A Morphology-Aware Contrastive Learning Network for Detecting a Potential Subtype of Knee Osteoarthritis

Zixuan Wang[1], Hu Li[3], Hui Li[3], Kai Niu[1,2], Jianhao Lin[3], Zhiqiang He[1,2,†]
[1]*Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, China*
[2]*BUPT Hainan Advanced Digital Technology and Systems Laboratory, Hainan, China*
[3]*Arthritis Clinic and Research Center, Peking University People's Hospital, Beijing, China*
[†]Corresponding author: hezq@bupt.edu.cn

***Abstract*—During a recent knee osteoarthritis (KOA) screening in the Tibetan population, clinicians identified a rare subtype characterized by subtle yet distinctive morphological abnormalities. In addition to proximal tibial and distal femoral enlargement with diaphyseal narrowing, this subtype exhibits fine-grained edge irregularities and micro-deformations along bone contours. These morphological cues are critical for clinical diagnosis but are difficult to capture using conventional convolutional neural networks (CNNs), which tend to overfit texture patterns and struggle under extreme data scarcity. In clinical practice, diagnosis relies primarily on careful inspection of bone shape and contour rather than internal texture appearance. Motivated by this clinical observation, we propose a Morphology-Aware Contrastive Learning Network (MACLNet) for automated detection of this KOA subtype. Specifically, MACLNet introduces a morphology-aware enhancement mechanism that employs a wide-edge stream as explicit shape guidance, enforcing the network to prioritize structural morphology over texture-dominant cues. Furthermore, a momentum-based supervised contrastive learning strategy is incorporated to enlarge subtle inter-class differences in the feature space, effectively alleviating overfitting and representation collapse under limited and imbalanced data conditions. Experimental results indicate promising detection performance, suggesting potential utility for broader screening and characterization of this subtype.**

***Index Terms*—Knee Osteoarthritis (KOA), Supervised Learning, Contrastive Learning**

I. INTRODUCTION

This study focuses on a unique type of knee osteoarthritis (KOA) that has been discovered by orthopedic experts from Peking University People's Hospital during a screening for osteoarthritic conditions in Nagqu, Tibet, China. This condition manifests in the knee and exhibits distinct radiographic features compared to other samples of KOA. Radiographic examination demonstrates a distinct morphological alteration of the femur, characterized by proximal shaft narrowing and marked epiphyseal broadening at the articular region. The articular surface exhibits a cup-shaped deformity with irregular articular margins and undulating contours, indicative of advanced degenerative changes. Similarly, common KOA and rheumatoid arthritis also exhibits irregular articular margins, often accompanied by joint space narrowing [1]. While each condition has its own distinguishing features, these shared characteristics make distinguishing the subtype from other knee conditions more challenging.

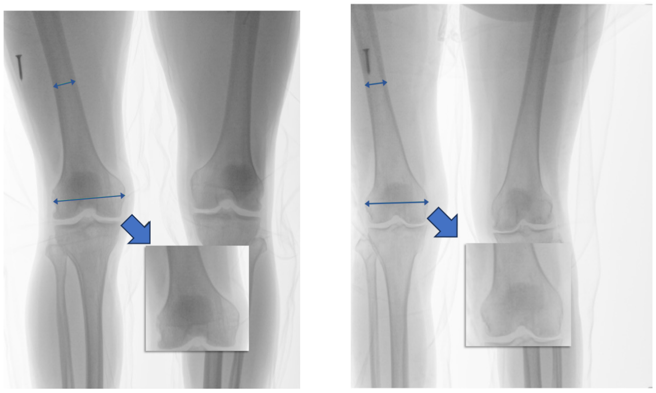

(a) Subtype of KOA (b) Normal Knees

Fig. 1. Comparative radiographs of special knee osteoarthritis subtypes. (a) Demonstrates characteristic findings of the Subtype, including proximal femoral shaft narrowing, marked epiphyseal broadening with cup-shaped deformity, and irregular articular margins with undulating contours. (b) Shows typical Normal knee presentation with preserved femoral morphology and smooth articular surfaces.

Current identification of this KOA subtype relies on manual radiographic assessment by experienced clinicians. Given its distinct morphological patterns, automated methods have the potential to enable efficient large-scale screening, particularly in resource-limited remote regions such as Tibet. The development of effective detection methods for this subtype faces two primary challenges. First, its radiographic features show significant overlap with other knee conditions, making visual differentiation particularly difficult. Second, the dataset exhibits severe imbalance- in a screening cohort of over 1,500 cases, only approximately 100 exhibited distinct characteristics of this subtype. This limited representation in the dataset poses significant obstacles for developing robust detection algorithms.

In this paper, we propose a morphology-aware contrastive learning framework for detecting a potential subtype of knee osteoarthritis (KOA), characterized by subtle morphological abnormalities such as proximal femoral shaft narrowing, epiphyseal broadening with cup-shaped deformity, and irregular articular margins, where diagnosis relies primarily on bone shape rather than internal texture. Conventional convolutional neural networks (CNNs) are biased toward texture cues and prone to overfitting under extreme data scarcity, limiting their ability to capture fine-grained structural deformations. To address this, we introduce a morphology-aware enhancement mechanism using a wide-edge representation as explicit shape guidance, coupled with a dual-stream encoder and attention-based feature fusion to integrate global context and localized boundary information. Supervised contrastive learning, together with a momentum encoder, stabilizes feature representations, enlarges subtle inter-class differences, and mitigates overfitting and feature space collapse in small and imbalanced datasets. By integrating clinical priors with contrastive constraints, our framework provides a robust, morphology-aware approach for automated KOA subtype analysis, capable of handling data scarcity. Experimental results demonstrate superior diagnostic performance, highlighting its effectiveness for automated KOA subtype screening.

In summary, we propose MACLNet, a morphology-aware feature differentiation network integrated with supervised contrastive learning for detecting a potential subtype of KOA. The specific contributions are as follows:

1) Hard-Attention Mechanism based on Clinical Prior: We identify lesion-prone bone joint edge regions according to prior clinical knowledge, focusing on irregular joint margins and epiphyseal enlargement. The Morphology-aware enhancement module segments these regions and generates a wide-edge representation that preserves a contiguous band around the bone contours. By providing explicit shape guidance, the model robustly captures subtle and spatially morphological deviations.
2) We incorporate supervised contrastive training loss jointly optimized with focal loss. This training strategy enables the model to simultaneously respond to gradients from both loss functions, where the supervised contrastive loss reduces intra-class variance and enlarges inter-class distances, while the focal loss adaptively emphasizes hard-to-classify and minority samples, mitigating class imbalance and guiding a more robust and structured organization of the feature space.
3) We introduce a momentum-based parameter update mechanism in MACLNet, employing dual encoders with identical architectures to extract key features. The momentum encoder progressively aligns with the primary encoder during training, stabilizing feature representations and mitigating overfitting and feature space collapse commonly observed in small-sample and imbalanced datasets. Given that the morphological deviations distinguishing this KOA subtype are subtle and spatially distributed, the momentum-based contrastive mechanism explicitly enlarges these fine-grained inter-class differences in the feature space. This approach enables the model to robustly capture discriminative subtype-specific features and overcome the challenges of extreme data scarcity.

II. RELATED WORKS

In recent years, deep learning methods are increasingly applied in orthopedic disease auxiliary diagnosis. The current diagnosis of orthopedic diseases relies primarily on the subjective assessment of clinical symptoms and imaging features. Therefore, deep learning methods can serve as a valuable auxiliary tool to support human specialists in making more accurate and objective diagnoses. Detection and classification techniques are used to differentiate disease categories, such as diagnosis of fractures [2]–[4], arthritis [5], and bone tumors [6], with performance varying based on imaging modality and disease complexity [7]. In the specific area of knee joint image analysis, multiple machine learning and deep learning models are developed for predicting knee osteoarthritis progression and diagnosing related conditions.H. Byeon utilized the ResNeXt model to automatically classify Kellgren-Lawrence grades, demonstrating the potential of deep learning in the automatic diagnosis of KOA [8]. D. Dai and P. Tang proposed the HCGN model, which combines convolutional neural networks and graph neural networks to enable early prediction and severity assessment of knee osteoarthritis [9]. V. Arunachalam and N. Kumareshan introduced a modified Elman Recurrent Neural Network (MERNN) for muscle segmentation, contributing to a comprehensive diagnosis of knee joint diseases [10]. While the aforementioned studies have significantly advanced the application of deep learning in knee osteoarthritis assessment through grading classification, progression prediction, and multi-modal analysis, these methods are primarily designed for conventional knee disease diagnostic tasks. This presents a fundamental technical challenge: developing feature representation strategies that can disentangle subtype-specific manifestations- a prerequisite for enabling reliable computational differentiation and supporting population-level phenotyping studies.

III. METHOD

This part provides a thorough explanation to the modules and design rationale of the MACLNet, which is used for detection of the KOA subtype. Firstly, we give an overall introduction to the framework of MACLNet. Subsequently, we will explain the principles of each module in separate subsections

A. Framework of MACLNet

As illustrated in Fig.2, MACLNet is a morphology-aware contrastive learning framework designed to detect a potential KOA subtype characterized by subtle morphological abnormalities such as edge irregularities and epiphyseal enlargement. The network emphasizes subtle differences in lesion features between the target subtype and morphologically similar

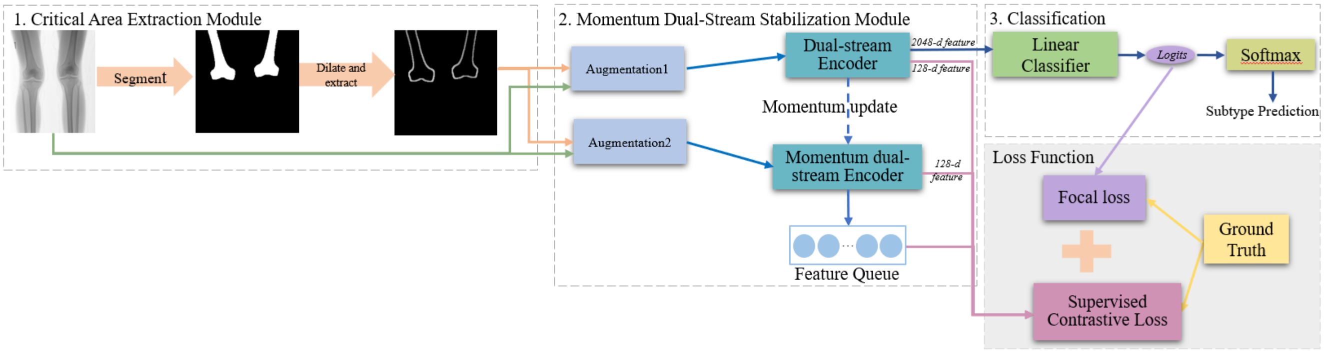


Fig. 2. Overall architecture of Contrastive Learning-based Differentiation Network (MACLNet)

conditions, enabling more accurate classification. MACLNet consists of three functional modules: the Morphology-aware enhancement module, the Momentum Dual-Stream Stabilization Module, and the Classification Module.

The Morphology-aware enhancement module segments femoral contours and expands boundary regions to capture areas prone to subtype lesions, thereby producing paired inputs consisting of the original radiograph and its corresponding wide-edge representation. During training, the paired inputs are subjected to two independent stochastic augmentations, generating two augmented views of the same sample. Each augmented view contains both the original radiograph and the wide-edge representation.

Each augmented view is subsequently processed by a dual-stream encoder, in which modality-specific feature extractors are used to capture global anatomical structure from the original radiograph and local boundary information from the wide-edge representation, respectively. An attention-based cross-feature fusion module is employed to integrate the two feature streams into a unified representation. Supervised contrastive learning is subsequently applied to the fused features, where one augmented view is encoded by a primary encoder and the other by a momentum encoder to stabilize representation learning.

Finally, subtype labels are predicted by the Classification Module using a linear classifier. The network is trained using a combined loss function, where Focal Loss emphasizes hard-to-classify and minority samples, while Supervised Contrastive Loss encourages feature separability. This joint optimization enables MACLNet to learn discriminative and semantically meaningful representations for accurate KOA subtype detection.

B. Morphology-Aware Enhancement Module

The Morphology-aware enhancement module aims to localize lesion-sensitive regions in knee X-ray images based on prior clinical knowledge. As summarized in Table I, the most discriminative imaging characteristics between knee osteoarthritis (KOA) and its potential subtype primarily involve abnormalities in bone margins and bone morphology. Specifically, the susceptible regions mainly include the distal femoral contour, where irregular margins and cup-shaped deformities are most discriminative for subtype identification. Although proximal tibial enlargement and diaphyseal narrowing may co-occur in this subtype, these tibial alterations are frequently observed in non-subtype individuals within the screened population and are therefore considered non-specific background findings that do not reliably distinguish the true subtype from other conditions. Accordingly, our morphology-aware modeling focuses on the femoral region as the primary anatomical site of diagnostic relevance.

To accurately extract these regions, our approach first segments the femoral bone using a deep learning-based segmentation network. We manually annotate femoral contours to train a U-Net++ model, which can automatically generate precise binary masks $M(x, y)$ representing the bone region:

$$M(x, y) = \begin{cases} 1, & \text{bone region} \\ 0, & \text{background} \end{cases} \tag{1}$$

To extract anatomically consistent lesion-prone regions, the segmentation network was trained on an independently annotated subset consisting of 120 radiographs with expert-labeled femoral masks, achieving an average Dice coefficient of 0.93 on a held-out validation set. During MACLNet training, the segmentation module is treated as a fixed preprocessing component. The segmentation network remains frozen, and no gradients are back-propagated through the segmentation branch.

Based on the segmented mask, we further extract a wide-edge region $W(x, y)$ through morphological dilation:

$$W(x, y) = (M \oplus B) - M \tag{2}$$

where B denotes the structuring element (e.g., circular or square kernel) and $\oplus$ represents the dilation operation. The wide-edge image captures detailed structural information around the femoral edges, including irregularities in the bone margin and subtle texture changes, which are important for distinguishing potential subtypes.

The extracted wide-edge image serves as an additional input modality for the subsequent classification and contrastive learning modules. By focusing on these lesion-prone regions,

TABLE I
Comparison of Imaging Features Between Knee Osteoarthritis Subtypes and KOA

Knee Imaging Feature	**Knee Subtype**	**Knee Osteoarthritis**	**Discrimination Method**
Bone Margin Abnormality	Irregular and rough joint surface edges; serrated appearance; unclear joint contour; subchondral bone destruction	Osteophyte formation at joint edges (irregular white protrusions or bone spurs)	Extract bone edge details: irregular for subtypes, regular or osteophytes for KOA
Bone Morphology Changes	Enlargement of proximal tibia and distal femur; narrowing at the diaphysis of femur and tibia	None	Morphological change indicates subtype
Subchondral Sclerosis	None	Subchondral sclerosis (white band) and/or cystic changes (black dots within white band)	Presence of sclerosis helps exclude subtype
Bone Absorption	Radiolucent areas along subchondral bone adjacent to joint line, often with eroded cortical margins	None	Extract texture features within joint edges
Joint Space Change	None	Asymmetric narrowing of joint space	Joint space narrowing helps exclude subtype

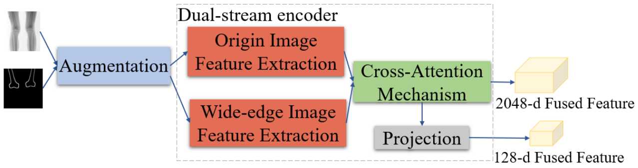


Fig. 3. Dual-stream encoder of MACLNet

the module enhances the model's ability to detect subtle morphological differences while mitigating the effects of irrelevant background and imaging noise.

C. Contrastive Feature Enhancement Module

1) Momentum Dual-Stream Stabilization Module: Inspired by MoCo [11], supervised contrastive learning is extended to a dual-input, region-aware framework tailored for knee osteoarthritis subtype identification. The Dual-Stream Encoder and the Momentum Dual-Stream Encoder share the same architecture, ensuring consistent feature extraction and fusion for both augmented views.. As illustrated in Fig.3, each training sample is augmented to generate paired inputs consisting of an original knee radiograph and its corresponding wide-edge image, which are treated as two complementary modalities.

After augmentation, the original and wide-edge images are separately processed by two parallel feature extraction branches, namely the Origin Image Feature Extraction branch and the Wide-edge Image Feature Extraction branch. Each branch encodes complementary information: the original image branch captures global anatomical appearance, while the wide-edge branch emphasizes edge-enhanced structural and boundary-related features associated with subtype lesions. Both branches employ a shared backbone and output 2048-dimensional feature representations.

To effectively model interactions between these complementary features, we introduce a cross-attention fusion mechanism, which adaptively highlights discriminative bone-edge patterns relevant to subtype differentiation. The resulting fused representation preserves rich semantic and structural cues and is retained as a 2048-dimensional feature vector for downstream classification.

In parallel, a lightweight projection head maps the fused feature into a 128-dimensional embedding space, which is used exclusively for supervised contrastive learning. During training, a momentum-based feature updating strategy is employed to maintain a stable representation space, enabling consistent supervision across augmented input pairs while preventing representation collapse.

This design balances feature adaptability and stability, leading to robust and discriminative representations. We adopt ResNet50 [12] as the backbone network, which outputs a 32×2048 feature tensor.

2) Loss Function: To jointly enhance feature representation learning and classification performance, we adopt a combined optimization strategy integrating Supervised Contrastive Loss (SupConLoss) and Focal Loss.

SupConLoss is applied to the feature embeddings obtained from the MoCo framework, where representations produced by both the query and key encoders are jointly considered within a mini-batch to construct positive and negative sample pairs. By leveraging label supervision, SupConLoss explicitly pulls together features from samples of the same class while pushing apart those from different classes, thereby reducing intra-class variance and enlarging inter-class separability. This property is particularly beneficial for capturing subtle morphological differences and improving robustness under class-imbalanced conditions.

The Supervised Contrastive Loss is formulated as:

$$\mathcal{L}_{SC} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|P(i)|}\sum_{p\in P(i)} -\log \frac{\exp(\mathbf{z}_i\cdot\mathbf{z}_p/\tau)}{\sum_{a=1}^{N}\mathbf{1}_{[a\neq i]}\exp(\mathbf{z}_i\cdot\mathbf{z}_a/\tau)} \tag{1}$$

where N denotes the batch size, $\mathbf{z}_i$ represents the normalized feature embedding of the i-th sample, τ is a temperature scaling parameter, $P(i)$ denotes the set of indices of samples sharing the same class label as sample i, and $\mathbf{1}_{[\cdot]}$ is an indicator function.

TABLE II

Performance comparison of different models using five-fold cross-validation (mean ± std).

Model	Accuracy	Precision	Recall	F1-score	AUC	Specificity
ResNet50	0.778 ± 0.021	0.610 ± 0.028	0.390 ± 0.035	0.455 ± 0.032	0.75 ± 0.03	0.85 ± 0.02
ViT	0.828 ± 0.019	0.647 ± 0.026	0.550 ± 0.031	0.595 ± 0.029	0.82 ± 0.02	0.88 ± 0.02
DenseNet121	0.801 ± 0.023	0.631 ± 0.027	0.450 ± 0.034	0.524 ± 0.030	0.78 ± 0.03	0.86 ± 0.02
Moco v2	0.780 ± 0.024	0.578 ± 0.029	0.435 ± 0.036	0.498 ± 0.031	0.76 ± 0.03	0.83 ± 0.03
SimCLR	0.768 ± 0.025	0.565 ± 0.030	0.410 ± 0.038	0.475 ± 0.033	0.74 ± 0.03	0.84 ± 0.03
MACLNet	0.927 ± 0.015	0.845 ± 0.020	0.893 ± 0.018	0.862 ± 0.019	0.94 ± 0.01	0.91 ± 0.02

For the classification task, we employ Focal Loss instead of standard cross-entropy loss to address class imbalance and emphasize hard-to-classify samples. Focal Loss dynamically down-weights easy examples while assigning higher importance to misclassified and minority-class samples, thereby improving classification robustness in imbalanced subtype detection scenarios.

The final training objective is defined as the sum of the two loss terms:

$$\mathcal{L} = \mathcal{L}_{SC} + \mathcal{L}_{FL}. \tag{2}$$

This joint optimization strategy encourages discriminative feature clustering while simultaneously guiding the classifier to focus on challenging cases, resulting in improved subtype differentiation performance.

IV. EXPRIMENTS

A. Dataset

We collected and processed knee X-ray images from the epidemic area in Nagqu, Tibet, China to construct a joint diagnostic dataset for knee osteoarthritis, aiming to validate and evaluate the performance of our algorithm. Individuals from several endemic villages in Nagqu, Tibet were screened, and knee X-ray images were collected from 346 individuals, including 91 patients diagnosed with a knee osteoarthritis subtype and 255 non-subtype cases. The disease markers for each X-ray were annotated by a professional radiologist. At the same time, we manually annotate the femur contour in X-ray images to provide ground truth labels for the segmentation method, enabling automatic femur segmentation. Given the small sample size and class imbalance in the dataset, we applied 5-fold cross-validation to evaluate the model.

B. Implementation Details

The proposed method was implemented using PyTorch 2.1.0 and Python 3.8. The image size of the input network is 224 * 224, the optimizer uses SGD, the initial learning rate is 0.001, 100 epochs were trained, and the batch size is 32. We performed data augmentation on the dataset, including resizing to a predefined input size, random horizontal flipping, random gray-scale conversion with a probability of 0.2, and random color jittering with brightness, contrast, saturation, and hue adjustments of up to 0.4, applied with a probability of 0.8. Additionally, we applied normalization and a two-crop transformation to generate multiple augmented views of the same image.

TABLE III

Ablation study of different component combinations.

	Choice		
Wide-edge image	×	✓	✓
Focal loss	✓	✓	×
Our joint loss	✓	×	✓
Accuracy	0.8288	0.9010	**0.9267**
F1-score	0.7100	0.8312	**0.8616**
AUC-ROC	0.83	0.85	**0.94**

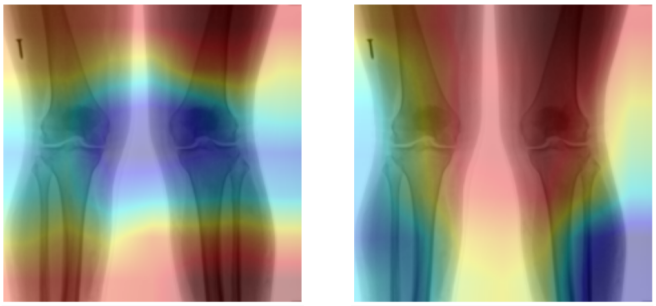
C. Results and Discussion

We evaluated the effectiveness of the proposed MACLNet in differentiating a potential KOA subtype under a small-sample and imbalanced setting. To ensure robust and reliable evaluation, all experiments were conducted using a five-fold cross-validation strategy, and the mean performance along with standard deviation is reported.

Table II presents a comprehensive comparison between MACLNet and several baseline approaches, including conventional supervised classifiers and contrastive learning-based methods. In particular, representative contrastive learning baselines, including MoCo v2 and SimCLR, were evaluated without the proposed morphology-aware enhancement to ensure a fair comparison.To further assess the reliability of these improvements, we conducted a paired t-test based on the five-fold cross-validation results. The proposed MACLNet achieves statistically significant improvements over all compared models, with all p-values less than 0.05.

Overall, MACLNet consistently achieves the best performance across all evaluation metrics, including accuracy, F1-score, and AUC-ROC, while maintaining a favorable balance between sensitivity and specificity. These results indicate that the proposed dual-stream architecture with cross-attention fusion effectively enhances discriminative feature learning. Moreover, supervised contrastive learning further reduces intra-class variance and enlarges inter-class separability, which is particularly beneficial for screening tasks involving subtle morphological differences and class imbalance.

We conducted ablation studies to evaluate the contributions of key components, as summarized in Table III. Removing the morphology-aware enhanced (wide-edge) input leads to a substantial decrease in F1-score from 0.8616 to 0.7100, whereas removing the supervised contrastive loss or focal loss also degrades performance. These findings confirm that all components contribute to the overall performance, with morphology-aware enhancement being particularly crucial for



(a) ResNet50 (b) MACLNet

Fig. 4. Comparison of model attention heatmaps for KOA subtype detection. Red indicates areas of highest attention, yellow denotes moderately high attention, green represents medium attention, and blue reflects the lowest attention.

capturing subtle and spatially extended structural patterns characteristic of this KOA subtype. Traditional edge detectors, such as Canny, generate very thin edges that are sensitive to noise and often discard contextual information around bone margins. In contrast, the proposed morphology-aware enhancement expands the bone boundary to include adjacent regions, providing a contiguous structural neighborhood. This design enables the network to capture clinically relevant macrostructures, such as cup-shaped deformities of the femoral condyle, which are essential for distinguishing this KOA subtype.

The qualitative effectiveness of this design is further illustrated by the attention heatmaps shown in Fig. 4. The ResNet50 baseline exhibits diffuse attention across background and non-diagnostic regions, whereas MACLNet concentrates attention on anatomically meaningful areas, particularly along the distal femoral contours. This visualization demonstrates that MACLNet successfully shifts model focus from texture-dominated cues toward clinically relevant morphological structures, providing intuitive interpretability that complements the quantitative performance gains.

Overall, these results demonstrate that MACLNet effectively leverages morphology-aware enhancement, dual-input cross-attention fusion, and supervised contrastive learning to achieve superior classification performance and robust feature differentiation compared with standard baselines and ablated variants.

V. CONCLUSION

In this work, we proposed MACLNet, a morphology-aware contrastive learning network for detecting a potential KOA subtype. The network integrates original radiographs with morphology-enhanced inputs and employs a dual-stream encoder with cross-attention to fuse global and edge-focused features. A momentum-based encoder and supervised contrastive learning enable robust feature representation, enlarging subtle inter-class differences while mitigating overfitting and class imbalance. Extensive experiments and ablation studies confirm that MACLNet effectively captures clinically relevant morphological deviations, such as edge irregularities and epiphyseal enlargement, and outperforms standard backbones in small-sample and imbalanced settings. Overall, MACLNet provides a robust and clinically guided framework for automated detection of subtle KOA subtypes, supporting improved diagnosis and patient management. Future work will investigate additional imaging modalities and temporal data to further enhance sensitivity to early-stage and progressive morphological changes. Our approach holds significant clinical application value and is of great importance for large-scale assisted diagnosis of KOA subtype.

ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (2025YFC2511800).

REFERENCES

[1] J. Pan, Y. Wu, Z. Tang, K. Sun, M. Li, J. Sun, J. Liu, J. Tian, and B. Shen, "Automatic knee osteoarthritis severity grading based on X-ray images using a hierarchical classification method," *Arthritis Research & Therapy*, vol. 26, no. 1, p. 203, 2024.

[2] N. Aghnia Farda, J.-Y. Lai, J.-C. Wang, P.-Y. Lee, J.-W. Liu, and I.-H. Hsieh, "Sanders classification of calcaneal fractures in CT images with deep learning and differential data augmentation techniques," *Injury*, vol. 52, no. 3, pp. 616–624, 2021.

[3] L. Shen, C. Gao, S. Hu, D. Kang, Z. Zhang, D. Xia, Y. Xu, S. Xiang, Q. Zhu, G. Xu, F. Tang, H. Yue, W. Yu, and Z. Zhang, "Using artificial intelligence to diagnose osteoporotic vertebral fractures on plain radiographs," *J. Bone Miner. Res.*, vol. 38, no. 9, pp. 1278–1287, 2023.

[4] L.-W. Cheng, H.-H. Chou, Y.-X. Cai, K.-Y. Huang, C.-C. Hsieh, P.-L. Chu, I.-S. Cheng, and S.-Y. Hsieh, "Automated detection of vertebral fractures from X-ray images: A novel machine learning model and survey of the field," *Neurocomputing*, vol. 566, p. 126946, 2024.

[5] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala, "Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach," *Sci. Rep.*, vol. 8, no. 1, p. 1727, 2018.

[6] J. Li, S. Li, X. Li, S. Miao, C. Dong, C. Gao, X. Liu, D. Hao, W. Xu, M. Huang, and J. Cui, "Primary bone tumor detection and classification in full-field bone radiographs via YOLO deep learning model," *Eur. Radiol.*, vol. 33, no. 7, pp. 4237–4248, 2023.

[7] J. Lee and S. W. Chung, "Deep learning for orthopedic disease based on medical image analysis: Present and future," *Appl. Sci.*, vol. 12, no. 2, p. 681, 2022.

[8] H. Byeon, "Automatic diagnosis of knee osteoarthritis based on Kellgren–Lawrence grades using the ResNeXt model," *Nanotechnol. Percept.*, vol. 20, pp. 1650–1662, 2024.

[9] X. Yang, P. Tang, K. Zou, and D. Dai, "HCGN: Hierarchical convolution and graph network for predicting knee osteoarthritis," in *Proc. IEEE Int. Conf. Med. Artificial Intelligence (MedAI)*, 2024, pp. 81–90.

[10] V. Arunachalam and N. Kumareshan, "An effective satin bowerbird optimization based modified Elman recurrent neural network for thigh muscle segmentation and quantification in patients undergoing total knee replacement (TKR)," *Iran. J. Sci. Technol. Trans. Electr. Eng.*, 2024.

[11] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 9729–9738.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.