A GRAPH-DRIVEN APPROACH TO KNEE OSTEOARTHRITIS SEVERITY CLASSIFICATION

Marouane Tliba[1,3], *Yassine Nasser*[1], *Mohamed Amine Kerkouri*[2], *Aladine Chetoauni*[3], *Rachid Jennane*[1]

[1]University of Orleans, Orleans, France
[2]f-initiatives, Paris, France
[3]Université Sorbonne Paris Nord, Villetaneuse, France

ABSTRACT

Knee Osteoarthritis (KOA) is a prevalent degenerative joint disease, with diagnosis and management predominantly reliant on radiographic analysis, particularly the Kellgren–Lawrence (KL) scale. Current deep learning methodologies for the automatic diagnosis of KOA primarily utilize raw X-ray images, which are often compromised by noisy texture artifacts. These artifacts can obscure critical shape cues that clinicians prioritize to determine the KL grading system. In this study, we present a novel morphological approach that emphasizes the geometric structure of the knee joint, thereby simplifying the model's inductive bias and directing training towards meaningful bone shapes. Our method employs the segmentation of input images to automatically extract precise anatomical landmarks on the femur and tibia, utilizing the Segment Anything Model (SAM). These landmarks are represented as nodes within a graph, with edges defined by geometric relationships, such as distances, thereby encoding both local and global structural information of the joint. An EdgeConv-based Graph Neural Network (GNN) classifier is subsequently employed to process this graph representation, effectively capturing inter- and intra-bone relationships to predict the KL grades of KOA severity. By adopting a graph-centric framework, our approach exhibits inherent invariance to translation, rotation, and scaling, while robustly representing subtle morphological changes that are critical for accurate diagnosis. To the best of our knowledge, this represents the first deep learning method that leverages morphological features for KOA assessment. Validation experiments conducted across various KL grade configurations (including all classes, 0–1, and 0–2) consistently demonstrate state-of-the-art classification performance.

Index Terms— Graph Neural Networks, Biomedical Image Analysis, Depp Learning, Knee Osteoarthritis

1. INTRODUCTION

Knee osteoarthritis (KOA) is a degenerative disease characterized by cartilage erosion, osteophyte formation, and subchondral bone remodeling, often assessed via the Kellgren–Lawrence (KL) grading system [1, 2]. While deep learning (DL) has shown promise in automating KOA severity classification [3, 4, 5, 6], prevailing methods predominantly analyze raw pixel intensities or texture patterns in X-rays and often misaligned with the anatomical structures central to KL grading. In [3], Nasser et al. introduced a discriminative shape-texture CNN block to capture texture features alongside shape descriptors, while Nguyen et al.[7] employed a semi-supervised framework using mixup to counter distributional gaps in limited training data. In [8, 4, 9], the authors proposed vision transformer-based methods to enhance KOA classification performance. Despite such advances, these methods often emphasize local textures (e.g., trabecular patterns) or global intensities without explicitly modeling the underlying bone morphology, risking an overfocus on incidental features rather than clinically validated anatomical structures.

In essence, although contemporary CNNs and Transformers demonstrate proficiency in detecting local intensity gradients and fine-grained textural details[10, 11], they frequently overlook crucial macro-level geometric relationships that form the cornerstone of clinical KOA assessment. These critical morphological indicators; including femoral–tibial alignment, progressive joint space narrowing, and the precise configuration and extent of osteophyte formation-represent the primary visual markers that experienced radiologists prioritize when performing reliable KOA staging. This disconnect between algorithmic focus and clinical practice creates a fundamental limitation: purely pixel-driven representations may effectively capture incidental textural artifacts and localized intensity variations, but they inherently lack the structural framework necessary to interpret and quantify the global shape transformations that characterize morphological disease progression.

To bridge this gap, we propose a geometry-centric framework that explicitly models bone morphology through graph neural networks (*GNN*s). Our approach mirrors the radiologist's workflow by first segmenting femoral and tibial contours using the Segment Anything Model (SAM) [12], then constructing a graph where nodes represent anatomical landmarks and edges encode geometric relationships (e.g., distances). An EdgeConv-based classifier [13] processes this graph to capture both local (e.g., osteophyte curvature) and global (e.g., femoral-tibial alignment) structural cues, directly aligning with

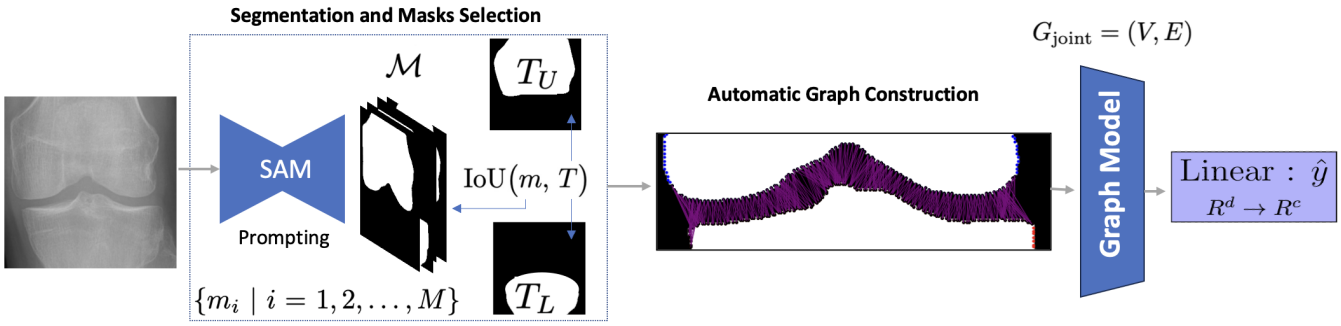

Fig. 1: Overview of our pipeline: First, SAM is prompted to generate candidate masks $\mathcal{M}$. The best mask m^* is chosen based on IoU with upper and lower bone templates (T_U, T_L). Next, a morphological graph G_{joint} is constructed from the joint boundary and processed by the proposed graph model $f_{\text{graph}}(\cdot; \Theta)$. Finally, a linear layer produces the KOA severity logits $\hat{y}$.

KL grading clinical criteria. By design, our method is invariant to translation, rotation, and scaling—critical for handling real-world X-ray variability—while avoiding texture-driven biases that hinder existing models.

Our main contributions are presented as what follows:

- **Automatic Segmentation & Graph Construction.** We employ the Segment Anything Model (SAM) to isolate precise bone contours and sample key landmarks, forming a morphological graph that explicitly encodes shape and spatial relationships crucial for KOA grading.
- **Morphology-Centric Network.** An EdgeConv-based architecture that captures both local and global structural cues, avoiding the texture biases common in radiograph-only classifiers.
- **Clinically Aligned Representation.** By shifting attention to geometry-driven features, our approach more faithfully reflects radiological biomarkers—improving interpretability and enabling a robust KOA severity classification that resonates with clinical insights.

2. METHODOLOGY

We introduce a graph-based strategy that explicitly encodes the joint's anatomical structure to address the texture-centric bias common in radiographic KOA classification. Figure 1 depicts our overall pipeline, which first derives high-fidelity joint masks via SAM, then constructs a morphological graph capturing the femoral–tibial geometry, and finally applies an EdgeConv-based network to classify KOA severity.

2.1. Segmentation and Landmark Extraction

Automatic Mask Generation. Given a radiograph $X \in \mathbb{R}^{H \times W \times C}$, we prompt the Segment Anything Model (SAM) [12] with dense grids of points, producing a set of candidate masks

$$\mathcal{M} = \{\, m_i \mid i = 1, \ldots, M \}.$$

Each candidate mask is then evaluated against predefined bone templates (T_U, T_L) by computing intersection-over-union (IoU). The mask m^* with the highest IoU is assumed to be the most accurate delineation of the joint area.

Graph Construction. From the boundary of m^*, we uniformly sample N landmarks $\{p_i\}_{i=1}^{N}$, where $p_i = (x_i, y_i) \in \mathbb{R}^2$, ensuring consistent spacing along the bone contours. These points form the vertex set $V = \{p_i\}$ in an undirected graph:

$$G_{\text{joint}} = (V, E).$$

To define E, we apply a k-nearest neighbor search under a distance threshold τ. More concretely, for each vertex $p_i \in V$,

$$\mathcal{N}(p_i) = \{p_j \mid \|p_i - p_j\|_2 \le \tau\} \cup \kappa_k(p_i),$$

where $\kappa_k(\cdot)$ returns the k-nearest neighbors. We then set

$$E = \big\{(p_i,\, p_j) \mid p_j \in \mathcal{N}(p_i)\big\}.$$

If G_{joint} is initially disconnected, τ is iteratively increased until the graph spans the joint region. This morphological representation prioritizes macro-level bone geometry, avoiding the subtle textural noise seen in raw radiographs.

2.2. Graph Model Classification

Morphological Classification via EdgeConv. Let $\mathbf{X}^{(\ell)} \in \mathbb{R}^{|V| \times d_\ell}$ represent node features at layer ℓ. At each layer, we define *edge features* through an asymmetric mapping

$$\mathbf{e}_{ij} \;=\; \phi_\Theta\big(\mathbf{x}_i,\; \mathbf{x}_i - \mathbf{x}_j\big),$$

where $\mathbf{x}_j$ is a neighbor of the central node $\mathbf{x}_i$, and $\phi_\Theta : \mathbb{R}^{d_\ell} \times \mathbb{R}^{d_\ell} \to \mathbb{R}^{d_{\ell+1}}$ is a nonlinear embedding function (e.g., an MLP) with learnable parameters Θ, as depicted in Fig.2. This adopted *GNN* convolutional operation, refereed as *EdgeConv* [13], has been widely adopted in *GNN*-based shape classification, particularly in 3D point-cloud tasks due to its strong capacity for modeling local geometric structure.

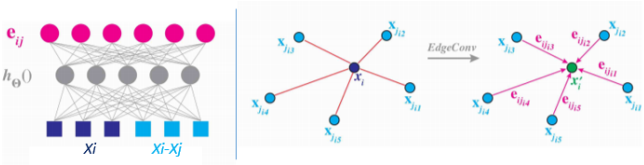

Fig. 2: **Left**: Computing an edge feature e_{ij}, **Right**: The EdgeConv operation [13].

Concretely, for each node i with neighbor set $\mathcal{N}(i)$, the EdgeConv layer produces updated features by aggregating local subgraphs:

$$\mathbf{x}_i^{(\ell+1)} = \max_{j \in \mathcal{N}(i)} \phi_\Theta\Big([\mathbf{x}_i^{(\ell)},\ \mathbf{x}_j^{(\ell)} - \mathbf{x}_i^{(\ell)}] \Big),$$

where $[\cdot,\cdot]$ denotes feature concatenation. Stacking multiple EdgeConv layers (three in our case) systematically increases the dimensionality of ϕ_Θ and captures deeper relationships among bone landmarks, essential for distinguishing subtle KOA-associated morphology.

Graph Normalization. After each EdgeConv layer, we employ GraphNorm [14] to maintain feature stability. For each node's updated feature $\mathbf{x}_i^{(\ell+1)}$,

$$\mathbf{g}_i^{(\ell+1)} = \gamma \, \frac{\mathbf{x}_i^{(\ell+1)} - \mu\big(\mathbf{x}^{(\ell+1)}\big)}{\sigma\big(\mathbf{x}^{(\ell+1)}\big) + \epsilon} + \beta,$$

where μ and σ respectively compute mean and standard deviation over node features, and γ, β are trainable parameters. This normalization encourages consistency within each graph and aids convergence.

Output Layer. We aggregate each node's feature via global mean and max pooling to form a single morphology-rich vector $\mathbf{z}$, which a linear layer maps into KOA severity logits: $\hat{\mathbf{y}} \in \mathbb{R}^C$, where C is the number of discrete KOA grades. Training proceeds with a standard cross-entropy objective, allowing the Graph Model to learn shape-relevant embeddings that better align with clinical perceptions of osteoarthritis.

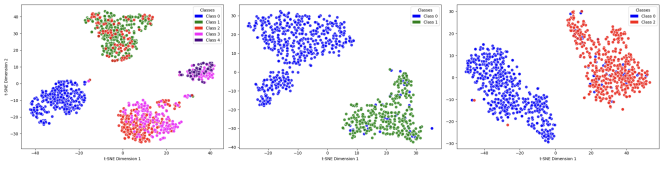

Fig. 3: T-SNE visualizations of learned embeddings for three KOA classification tasks, from left to right: full multi-class (KLG 0–4), binary classification KLG 0–1, and binary classification KLG 0–2.

3. RESULTS ANALYSIS

We evaluated our proposed *GNN*-based approach across three distinct KOA classification setups: (1) binary KLG-0 vs. KLG-1, (2) binary KLG-0 vs. KLG-2, and (3) the full multi-class scenario with KLG-0–4 on the OAI Dataset[15]. For each task we train the model and validate it based on the corresponding classes labels, we adopt the same Train-Test split as other methods in the literature.

3.1. General Observations

Table 3 presents a summary of state-of-the-art (SOTA) approaches for each classification task, indicating both accuracy and F1-Score. Our graph-only solution surpasses all listed baselines in each task, emphasizing the utility of explicit bone morphology for KOA grading. This advantage remains consistent even when moving from simpler binary tasks (KLG-0 vs. KLG-1 or KLG-0 vs. KLG-2) to the more challenging multi-class (0–4) classification. Notably, a purely radiograph-based model often struggles with subtle anatomical changes, while our graph-driven design highlights clinically relevant cues such as femoral–tibial alignment and osteophyte formation.

3.2. Binary Classification Performance

KLG-0 vs. KLG-1. Table 2 (KLG-0 vs. KLG-1) reveals that existing methods typically attain accuracies around 70–75% [4, 16]. In contrast, our model achieves 85.66% accuracy and 86.15% F1-Score, underscoring the model's ability to handle even mild or doubtful KOA stages through precise bone-structure embeddings.

KLG-0 vs. KLG-2. For the second binary classification setting, Table 1 (KLG-0 vs. KLG-2) shows prior works [8, 17, 18, 19, 20, 16] remain below 90%. Our model keep competitive with other methods, achieving 88.12% accuracy and 88.20% F1-Score—outperforming the best methods in this later one, despite the complexity inherent in detecting moderate KOA changes.

3.3. Multi-Class (KLG-0–4) Classification

Finally, Table 3 (All Classes 0–4) compares our approach to the most prominent multi-class KOA classifiers [20, 21, 22, 23, 4, 9]. The graph-based method achieves 74.94% accuracy and 75.19% F1-Score, outperforming previous SOTA results by a margin of roughly 2–5%. In this more challenging scenario, capturing global morphological changes proves instrumental in distinguishing borderline cases and advanced KOA stages.

3.4. T-SNE Visualization

Figure 3 provides a T-SNE plot for each classification setting, illustrating how samples cluster based on the learned embeddings. Across all three tasks (binary and multi-class), our graph-centric representation shows a distinct class separation. thus, demonstrating that explicit anatomical structure helps the model better discriminate subtle morphological variations.

3.5. Discussion

Overall, these results validate that a purely morphological approach can effectively capture the structural cues that radiolo-

Table 1: State of the art Comparison on OAI Test set (KLG-0 vs. KLG-2) [24]

Model	Accuracy (%)	F1-Score (%)
Wang et al. 2024 [8]	**89.80**	87.66
Wang et al. 2023 [17]	**89.14**	86.78
Wang et al. 2023 [18]	**89.54**	87,62
Tiuplin et al. 2018 [19]	87.33	84.82
Antony et al. 2017 [20]	85.50	81.31
Nasser et al. 2020 [16]	82.53	83.48
Our - Graph Model	**88.12**	**88.20**

gists rely on for KOA diagnosis. Moreover, from a deep learning standpoint, mapping radiographic data onto an anatomically grounded graph addresses two pivotal challenges: (i) the high-dimensional complexity of raw images, and (ii) the tendency of standard convolutional layers to focus on local intensity gradients rather than global bone geometry. Our strategy counters these issues by sampling anatomically salient points along the joint boundary and preserving spatial adjacency in a learnable graph. Consequently, the *GNN* layers handle the data in a domain where geometric relationships are explicit, highlighting osteophytes, joint space narrowing, and other structural cues integral to KOA diagnosis.

Table 2: State of the art Comparison on OAI Test set (KLG-0 vs. KLG-1) [24]

Model	Accuracy (%)	F1-Score (%)
Wang et al. 2024 [25]	70.21	-
Nasser et al. 2023 [4]	74.08	69.01
Nasser et al. 2020 [16]	69.83	70.95
Our - Graph Model	**85.66**	**86.15**

Table 3: State of the art Comparison on OAI Test set (All Classes 0-4) [24]

Model	Accuracy	F1-Score
Antony et al. 2016 [20]	53.40	43.00
Antony et al. 2017 [20]	63.60	59.00
Tiulpin et al. 2018 [21]	66.71	-
Chen et al. 2019 [22]	69.60	-
Wang et al. 2022 [23]	69.18	-
Sekhri et al. 2023 [4]	70.17	67.00
Sekhri et al. 2024 [9]	72.40	70.00
Our - Graph Model	**74.94**	**75.19**

3.6. Ablation Study

We evaluate how variations in the Graph model architecture influence classification performance on the OAI dataset [24].

Embedding Layer Depth. Table 4 examines three embedding dimensions ($D = \{32, 64, 128\}$). A moderate dimension of $D = 64$ achieves the best overall metrics, including 74.94% accuracy and 75.19% F1. We surmise that a smaller dimension ($D = 32$) underfits the morphological complexity, whereas a larger dimension ($D = 128$) may introduce unnecessary parameter overhead, impairing generalization.

Table 4: Ablation: Embedding Layer Depths on OAI Test set (All Classes 0-4) [24]

Depth	Accuracy	Precision	Recall	F1
D = 32	60.57	65.37	60.57	56.87
D = 64	**74.94**	**79.31**	**74.94**	**75.19**
D = 128	71.26	76.80	71.26	70.54

Table 5: Ablation: Number of Graph Layers on OAI Test set (All Classes 0-4)[24]

Layers	Accuracy	Precision	Recall	F1
2 Layers	46.49	37.64	46.49	38.04
3 Layers	**74.94**	**79.31**	**74.94**	**75.19**

Number of Graph Layers. Table 5 compares two- versus three-layer configurations of EdgeConv blocks. Employing only two layers severely hinders performance (46.49% accuracy), indicating insufficient depth for capturing higher-level geometric relationships. By contrast, three layers yield 74.94% accuracy and 75.19% F1, underscoring the importance of deeper graph-based representations for modeling bone morphology in KOA.

4. CONCLUSION

We presented a learnable *GNN*-based approach for KOA severity classification, departing from conventional radiographic pipelines that often emphasize local texture at the expense of global bone geometry. By sampling key joint landmarks and modeling them through a specialized graph neural network (EdgeConv), we introduce a morphology-centric inductive bias better aligned with clinical criteria. This design addresses the high-dimensional nature of raw radiographs, redirects the model away from incidental intensity artifacts, and highlights macroscopic structural changes central to KOA diagnosis. Our findings, emphasized by the state-of-the-art performance on the complex OAI dataset [24], substantiate our hypothesis that focusing on morphological cues—rather than purely pixel-level details—offers significant gains in both interpretability and performance for challenging KOA tasks.

5. ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the French National Research Agency (ANR) through the ANR-20-CE45 0013-01 project. This manuscript was prepared using data from the OAI and does not necessarily represent the views of the OAI investigators, the NIH, or private funding partners. The authors extend their sincere thanks to the study participants, clinical staff, and the coordinating center at UCSF.

6. REFERENCES

[1] J. H. Kellgren and J. Lawrence, "Radiological assessment of osteo-arthrosis," *Annals of the Rheumatic Diseases*, vol. 16, no. 4, pp. 494–502, 1957.

[2] D. Hayashi et al., "Imaging for osteoarthritis," *Annals of Physical and Rehabilitation Medicine*, vol. 59, no. 3, pp. 161–169, 2016.

[3] Yassine Nasser et al., "A discriminative shape-texture convolutional neural network for early diagnosis of knee osteoarthritis from x-ray images," *Physical and Engineering Sciences in Medicine*, vol. 46, no. 2, pp. 827–837, 2023.

[4] Aymen Sekhri et al., "Automatic diagnosis of knee osteoarthritis severity using swin transformer," in *Proceedings of the 20th International Conference on Content-Based Multimedia Indexing*, 2023.

[5] Mushrat Jahan et al., "Koa-cctnet: An enhanced knee osteoarthritis grade assessment framework using modified compact convolutional transformer model," *IEEE Access*, 2024.

[6] Richard Kijowski et al., "Deep learning applications in osteoarthritis imaging," *Skeletal Radiology*, vol. 52, no. 11, pp. 2225–2238, 2023.

[7] Huy Hoang Nguyen et al., "Semixup: in-and out-of-manifold regularization for deep semi-supervised knee osteoarthritis severity grading from plain radiographs," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4346–4356, 2020.

[8] Zhe Wang et al., "Transformer with selective shuffled position embedding and key-patch exchange strategy for early detection of knee osteoarthritis," *Expert Systems with Applications*, vol. 255, pp. 124614, 2024.

[9] Aymen Sekhri et al., "Shifting focus: From global semantics to local prominent features in swin-transformer for knee osteoarthritis severity assessment," in *2024 32nd EUSIPCO*, Lyon, France, 2024, pp. 1686–1690.

[10] Marouane Tliba et al., "Satsal: A multi-level self-attention based architecture for visual saliency prediction," *IEEE Access*, vol. 10, pp. 20701–20713, 2022.

[11] Marouane Tliba et al., "Deep-based quality assessment of medical images through domain adaptation," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 3692–3696.

[12] Alexander Kirillov et al., "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[13] Yue Wang et al., "Dynamic graph cnn for learning on point clouds," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 5, pp. 1–12, 2019.

[14] Tianle Cai et al., "Graphnorm: A principled approach to accelerating graph neural network training," in *International Conference on Machine Learning (ICML)*, 2020, Available at: `https://api.semanticscholar.org/CorpusID:221516279`.

[15] Osteoarthritis Initiative Investigators, "Osteoarthritis initiative (oai) data," 2006, Dataset.

[16] Yassine Nasser et al., "Discriminative regularized auto-encoder for early detection of knee osteoarthritis: Data from the osteoarthritis initiative," *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2976–2984, 2020.

[17] Zhe Wang, Aladine Chetouani, and Rachid Jennane, "A confident labelling strategy based on deep learning for improving early detection of knee osteoarthritis," 2023, arXiv:2303.13203.

[18] Zhe Wang, Aladine Chetouani, and Rachid Jennane, "Key-exchange convolutional auto-encoder for data augmentation in early knee osteoarthritis classification," 2023, arXiv:2302.13336.

[19] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala, "Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach," *Scientific Reports*, 2018.

[20] Joseph Antony et al., "Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks," in *MLDM 2017*, New York, NY, USA, 2017, pp. 376–390, Springer.

[21] Aleksei Tiulpin and Simo Saarakkala, "Automatic grading of in-dividual knee osteoarthritis features in plain radiographs using deep convolutional neural networks," *Diagnostics*, vol. 10, no. 11, pp. 932, 2020.

[22] Pingjun Chen et al., "Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 84–92, 2019.

[23] Zhe Wang et al., "Siamese-gap network for early detection of knee osteoarthritis," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, 2022, pp. 1–4.

[24] Osteoarthritis Initiative Investigators, "Osteoarthritis initiative (oai) data," 2006.

[25] Zhe Wang et al., "Temporal evolution of knee osteoarthritis: A diffusion-based morphing model for x-ray medical image synthesis," 2024, arXiv:2408.00891.