Journal of Science and Technology – Smart Systems and Devices
Volume 36, Issue 3, Month 2026, 000–000

Dual-stream ResNet-18 with Cross-Attention for Enhanced Knee Osteoarthritis Classification from X-ray Images

Pham Thi Viet Huong[1], Nguyen Minh Anh[2], Vuong Thi Hang[2], Tran Anh Vu[2*], Hoang Quang Huy[2]
1 International School, Vietnam National University, Ha Noi, Vietnam
2 Hanoi University of Science and Technology, Ha Noi, Vietnam
** Corresponding author email: vu.trananh@hust.edu.vn*

Abstract

*Knee osteoarthritis (KOA) is a prevalent musculoskeletal disorder that significantly affects mobility and quality of life. This study proposes a dual-stream ResNet-18 deep learning model integrated with a cross-attention mechanism to classify KOA severity according to the 5-level Kellgren–Lawrence (KL) grading scale from X-ray images. The dataset consists of 9,786 public knee X-ray images from the Osteoarthritis Initiative (OAI). The model processes both raw and CLAHE-enhanced images to capture structural features and high-contrast details. To address class imbalance, selective class-specific data augmentation and Weighted Focal Loss are employed. The architecture uses a shared ResNet-18 backbone combined with cross-attention and attention fusion modules, along with a multi-task learning strategy featuring a primary 5-class branch and auxiliary branches focusing on earlystage classification (KL 0–1–2). Experimental results show an overall accuracy of 72.06%, improving by 6.96% over the baseline. Notably, the model achieves a high sensitivity (Recall = 0.7843) for severe cases (KL-4), despite representing only 3.02% of the dataset, highlighting its potential for clinical decision support.*Keywords: include three to five different keywords, phrases in alphabetical order, separated by commas.

Keywords: Attention mechanism, CLAHE, deep learning, dual-stream model, knee osteoarthritis, X-ray images.

1. Introduction

Knee osteoarthritis (KOA) is a common chronic disease particularly affecting middle-aged and elderly individuals [1]. This condition occurs when the protective cartilage layer wears down, leading to direct bone contact, pain, swelling, restricted mobility, and sometimes osteophytes (bone spurs) [1]. KOA has emerged as a critical public health challenge, with a pooled global prevalence of 16.0% in individuals aged 15 and older [2], and approximately 14 million people suffering from symptomatic forms in the United States alone [3].

Various diagnostic methods have been applied in practice. X-ray imaging is the most common due to its low cost, rapid acquisition, and ability to provide crucial information about bone structure changes [4] [5]. However, it has limitations in detecting early-stage OA and lacks complete information about articular cartilage and soft tissues. In contrast, Magnetic Resonance Imaging and Computed Tomography offer detailed images of bone and soft tissue but come with higher costs and extended procedure times, making them less common in routine clinical practice [6].

In recent years, machine learning methods have been widely applied to diagnose KOA from X-ray images. Traditional algorithms like logistic regression, K-nearest neighbors, and support vector machines have been implemented for KOA classification, with KNN showing more effectiveness than other models. However, these methods are limited by their reliance on manual feature extraction and often do not achieve high accuracy [7] [8].

Deep learning models like Convolutional Neural Networks (CNN) automatically extract features and classify KOA severity. CNNs demonstrate significant potential in improving accuracy but face challenges such as requiring large amounts of data, long training times, and difficulties in generalization [9]. Recently, ResNet-18 has been widely used in medical computer vision due to its shallow structure, high computational efficiency, and strong feature extraction capabilities [10] [11], showing reliable performance and potential for supporting KOA diagnosis [12]. Initial studies applying ResNet-18 still face challenges like data shift, limited dataset size, and the need for data augmentation [10] [13], leaving a lack of in-depth research on fine-tuning it for smaller medical datasets common in clinical practice [14]. Other deep learning approaches include Siamese CNN models to detect progressive changes over time (focusing only on two main classes) [15], OsteoHRNet combined with Grad-CAM to identify diagnostic focus regions [16], and EfficientNet combined with CLAHE enhancement to improve classification in severe

p-ISSN 3093-3285
e-ISSN 3093-3315
https://doi.org/10.51316/jst.192.ssad.2026.36.3.x
Received: Feb 28, 2026; Revised: Apr 24, 2026;
Accepted: May 18, 2026; Online: Jun 21, 2026

classes [17]. However, existing research has not fully leveraged information from multiple transformed images, nor have they learned the correlation between different image representations.

This study focuses on applying a ResNet-18-based model for KOA diagnosis from X-ray images to classify pathological features with higher accuracy. This research contributes a novel dual-stream network architecture based on ResNet-18. The first branch extracts structural features from raw images, while the second branch focuses on high-contrast details through CLAHE-processed images. A Cross-Attention mechanism is embedded to synchronize information between the two streams, helping the model accurately detect the smallest pathological signs, such as osteophytes and joint space narrowing, on the standardized OAI dataset.

2. Methodology

2.1. Dataset Description

In this study, knee X-ray images were sourced from a public dataset for knee osteoarthritis severity classification [18], organized according to the Osteoarthritis Initiative (OAI). The 9,786 images (224 × 224 pixels) are divided into 5 expert-validated Kellgren–Lawrence (KL) severity levels (0 to 4), as shown in Fig. 1. The dataset exhibits significant imbalance: Healthy (40%), Doubtful (18%), Minimal (26%), Moderate (13%), and Severe (just over 3%). The KL-1 class is the most confusing due to subtle visual characteristics between KL-0 and KL-2. Conversely, the distinct KL-4 class accounts for only 3.02%, making feature learning challenging and necessitating a handling strategy at the loss function level rather than simple sample augmentation.

2.2. Data Preprocessing

Raw X-ray images often contain significant irrelevant noise. We apply a central region cropping algorithm based on pixel density to extract the knee joint space – the area containing key diagnostic signs. Subsequently, images are normalized to a size of 224 x 224 pixels.

To overcome uneven exposure and low contrast commonly found in X-ray images, the CLAHE (*Contrast Limited Adaptive Histogram Equalization*) technique is applied. CLAHE helps to highlight osteophytes and subchondral sclerosis, aiding the raw image branch in identifying subtle features – critical characteristics for distinguishing between KL-0 and KL-1.

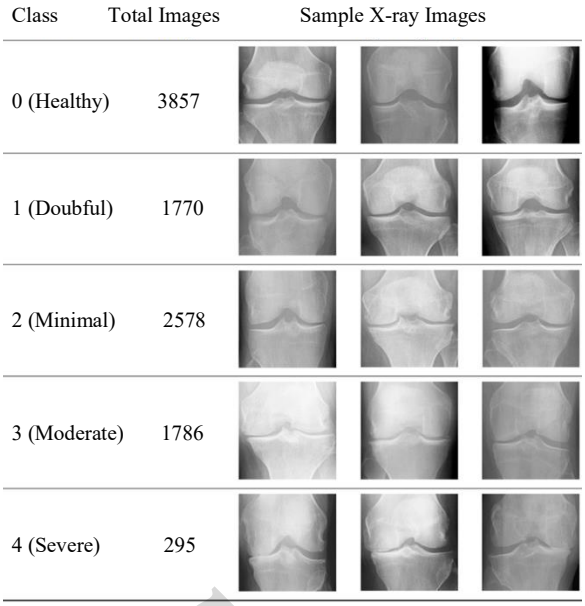

Fig. 1. Sample images for each class from the OAI dataset

The CLAHE-enhanced image therefore provides improved visibility of joint-space narrowing and osteophyte boundaries, serving as a complementary representation to the raw X-ray image in the dual-stream architecture. Fig. 2 and Fig. 3 present the original images and images after CLAHE.

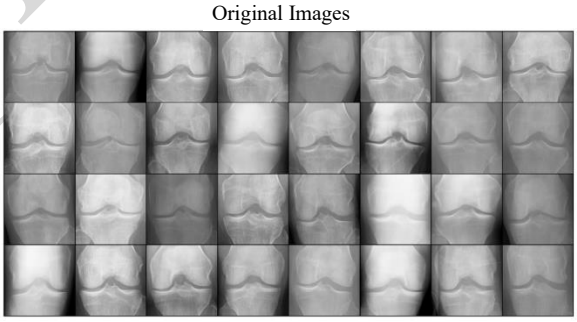

Fig. 2. Original knee osteoarthritis images

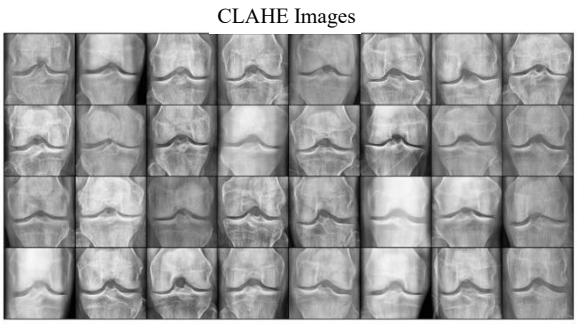

Fig. 3. Knee osteoarthritis images after CLAHE

For data spliting, we implemented a rigorous process of splitting the original dataset into three independent

subsets: Training, Validation, and Test. This procedure was carried out strictly to ensure objectivity and accurate evaluation of the model's generalization capability on new patient data. Data was loaded from pre-labeled categories according to 5 levels of degeneration from KL-0 to KL-4. The total number of samples after processing and experimentation was 8,260 images. The specific distribution of sample counts is presented in Table 1.

Table 1: Detailed statistics on the number of images after splitting

KL level	Train	Validation	Test	Total
KL-0 (Healthy)	2.286	328	639	3.253
KL-1 (Doubtful)	1.046	153	296	1.495
KL-2 (Minimal)	1.516	212	447	2.175
KL-3 (Moderate)	757	106	223	1.086
KL-4 (Severe)	173	27	51	251
Total	5.778 (~70%)	826 (~10%)	1.656 (~20%)	8.260 (100%)

2.3. Data Augmentation and Label Balancing

Due to significant data imbalance, particularly for doubtful and severe classes, we implemented a selective online data augmentation strategy via Albumentations and utilized Weighted Focal Loss to enhance training effectiveness without increasing storage size.

a. Selective data augmentation (Class-specific data augmentation)

Even with 9,786 physical images, at each epoch, the model encounters a new morphological variant, extending the decision boundary without causing overfitting.

- Class 1 (KL_1 Doubtful): Applied strong transformations like Shift Scale Rotate and Random Brightness Contrast (rotation ±10°, horizontal flip, geometric distortions) to help the model learn subtle features easily confused between KL-0 and KL-2 and handle varied X-ray conditions like subchondral sclerosis.
- Classes 3 and 4 (KL_3 Moderate and KL_4 Severe): Used GridDistortion and RandomGamma to simulate anatomical deformations of the joint space and bone surface/axis changes commonly seen in severe patients.
- Classes 0 and 2 (KL_0 Healthy and KL_2 Minimal): Applied basic techniques (Resize, RandomCrop, Normalize) to avoid introducing noise to stable classes.

b. Label balancing with weighted focal loss

To thoroughly address the imbalance without replicating data, we adjusted "mathematical weights" through the *Weighted Focal Loss* function (L_{FL}), impacting the gradient update process directly.

$$L_{FL} = -\alpha_i(1 - p_i)^{\gamma} log(p_i)$$

Here, p_i is the predicted probability of the model for the class i^{th}; α_i is the class balancing weight based on the frequency; γ is the focusing parameter, set to 2 to adjust the decay rate for easy samples.

The use of the Weighted Focal Loss function affects the classification of well-classified and hard-classified samples as follows:

- *Well-classified examples:* When the model correctly predicts a KL-0 sample with high probability ($p_i \to 1$), the factor $(1-p_i)^{\gamma}$ approaches 0, preventing the gradient from being overwhelmed by a large number of normal images.
- *Hard-classified examples:* For KL-1 or KL-3 casesthat are incorrectly predicted (low p_i), the factor $(1-p_i)^{\gamma}$will be close to. Combined with a high coefficient α, the model is "forced" to concentrate on updating weights to correct these cases.

The key hyperparameters in this study were selected through empirical evaluation to ensure a balance between model performance and training stability.

For the CLAHE preprocessing step, the clip limit parameter was tested with values β = 1.0, 2.0, and 3.0. A lower value (β = 1.0) resulted in insufficient contrast enhancement, making joint space boundaries less distinguishable. Conversely, higher values ($\beta \geq 3.0$) introduced noticeable noise amplification in homogeneous regions. Therefore, β = 2.0 was selected as it provided the best visual quality and contributed to more stable feature extraction.

For the Weighted Focal Loss, the focusing parameter γ was evaluated in the range from 1 to 5. Lower values (γ = 1) reduced the emphasis on hard-toclassify samples, while higher values ($\gamma \geq 3$) led to unstable training and reduced performance on majority classes. The choice γ = 2 achieved the best trade-off, improving the Macro F1-score by approximately 1%–2% compared to other configurations.

For the Dropout layer, probabilities in the range p = 0.2 to 0.5 were tested. A lower dropout rate (p = 0.2) resulted in mild overfitting, whereas higher values ($p \geq 0.5$) reduced model capacity and negatively impacted performance. The selected value p = 0.4

provided the most stable generalization performance on the validation set.

Table 2 demonstrates that adjusting the loss function using Weighted Focal Loss significantly shifts the mathematical gradient influence, effectively reducing the impact of easily diagnosable samples (KL-0 decreases from 39.41% to ~19.5%) while substantially forcing the model to concentrate on the severe minority class (KL-4 increases from 3.02% to 37.2%).

Table 2. Change in Effective Influence Share of Classes Before and After Weight Adjustment

KL level	Count	Percentage	The ratio α	Effective Gradient Influence Share
0 - Normal	3.857	39,41 %	0.133	~19.5%
1 - Doubtful	1.770	18,09 %	0.267	~18.0%
2 - Minimal	2.578	26,34 %	0.133	~12.8%
3 - Moderate	1.286	13,14 %	0.133	~6.5%
4 - Severe	295	3,02%	0.333	~37.2%

The combination of online augmentation and Weighted Focal Loss allows the model to focus on learning from severe and challenging cases, equivalent to having a perfectly balanced dataset in terms of quantity without losing the natural properties of the original data.

2.4. Proposed Model Architecture

Each knee X-ray image is processed through two parallel input branches: the raw image and the CLAHE-processed image. The original grayscale images are converted to 3-channel RGB by duplicating channel values to leverage spatial filters pre-trained on ImageNet without altering pixel intensity distributions. This Dual-stream architecture incorporates an inter-stream attention mechanism to process these complementary information sources simultaneously, as illustrated in the model architecture diagram (Fig. 4).

a. Dual-stream backbone (Shared-weight ResNet-18)

While the raw stream preserves natural anatomy, the CLAHE stream amplifies contrast at cartilage and soft tissue boundaries to highlight subtle osteophytes. Both streams utilize a shared ResNet-18 backbone, ensuring consistency in the extracted feature space and mitigating overfitting on medium-sized medical datasets by significantly reducing trainable parameters.

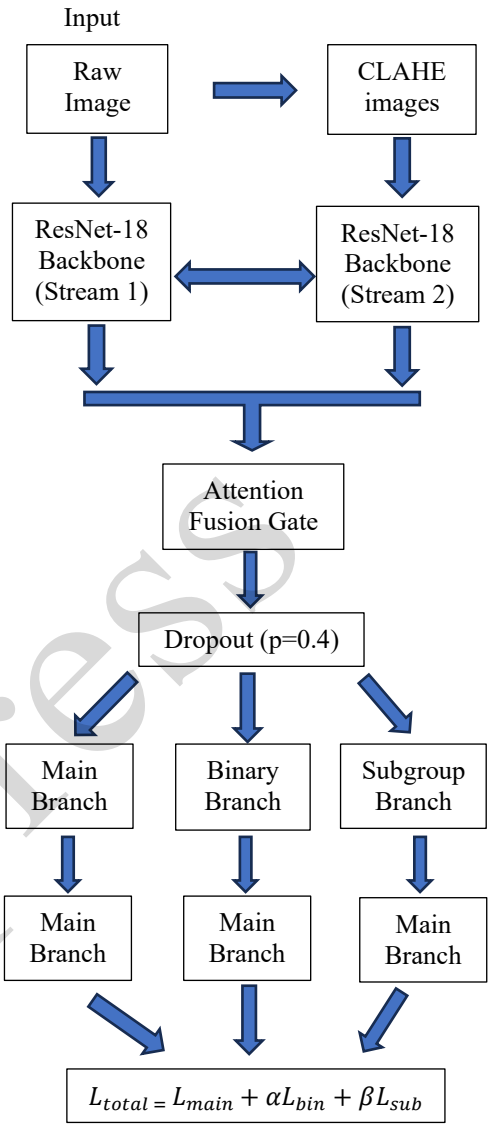

Fig. 4. Model architecture diagram

b. Cross-attention between streams

Feature streams from raw and CLAHE images pass through a Cross-Attention block to exchange information. The raw branch attends to high-contrast regions from the CLAHE branch to refine its representation, while the CLAHE branch references the structural framework from the raw branch to maintain anatomical authenticity, enhancing discriminability in early stages (KL-1, KL-2)

c. Attention fusion

An Attention Fusion block integrates a small MLP network to learn a Gating vector, generating adaptive weights: w_{raw}, w_{clahe} , in which $w_{raw} + w_{clahe} = 1$. The merged feature F_{fused} is calculated as:

$$F_{\text{fused}} = w_{raw} . F_{raw} + w_{clahe} . F_{clahe}$$

This gate automatically determines the contribution ratio, prioritizing the CLAHE branch for subtle boundaries and the raw branch for stable bone structure.

3. Results and Discussion

3.1. Classification Results

a. Model Performance with of single-stream raw input

Table 3 shows the evaluation results of the single-stream raw image model. Class 0 was best recognized (F1 = 0.7477, Recall = 0.8858) due to a larger number of samples and clear characteristics. Class 4, despite having fewer samples, maintained high sensitivity (Recall = 0.8039). However, Classes 1 and 2 had low F1-scores due to subtle features easily confused with neighboring classes , and Class 3 missed many moderate cases (low recall). Overall accuracy reached only ~65%, indicating that using raw images alone is not sufficiently robust. The confusion matrix (Fig. 5) confirms high confusion between Classes 1, 2, and 3.

Table 3. Performance evaluation results of single-stream raw image model

Class	Precision	Recall	F1-score	Support
0	0.6469	0.8858	0.7477	639
1	0.6805	0.5541	0.6108	296
2	0.6065	0.5034	0.5501	447
3	0.9111	0.3677	0.5240	223
4	0.5190	0.8039	0.6308	51
Accuracy			0.6510	1656
Macro Avg	0.6728	0.6230	0.6127	1656
Weighted Avg	0.6736	0.6510	0.6362	1656

b. Model performance with CLAHE-processed image input

The proposed dual-stream model achieved an overall accuracy of 72.06% on the test set. As shown in Table 4, Class 0 achieved the highest F1-score with a Recall of 0.8607 due to its clear image characteristics, demonstrating excellent ability to detect normal cases. Class 1 (F1 = 0.6679, Recall = 0.5912) and Class 2 (F1 = 0.6814, Recall = 0.6897) still faced challenges due to feature overlap and unclear early-stage characteristics. Class 3 remained difficult to identify accurately, with over 40% of samples misclassified as Class 2 or Class 4. Notably, Class 4 achieved a high Recall (0.7843), correctly classifying 40 out of 51 severe cases. This high sensitivity to severe characteristics (joint space narrowing and osteophytes) is crucial for screening patients requiring immediate surgery.

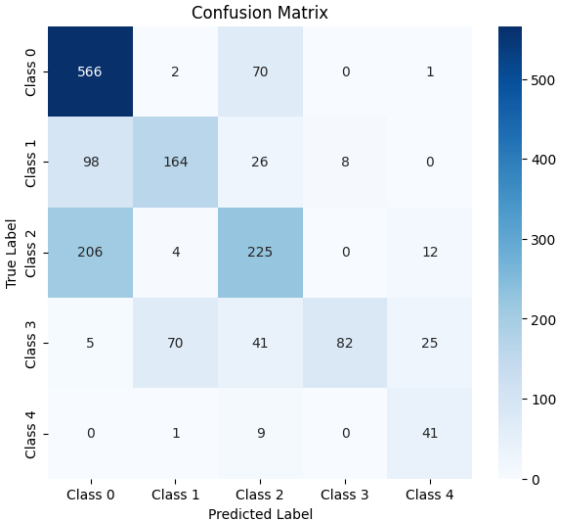

Fig. 5. Confusion matrix of the single-stream raw image model

Table 4. Performance Evaluation Results of Dual-stream Model

Class	Precision	Recall	F1-score	Support
0	0.7607	0.8607	0.8076	639
1	0.7675	0.5912	0.6679	296
2	0.6732	0.6897	0.6814	448
3	0.7643	0.5381	0.6316	223
4	0.4444	0.7843	0.5674	51
Accuracy			0.7206	1657
Macro Avg	0.6820	0.6928	0.6712	1657
Weighted Avg	0.7290	0.7206	0.7175	1657

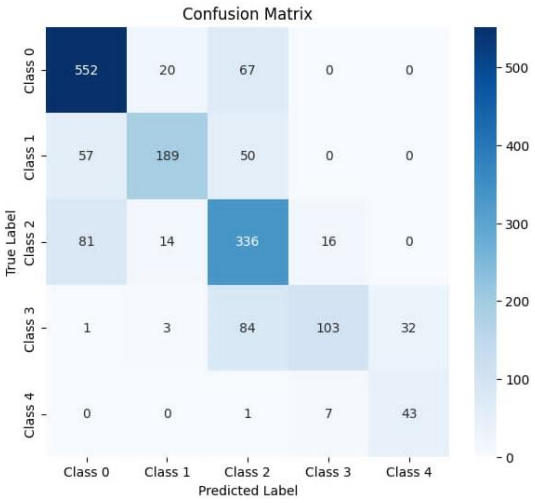

Fig. 6. Confusion matrix of the dual-stream model

The confusion matrix (Fig. 6) indicates the model classifies distinctive classes well but struggles with intermediate classes having ambiguous boundaries.

- Class 0: The majority of samples are correctly classified with minor confusion with Classes 1 and 2,

demonstrating high sensitivity in identifying non-degenerative cases.

- Class 1: Frequently confused with Classes 0 and 2 because its indistinct characteristics easily overlap with milder or more severe levels.

- Class 2: Tends to be confused with Classes 0 and 3, reflecting the difficulty of establishing precise boundaries for this intermediate stage.

- Class 3 The most frequently confused class, with over 40% of samples misclassified as Class 2 or 4, showing limitations in accurately identifying this intermediate phase.

- Class 4: Despite having only 51 samples, the model correctly classified 40 cases (misclassifying only 11), proving highly sensitive to severe features like prominent joint space narrowing and obvious osteophytes.

Experimental results show that KL-3 class has low Recall. Through confusion matrix analysis, we observed a large morphological overlap between KL-2 and KL-3. The inconsistency in expert labeling within the original OAI dataset for these boundary cases also contributed to the model's error. However, maintaining high Recall in the KL-4 class indicates that the model still ensures safety by not overlooking cases requiring immediate surgical intervention.

To investigate the performance bottleneck in intermediate stages, Grad-CAM was utilized (Fig. 7). For misclassified boundary cases, the model's output probabilities were nearly uniform, indicating high uncertainty. Heatmaps revealed a significant 'feature conflict' within the same view: for Top-1 (Class 0), attention localized on the central joint space, whereas for Top-5 (Class 3), attention shifted toward peripheral margins to identify subtle osteophytes. This dispersion confirms that morphological overlap and anatomical complexity are the primary causes of misclassification rather than image quality.

3.2. Ablation Study

To evaluate the contribution of each component, an ablation study was conducted with four configurations (Table 5):

- The Raw-only model outperformed the CLAHE-only model by 3.26%, indicating that original structural information is critical, while CLAHE alone may introduce noise.

- The dual-stream configuration improved performance significantly (+5.37% over Raw-only), confirming that Raw and CLAHE representations are complementary.

- The addition of the cross-attention mechanism further improved accuracy by +1.59%, demonstrating its effectiveness in adaptively fusing features.

Overall, these results prove that both CLAHE preprocessing and the dual-stream cross-attention design play essential, complementary roles

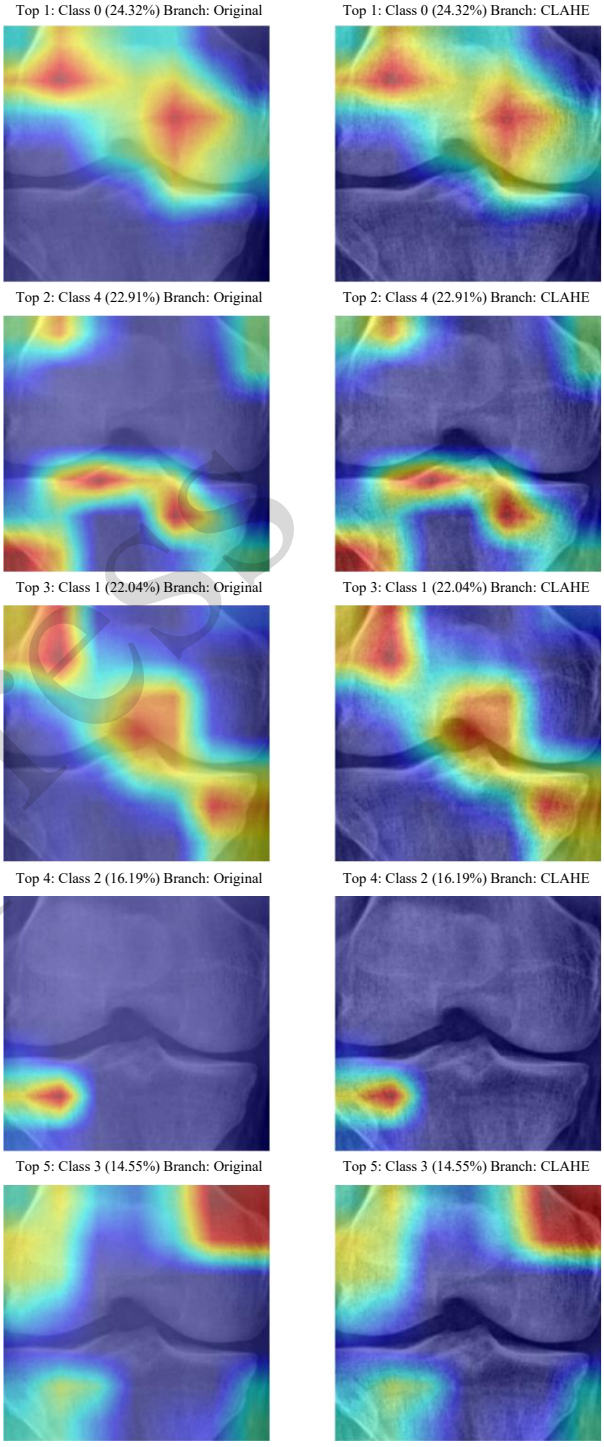

Fig. 7. Grad-CAM visualization

Table 5. Performance Comparison of Different Model Configurations (Ablation Study)

Configuration	Test Accuracy	Improvement
Single-stream (CLAHE only)	61.84%	-
Single-stream (Raw only)	65.10%	+3.26%
Dual-stream (Raw + CLAHE, No Attention)	70.47%	+8.63%
Proposed Model (Dual-stream + Attention)	72.06%	+10.22%

3.3. Comparison with Existing Methods

As presented in Table 7, the proposed model achieved the highest classification accuracy (72.06%), outperforming strong baselines like VGG-19 (69.7%), ResNet-101 (69.0%), InceptionV3 (67.0%), and DenseNet-121 (71.0%). Although its F1-score (0.67) is slightly lower than DenseNet-121 (0.70) due to class imbalance, the higher accuracy suggests more consistent overall predictions.

Table 7. Comparison with State-of-the-Art Methods

Model Architecture	Accuracy (%)	F1-score
VGG-19 + Ordinal Loss [19]	69.7	0.68
ResNet-101 [20]	69.0	0.65
InceptionV3 [21]	67.0	0.66
DenseNet-121 [22]	71.0	0.70
Proposed (Dual + Cross-Attention)	72.06	0.67

Compared to existing approaches, the proposed method introduces key methodological advances: a dual-stream architecture for complementary feature extraction , a cross-attention mechanism for adaptive stream interaction , and a multi-task learning framework to enhance generalization. Furthermore, while competing methods employ deeper, computationally expensive networks, our model utilizes a lightweight ResNet-18 backbone. This achieves a highly favorable trade-off between accuracy and computational efficiency, which is crucial for practical and clinical deployment

4. Conclusion

This study presents a dual-stream deep learning architecture for automatic classification of knee osteoarthritis severity from X-ray images. The proposed framework integrates raw radiographs and CLAHE-enhanced images through a shared ResNet-18 backbone combined with a Cross-Attention mechanism and an Attention Fusion block. This design enables the model to learn complementary structural and contrast-enhanced features, improving the representation of subtle morphological variations in early osteoarthritis stages.

To further enhance discrimination among clinically ambiguous classes, a multi-head learning strategy is employed, consisting of a primary 5-class classification branch and two auxiliary branches focusing on early-stage KL levels (0–1–2). The auxiliary binary and subgroup tasks are optimized using a label-masking mechanism during training, allowing the network to concentrate on subtle decision boundaries without interfering with higher-severity classes. A Dropout layer with $p = 0.4$is applied after feature fusion to improve generalization and prevent overfitting.

Experimental results demonstrate that the proposed dual-stream architecture significantly outperforms the single-stream baseline. The overall classification accuracy increases from 65.10% to 72.06%, representing a 6.96% improvement. The Weighted F1-score also improves from 0.6127 to 0.6712, indicating more balanced classification performance across severity levels. Notably, F1-scores for KL-0, KL-1, KL-2, and KL-3 show consistent improvement, with the most substantial gain observed in KL-2. Although KL-4 exhibits a slight decrease in F1-score, its recall remains high, suggesting that the model maintains strong sensitivity to severe osteoarthritis cases. Additionally, confusion between adjacent KL grades is reduced due to the complementary feature learning enabled by the dual- stream architecture and auxiliary supervision.

Overall, the results confirm that combining contrast enhancement, attention-based feature interaction, and multi-task learning provides a robust and effective framework for improving automated knee osteoarthritis severity classification from X-ray images.

References

[1] Y. Zhang and J. M. Jordan, Epidemiology of osteoarthritis, Clin Geriatr Med., vol. 26, iss. 3, pp. 355-369, Aug. 2010. https://doi.org/10.1016/j.cger.2010.03.001

[2] G. Cui, H. Li, D. Wang, J. Zhong, Y. Chen, and H. Lu, Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies, EclinicalMedicine., Nov. 2020, Art. no. 100587. https://doi.org/10.1016/j.eclinm.2020.100587

[3] W Novelyn and R. V. Munthe, Relationship between body mass index and knee osteoarthritis at UKI general hospital in 2017, Journal of Drug Delivery

and Therapeutics, vol. 12, iss. U-S, p. 154, Aug. 2022. https://doi.org/10.22270/jddt.v12i4-S.5575

[4] J. Khader, A. Zyout, and A. A. Fahoum, Combining enhanced spectral resolution of EMG and a deep learning approach for knee pathology diagnosis, PLoS One., vol. 19, iss. 5, May. 2024, Art. no. e0302707. https://doi.org/10.1371/journal.pone.0302707

[5] J. Lopes, A. Vilaça, C. Rocha, and J. Mendes, Knee positioning systems for X-ray environment: a literature review, Physical and Engineering Sciences in Medicine., vol. 46, iss. 1. p. 45, Jan. 2023. https://doi.org/10.1007/s13246-023-01221-y

[6] E. J. Lu, F. U. Hassan, S. Vijayanathan, and G. Gnanasegaran, Radionuclide bone SPECT/CT in the evaluation of knee pain: comparing two-phase bone scintigraphy, SPECT and SPECT/CT. Br J Radiol., Oct. 2018. https://doi.org/10.1259/bjr.20180168

[7] E. Cui, R. Liu, Y. Jing, J. Fu, and Jiying Chen, Development of machine learning models aiming at knee osteoarthritis diagnosing: an MRI radiomics analysis, Journal of Orthopaedic Surgery and Research., vol. 18, iss. 1, May.2023, Art. no.375. https://doi.org/10.1186/s13018-023-03837-y

[8] J. Mahum et.al. "A Novel Hybrid Approach Based on Deep CNN Features to Detect Knee Osteoarthritis." Sensors. 2021; 21(18):6189. https://doi.org/10.3390/s21186189

[9] E. Olsson, E. Akbarian, A. Lind, A. S. Razavian, and M. Gordon, Automating classification of osteoarthritis according to kellgren-lawrence in the knee using deep learning in an unfiltered adult population, BMC Musculoskeletal Disorders., Oct. 2021. https://doi.org/10.1186/s12891-021-04722-7

[10] J. Niu, J. Pan, Z. Qin, F. Huang, and H. Qin, Small-sample bearings fault diagnosis based on ResNet18 with pre-trained and fine-tuned method, Applied Science., vol. 14, iss. 12, Jun. 2024, Art. no. 5360. https://doi.org/10.3390/app14125360

[11] S. Ayyachamy, V. Alex, M. Khened, and G. Krishnamurthi, "Medical image retrieval using Resnet-18, vol. 10954, pp. 233–241, Mar. 2019, https://doi.org/10.1117/12.2515588

[12] A. S. Mohammed, A. A. Hasanaath, G. Latif, and A. Bashar, Knee osteoarthritis detection and severity classification using tesidual neural networks on nreprocessed X-ray images, Diagnostics, vol. 13, iss. 8, p. 1380, Apr. 2023. https://doi.org/10.3390/diagnostics13081380

[13] P. S. Q. Yeoh, K. W. Lai, S. L. Goh, K. Hasikin, X. Wu, and P. Li, Transfer learning- assisted 3D deep learning models for knee osteoarthritis detection: Data from the osteoarthritis initiative, Frontiers in Bioengineering and Biotechnology., vol. 11, Apr. 2023, Art. no. 1164655. https://doi.org/10.3389/fbioe.2023.1164655

[14] S. B. Francis and J. P. Verma, Deep CNN ResNet-18 based model with attention and transfer learning for Alzheimer's disease detection, Frontiers in Neuroinformatics, vol. 18, Jan. 2024, Art. no. 1507217, https://doi.org/10.3389/fninf.2024.1507217

[15] M. S. Islam and M. A. T. Rony, CDK: A novel high-performance transfer feature technique for early detection of osteoarthritis, Journal of Pathology Informatics., vol. 15, Dec. 2024, Art. no. 100382. https://doi.org/10.1016/j.jpi.2024.100382

[16] R. K. Jain, P. K. Sharma, S. Gaj, A. Sur, and P. Ghosh, Knee osteoarthritis severity prediction using an attentive multi-scale deep convolutional neural network, Multimedia Tools and Applications., vol. 83, iss. 3, pp. 6925–6942, Jun. 2023, https://doi.org/10.1007/s11042-023-15484-w

[17] P. N. Chowdary, G. V. N. S. L. V. Vardhan, M. S. Akshay, M. S. Aashish, V. S. Aravind, G. V. K. Rayalu, and P. Aswathy, Enhancing knee osteoarthritis severity level classification using diffusion augmented images, pp. 266–274, Sep. 2023. https://doi.org/10.2991/978-94-6463-314-6_27

[18] P. Chen, Knee osteoarthritis severity grading dataset, Sep. 2018. https://doi.org/10.17632/56rmx5bjcr.1

[19] P. Chen, L. Gao, X. Shi, K. Allen, and L. Yang, Fully Automatic Knee Osteoarthritis Severity Grading Using Deep Neural Networks with a Novel Ordinal Loss, Computerized Medical Imaging and Graphics., vol. 75, p. 84, Jul. 2019, https://doi.org/10.1016/j.compmedimag.2019.06.002

[20] K. A. Thomas, Ł. Kidziński, E. Halilaj, S. Fleming, G. R. Venkataraman, E. H. G. Oei, G. E. Gold, and S. L. Delp, Automated classification of radiographic knee osteoarthritis severity using deep neural networks, Radiol. Artif. Intell., vol. 2, iss. 2, Mar. 2020, Art. No. e190065. https://doi.org/10.1148/ryai.2020190065

[21] A. S. Mohammed, A. A. Hasanaath, G. Latif, and A. Bashar, Knee Osteoarthritis Detection and Severity Classification Using Residual Neural Networks on Preprocessed X-ray Images, Diagnostics 2023, vol. 13, iss. 8, Page 1380, Apr. 2023. https://doi.org/10.3390/diagnostics13081380

[22] S. Kinger, Deep learning for automatic knee osteoarthritis severity grading and classification, Indian Journal of Orthopaedics 2024, vol. 58, iss. 1, pp. 1458–1473, Sep. 2024. https://doi.org/10.1007/s43465-024-01259-4