces the number of redundant features learnt and needs fewer trainable parameters to achieve high classification accuracy[41]. DenseNet121 is used as the backbone feature extractor in the proposed Federated Attention Based DL Framework to learn discriminative representations from preprocessed knee x-ray images. The features extracted are clinically relevant including joint space narrowing, osteophyte formation, and

subchondral bone abnormalities, and are used to generate informative feature maps for the next attention mechanism and classification of the severity of KOA.
The output of the $l^{th}$dense layer is represented as

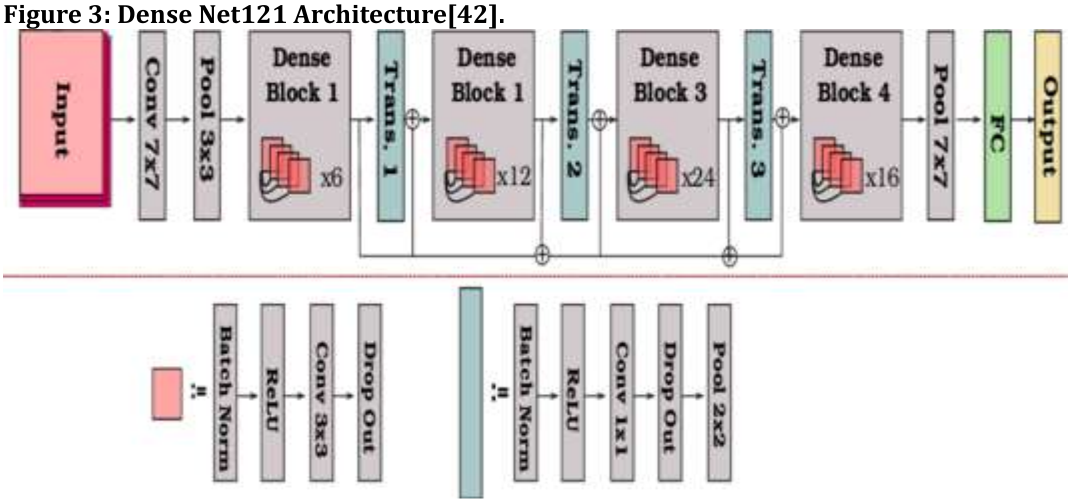
$$x_l = H_l([x_0, x_1, x_2, \ldots, x_{l-1}]) \qquad (6)$$

where $x_l$denotes the output feature map of the $l^{th}$layer, $H_l(\cdot)$represents the composite operation consisting of Batch Normalization (BN), ReLU activation, and convolution, while $[x_0, x_1, \ldots, x_{l-1}]$indicates the concatenation of feature maps generated by all preceding layers.
The convolution operation used for feature extraction is expressed as

$$F(x, y) = I(x, y) * K(x, y) \qquad (7)$$

where $I(x, y)$denotes the input knee radiograph, $K(x, y)$represents the convolution kernel, and $F(x, y)$is the extracted feature map used for subsequent attention-guided classification. Figure 3 illustrates the DenseNet121 architecture comprising dense blocks, transition layers, pooling, and fully connected classification layers.

**Figure 3: Dense Net121 Architecture[42].**




### 3.3.2 Attention Mechanism

The attention mechanism is embedded in the proposed framework to improve the model's ability to selectively attend to clinically relevant regions of knee radiographs and ignore irrelevant background information. DenseNet121 extracts rich hierarchical features but not all the extracted features are equal for the diagnosis of KOA. The attention module puts more weight on the structures of the body that are related to the disease, like joint space narrowing, osteophytes, cartilage damage, and changes in the subchondral bone. The model focuses on these informative areas, gaining more discriminative feature representations and subsequently enhancing the classification accuracy and robustness in heterogeneous radiographic images obtained from various healthcare institutions. It is especially useful in federated learning settings with imaging characteristics varying between medical centers, allowing a model to accurately detect salient pathological features across all medical centers.
The attention weights are computed as

$$\alpha = \frac{exp(s_i)}{\sum_{j=1}^{N} exp(s_j)} \qquad (8)$$

where $\alpha$denotes the normalized attention weight, $s_i$represents the attention score of the $i^{th}$feature, and $N$is the total number of extracted features.
The refined feature representation is obtained by applying the learned attention weights to the extracted feature maps:

$$F_{att} = \alpha \odot F \qquad (9)$$

where $F$denotes the feature map extracted by DenseNet121, $\alpha$represents the attention weight matrix, $\odot$ indicates element-wise multiplication, and $F_{att}$is the attention-refined feature map. The attention-refined features are subsequently forwarded to the federated learning framework for collaborative model optimization and KOA severity classification.

### 3.4 Federated Model Aggregation

Federated model aggregation is one of the proposed Federated Attention-Based DL Framework, which allows multiple healthcare institutions to jointly train a global KOA diagnosis model without sharing sensitive patient information. Once local feature learning and model optimization has been performed, each hospital only sends the modified model parameters to the central federated server, with the original knee radiographs being safely stored locally at each hospital. This approach to learning is decentralized, which helps to maintain patient privacy, adhere to healthcare data protection laws, and facilitates knowledge sharing among distributed healthcare centers. A common global model is initialized on the federated server and then sent to all participating clients before the first communication round can begin in order to initiate the collaborative learning process.

The initial global model is represented as

$$w^0 = \frac{1}{K}\sum_{k=1}^{K} w_k^0 \quad (10)$$

where $w^0$ denotes the initial global model, $w_k^0$ represents the initial model at the $k^{th}$ participating client, and $K$ is the total number of healthcare institutions involved in federated training.

In local training, the parameters of the model are updated by each client using gradient descent to minimize its local loss function. The local model update is shown as:

$$w_k^{(t+1)} = w^t - \eta \nabla L_k(w^t) \quad (11)$$

where $w^t$ denotes the global model parameters at communication round $t$, $w_k^{(t+1)}$ represents the updated model of the $k^{th}$ client, $\eta$ is the learning rate, and $L_k$ denotes the local loss function.

After receiving model updates from all participating hospitals, the federated server performs weighted aggregation using the Federated Averaging (FedAvg) algorithm:

$$w^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{N} w_k^{(t+1)} \quad (12)$$

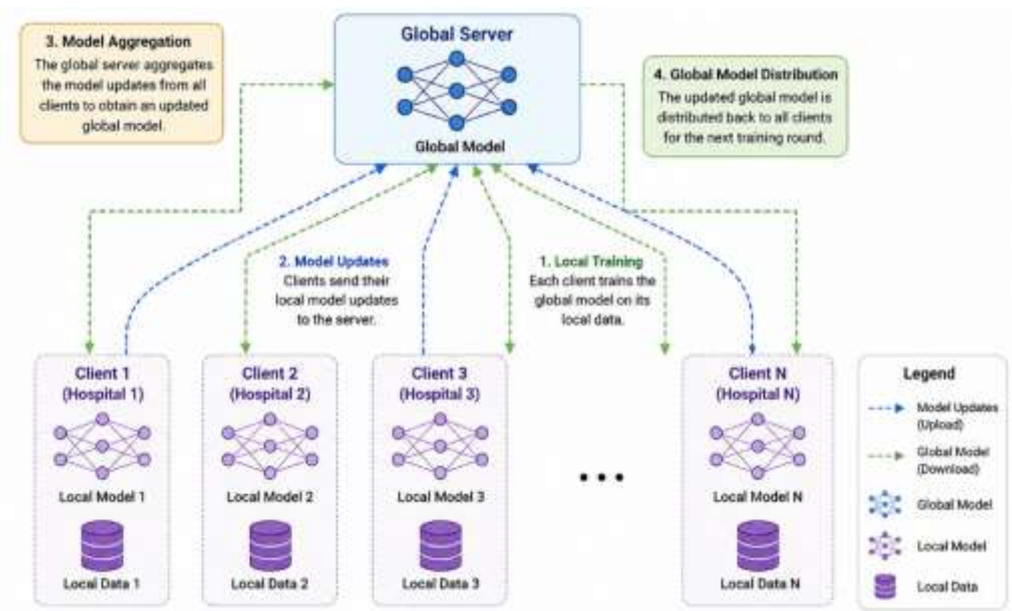
where $K$ is the total number of participating clients, $n_k$ represents the number of training samples at the $k^{th}$ client, $N = \sum_{k=1}^{K} n_k$ is the total number of training samples across all clients, and $w^{(t+1)}$ is the updated global model.

The overall optimization objective of federated learning is defined as

$$L(w) = \sum_{k=1}^{K} \frac{n_k}{N} L_k(w) \quad (13)$$

where $L(w)$ represents the global loss obtained by combining the weighted local losses of all participating clients. The next time the aggregated global model is then redistributed to each healthcare institution for the next round of communications, and the iterative training is repeated until the models converge. The proposed federated framework not only mitigates the data privacy concerns of patients but also reduces the communication burden and enhances the generalization ability of the global model over heterogeneous knee radiographic datasets across multiple healthcare providers. Figure 4 shows collaborative federated model aggregation through local training, parameter sharing, global aggregation, and model redistribution.

**Figure 4: Federated Model Aggregation Process**




## 3.5 Global Model Distribution

Federated model aggregation results in the distribution of the updated global model to all healthcare institutions in a secure way for the next communication round. After producing the new global model, each client takes the new global model and keeps training with its own local knee radiograph dataset. This repeats and synchronizes the use of each institution's collective knowledge to each other without providing access to raw patient information. Therefore, the global model gradually enhances its generalization ability on various medical data sets, while maintaining patient privacy and data confidentiality. The distribution process is repeated until the model reaches convergence and the diagnostic power of the classification of KOA becomes stable.

The global model distribution to each client is represented as

$$w_k^{(t+1)} \leftarrow w^{(t+1)},\ k = 1,2,\dots,K \tag{14}$$

where $w^{(t+1)}$denotes the aggregated global model, $w_k^{(t+1)}$represents the updated model distributed to the $k^{th}$client, and $K$is the total number of participating healthcare institutions.

The convergence criterion is expressed as

$$\| w^{(t+1)} - w^{(t)} \| < \varepsilon \tag{15}$$

where $w^{(t)}$and $w^{(t+1)}$denote the global model parameters of two consecutive communication rounds, and $\varepsilon$represents the predefined convergence threshold. In this case, the federated training is stopped and the final global model is used for the classification of severity of KOA. The synchronized global model allows all healthcare institutions involved to take training from the same parameters from the model, thereby guaranteeing consistency across communication rounds.

### 3.6 KOA Severity Classification

The global model is used after global convergence to classify knee radiographic image as KL grade (from 0 to 4), which indicates the severity of KOA. DenseNet121 is used to extract the attention-refined features from images and these features are further passed through a fully connected network layer, and then a Softmax classifier to estimate the probability of different severity classes of KOA. The class with the most confidence is chosen as the final prediction based on these probabilities. This classification task helps to reliably and precisely identify the severity of the disease without compromising diagnostic accuracy for heterogeneous radiograph sets from various healthcare institutions.

The Softmax function is defined as

$$P(y = i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{16}$$

where $P(y = i)$denotes the probability of class $i$, $z_i$is the output score of class $i$, and $C$represents the total number of KOA severity classes.

The final predicted class is determined using the ArgMax function:

$$\hat{y} = arg\ \max_i P(y = i) \tag{17}$$

where $\hat{y}$denotes the predicted KOA severity class corresponding to the highest probability generated by the Softmax classifier.

The classification loss is computed using the Cross-Entropy loss function:

$$L = -\sum_{i=1}^{C} y_i\ log(\hat{y}_i) \tag{18}$$

where $y_i$is the true class label and $\hat{y}_i$represents the predicted probability for class $i$. The minimisation of this loss during the classification of the severity of KOA, ensures and enhances the accuracy of the classification and convergence of the proposed Federated Attention-Based DL Framework.

### 3.7 Performance Evaluation

The effectiveness of the proposed Federated Attention-Based DL Framework is demonstrated by the common classification metrics in diagnosing KOA. Classification accuracy, precision, recall, and F1-score are calculated by comparing the predicted results with the ground truth labels. These assessment parameters allow an overall evaluation of the diagnostic potential, reliability, and generalization ability of the model on heterogeneous knee radiographic datasets.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{19}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{20}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{21}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{22}$$

where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively.A confusion matrix is also used to display the classification accuracy of the proposed model by comparing the predicted and actual KOA severity classes.

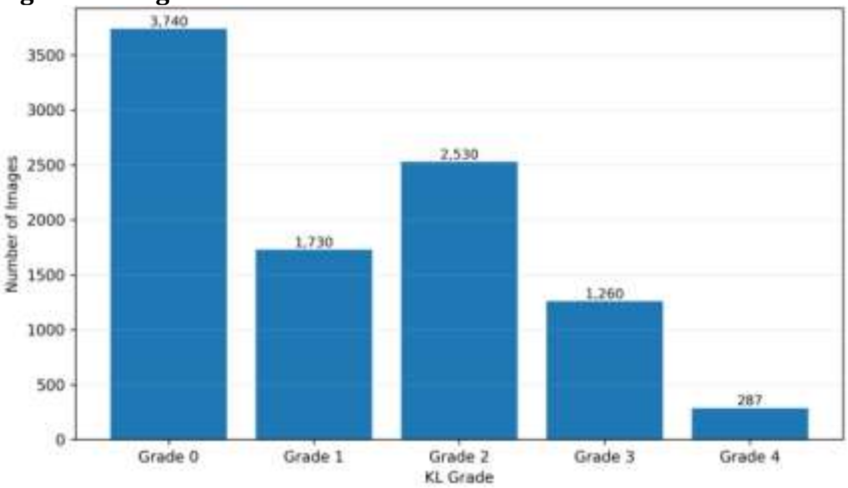
## 4. Result & Discussion

The results showed a difference in the number of classes. KL Grade 0 made up 39.17%. Kl Grade 4 only 3.01%. The framework that was suggested works well for determining how bad the condition is. The framework also keeps information safe. This happens when different places, with different systems are involved.

### 4.1 KL Grade-wise Image Distribution

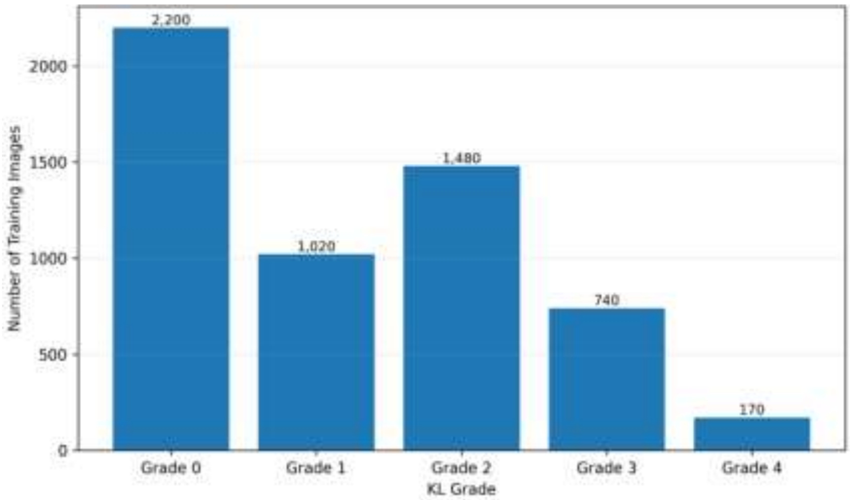
The figure 5, the frequency of radiographic images of the KOA according to Kellgren–Lawrence (KL) grading. This dataset has 9,547 images. The largest class is KL Grade 0 (Healthy) which represents 3,740 images (39.17%). KL Grade 2 (Minimal) contains 2,530 images (26.50%), followed by KL Grade 1 (Doubtful) with 1,730 images (18.12%). KL Grade 3 (Moderate) accounts for 1,260 images (13.20%), while KL Grade 4 (Severe) has only 287 images (3.01%). This suggests significant imbalances in the classes, especially in the case of severe KOA. The imbalanced attitude is crucial for the proposed Federated Attention-Based Deep Learning Framework since the proposed model is required to capture meaningful features from both the highly represented classes and the minority severity classes.

**Figure 5: KL grade distribution.**



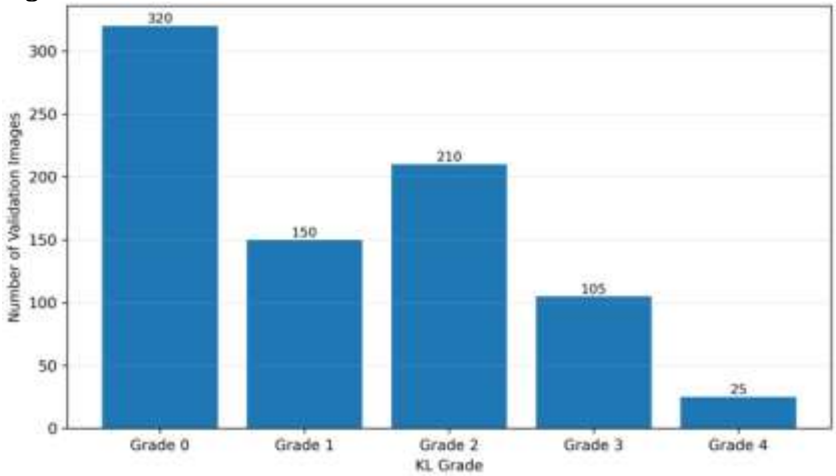

In figure 6, the 5,610 training images were distributed over the 5 KL severity grades as illustrated in the graph. The highest number of images constitutes Grade0 with 2200 images, around 39.22% of the training set. Grade 2 follows with 1,480 images (26.38%), while Grade 1 contains 1,020 images (18.18%). Grade 3 contributes 740 images (13.19%), and Grade 4 contains only 170 images (3.03%). The results show a distinct difference between healthy/minimal cases and severe KOA cases. The class distribution could affect both the local optimization and the FedAvg aggregation performed on these training images in the participating healthcare institutions, which is why local model learning is performed with these training images. Local model learning is performed with these training images, which means that the class distribution could impact local optimization and FedAvg aggregation when these are performed in the participating healthcare institutions.

**Figure 6: Training class distribution.**




The following figure 7, displays the number of validation images for each grade level and all KL grades 0–4. The number of images in Grade 0 is 320 (about 39.51% of the validation set). Grade 2 has 210 images (25.93%), while Grade 1 contains 150 images (18.52%). Grade 3 contributes 105 images (12.96%), and Grade 4 contains only 25 images (3.09%). The distribution of the data is similar to the training data, with the Grid 0 being the largest proportion and the Grid 4 being the smallest. Keeping the same proportion of the classes in training and validation sets could be used to check the consistency of the federated model's generalization levels. As the model is being developed, the performance of the model could be monitored in the process of convergence and identify possible overfitting. With a comparatively low number of images in Grade 4, it is important to take into account the results of individual classes, and not just the average image score. Precision, recall and F1-score are thus very important for evaluating minority-class performance.

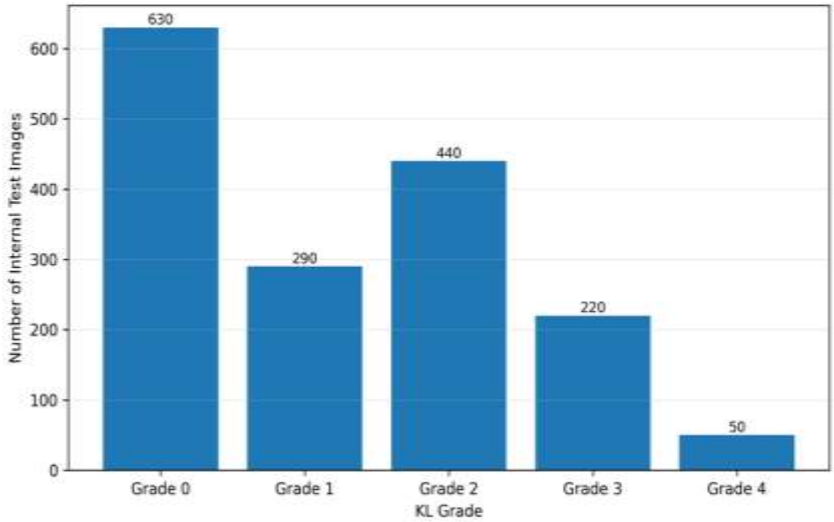
**Figure 7: Validation class distribution.**




## 4.2 Internal & External Test Images by KL Grade

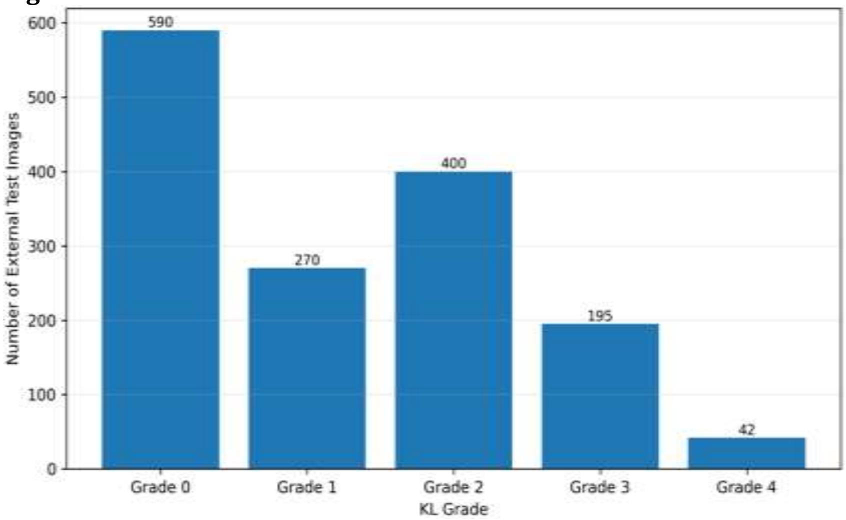
This figure 8, shows the number of internal test images in each of the 5 severity classes in KOA. Grade 0 has the largest number of images, having 630 samples, which represents ~38.65% of the internal test set. Grade 2 contains 440 images (26.99%), followed by Grade 1 with 290 images (17.79%). Grade 3 contains 220 images (13.50%), whereas Grade 4 has only 50 images (3.07%). The internal test distribution is generally representative of the classes of the training and test sets. This enables the final global model to be been tested for a set of data with distribution comparable to the data used when learning. The DenseNet121-attention architecture should be able to recognize useful radiological features and could be used to classify them according to KL Grades 0–4. Accuracy is an overall measure, while precision, recall and F1 score provide more meaningful evidence for performance for specific severity classes, particularly the relatively small severe KOA class.

**Figure 8: Internal test distribution.**



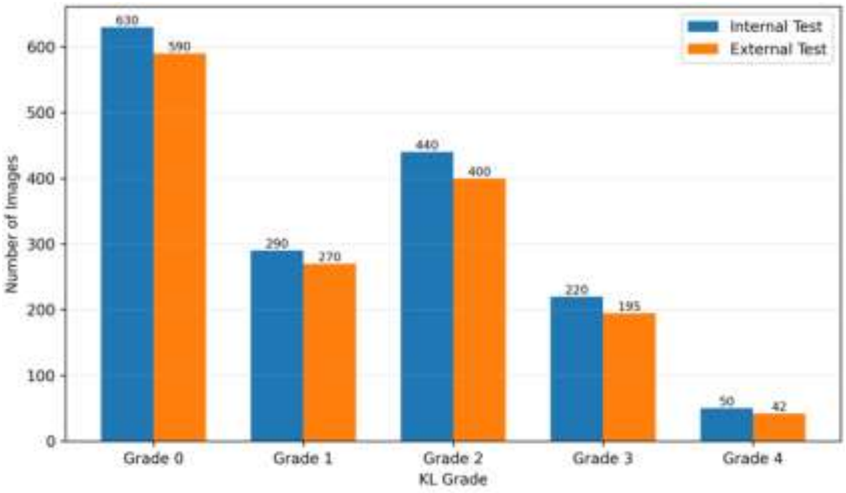

The figure 9, shows the distribution of 1497 external test images used for the analysis of the generalization capability of the proposed federated framework. The highest number of images, 590 (39.41%), falls under grade 0. Grade 2 contains 400 images (26.72%), followed by Grade 1 with 270 images (18.04%). Grade 3 contains 195 images (13.03%), whereas Grade 4 includes only 42 images (2.81%). The external test data set is crucial because it offers an evaluation environment and setting that is different from what was used for the training of the local model. This assessment could determine if the federated global model is able to accommodate a variety of radiographic properties from various sources. The attention mechanism and DenseNet121 backbone are used to highlight structures related to the disease, and federated learning allows for knowledge sharing between healthcare organizations without the exchange of the original radiograph of patients.

**Figure 9: External test distribution.**



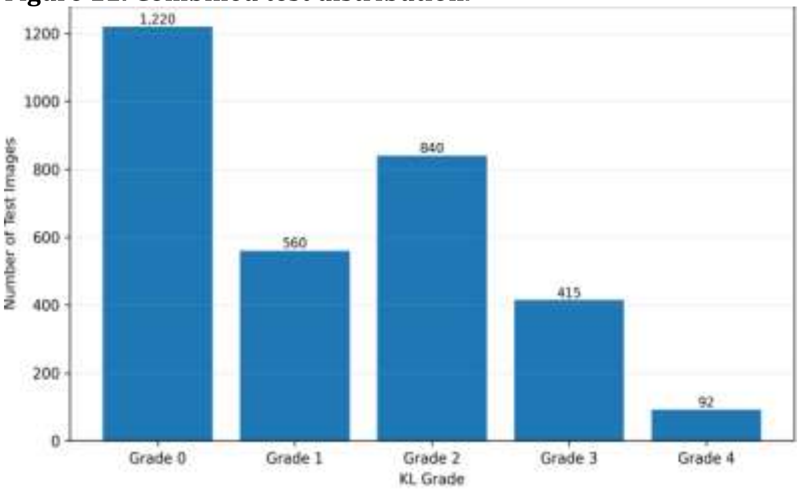

In the figure 10, the distributions of the internal test images and external test images are plotted as classes. There are 1630 images in the internal test set, and 1497 images in the external test set. The total number of images included for Grade 0 are 630 internal images, and 590 external images. In Grade 1, there are 290 images and 270 images and in Grade 2 there are 440 images and 400 images. There are 220 internal images and 195 external images in Grade 3, and 50 and 42 images, respectively, in Grade 4. The similar distributions serve as an effective means of comparing the performance of the federated global model in terms of various test conditions. The external test set is especially significant for the proposed methodology, since it evaluates the external generalization to heterogeneous images from radiography.

**Figure 10: Internal versus external test distribution.**



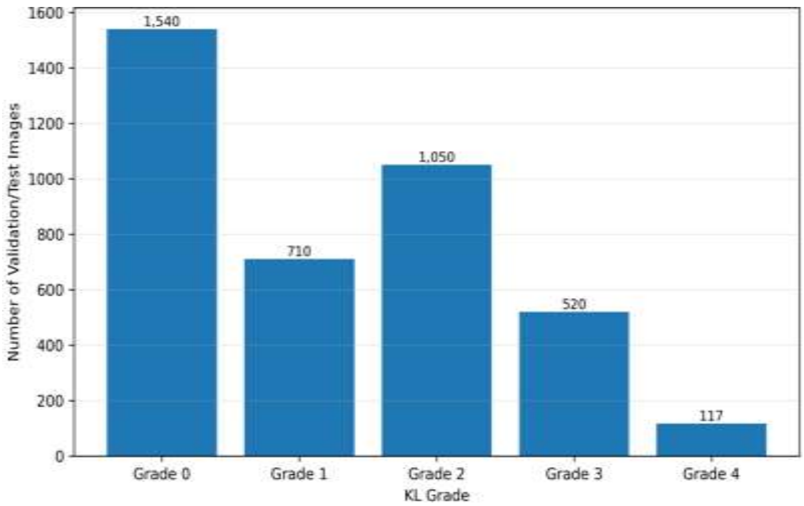

The figure 11, is created by mixing the internal and external test datasets together to reveal the overall distribution of the testing data based on KL grade. The grade 0 test images come from 630 internal and 590 external sample images. A total of 560 images exist for Grade 1 of which 290 are internal images and 270 are external images. The highest number of diseased images are present in grade 2, where 440 of the 840 images are internal and 400 of the 840 images are external. Grade 3 has 415 images and Grade 4 has 92 images. The combined test datasets have a total of 3,127 images. The results demonstrate that Grade 0 is still the largest class and that there is very little representation in grade 4. This distribution has impact on the final performance of the global federated model, and is important when interpreting the model.

**Figure 11: Combined test distribution.**




The figure 12, shows the distribution of non-training images in the 5 KL grades from validation, internal test and external test images. The 1,540 images in Grade 0 are made up of 320 validation images, 630 internal test images and 590 external test images. There are 710 images in Grade 1 and 1,050 images in Grade 2. Grade 3 has 520 images and Grade 4 has just 117 images. In total, these partitions have 3,937 images that are not in the training set. The distribution shows that the most common grade is 0 and that the rarest grade is 4. This assessment system will facilitate post-convergence independent evaluation of the federated model on the global level. Monitoring could be used on the validation set during training, classification and generalization testing could be done on an internal and external test set.

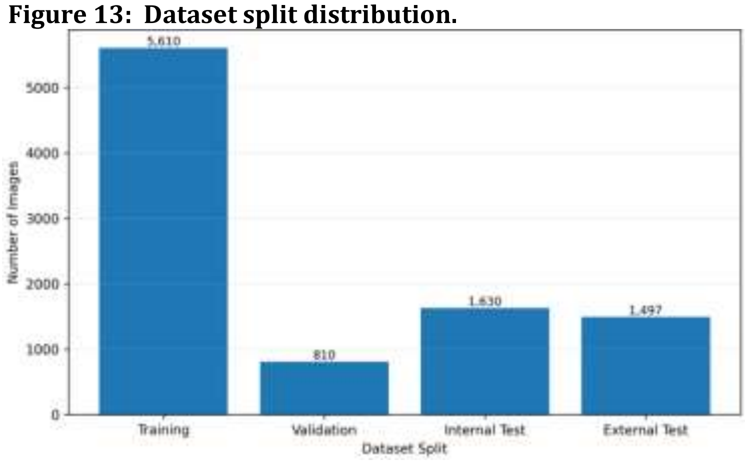
**Figure 12: Validation and test distribution.**




## 4.3 Dataset Split Distribution

This figure 13, is used to compare the four major dataset-partitions used in the proposed framework. The largest part of the dataset is the training set (5,610 images). There are 810, 1630 and 1497 images in the validation set and internal test set and external test sets respectively. The four partitions together contain 9547 X-ray pictures. The training set comprises about 58.76% of the data, and the other sets are respectively the internal test set (17.07%), the external test set (15.68%), and the validation set (8.48%). The proportion of the training set is relatively large enough for local Federated Optimization. The validation set is used for monitoring the model while the internal test set could be used for data conditions that are similar to those under which the model was trained.

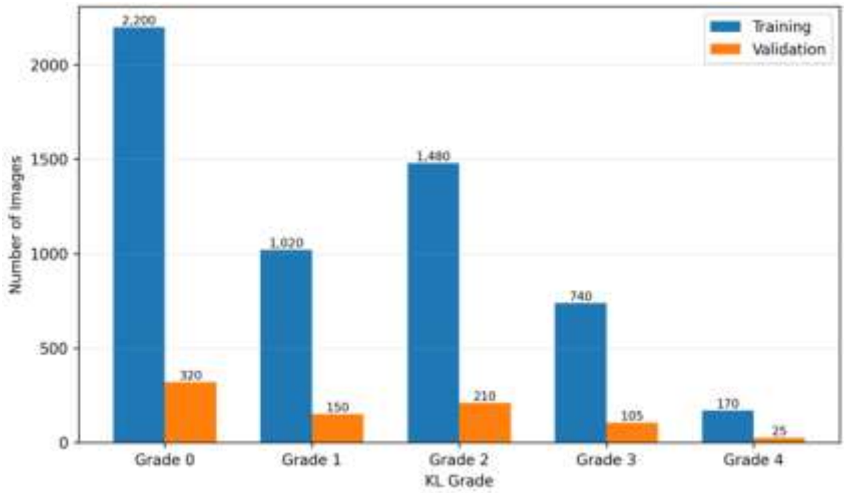
**Figure 13: Dataset split distribution.**




## 4.4 Training and Validation Distribution

The figure 14, depicts a comparison between the images in the train and validate sets for every KL severity grade. This training set has a size of 5,610 images; the validation set has a size of 810 images. The largest class in both partitions is Grade 0 which contains 2200 training images and 320 validation images. There are 1,020 images in the training set and 150 images in the validation set for Grade 1, and 1480 and 210 images in the training and validation sets for Grade 2, respectively. There are 740 training and 105 validation images in Grade 3 and 170 training and 25 validation images in Grade 4. By comparing these two datasets,It was observed that the values were proportional, facilitating consistent model developmentand validation. The relatively low number of Grade 4 children, however, is a potential source of class imbalance. The proposed augmentation of the data could help make the training process more diverse for each local; and feature extraction using DenseNet121 could provide attention to clinically relevant features.

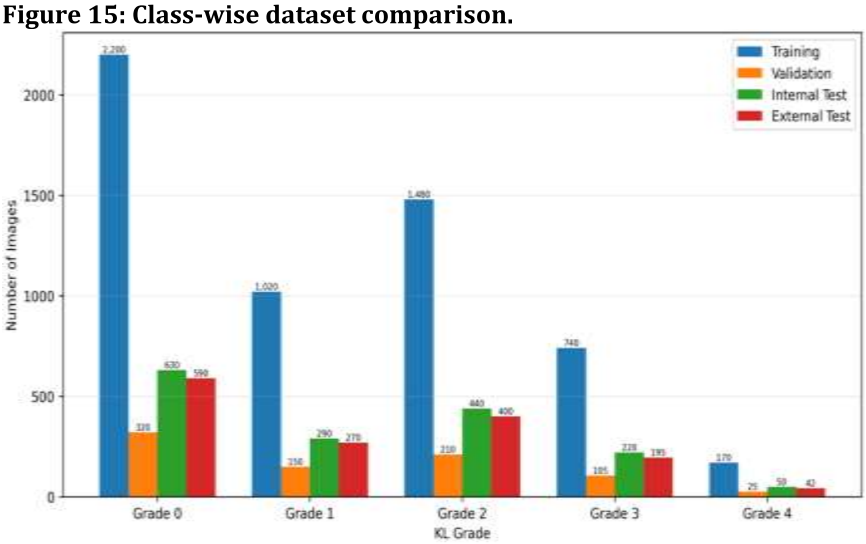
**Figure 14: Training versus validation distribution.**




## 4.5 Complete Class-wise Dataset Comparison

The figure 15, shows the complete comparison of all images (train, validation, internal test, external test) on all KL grades. The Grade 0 has 2200 training, 320 validation, 630 internal test and 590 external test images. Grade 1 contains 1,020, 150, 290, and 270 images, respectively. Grade 2 includes 1,480, 210, 440, and 400 images, while Grade 3 contains 740, 105, 220, and 195 images. Grade 4 contains the least amount of images with 170 training, 25 validation, 50 internal test and 42 external test images. In all partitions, the class of Grade 0 dominates while the class of Grade 4 is always the minority class. The consistent distribution shows that the experimental data is imbalanced in terms of the distribution of classes. Therefore, more than one classification metric should be applied to the proposed framework. Coordinated attention-based feature learning, augmentation and federated optimization could aid in improving the recognition of severity categories in minority classes of KOA.

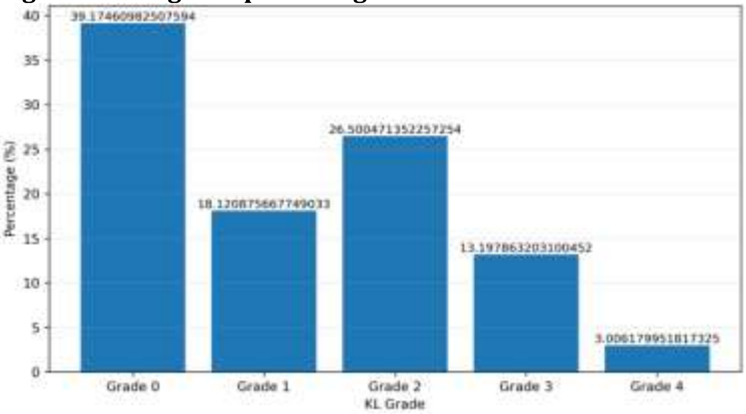
**Figure 15: Class-wise dataset comparison.**




## 4.6 Percentage Distribution of KL Grades

In the figure 16, the overall KL-grade distribution is expressed in percentage of the 9,547 available radiographic images. The largest proportion of the data (39.17%) is made up of Grade 0. Grade 2 represents 26.50%, followed by Grade 1 at 18.12%. Grade 3 makes up 13.20% and Grade 4 only 3.01%. So, 65.67% of the total data are for Grade 0 and Grade 2, and 16.21% for grade 3 and 4. This distribution shows a significant difference in the number of cases in the lower and higher KOA severity categories. This imbalance may impact on the optimization of a model, and could cause a classifier to overlook other classes to favor dominant classes. The augmentation and the preprocessing procedures suggested could add more variability to the minority class training data. Furthermore, DenseNet121 feature extraction based on attention enables to enhance the representation of disease-specific radiographic patterns. Therefore, along with overall accuracy, evaluation using precision, recall and F1-score on a class-wise basis is also necessary.

**Figure 16: KL grade percentage distribution.**




## Conclusion

The proposed Federated Attention-Based Deep Learning Framework offers a privacy-preserving method of automated multi-center diagnosis of knee osteoarthritis (KOA) from heterogeneous radiographic images. The framework combines image preprocessing, DenseNet121 based feature extraction, attention mechanism, federated model aggregation and KL-grade severity classification. The study employs 9547 knee radiographs ranging from KL Grades 0-4, of which 5610 are allocated to training, 810 are allocated to validation, 1630 are allocated to internal testing and 1497 are allocated to external testing. The class distribution of the dataset is highly skewed, with the Grade 0 class making up 39.17% of the images and the Grade 4 class limited to 3.01% of the images. The pre-processing pipeline consists of various steps to minimize the radiographic variations between different health care centers, such as resizing, intensity normalization, contrast enhancement, noise reduction, ROI extraction, augmentation, and client level standardization. There are improvements of the clinically relevant features on DenseNet121 and attention based feature learning, such as joint-space narrowing, and osteophytes. Collaborative model training in federated learning allows for avoiding sharing of raw patient radiographs, preserving data privacy and confidentiality. Incorporation of in-house and outside testing offers a foundation for evaluating generalization across diverse imaging situations. In general, the proposed framework provides a scalable and secure structure to support reliable classification of the severity of KOA in a distributed healthcare system.

## References

1. Geng, Ruizhi, Jiayi Li, Chen Yu, Chaoqun Zhang, Fei Chen, Jie Chen, Haonan Ni et al. "Knee osteoarthritis: Current status and research progress in treatment." Experimental and therapeutic medicine 26, no. 4 (2023): 481.
2. ur Rehman, Saif, Safdar Iqbal, Muhammad Umair Shahid, Muhammad Soman Jahangir, and Adnan Latif Malik. "Cartilage: structure, function, and the pathogenesis of osteoarthritis." Advancements in Synovial Joint Science-Structure, Function, and Beyond (2024): 1.
3. Xu, Jiaxian, Mingming Lei, and Dandan Xu. "Global, regional, and national burden of knee osteoarthritis attributable to high BMI: a systematic analysis from 1990 to 2021 and projections to 2050." Frontiers in Public Health 13 (2025): 1668257.
4. Piccolo, Claudia Lucia, Carlo Augusto Mallio, Federica Vaccarino, Rosario Francesco Grasso, and Bruno Beomonte Zobel. "Imaging of knee osteoarthritis: a review of multimodal diagnostic approach." Quantitative imaging in medicine and surgery 13, no. 11 (2023): 7582.
5. Ahmed, Sozan Mohammed, and Ramadhan J. Mstafa. "Identifying severity grading of knee osteoarthritis from x-ray images using an efficient mixture of deep learning and machine learning models." Diagnostics 12, no. 12 (2022): 2939.
6. Li, Mengfang, Yuanyuan Jiang, Yanzhou Zhang, and Haisheng Zhu. "Medical image analysis using deep learning algorithms." Frontiers in public health 11 (2023): 1273253.
7. Rawat, Saumya, Ved Prakash Chaturvedi, Binit Vaidya, Hemalatha Shanmugam, Abhilasha Shah, and Lavanya Airen. "Artificial intelligence in Kellgren–Lawrence grading of knee osteoarthritis: bridging radiographic tradition with algorithmic precision." Therapeutic Advances in Musculoskeletal Disease 18 (2026): 1759720X261442408.

8. Zhang, Di, Yuting Dong, Yao Xu, Junhui Qian, Miaoyu Ye, Qiang Yuan, and Jian Luo. "Enhancing knee osteoarthritis diagnosis with DMS: a novel dense multi-scale convolutional neural network approach." Journal of Orthopaedic Surgery and Research 19, no. 1 (2024): 851.
9. Qureshi, Ramsha, and Insoo Koo. "A Comprehensive Survey of Cybersecurity Threats and Data Privacy Issues in Healthcare Systems." Applied Sciences 16, no. 3 (2026): 1511.
10. Sood, Anubhuti, TanuragPattnaik, Rewa Malhotra, Vivek Nayyar, Bhaskar Narayan, Deepika Mishra, and Varun Surya. "Medical imaging privacy: a systematic scoping review of key parameters in dataset construction and data protection." Journal of Medical Imaging and Radiation Sciences 56, no. 5 (2025): 101914.
11. Lu, Chenming, Jiayin Zhang, and Ren Liu. "Deep learning-based image classification for integrating pathology and radiology in AI-assisted medical imaging." Scientific Reports 15, no. 1 (2025): 27029.
12. Ying, Haoran, Yijun Lia, and Zedong Fu. "Domain adaptation and generalization using foundation models in healthcare imaging." Available at SSRN 5345726 (2025).
13. Rauniyar, Ashish, Desta Haileselassie Hagos, Debesh Jha, Jan Erik Håkegård, Ulas Bagci, Danda B. Rawat, and Vladimir Vlassov. "Federated learning for medical applications: A taxonomy, current trends, challenges, and future research directions." IEEE Internet of Things Journal 11, no. 5 (2023): 7374-7398.
14. Nguyen, Dinh C., Quoc-Viet Pham, Pubudu N. Pathirana, Ming Ding, Aruna Seneviratne, Zihuai Lin, Octavia Dobre, and Won-Joo Hwang. "Federated learning for smart healthcare: A survey." ACM Computing Surveys (Csur) 55, no. 3 (2022): 1-37.
15. Ali, Md Shahin, Md Manjurul Ahsan, Lamia Tasnim, Sadia Afrin, Koushik Biswas, Md Maruf Hossain, Md Mahfuz Ahmed, Ronok Hashan, Md Khairul Islam, and Shivakumar Raman. "Federated learning in healthcare: Model misconducts, security, challenges, applications, and future research directions--A systematic review." arXiv preprint arXiv:2405.13832 (2024).
16. Abbas, Syed Raza, Zeeshan Abbas, Arifa Zahir, and Seung Won Lee. "Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration." In Healthcare, vol. 12, no. 24, p. 2587. MDPI, 2024.
17. Kwak, Lily, and Harrison Bai. "The role of federated learning models in medical imaging." Radiology: Artificial Intelligence 5, no. 3 (2023): e230136.
18. Efthymiadis, Filippos, Aristeidis Karras, Christos Karras, and Spyros Sioutas. "Advanced optimization techniques for federated learning on non-iid data." Future Internet 16, no. 10 (2024): 370.
19. Wang, Xu, Tianpeng Wang, and Zhanguo Su. "Attention-enhanced deep learning and machine learning framework for knee osteoarthritis severity detection in football players using X-ray images." Journal of Radiation Research and Applied Sciences 18, no. 2 (2025): 101428.
20. Zhang, Lirong, Hang Yu, and Yating Yang. "Large Separable Kernel Attention–Driven Multidimensional Feature Cross-Level Fusion Classification Network of Knee Cartilage Injury: Algorithm Development and Validation." JMIR Medical Informatics 13 (2025): e79748.
21. Pugazharasi, K., Bhuvaneswari Srinivasan, Anandh Nagarajan, and I. Govindharaj. "RFF-JSS: A Radiomic Feature Fusion and Joint Space Segmentation Framework for Automated Knee Osteoarthritis Assessment." Journal of Orthopaedic Reports (2026): 100899.
22. Pi, Sun-Woo, Byoung-Dai Lee, Mu Sook Lee, and Hae Jeong Lee. "Ensemble deep-learning networks for automated osteoarthritis grading in knee X-ray images." Scientific Reports 13, no. 1 (2023): 22887.
23. Yasin, Parhat, Yasen Yimit, Abuduainijiang Abulimiti, Haopeng Luan, Cong Peng, MaihemutiYakufu, and Xinghua Song. "Dual-center study on AI-driven multi-label deep learning for X-ray screening of knee abnormalities." Scientific Reports 15, no. 1 (2025): 38014.
24. Sudha, K., and A. Rajiv Kannan. "Enhanced swin transformer with dual attention for knee osteoarthritis severity grading from X-ray images." Scientific Reports (2026).
25. Tiwari, Anjali, Murali Poduval, and Vaibhav Bagaria. "Evaluation of artificial intelligence models for osteoarthritis of the knee using deep learning algorithms for orthopedic radiographs." World Journal of Orthopedics 13, no. 6 (2022): 603.
26. Nguyen, Dinh T., Duy K. Hoang, Linh N. Luu, An T. Truong, Hien Q. Kha, Huy T. Tran, Phu D. Ho et al. "An integrated deep learning and large language model for automated knee osteoarthritis grading and reporting from the Vietnam Osteoporosis Study." Osteoporosis and Sarcopenia (2026).
27. Ahmad, Fuzail, Mufda Jameel Abdalhadi Alrawashdeh, Adel Mefleh Widyan, Mujeeb Ahmed Shaikh, Pankaj Kumar, Ghanshyam G. Tejani, and Sunil Kumar Sharma. "RAD-OAEnsemble: Residual attention DenseNet with radiograph-EHR fusion and dual-stream deep feature embedding for knee osteoarthritis risk prediction." Egyptian Informatics Journal 34 (2026): 100929.
28. Rifat, Rakib Hossain, Abanti Chakraborty Shruti, Marufa Kamal, and Md Golam Rabiul Alam. "Privacy-Preserving Knee Osteoarthritis Classification: A Federated Learning Approach with GradCAM

Visualization." In 2023 26th International Conference on Computer and Information Technology (ICCIT), pp. 1-6. IEEE, 2023.

29. Alkhatatbeh, Tariq, Ahmad Alkhatatbeh, Hao Liao, Yan Liao, Zhilin Zhang, Hang Fang, Weidong Chen, Dongfeng Wu, and Rongkai Zhang. "Standard knee radiographs enable deep learning inference of MRI-defined cartilage and meniscal damage in early knee osteoarthritis: a study using the osteoarthritis initiative database." Frontiers in Physiology 17 (2026): 1858407.
30. Prezja, Fabi, Juha Paloneva, Ilkka Pölönen, Esko Niinimäki, and Sami Äyrämö. "DeepFake knee osteoarthritis X-rays from generative adversarial neural networks deceive medical experts and offer augmentation potential to automatic classification." Scientific Reports 12, no. 1 (2022): 18573.
31. Rawat, Saumya, Ved Prakash Chaturvedi, Binit Vaidya, Hemalatha Shanmugam, Abhilasha Shah, and Lavanya Airen. "Artificial intelligence in Kellgren–Lawrence grading of knee osteoarthritis: bridging radiographic tradition with algorithmic precision." Therapeutic Advances in Musculoskeletal Disease 18 (2026): 1759720X261442408.
32. Sundaramurthy, Arivoli, Murugesan Arumugam, Prabhu Rathinam, Adhithya Kathiravan, Dhana Shree Amirtharajan, and Yagnashri Nellutla. "Knee Osteoarthritis: A Diagnosis from Radiographs-A Hybrid Quantum-Based Deep Learning Approach."
33. Alkhatatbeh, Tariq, Ahmad Alkhatatbeh, Yan Liao, Zhilin Zhang, Hang Fang, Weidong Chen, and Rongkai Zhang. "Cross-Institutional Five-Class Kellgren–Lawrence Grading of Knee Osteoarthritis via Multitask Deep Learning." Annals of the New York Academy of Sciences 1557, no. 1 (2026): e70254.
34. Owida, Hamza Abu, Areen Arabiat, and Suhaila Abuowaida. "AI enhanced early diagnosis of knee osteoarthritis using X-ray imaging and machine learning approaches." Discover Applied Sciences (2026).
35. Abd Alnabi, Delveen Luqman, Shereen Sh Ahmed, and Nisreen Luqman Abd Alnabi. "Knee osteoarthritis stage classification based on hybrid fusion deep learning framework." Science Journal of University of Zakho 13, no. 2 (2025): 262-278.
36. Abdusalomov, Akmalbek, Sanjar Mirzakhalilov, Sabina Umirzakova, Otabek Ismailov, Djamshid Sultanov, Rashid Nasimov, and Young Im Cho. "Lightweight early detection of knee osteoarthritis in athletes." Scientific Reports 15, no. 1 (2025): 31413.
37. https://www.kaggle.com/datasets/jeftaadriel/osteoarthritis-initiative-oai-dataset
38. Chen, Pingjun. "Knee osteoarthritis severity grading dataset." Mendeley Data 1, no. 10.17632 (2018): 30784984.
39. Zhou, Tao, XinYu Ye, HuiLing Lu, Xiaomin Zheng, Shi Qiu, and YunCan Liu. "Dense convolutional network and its application in medical image analysis." BioMed Research International 2022, no. 1 (2022): 2384830.
40. Alzubaidi, Laith, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions." Journal of big Data 8, no. 1 (2021).
41. Cong, Shuang, and Yang Zhou. "A review of convolutional neural network architectures and their optimizations." Artificial Intelligence Review 56, no. 3 (2023): 1905-1969.
42. Chetan, R., D. V. Ashoka, and B. V. Ajay Prakash. "HybridTransferNet: Advancing soil image classification through comprehensive evaluation of hybrid transfer learning." (2023).