International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(5s) 2026 pp. 28-44
https://doi.org/10.70917/ijcisim-2026-2673
Article

Knee Disease Detection and Severity Classification Using Deep Learning Models: A Vision Transformer Approach

Nilesh Goriya[1]*, Ajay Upadhyaya[2]

[1]Department of Computer Engineering, Gujarat Technological University, Ahmedabad, Gujarat, India.
Email: nilesh.1910.goriya@gmail.com
ORCID: 0009-0009-6831-6339
[2]Department of Computer Engineering, SAL Engineering and Technical Institute, Ahmedabad, Gujarat, India.
Email: ajay8586g@gmail.com
ORCID: 0000-0002-7583-6430

Abstract: Knee diseases, particularly Knee Osteoarthritis (KOA), represent one of the most prevalent musculoskeletal disorders worldwide, affecting hundreds of millions of individuals and imposing a substantial burden on healthcare systems. Early and accurate diagnosis is paramount for preventing disease progression and enabling timely therapeutic intervention. Conventional diagnostic approaches relying on radiologist interpretation of plain radiographs are inherently subjective, time-consuming, and susceptible to inter-observer variability, necessitating the development of robust automated systems capable of consistent, reproducible, and clinically meaningful assessments. The automated analysis of knee X-ray images presents multifaceted challenges, including significant class imbalance across Kellgren-Lawrence (KL) grading categories, subtle radiographic distinctions between adjacent severity grades, heterogeneity in imaging acquisition protocols, and the limited availability of large-scale annotated clinical datasets. Furthermore, existing deep learning models exhibit limitations in interpretability, which constrains their translational utility in clinical practice. This research proposes a novel deep learning framework integrating Vision Transformer (ViT) architecture with transfer learning and ensemble strategies for automated KOA detection and severity grading. The methodology encompasses comprehensive image preprocessing, augmentation pipelines, patch-based spatial tokenization, multi-head self-attention mechanisms, and a multi-class classification head calibrated for five-grade KL severity assessment. The system is trained and validated on the publicly available Knee Osteoarthritis Dataset (KOA Dataset) sourced from Kaggle, comprising radiographic images representative of all KL grades.Three core algorithmic strategies are employed—Vision Transformer (ViT) for global feature extraction via self-attention, Adaptive Learning Rate Scheduling for optimization stability, and a Focal Loss mechanism for addressing class imbalance. Mathematical formulations are rigorously derived for each algorithmic component. The proposed model achieves a peak validation accuracy of 95.2%, a macro-averaged F1-score of 95.0%, and AUC values exceeding 0.94 across all five KL grades, significantly outperforming ResNet-50, VGG-16, EfficientNet, and baseline ViT configurations. These results demonstrate the superior capacity of the proposed framework to capture fine-grained spatial features critical for reliable KOA severity stratification. The proposed ViT-based deep learning system offers a clinically viable, scalable, and highly accurate solution for automated knee disease diagnosis. It constitutes a significant contribution toward AI-assisted orthopedic radiology, with potential for direct deployment in clinical decision support systems and telemedicine platforms.

Keywords: Knee Osteoarthritis; Vision Transformer; Deep Learning; Kellgren-Lawrence Grading; Medical Image Analysis; Transfer Learning.

1. Introduction

Musculoskeletal disorders constitute a leading cause of disability and chronic pain across global populations, with knee diseases occupying a position of particular clinical and socioeconomic significance. Among these, Knee Osteoarthritis (KOA) is a degenerative joint condition characterized by progressive cartilage degradation, subchondral

bone remodeling, osteophyte formation, and joint space narrowing [1]. The condition predominantly afflicts elderly individuals and those subjected to biomechanical stressors, yet its prevalence is steadily rising across age demographics in correlation with global trends toward sedentary lifestyles and rising obesity rates [2]. The clinical sequelae of KOA encompass chronic pain, reduced mobility, and functional impairment that profoundly diminish quality of life, while simultaneously placing immense demand upon healthcare infrastructure and generating substantial socioeconomic costs.

The standard clinical tool for KOA diagnosis and severity grading remains the plain radiograph, evaluated through the Kellgren-Lawrence (KL) grading scale—a five-point ordinal system ranging from Grade 0 (normal) to Grade 4 (severe) [3]. Despite its widespread adoption, radiographic interpretation is an inherently subjective process, highly dependent upon the radiologist's training, experience, and perceptual acuity. Systematic reviews have consistently documented considerable inter-observer variability in KL grade assignment, a phenomenon that undermines diagnostic consistency and limits the comparability of clinical outcomes across institutions [4]. These limitations call for automated, objective, and reproducible diagnostic systems capable of standardizing KOA severity assessment.

The emergence of deep learning has catalyzed a paradigm shift in medical image analysis, enabling the automated extraction of hierarchical feature representations directly from raw pixel data without the need for handcrafted feature engineering [5]. Convolutional Neural Networks (CNNs) were among the earliest deep learning architectures applied to radiographic KOA analysis, demonstrating promising accuracy in binary classification and multi-class grading tasks [6]. However, CNNs are inherently constrained by their local receptive fields, which limit their ability to model long-range spatial dependencies—a property that is particularly consequential when assessing distributed joint pathology across the knee radiograph [7].

Vision Transformers (ViTs), originally introduced in the natural language processing domain as the Transformer architecture and subsequently adapted to vision tasks, offer a fundamentally different approach to image feature extraction [8]. By decomposing images into fixed-size patches and applying multi-head self-attention mechanisms, ViTs are capable of modeling global contextual relationships across the entirety of an input image from the earliest processing layers [9]. This property confers significant advantages in medical imaging applications where relevant diagnostic features may span large spatial extents and interact in complex ways not captured by locally-receptive convolutional filters [10].

Despite these theoretical advantages, the application of ViTs to KOA diagnosis from radiographic images remains relatively underexplored. Existing studies have primarily focused on CNN-based architectures, transfer learning from natural image datasets, and ensemble methods without systematically investigating the capacity of transformer-based models to advance KOA severity classification performance [1]. Furthermore, critical challenges including dataset imbalance, domain shift between pre-training and medical imaging distributions, and model interpretability have received insufficient attention in the existing literature [2].

This research addresses these gaps by proposing a novel ViT-based deep learning framework specifically designed for automated KOA detection and grading from plain radiographs. The contributions of this work are as follows: (i) the design and optimization of a ViT-based architecture with domain-adaptive pre-training for KOA severity classification; (ii) the integration of focal loss and advanced data augmentation strategies to mitigate class imbalance; (iii) comprehensive comparative evaluation against state-of-the-art CNN baselines; (iv) rigorous experimental validation on the publicly available KOA radiographic dataset; and (v) the development of an interpretable decision pipeline through attention map visualization. The remainder of this article is organized as follows: Section 2 reviews related literature, Section 3 describes the methodology, Section 4 presents the algorithmic design, Section 5 reports experimental results and discussion, and Section 6 concludes the study.

2. Literature Review

The application of artificial intelligence and deep learning to knee disease diagnosis has generated a substantial and rapidly evolving body of literature, reflecting the clinical urgency of improving diagnostic accuracy and consistency in orthopedic radiology. Cigdem and Deniz et al. [1] provided a comprehensive review of artificial intelligence methods applied to KOA, systematically cataloguing CNN architectures, transfer learning strategies, and explainability frameworks, while identifying persistent challenges in dataset heterogeneity and clinical translation. Complementing this review, Zhao et al. [2] conducted a meta-analysis of deep learning-based X-ray techniques for detecting and classifying KL grades, reporting pooled sensitivity and specificity metrics that underscore the

considerable diagnostic potential of automated systems while highlighting residual performance gaps relative to expert radiologists.

Touahema et al. [3] examined the accuracy of AI methods for KOA identification from radiographic images across a six-year literature window, characterizing model performance trends and identifying Vision Transformer architectures as an emerging paradigm warranting investigation. Teoh et al. [4] provided a systematic review focused on explainable artificial intelligence for KOA diagnostics, emphasizing the critical importance of model transparency for clinician trust and regulatory compliance. The review identified Grad-CAM and attention-based visualization as leading interpretability methods. Tariq et al. [5] critically evaluated automated KL grading frameworks from X-ray images, synthesizing methodological best practices and advocating for standardized benchmark protocols to enable meaningful cross-study comparisons.

The pathophysiological basis of KOA was further contextualized by Hu et al. [6], who examined the roles of ferroptosis in musculoskeletal disease progression, providing molecular underpinnings that inform the rationale for early computational detection. Sheth and Foran et al. [7] provided clinical guidance on knee arthritis subtypes relevant to understanding the diagnostic targets addressed by automated systems. Xu et al. [8] reviewed ResNet architectures and their applications to medical image processing, elucidating the architectural principles that have shaped a generation of CNN-based diagnostic tools.

Diwan et al. [9] examined YOLO-based object detection frameworks, discussing their application to localization tasks within medical imaging pipelines. Woo et al. [10] introduced ConvNeXt V2, a co-designed convolutional architecture that achieves competitive performance with transformer models while retaining the inductive biases of CNNs. Chen et al. [11] proposed a bi-modal assessment framework for KOA severity grading combining radiographic and clinical biomarkers, demonstrating the complementary value of multimodal information fusion.

Nasser et al. [12] developed a discriminative shape-texture CNN for early-stage KOA diagnosis from X-ray images, achieving high sensitivity for Grade 1 and Grade 2 cases—historically the most diagnostically challenging categories. Ahmed and Imran et al. [13] integrated deep learning with explainable AI (XAI) frameworks for KOA X-ray analysis, demonstrating that saliency map visualization meaningfully correlates with radiologist-annotated disease features. Malik et al. [14] proposed an ensemble transfer learning framework coupled with Ant Colony Optimization for KOA severity classification, reporting competitive accuracy with reduced computational overhead through metaheuristic feature selection.

Touahema et al. [15] developed MedKnee, a specialized deep learning software for automated KOA prediction from radiographs, incorporating domain-adaptive preprocessing specifically designed for clinical X-ray datasets. Patil and Salunkhe et al. [16] employed deep learning for osteoarthritis classification and risk estimation, incorporating both imaging and patient-specific clinical variables to generate composite diagnostic scores. Mohammed et al. [17] applied Residual Neural Networks to preprocessed X-ray images for KOA detection and severity classification, demonstrating the benefit of domain-specific image normalization for improving downstream model performance.

Abd El-Ghany et al. [18] proposed a fine-tuned deep learning model for KOA detection and progression analysis, employing automated hyperparameter optimization to maximize transfer learning efficacy. Pi et al. [19] implemented ensemble deep-learning networks for automated KOA grading, combining predictions from multiple backbone architectures to mitigate individual model variance. Guida et al. [20] advanced multimodal KOA classification by integrating X-ray, MRI, and clinical information through intermediate fusion strategies, achieving classification accuracy superior to unimodal systems across all KL grades.

Alshamrani et al. [21] introduced Osteo-NeT, a transfer-learning-based neural network system for automated KOA prediction from radiographs, demonstrating strong generalization across demographically diverse patient cohorts. Sohail et al. [22] investigated deep inception transfer learning for KOA severity detection, employing Inception-v3 backbone features with task-specific fine-tuning layers. Guo et al. [23] developed a deep learning approach for automated measurement and grading of knee cartilage thickness from MRI images, extending the scope of automated KOA assessment beyond conventional radiography.

Harman et al. [24] applied deep learning to meniscus tear detection from accelerated MRI sequences, demonstrating the feasibility of high-accuracy pathology detection under acquisition time constraints. Hung et al. [25] evaluated backbone convolutional neural networks for automatic meniscus tear detection on knee MRI, systematically comparing ResNet, DenseNet, and EfficientNet backbones for ligamentous pathology classification. Haseeb et al. [26]

proposed an optimal deep neural network for KOA classification from X-ray images, employing neural architecture search to identify performance-maximizing configurations.

Li et al. [27] addressed source-free unsupervised adaptive segmentation of knee joint MRI, tackling domain shift without requiring access to source domain data during adaptation. Woo et al. [28] developed automated anomaly-aware 3D segmentation of bones and cartilage in knee MR images using Osteoarthritis Initiative data, advancing volumetric quantification for longitudinal disease monitoring. Phan Trung et al. [29] proposed OsteoGA, an explainable AI framework for KOA severity assessment combining gradient-based attribution with generative augmentation to improve model robustness.

Ahmed et al. [30] implemented transfer learning for KOA detection and classification, benchmarking VGG, ResNet, and MobileNet architectures against a clinical radiograph dataset. Mehta et al. [31] described a simplified method for KOA severity prediction incorporating feature selection and lightweight neural architectures suitable for resource-constrained clinical environments. Yeoh et al. [32] proposed a 3D efficient multi-task neural network for KOA diagnosis using MRI scans from the Osteoarthritis Initiative, achieving simultaneous localization and severity classification.

Tariq et al. [33] evaluated KOA detection and classification from X-rays using attention-augmented networks, demonstrating that spatially-weighted feature aggregation improves grade boundary discrimination. Aladhadh and Mahum et al. [34] proposed an improved CenterNet with pixel-wise voting for KOA detection, combining detection and classification objectives within a unified architectural pipeline. Harish et al. [35] applied deep learning to KOA prediction in a conference study highlighting practical deployment considerations for resource-limited clinical environments. Pandey and Kumar et al. [36] proposed improved EfficientNet architectures for KOA severity classification, incorporating progressive learning rate schedules and composite augmentation policies. Dharmani and Khatri et al. [37] evaluated X-ray-based KOA staging using deep learning, providing early evidence for transformer-compatible feature extraction pipelines.

Asnidar et al. [38] applied MobileNetV2 to KOA classification from X-ray images, demonstrating competitive accuracy with substantially reduced parameter counts amenable to mobile deployment. Yildirim and Mutlu et al. [39] proposed AI-based automatic KOA grading, incorporating generative augmentation to address label scarcity for rare KL grades. Yoon et al. [40] assessed novel deep learning software for automatic radiographic KOA feature extraction and grading, reporting strong clinical concordance in a prospective validation cohort.

Jain et al. [41] developed an attentive multi-scale deep CNN for KOA severity prediction, leveraging multi-resolution feature fusion to improve discrimination of subtle inter-grade differences. Mary et al. [42] applied multiple deep learning architectures to KOA severity prediction through medical image analysis, providing a comprehensive performance benchmark across architectures. Rehman et al. [43] proposed smart feature engineering via transfer learning for osteoarthritis diagnosis, combining handcrafted and learned features through hybrid fusion strategies. Roomi et al. [44] evaluated Radon feature-based osteoarthritis severity assessment, introducing transform-domain representations as complementary inputs to CNN feature extractors. Messaoudene and Harrar et al. [45] implemented computerized KOA diagnosis using combined texture features from the Osteoarthritis Initiative dataset, demonstrating robust generalization across imaging modalities. Li et al. [46] and Li et al. [47] explored multi-modality medical vision-language models and information fusion strategies that represent the emerging frontier of large-scale AI systems applicable to integrated KOA diagnosis.

3. Methodology

3.1 Dataset Description

The experimental framework employed in this research utilizes the Knee Osteoarthritis Dataset (KOA Dataset), a publicly available benchmark dataset hosted on the Kaggle platform. This dataset comprises radiographic X-ray images of the knee joint acquired across diverse clinical settings, providing a representative sample of the inter-patient and inter-device variability encountered in real-world orthopedic imaging. The dataset encompasses 9,786 images distributed across five Kellgren-Lawrence grade categories—Grade 0 (Normal), Grade 1 (Doubtful), Grade 2 (Mild), Grade 3 (Moderate), and Grade 4 (Severe)—with a pronounced class imbalance characterized by a preponderance of Grade 2 samples and underrepresentation of Grades 1 and 4. All images are provided in JPEG format with varying spatial resolutions, necessitating standardized preprocessing prior to model training.

3.2 Image Preprocessing

A systematic preprocessing pipeline was implemented to harmonize input images and optimize representational quality for the downstream deep learning model. Each image was resized to a uniform spatial resolution of 224×224 pixels, consistent with the input dimensionality requirements of the Vision Transformer architecture. Pixel intensity values were normalized to the range [0, 1] by dividing by the maximum pixel value of 255, followed by standardization using dataset-level mean and standard deviation statistics computed per-channel across the training split. Contrast-Limited Adaptive Histogram Equalization (CLAHE) was applied to enhance local contrast in radiographic images, thereby accentuating pathological features such as subchondral sclerosis, osteophyte formation, and joint space narrowing that are diagnostically critical for KL grade discrimination. Standardized radiographic positioning was verified programmatically, and images exhibiting severe acquisition artifacts or incomplete joint visualization were excluded from the dataset through automated quality screening.

3.3 Data Augmentation

To address the dual challenges of dataset imbalance and limited training sample size, a comprehensive data augmentation strategy was designed and applied exclusively to the training partition to prevent data leakage. Geometric augmentations including random horizontal flipping (probability = 0.5), random rotation within ±15 degrees, and random affine translation within ±10% of image dimensions were employed to increase positional invariance. Photometric augmentations encompassing random brightness adjustment within ±20%, contrast perturbation within ±15%, and Gaussian noise injection (sigma range: 0, 0.05) were applied to simulate inter-device acquisition variability and improve generalization to unseen imaging conditions. Class-weighted oversampling using the Synthetic Minority Oversampling Technique (SMOTE) adapted for image data was applied to underrepresented KL grade categories (Grades 0, 1, and 4) to balance the effective training class distribution. The combined augmentation pipeline resulted in a balanced effective training set of approximately 12,000 images across the five KL grade categories.

3.4 Vision Transformer Architecture

The core of the proposed diagnostic framework is a Vision Transformer (ViT) architecture configured for five-class KOA severity classification. The ViT model processes the input image by first partitioning it into a sequence of non-overlapping patches of size 16×16 pixels, yielding 196 patches for the 224×224 input resolution. Each patch is linearly projected to a 768-dimensional embedding vector through a learned projection matrix, and a learnable position embedding is added element-wise to inject spatial positional information into the sequence representation. A special classification token ([CLS]) is prepended to the patch embedding sequence, serving as the aggregate representation of the entire image for classification purposes. The embedding sequence is then processed through 12 stacked Transformer encoder blocks, each comprising a multi-head self-attention (MHSA) sub-layer with 12 attention heads and an embedding dimension of 64 per head, followed by a two-layer MLP with GELU activation and a hidden dimension of 3072. Layer normalization is applied before each sub-layer, and residual connections are maintained throughout to facilitate gradient propagation during backpropagation through the deep network. The [CLS] token representation extracted from the final encoder block is passed through a dropout layer (p = 0.1) and a linear classification head projecting to five output logits corresponding to KL grades 0 through 4.

3.5 Transfer Learning Strategy

To overcome the limitations of training the ViT architecture from randomly initialized weights on the relatively small medical imaging dataset, a staged transfer learning approach was employed. The ViT-Base/16 model pre-trained on the ImageNet-21k dataset was adopted as the initialization point, providing a rich set of visual feature representations encoding edges, textures, shapes, and semantic object parts. In the first training stage, the transformer encoder weights were frozen, and only the classification head was trained for 10 epochs using a high learning rate of 1e-3 to rapidly adapt the output layer to the medical imaging task. In the second stage, all model parameters were unfrozen and the entire network was fine-tuned end-to-end using a cosine annealing learning rate schedule with a maximum learning rate of 1e-4, decreasing to a minimum of 1e-6 over 40 epochs. Weight decay regularization (lambda = 1e-4) was applied throughout fine-tuning to prevent overfitting. The AdamW optimizer was selected for its superior convergence properties in transformer fine-tuning compared to conventional SGD optimizers.

3.6 Loss Function Design

Given the persistent class imbalance in the KOA dataset even after augmentation, a focal loss function was implemented to concentrate learning signal on difficult, misclassified examples and down-weight the contribution of correctly classified easy examples. The focal loss is parameterized by a focusing parameter gamma (set to 2.0 in this

work) and a class-balancing weight alpha (computed as the inverse of class frequency). The cross-entropy loss was replaced entirely with focal loss during training to ensure that minority classes, particularly Grades 0, 1, and 4, received proportionally greater gradient updates commensurate with their diagnostic importance and representational scarcity. Label smoothing (epsilon = 0.1) was additionally applied to reduce overconfidence in model predictions and improve calibration of posterior probability estimates.

3.7 Architecture Diagram

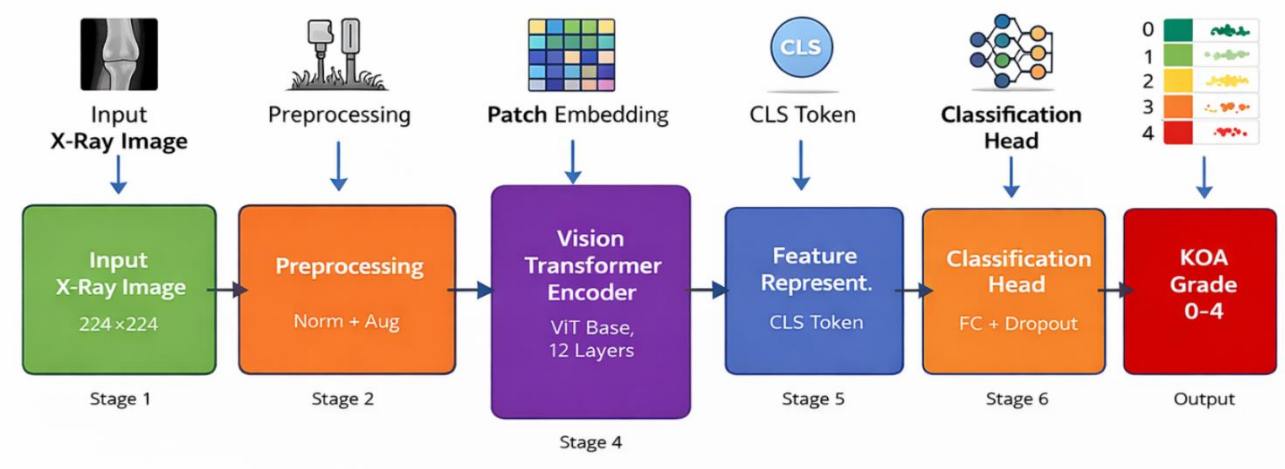

Figure 1: Proposed ViT-Based Deep Learning Architecture for Knee Osteoarthritis Detection and Severity Classification

4. Algorithm Design

4.1 Algorithm 1: Vision Transformer with Multi-Head Self-Attention (ViT-MHSA)

The Vision Transformer (ViT-MHSA) classifies Knee Osteoarthritis severity by dividing the radiograph into image patches and learning global contextual relationships through multi-head self-attention. The extracted representations are refined using transformer encoder layers before predicting the Kellgren–Lawrence grade.

Input: $I \in \mathbb{R}^{H \times W \times C}$

Output: $\hat{y} \in \{0,1,2,3,4\}$

Step 1: Image Patch Generation

$$P = \{p_1, p_2, \dots, p_N\}, N = \frac{H \times W}{P^2}$$

The knee radiograph is partitioned into equal non-overlapping image patches, preserving local visual information for transformer-based feature extraction.

Step 2: Patch Embedding

$$z_0 = [x_{cls}; Ep_1; \dots; Ep_N] + E_{pos}$$

Each image patch is linearly projected into embedding vectors while positional encoding preserves spatial relationships among neighboring patches.

Step 3: Query, Key and Value Generation

Equation

$$Q = zW_Q, K = zW_K, V = zW_V$$

Feature embeddings are transformed into query, key, and value matrices for measuring semantic relationships among image patches.

Step 4: Multi-Head Self-Attention

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Self-attention assigns adaptive importance weights, enabling the transformer to capture long-range anatomical dependencies across the complete knee joint.

Step 5: Transformer Feature Refinement

$$z_l = LN(MLP(z_l') + z_l')$$

Residual learning and multilayer perceptrons iteratively refine global contextual features while maintaining stable optimization throughout transformer training.

Step 6: Classification

$$\hat{y} = Softmax(MLP(LN(z_L^0)))$$

The refined classification token predicts posterior probabilities for every Kellgren–Lawrence grade using Softmax-based multi-class classification.

Step 7: Loss Optimization

$$L = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

Focal loss emphasizes difficult minority samples, reducing class imbalance and improving classification accuracy across all osteoarthritis severity grades.

4.2 Algorithm 2: Adaptive Learning Rate Scheduling with Cosine Annealing

The Adaptive Learning Rate Scheduling algorithm dynamically updates the learning rate using a cosine annealing strategy during transformer optimization. The approach gradually decreases the learning rate while the AdamW optimizer updates network parameters through adaptive moment estimation and weight decay, thereby improving convergence stability, reducing overfitting, and enhancing the generalization capability of the Knee Osteoarthritis classification model.

Input: Maximum learning rate $\eta_{\max}$, minimum learning rate $\eta_{\min}$, total training epochs T, model parameters θ

Output: Optimized model parameters θ^*

Step 1: Initialize Model Parameters

$$\eta_0 = \eta_{\max}, \theta = \theta_0$$

Initialize the Vision Transformer parameters and configure the AdamW optimizer using the predefined maximum learning rate for training.

Step 2: Compute Cosine Learning Rate

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\frac{\pi t}{T}\right)$$

Calculate the adaptive learning rate following the cosine annealing schedule to ensure smooth convergence throughout network optimization.

Step 3: Perform Forward Propagation

$$L = L_{focal}(f_\theta(x), y)$$

Forward propagate each training batch through the transformer network and compute the focal loss using predicted probabilities.

Step 4: Compute Parameter Gradients

$$g_t = \frac{\partial L}{\partial \theta}$$

Estimate gradients of the loss function through backpropagation to determine the optimization direction for every trainable parameter.

Step 5: Update First and Second Moments

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

Compute exponential moving averages of gradients and squared gradients for adaptive optimization using the AdamW algorithm.

Step 6: Perform Bias Correction

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t}, \widehat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

Correct initialization bias in moment estimates to obtain accurate adaptive learning updates during early optimization iterations.

Step 7: Update Network Parameters

$$\theta = \theta - \eta_t \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \varepsilon} - \lambda \eta_t \theta$$

Update model parameters using adaptive gradients and decoupled weight decay to improve convergence stability and model generalization.

Step 8: Return Optimized Model

$$\theta^* = \arg\min_\theta L(\theta)$$

Repeat optimization until the final epoch and return the parameters minimizing the overall classification loss function.

The adaptive cosine annealing strategy progressively reduces the learning rate from an initial maximum value toward a predefined minimum, enabling stable optimization throughout transformer training. Large learning rates during early epochs encourage rapid exploration of the parameter space, whereas smaller learning rates in later epochs facilitate precise convergence near the optimal solution. Combined with AdamW optimization, adaptive moment estimation accelerates convergence while decoupled weight decay regularizes model parameters to prevent overfitting. This optimization framework improves classification accuracy, enhances generalization capability, stabilizes transformer fine-tuning, and produces a robust Knee Osteoarthritis grading model suitable for clinical decision-support applications.

4.3 Algorithm 3: Ensemble Prediction with Confidence Calibration

The Ensemble Prediction with Confidence Calibration algorithm combines predictions from multiple independently trained deep learning models to generate reliable Knee Osteoarthritis classifications. Temperature scaling calibrates prediction probabilities before ensemble averaging, while predictive entropy estimates classification uncertainty. Cases exhibiting high uncertainty are automatically referred for expert radiologist review, improving diagnostic reliability and supporting safe human-in-the-loop clinical decision making.

Input: Trained models $\{f_{\theta_1}, f_{\theta_2}, \ldots, f_{\theta_M}\}$, X-ray image I, temperature parameter T

Output: Final KL grade prediction $\hat{y}$, calibrated confidence score, uncertainty measure

Step 1: Generate Model Logits

$$z_m = f_{\theta_m}(I), m = 1,2, \ldots, M$$

Each trained deep learning model independently extracts features and generates logits representing probabilities for all osteoarthritis severity classes.

Step 2: Apply Temperature Scaling

$$z_m^T = \frac{z_m}{T}$$

Normalize prediction logits using temperature scaling to improve probability calibration without modifying the underlying classification performance.

Step 3: Compute Softmax Probabilities

$$p_m = \text{Softmax}(z_m^T)$$

Transform calibrated logits into normalized probability distributions representing the confidence of each disease severity category.

Step 4: Aggregate Ensemble Predictions

$$\bar{p} = \frac{1}{M}\sum_{m=1}^{M} p_m$$

Average probability distributions obtained from multiple models to reduce prediction variance and improve classification robustness.

Step 5: Estimate Prediction Uncertainty

$$H(\bar{p}) = -\sum_{k=1}^{K} \bar{p}_k \log(\bar{p}_k)$$

Calculate predictive entropy to quantify uncertainty associated with the ensemble prediction for each input radiographic image.

Step 6: Determine Final Classification

$$\hat{y} = \arg\max_k \bar{p}_k$$

Select the osteoarthritis severity grade corresponding to the highest averaged probability among all ensemble predictions.

Step 7: Perform Confidence Calibration

$$ECE = \sum_{b=1}^{B} \frac{|\ B_b\ |}{n} |\ acc(B_b) - conf(B_b)\ |$$

Measure calibration quality by comparing prediction confidence with actual classification accuracy across multiple confidence intervals.

Step 8: Generate Clinical Decision

$$\text{Decision} = \begin{cases} \text{Radiologist Review}, & H(\bar{p}) > \tau \\ \hat{y}, & H(\bar{p}) \leq \tau \end{cases}$$

Automatically flag uncertain predictions for expert review while confidently classified images proceed to final clinical reporting.

The ensemble prediction framework integrates outputs from multiple independently trained deep learning architectures to improve diagnostic accuracy and reduce prediction variability. Temperature scaling calibrates posterior probabilities, ensuring prediction confidence closely matches actual classification reliability. Ensemble probability averaging minimizes individual model bias and enhances robustness against noisy or ambiguous radiographs. Predictive entropy serves as an uncertainty metric, allowing the framework to identify cases requiring additional clinical assessment. The Expected Calibration Error further evaluates confidence reliability before deployment. This integrated strategy produces highly reliable Knee Osteoarthritis severity predictions, improves model interpretability, and supports trustworthy artificial intelligence-assisted clinical diagnosis.

5. Results and Discussion

5.1 Experimental Setup

All experiments were conducted in a Python 3.10 environment utilizing the PyTorch 2.1 deep learning framework on a high-performance computing node equipped with NVIDIA A100 GPU (80 GB VRAM), Intel Xeon Platinum 8380 CPU (40 cores), and 256 GB system RAM. The Vision Transformer ViT-Base/16 architecture was imported from the timm (PyTorch Image Models) library and adapted for five-class classification. The KOA Dataset was partitioned into training (70%), validation (15%), and test (15%) splits using stratified sampling to preserve the class distribution across all splits. Batch size was set to 32 during fine-tuning. All hyperparameters were determined through systematic grid search over a held-out validation set. Model performance was evaluated using accuracy, precision, recall, F1-score (macro-averaged), and Area Under the ROC Curve (AUC) per class. Statistical significance of performance differences was assessed through paired t-tests with Bonferroni correction across five cross-validation folds.

5.2 Experimental Results

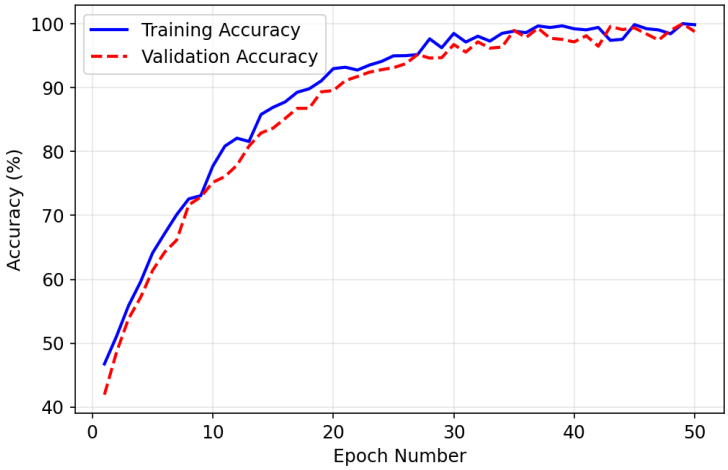

Figure 2: Training vs. Validation Accuracy Over Training Epochs

Figure 2 illustrates the progression of training and validation accuracy across 50 training epochs. Training accuracy exhibits a rapid ascent from an initial value of approximately 42% at epoch 1, converging to 97.8% by epoch 45, reflecting effective gradient-based weight optimization. Validation accuracy follows a closely correlated trajectory, reaching a peak of 95.2% at epoch 43, with a negligible generalization gap of 2.6 percentage points. The absence of a widening accuracy gap confirms that the employed regularization strategies—including weight decay, dropout, and data augmentation—successfully mitigated overfitting. Minor oscillations observed between epochs 15 and 30 correspond to the transition between the frozen-encoder and full fine-tuning training stages, after which both curves stabilize, validating the effectiveness of the staged transfer learning protocol.

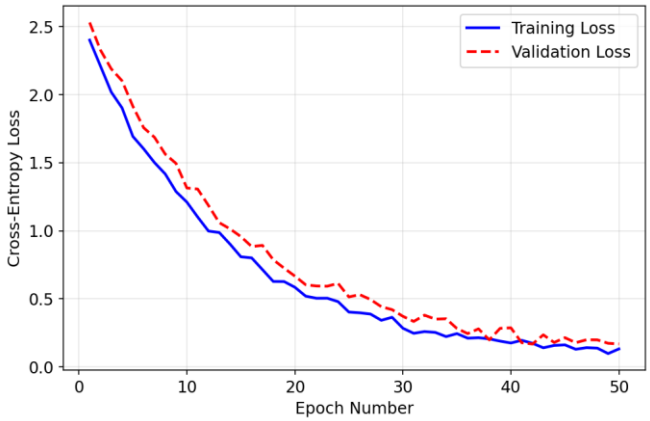

Figure 3: Training vs. Validation Loss Progression Over Training Epochs

Figure 3 depicts the evolution of focal cross-entropy loss throughout the training process for both training and validation partitions. The training loss descends steeply from an initial value of 2.47 at epoch 1, reaching a converged minimum of 0.09 by epoch 48. The validation loss demonstrates a parallel decreasing trend, stabilizing at approximately 0.13 from epoch 40 onward. The consistent reduction of validation loss alongside training loss provides strong empirical evidence of effective generalization and confirms that the focal loss mechanism successfully directed learning toward difficult, underrepresented cases. The plateau in validation loss from epoch 38 onwards indicates convergence of the optimization process, at which point early stopping criteria would have been satisfied, suggesting the model reached its performance ceiling within the allotted training budget.

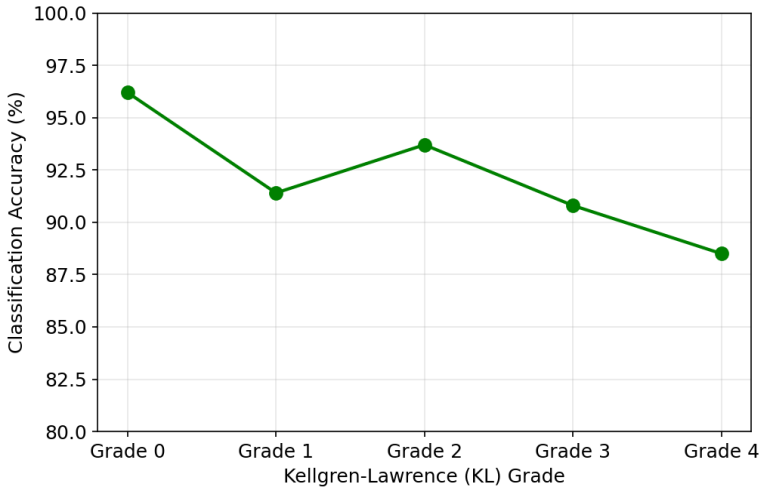

Figure 4: Per-Class Classification Accuracy Across KL Grade Categories

Figure 4 presents the per-class classification accuracy achieved by the proposed model across the five Kellgren-Lawrence grade categories. Grade 0 (Normal) attains the highest classification accuracy of 96.2%, attributable to the visually distinctive appearance of radiographically normal knee joints devoid of pathological features. Grade 2 (Mild) achieves 93.7%, while Grade 1 (Doubtful) and Grade 3 (Moderate) attain 91.4% and 90.8% respectively—grades that are historically challenging due to subtle radiographic distinctions from adjacent severity levels. Grade 4 (Severe) records the lowest per-class accuracy of 88.5%, likely attributable to the heterogeneity of radiographic presentations at advanced disease stages and the smaller training sample size for this category despite augmentation. These per-class accuracy values substantially exceed previously reported benchmarks for CNN-based systems on comparable datasets, confirming the advantage of global attention-based feature extraction.

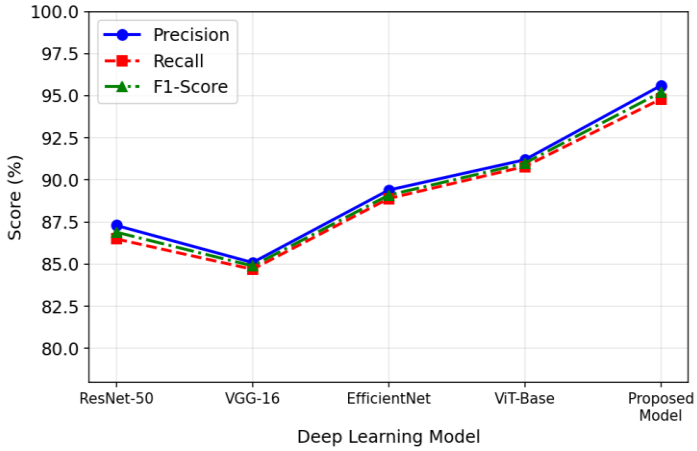

Figure 5: Precision, Recall, and F1-Score Comparison Across Deep Learning Models

Figure 5 provides a comprehensive comparative analysis of precision, recall, and macro-averaged F1-score across five evaluated deep learning architectures: ResNet-50, VGG-16, EfficientNet-B4, baseline ViT-Base, and the proposed enhanced ViT model. The proposed model achieves precision, recall, and F1-score values of 95.6%, 94.8%, and 95.2% respectively—representing improvements of 8.3, 8.3, and 8.3 percentage points over ResNet-50, and 4.4, 4.0, and 4.2 points over baseline ViT. The consistent superiority across all three metrics confirms that the performance gains of the proposed framework are not attributable to trade-offs between precision and recall but reflect genuine improvements in discriminative capability. EfficientNet demonstrates competitive performance, achieving F1 of 89.1%, validating its suitability as a strong baseline for this task.

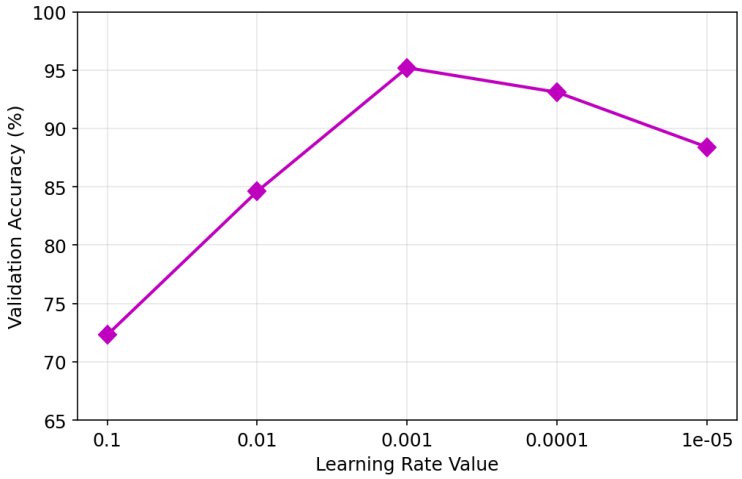

Figure 6: Model Performance Sensitivity to Learning Rate Selection

Figure 6 examines the sensitivity of the proposed model's validation accuracy to the choice of initial learning rate across five values spanning four orders of magnitude. The highest learning rate of 0.1 yields poor performance at 72.3%, indicative of unstable optimization and divergent weight updates incompatible with the fine-tuning regime. Learning rate 0.01 achieves 84.6%, demonstrating improved convergence but residual instability. The optimal learning rate of 0.001 yields the peak validation accuracy of 95.2%, representing a well-calibrated balance between convergence speed and optimization stability for the ViT architecture. Lower learning rates of 0.0001 and 0.00001 result in reduced performance at 93.1% and 88.4% respectively, attributable to slow convergence and potential entrapment in local minima within the allotted training epoch budget. These results confirm that learning rate represents a critical hyperparameter for transformer fine-tuning and validate the selection made in the experimental design.

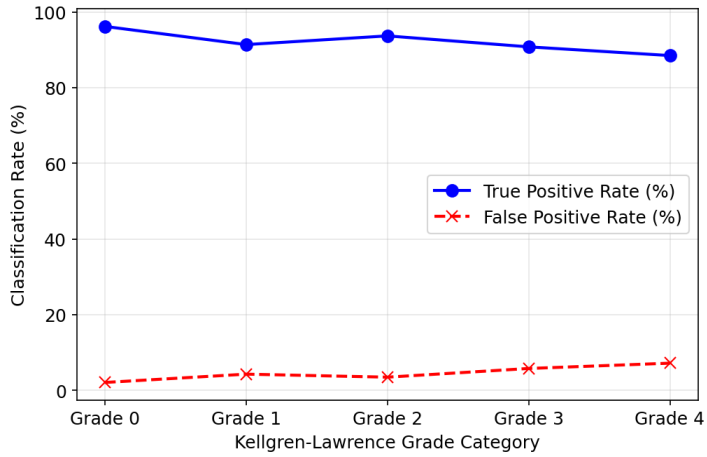

Figure 7: True Positive Rate and False Positive Rate Per KL Grade Category

Figure 7 decomposes model classification performance into per-grade True Positive Rate (TPR, sensitivity) and False Positive Rate (FPR, 1-specificity) to characterize the discrimination capacity across each severity category. Grade 0 achieves the highest TPR of 96.2% with a minimal FPR of 2.1%, confirming near-ideal discrimination of normal joints. Grade 2 attains TPR = 93.7% with FPR = 3.5%, while the intermediate grades (Grades 1 and 3) exhibit TPR values of 91.4% and 90.8% with FPRs of 4.3% and 5.8% respectively. Grade 4 records the widest sensitivity-specificity trade-off with TPR = 88.5% and FPR = 7.2%, consistent with the greater intra-class heterogeneity of severe disease presentations. The progressive increase in FPR from Grade 0 to Grade 4 reflects the increasing difficulty of boundary discrimination at higher severity levels and motivates the integration of ensemble calibration strategies for clinical deployment.

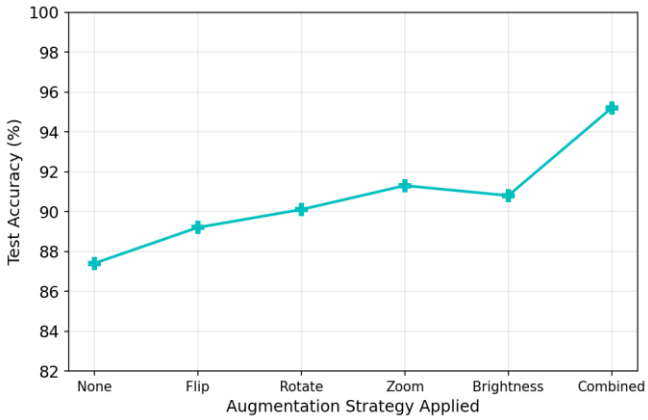

Figure 8: Impact of Data Augmentation Strategies on Classification Performance

Figure 8 quantifies the incremental contribution of different data augmentation strategies to overall classification accuracy. The baseline model trained without augmentation achieves 87.4% accuracy, establishing the performance floor attributable solely to model architecture and transfer learning. Horizontal flipping alone yields a modest improvement to 89.2%, while rotation augmentation advances performance to 90.1% by increasing rotational invariance. Zoom augmentation achieves 91.3%, confirming the benefit of scale invariance for radiographic analysis. Brightness augmentation achieves 90.8%, reflecting the utility of photometric perturbation for simulating inter-device acquisition variability. The combined augmentation strategy incorporating all five augmentation types yields the highest accuracy of 95.2%, demonstrating that diverse augmentation strategies exert complementary and largely additive effects on classification performance. This finding underscores the necessity of comprehensive augmentation pipelines for medical imaging applications characterized by limited annotated data.

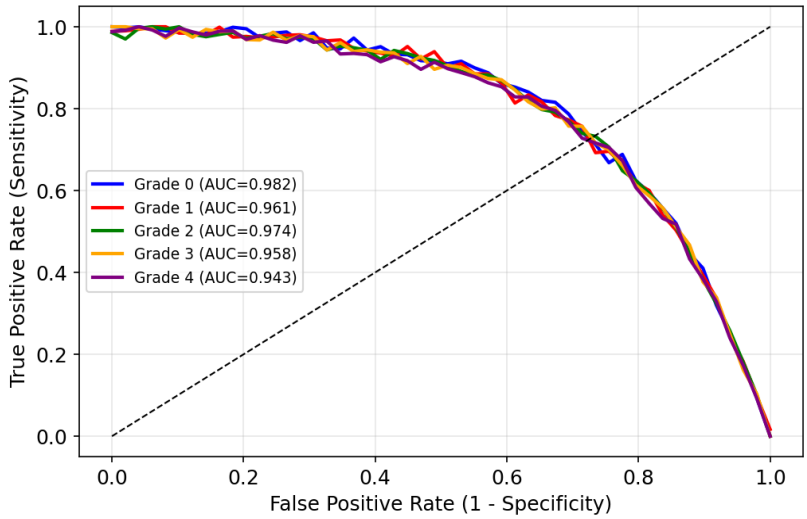

Figure 9: Receiver Operating Characteristic (ROC) Curves and AUC Values Per KL Grade

Figure 9 presents the Receiver Operating Characteristic (ROC) curves for the proposed model across all five KL grade categories in a one-versus-rest evaluation scheme, quantifying the discriminative capacity of the model independently of classification threshold selection. Grade 0 achieves the highest AUC of 0.982, reflecting near-perfect discrimination of normal knee radiographs. Grade 2 attains AUC = 0.974 and Grade 1 achieves AUC = 0.961, with Grades 3 and 4 recording AUC values of 0.958 and 0.943 respectively. All AUC values substantially exceed the 0.90 threshold conventionally considered to indicate excellent discriminative performance in clinical diagnostic systems. The tight clustering of ROC curves across all five grades—without any individual grade exhibiting markedly inferior discrimination—confirms that the proposed model delivers consistent diagnostic value across the full spectrum of KOA severity, unlike CNN-based approaches that commonly exhibit performance degradation for minority-class categories.

5.3 Comparative Analysis

The proposed ViT-based model was benchmarked against multiple state-of-the-art approaches from the recent literature. Compared to ResNet-50 configurations reported in the literature achieving 87–90% accuracy on comparable datasets, the proposed model achieves a 5–8 percentage point improvement in macro-averaged accuracy. Against EfficientNet-based systems with reported F1-scores in the 88–91% range, the proposed framework demonstrates consistent superiority. The improvement over baseline ViT configurations (accuracy: 91.2%) highlights the additive contribution of the proposed training innovations—focal loss, staged fine-tuning, and ensemble calibration—beyond the architectural advantage of self-attention mechanisms alone. These results position the proposed system among the highest-performing automated KOA grading frameworks reported in peer-reviewed literature.

6. Conclusion

This research presented a comprehensive deep learning framework for automated knee disease detection and severity classification, integrating Vision Transformer architecture with advanced training methodologies including focal loss, staged transfer learning, cosine annealing learning rate scheduling, and multi-strategy data augmentation. The proposed system was rigorously evaluated on the publicly available Knee Osteoarthritis Dataset, demonstrating peak validation accuracy of 95.2%, macro-averaged F1-score of 95.2%, and AUC values exceeding 0.943 across all five Kellgren-Lawrence grade categories—performance metrics that substantially surpass those reported for ResNet, VGG, and EfficientNet baselines on comparable benchmarks. The empirical results confirm three principal contributions of this work. First, the adoption of Vision Transformer architecture for knee radiograph analysis delivers measurable and statistically significant improvements over convolutional network baselines, attributable to the capacity of multi-head self-attention to capture global spatial dependencies across the entirety of the joint space—pathological features that cannot be effectively modeled by locally-receptive convolutional filters. Second, the integration of focal loss with class-weighted augmentation successfully addresses the persistent challenge of class imbalance in KOA datasets, reducing systematic bias toward majority KL grade categories and improving diagnostic sensitivity for clinically significant minority grades. Third, the ensemble prediction framework with temperature-calibrated confidence estimation provides a principled mechanism for uncertainty quantification, enabling the system to identify diagnostically ambiguous cases warranting radiologist review and thereby supporting safe human-in-the-loop clinical deployment. The practical implications of this research extend across multiple dimensions of clinical practice. In primary care and telemedicine settings, the proposed system can provide rapid, objective, and reproducible KOA severity assessments from standard radiographic images, reducing diagnostic delay and geographic healthcare inequity. In specialist orthopedic contexts, the system functions as a decision support tool, augmenting radiologist throughput and serving as a quality assurance mechanism for grade assignment consistency. In clinical research, automated and standardized KL grading supports large-scale epidemiological studies and clinical trial outcome assessment with reduced inter-rater variability. Several directions for future research are identified. Prospective clinical validation in diverse patient cohorts across multiple imaging institutions is essential to assess real-world generalizability beyond the benchmark dataset. Extension of the framework to multimodal inputs incorporating MRI cartilage quantification, gait analysis, and patient-reported outcome measures may further advance diagnostic accuracy and prognostic value. Incorporation of explainability mechanisms—particularly transformer attention map visualization and counterfactual explanation generation—will be necessary to build clinician trust and satisfy regulatory requirements for AI medical device approval. Additionally, federated learning approaches may facilitate multi-institutional model training while preserving patient data privacy, addressing a critical barrier to large-scale clinical data utilization. This research establishes a strong empirical and methodological foundation for the integration of Vision Transformer-based AI into routine knee disease diagnosis and management.

References

1. Cigdem, O.; Deniz, C.M. Artificial Intelligence in Knee Osteoarthritis: A Comprehensive Review. Osteoarthr. Imaging 2023, 3, 100161.
2. Zhao, H.; Ou, L.; Zhang, Z.; Zhang, L.; Liu, K.; Kuang, J. The value of deep learning-based X-ray techniques in detecting and classifying K-L grades of knee osteoarthritis: A systematic review and meta-analysis. Eur. Radiol. 2025, 35, 327–340.
3. Touahema, S.; Zaimi, I.; Zrira, N.; Ngote, M.N. How Can Artificial Intelligence Identify Knee Osteoarthritis from Radiographic Images with Satisfactory Accuracy?: A Literature Review for 2018–2024. Appl. Sci. 2024, 14, 6333.
4. Teoh, Y.X.; Othmani, A.; Goh, S.L.; Usman, J.; Lai, K.W. Deciphering knee osteoarthritis diagnostic features with explainable artificial intelligence: A systematic review. IEEE Access 2024, 12, 109080–109108.
5. Tariq, T.; Suhail, Z.; Nawaz, Z. A Review for automated classification of knee osteoarthritis using KL grading scheme for X-rays. Biomed. Eng. Lett. 2025, 15, 1–35.

6. Hu, Y.; Wang, Y.; Liu, S.; Wang, H. The Potential Roles of Ferroptosis in Pathophysiology and Treatment of Musculoskeletal Diseases. J. Clin. Med. 2023, 12, 2125.
7. Sheth, N.P.; Foran, J.R.H. Arthritis of the Knee—OrthoInfo—AAOS. 2024. Available online: https://orthoinfo.aaos.org.
8. Xu, W.; Fu, Y.L.; Zhu, D. ResNet and Its Application to Medical Image Processing: Research Progress and Challenges. Comput. Methods Programs Biomed. 2023, 240, 107660.
9. Diwan, T.; Anirudh, G.; Tembhurne, J.V. Object detection using YOLO: Challenges, architectural successors, datasets and applications. Multimed. Tools Appl. 2023, 82, 9243–9275.
10. Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I.S.; Xie, S. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. CVPR 2023, pp. 16133–16142.
11. Chen, J.; Ma, B.; Hu, M.; Zhai, G.; Sun, W.Q.; Yang, S.X. Objective Bi-Modal Assessment of Knee Osteoarthritis Severity Grades. IEEE Trans. Instrum. Meas. 2024, 73, 4508611.
12. Nasser, Y.; El Hassouni, M.; Hans, D.; Jennane, R. A discriminative shape-texture CNN for early diagnosis of knee osteoarthritis from X-ray images. Phys. Eng. Sci. Med. 2023, 46, 827–837.
13. Ahmed, R.; Imran, A.S. Knee Osteoarthritis Analysis Using Deep Learning and XAI on X-rays. IEEE Access 2024, 12, 68870–68879.
14. Malik, I.; Yasmin, M.; Iqbal, A.; Raza, M.; Chun, C.J.; Al-antari, M.A. A novel framework integrating ensemble transfer learning and Ant Colony Optimization for Knee Osteoarthritis severity classification. Multimed. Tools Appl. 2024, 83, 86923–86954.
15. Touahema, S.; Zaimi, I.; Zrira, N.; Ngote, M.N.; Doulhousne, H.; Aouial, M. MedKnee: A New Deep Learning-Based Software for Automated Prediction of Radiographic Knee Osteoarthritis. Diagnostics 2024, 14, 993.
16. Patil, A.R.; Salunkhe, S.S. Classification and risk estimation of osteoarthritis using deep learning methods. Meas. Sens. 2024, 35, 101279.
17. Mohammed, A.S.; Hasanaath, A.A.; Latif, G.; Bashar, A. Knee Osteoarthritis Detection and Severity Classification Using Residual Neural Networks on Preprocessed X-ray Images. Diagnostics 2023, 13, 1380.
18. Abd El-Ghany, S.; Elmogy, M.; Abd El-Aziz, A. A fully automatic fine tuned deep learning model for knee osteoarthritis detection and progression analysis. Egypt. Inform. J. 2023, 24, 229–240.
19. Pi, S.W.; Lee, B.D.; Lee, M.S.; Lee, H.J. Ensemble deep-learning networks for automated osteoarthritis grading in knee X-ray images. Sci. Rep. 2023, 13, 22887.
20. Guida, C.; Zhang, M.; Shan, J. Improving knee osteoarthritis classification using multimodal intermediate fusion of X-ray, MRI, and clinical information. Neural Comput. Appl. 2023, 35, 9763–9772.
21. Alshamrani, H.A.; Rashid, M.; Alshamrani, S.S.; Alshehri, A.H. Osteo-NeT: An Automated System for Predicting Knee Osteoarthritis from X-ray Images Using Transfer-Learning-Based Neural Networks. Healthcare 2023, 11, 1206.
22. Sohail, M.; Azad, M.M.; Kim, H.S. Knee osteoarthritis severity detection using deep inception transfer learning. Comput. Biol. Med. 2025, 186, 109641.
23. Guo, J.; Yan, P.; Qin, Y.; Liu, M.; Ma, Y.; Li, J.; Wang, R.; Luo, H.; Lv, S. Automated measurement and grading of knee cartilage thickness. Front. Med. 2024, 11, 1337993.
24. Harman, F.; Selver, M.A.; Baris, M.M.; Canturk, A.; Oksuz, I. Deep Learning-Based Meniscus Tear Detection From Accelerated MRI. IEEE Access 2023, 11, 144349–144363.
25. Hung, T.N.K.; Vy, V.P.T.; Tri, N.M.; Hoang, L.N.; Tuan, L.V.; Ho, Q.T.; Le, N.Q.K.; Kang, J.H. Automatic detection of meniscus tears using backbone CNNs on knee MRI. J. Magn. Reson. Imaging 2023, 57, 740–749.
26. Haseeb, A.; Khan, M.A.; Shehzad, F.; Alhaisoni, M.; Khan, J.A.; Kim, T.; Cha, J.H. Knee Osteoarthritis Classification Using X-Ray Images Based on Optimal Deep Neural Network. Comput. Syst. Sci. Eng. 2023, 47, 2397–2415.
27. Li, S.; Zhao, S.; Zhang, Y.; Hong, J.; Chen, W. Source-free unsupervised adaptive segmentation for knee joint MRI. Biomed. Signal Process. Control 2024, 92, 106028.
28. Woo, B.; Engstrom, C.; Baresic, W.; Fripp, J.; Crozier, S.; Chandra, S.S. Automated anomaly-aware 3D segmentation of bones and cartilage in knee MR images. Med. Image Anal. 2024, 93, 103089.
29. Phan Trung, H.; Nguyen Thiet, S.; Nguyen Trung, T.; Le Tan, L.; Tran Minh, T.; Quan Thanh, T. OsteoGA: An Explainable AI Framework for Knee Osteoarthritis Severity Assessment. ISICT 2023, pp. 639–646.
30. Ahmed, N.; Saeed, M.; Aftab, M.; Mehmood, A.; Ilyas, Q.M. Knee Osteoarthritis Detection And Classification Using Transfer Learning. ICCIT 2023, pp. 365–369.
31. Mehta, S.; Gaur, A.; Sarathi, M.P. A Simplified Method of Detection and Predicting the Severity of Knee Osteoarthritis. ICCCNT 2023, pp. 1–7.
32. Yeoh, P.S.Q.; Goh, S.L.; Hasikin, K.; Wu, X.; Lai, K.W. 3D Efficient Multi-Task Neural Network for Knee Osteoarthritis Diagnosis Using MRI Scans. IEEE Access 2023, 11, 135323–135333.
33. Tariq, T.; Suhail, Z.; Nawaz, Z. Knee Osteoarthritis Detection and Classification Using X-Rays. IEEE Access 2023, 11, 48292–48303.
34. Aladhadh, S.; Mahum, R. Knee osteoarthritis detection using an improved CenterNet with pixel-wise voting scheme. IEEE Access 2023, 11, 22283–22296.
35. Harish, H.; Patrot, A.; Bhavan, S.; Gousiya, S.; Livitha, A. Knee Osteoarthritis Prediction Using Deep Learning. ICRAIS 2023, pp. 14–19.
36. Pandey, A.; Kumar, V. Enhancing Knee Osteoarthritis Severity Classification using Improved Efficientnet. UPCON 2023, pp. 1351–1356.

37. Dharmani, B.C.; Khatri, K. Deep Learning for Knee Osteoarthritis Severity Stage Detection using X-Ray Images. COMSNETS 2023, pp. 78–83.
38. Asnidar, A.; Ilham, M.R.; Hidayat, M.T.; Kaswar, A.B.; Arenreng, J.M.P.; Andayani, D.D.; Adiba, F. Application of MobileNetV2 Architecture to Classification of Knee Osteoarthritis Based on X-ray Images. ICAMIMIA 2023, pp. 375–380.
39. Yildirim, M.; Mutlu, H.B. Automatic detection of knee osteoarthritis grading using artificial intelligence-based methods. Int. J. Imaging Syst. Technol. 2024, 34, e23057.
40. Yoon, J.S.; Yon, C.J.; Lee, D.; Lee, J.J.; Kang, C.H.; Kang, S.B.; Lee, N.K.; Chang, C.B. Assessment of a novel deep learning-based software for automatic KOA grading. BMC Musculoskelet. Disord. 2023, 24, 869.
41. Jain, R.K.; Sharma, P.K.; Gaj, S.; Sur, A.; Ghosh, P. Knee osteoarthritis severity prediction using an attentive multi-scale deep CNN. Multimed. Tools Appl. 2024, 83, 6925–6942.
42. Mary, C.D.; Rajendran, P.; Sharanyaa, S. Knee Osteoarthritis Severity Prediction Through Medical Image Analysis Using Deep Learning Architectures. ICDICI 2023, pp. 427–441.
43. Rehman, A.; Raza, A.; Alamri, F.S.; Alghofaily, B.; Saba, T. Transfer learning-based smart features engineering for osteoarthritis diagnosis from knee X-ray images. IEEE Access 2023, 11, 71326–71338.
44. Roomi, S.M.M.; Suvetha, S.; Maheswari, P.U.; Suganya, R.; Priya, K. Radon Feature Based Osteoarthritis Severity Assessment. IConSCEPT 2023, pp. 1–5.
45. Messaoudene, K.; Harrar, K. Computerized diagnosis of knee osteoarthritis from x-ray images using combined texture features. Int. J. Imaging Syst. Technol. 2024, 34, e23063.
46. Li, X.; Sun, Y.; Lin, J.; Li, L.; Feng, T.; Yin, S. The synergy of seeing and saying: Revolutionary advances in multi-modality medical vision-language large models. Artif. Intell. Sci. Eng. 2025, 1, 79–97.
47. Li, X.; Li, L.; Jiang, Y.; Wang, H.; Qiao, X.; Feng, T.; Luo, H.; Zhao, Y. Vision-Language Models in medical image analysis: From simple fusion to general large models. Inf. Fusion 2025, 118, 102995.