2025 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP 2025)
April 06 - 11, 2025 Hyderabad, India

MedCAM-OsteoCls: Medical Context Aware Multimodal Classification of Knee Osteoarthritis

Paper ID: 5387
Track: Biomedical signal and image processing

Presenter: Akshay Daydar, Research Scholar, IIT Guwahati

2025 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP 2025)

Paper Title: MedCAM-OsteoCls: Medical Context Aware Multimodal Classification of Knee Osteoarthritis

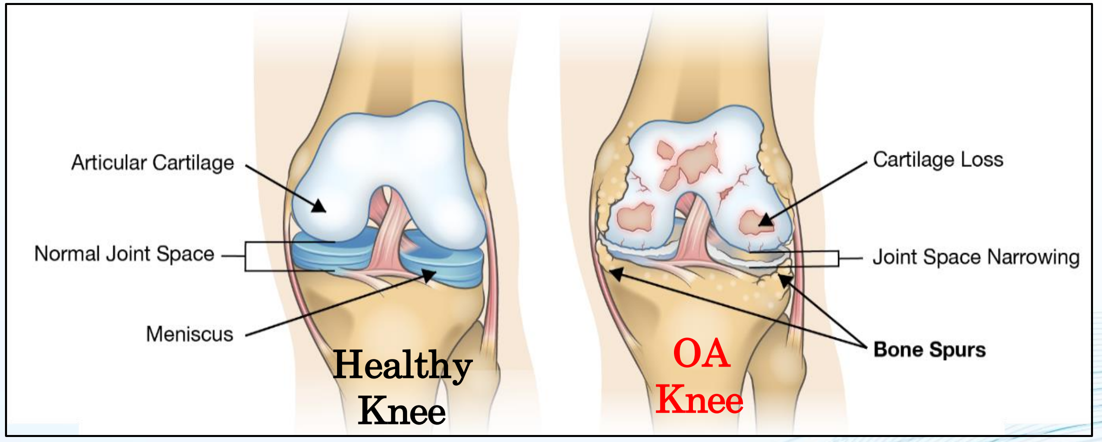
Introduction

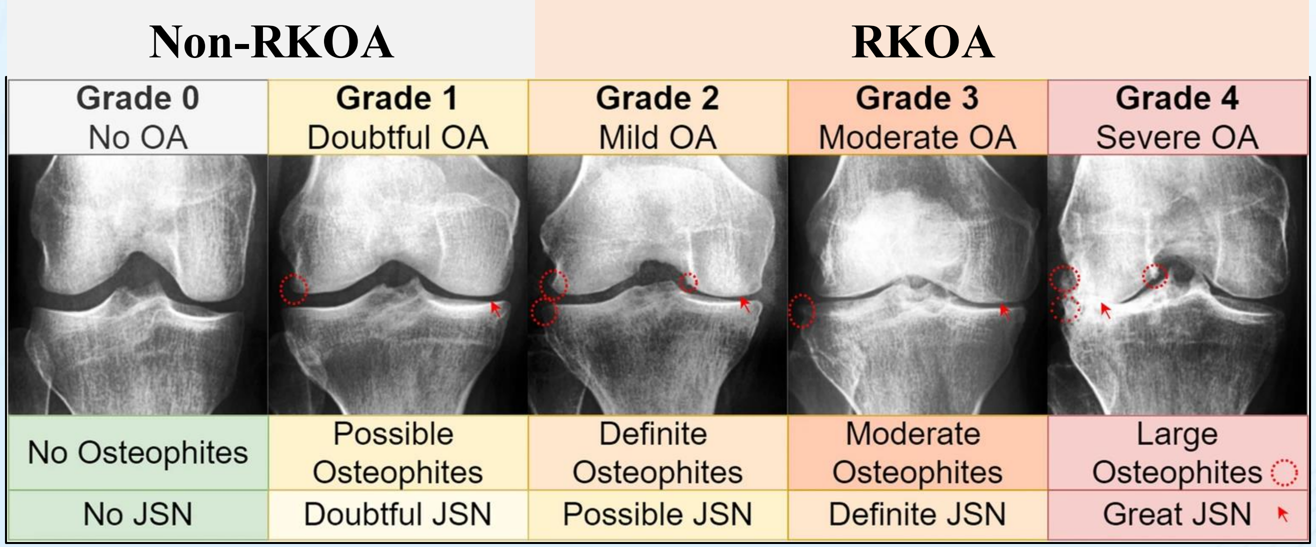
1.7 Billion MSD[a]

343 million OA [b]

47 million KOA in India [b]

Preceded by only low back pain and neck pain Among MSD





Multi-class KL grading: KL grades 0 to 5

Binary/Radiological KL grading: KL grades 0 and 1 → Non-RKOA

KL grades 2,3 and 4 → RKOA

Figure: Visual representation illustrating X-rays with Kellgren-Lawrence (KL) Grades ranging from 0 to 4, progressing from left to right.

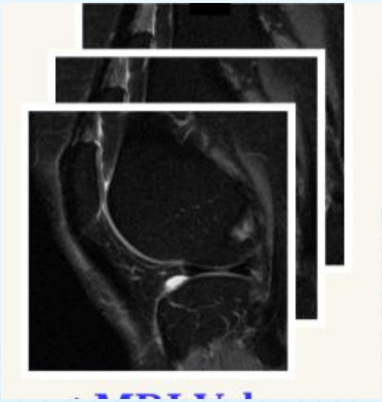
Challenges

1. **Heterogeneity in KL grading:** Less compactness within feature clusters

2. **Distinct nature of diagnostic modalities:** Hard to determine nature and location at which information exchange can take place across modalities

3. **Irregularities in the number of scans per MRI volume:** Difficulties in training due to sparse anatomical representation in MRI volume

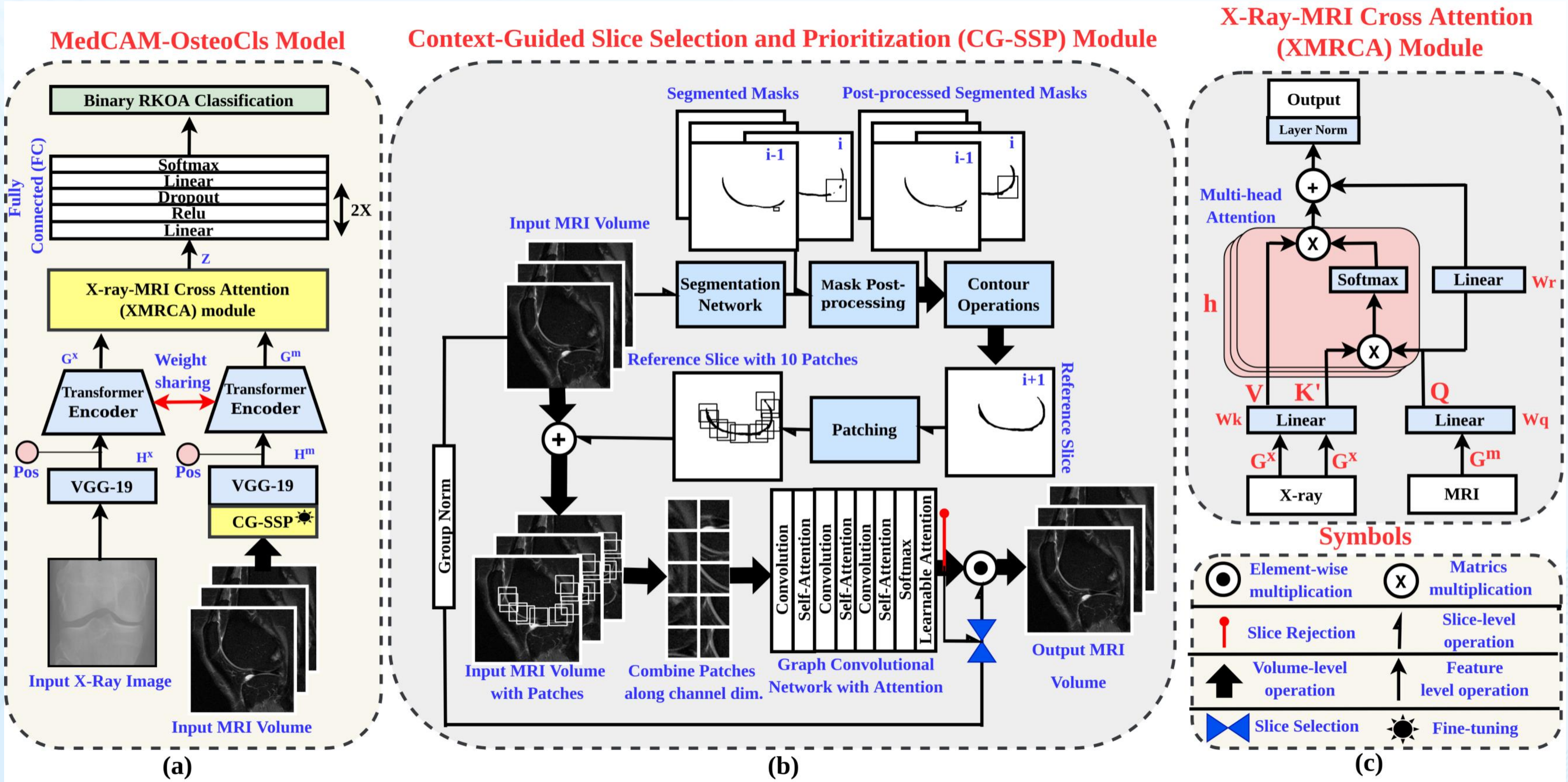
Knee X-ray image (AP view)



Knee MRI volume (Sagittal view)

Paper Title: MedCAM-OsteoCls: Medical Context Aware Multimodal Classification of Knee Osteoarthritis

Contributions



MedCAM-OsteoCls

=

Context-Guided Slice Selection and Prioritization (CG-SSP)

+

Xray-MRI Cross-Attention (XMRCA) modules

Figure: Overall schematic of the proposed MedCAM-OsteoCls model with (a) VGG-19-TE +Fully Connected (FC) Network, (b) the CG-SSP module, and (c) the XMRCA module.

The CG-SSP Module

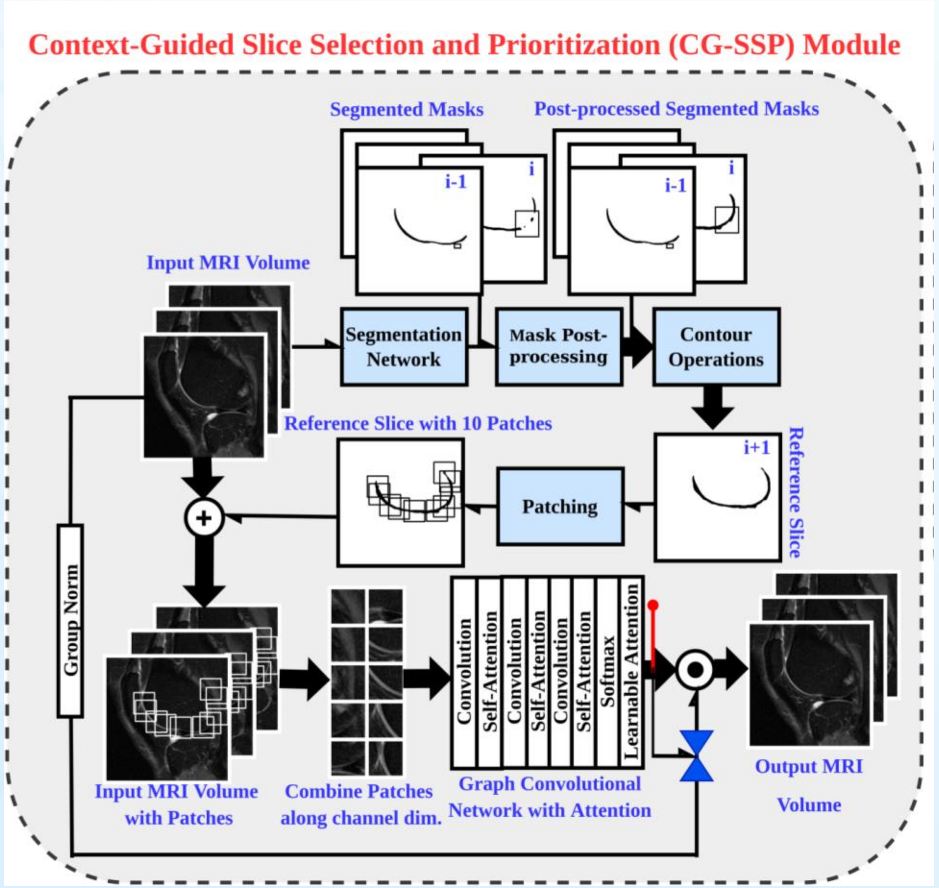


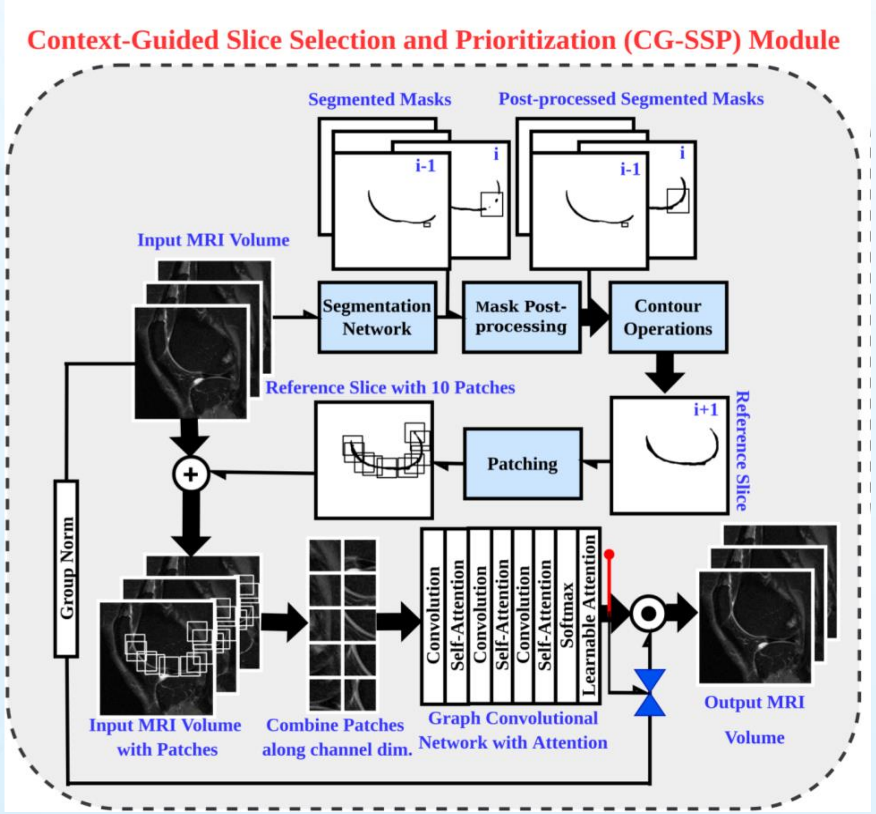
Figure: CG-SSP module

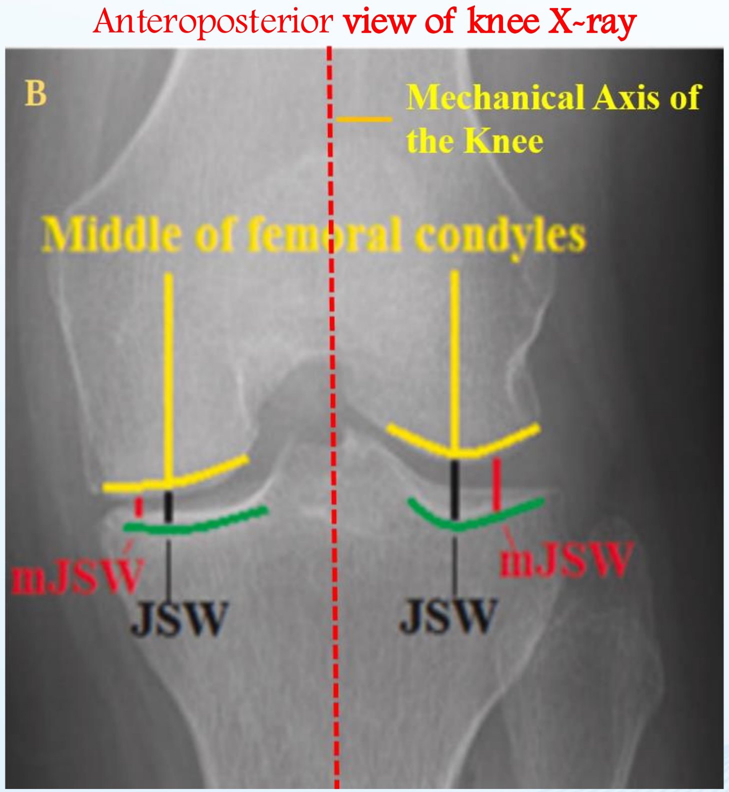
Procedure

1. MRI volume passed to **MtRA Segmentation Network** for **extraction of Femoral Cartilage (FC)**
2. FC Mask refinement by **Erosion and Dilation** followed by **Identification of reference mask** which has largest contour length per MRI volume
3. **Patch generation** on selected reference mask by **accessing the centroidal line** of Bone-cartilage and Cartilage-Synovium interface
4. **Extend the patches** to all slices within MRI volume and pass it in channel-wise manner to **trained-GCNT**
5. **Find the attention scores** from all the slices and **select top-14** slices per MRI volume

The CG-SSP Module

Paper Title: MedCAM-OsteoCls: Medical Context Aware Multimodal Classification of Knee Osteoarthritis





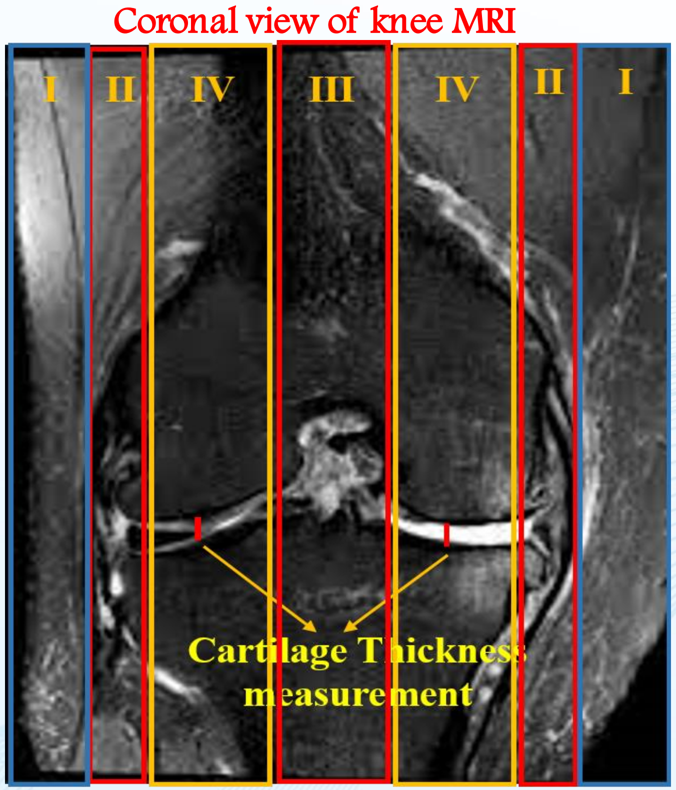


Figure: A) CG-SSP module, B) Radiograph of knee and, C) Coronal Sagittal view of MR slices with the depiction of (I) initial and ending slices (i.e. noise-only slices with no tissue information), (II) slices with fewer occupancy of femoral and tibial bone, and (III) slices with anterior and posterior cruciate ligaments (that contain no tibia cartilage information), and (IV) medial-lateral regions (selected for segmentation analysis)

The XMRCA Module

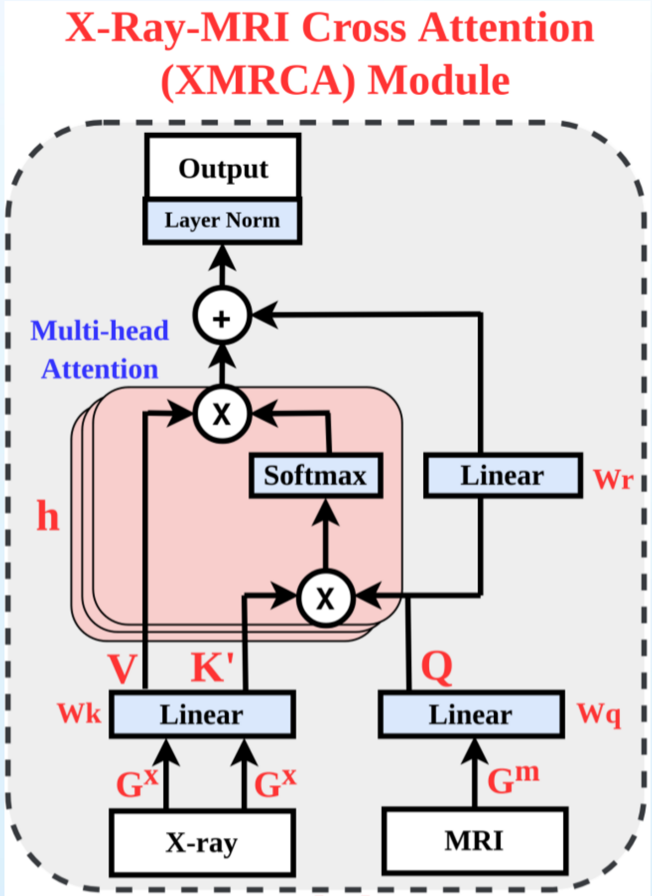


Figure: The Xray-MRI Cross-Attention (XMRCA) module

Formulations

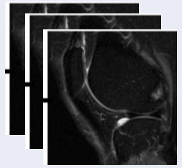
$$Q = G_m W_Q, \quad K = G_x W_K, \quad V = G_x W_V \tag{1}$$

$$\text{CrossAttention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k/h}}\right)V \tag{2}$$

$$z = \text{LayerNorm}(\text{CrossAttention}(Q, K, V) + G_x W_r)$$

Experimental Setup

Dataset	Data	View	Subjects	MRI sequence	Number of Slices	Train:Val:Test
OAI	X-ray+ MRI	Coronal, Sagittal	1695	SAGittal MRI Inversion Weighted Turbo Spin Echo with Fat Suppression technique (SAG IW TSE FS)	Min. 14 to Max. 31	70:10:20
SKM-TEA	MRI	Sagittal	68	3D Double Echo Steady-State (DESS)	160	90:10:NA



Parameter	Value
Input shape	300*300
Training Epochs & Batch Size	36
Learning Rate	$5*10^{-4}$
Weight Decay	$5*10^{-3}$
Loss Function	Cross entropy + Ordinal

$$\text{Adjustable Ordinal Loss}(\mathcal{L}_O) = \frac{1}{N}\sum^{N} \mathcal{L}_{O,\psi}$$

$$\text{Cross Entropy Loss}(\mathcal{L}_{CS}) = -\frac{1}{N}\sum_{\psi=1}^{N}\sum_{\phi=1}^{b} GT_{\psi,b}\log(p_{\psi,b})$$

$$\text{Loss}(\mathcal{L}) = \mathcal{L}_O + \mathcal{L}_{CS}$$

Experimental Results

Table: Comparison of the experimental results of the proposed MedCAM-OsteoCls with the unimodal and multimodal SOTA for the OAI dataset

	Architecture	Acc.(%)	Pre.	Recall	K	MAE	GPU Load#(GB)
	Unimodal						
X-ray	Chen et al. (P. Chen et al., 2019)	83.95	0.83	0.83	0.66	0.16	4.60
X-ray	Jain et al. (Jain et al., 2024)	82.5	0.83	0.80	0.62	0.18	2.79
X-ray	Guida et al. (Guida et al., 2023)	70.34	0.69	0.67	0.36	0.3	2.29
X-ray	Karim et al. (Karim et al., 2021)	85.25	0.85	0.84	0.68	0.15	5.27
X-ray	MedCAM-OsteoCls^{x}	85.58	0.85	0.84	0.69	0.14	4.05
MRI	Guida et al. (Guida et al., 2023)	67.75	0.66	0.63	0.27	0.32	53.79
MRI	Karim et al. (Karim et al., 2021)	64.02	0.62	0.56	0.14	0.36	8.39
MRI	Berrimi et al. (Berrimi et al., 2024)	61.00	0.31	0.50	NA	0.38	79.25
MRI	Bien et al. (Bien et al., 2018)	80.00	0.8	0.77	0.56	0.2	4.30
MRI	Tsai et al. (Tsai et al., 2020)	75.00	0.75	0.71	0.45	0.25	66.19*
MRI	MedCAM-OsteoCls^{m}	75.53	0.74	0.75	0.43	0.24	4.15
	Multimodal						
Xray+MRI	Guida et al. (Guida et al., 2023)	62.88	0.66	0.53	0.29	0.37	53.84
Xray+MRI	Karim et al. (Karim et al., 2021)	85.25	0.85	0.83	0.68	0.14	13.66
Xray+MRI	MedCAM-OsteoCls	87.52	0.87	0.86	0.73	0.12	4.42

• The best and second best results are denoted in red and blue colors, respectively.
* Calculated after 5 epochs due to memory usage overhead.
# Calculated for single batch size with image dimensions of (300,300).

Experimental Results

Paper Title: MedCAM-OsteoCls: Medical Context Aware Multimodal Classification of Knee Osteoarthritis

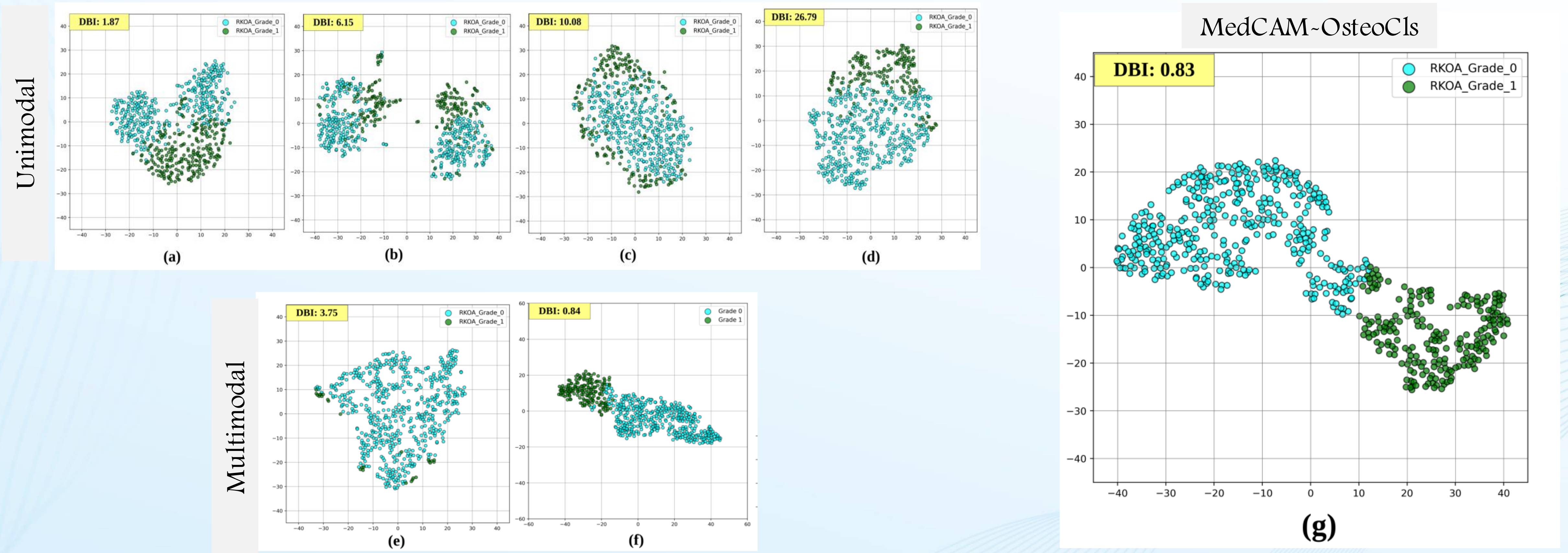


Figure: Visualization of tSNE plots for Unimodal (a) Chen et al. (P. Chen et al., 2019), Jain et al. (Jain et al.,2024) of X-ray, (c) Bien et al. (Bien et al.,2018), Tsai et al. (Tsai et al.,2020) of MRI and Multimodal schemes such as (d) Guida et al. (Guida et al.,2023) and Karim et al.(Karim et al.,2021).

Ablation Study

1. Effect of CG-SSP and XMRCA modules on OAI dataset

Ablation formulations

M1: Concat$(H_x, H_m) \rightarrow$ TE, **M2**: Concat$(G_x, G_m) \rightarrow$ FC,

M3: $XMRCA^{H_m=K,V}_{H_x=Q}(H_x, H_m) \rightarrow$ TE, **M4**: $XMRCA^{G_m=K,V}_{G_x=Q}(G_x, G_m) \rightarrow$ FC

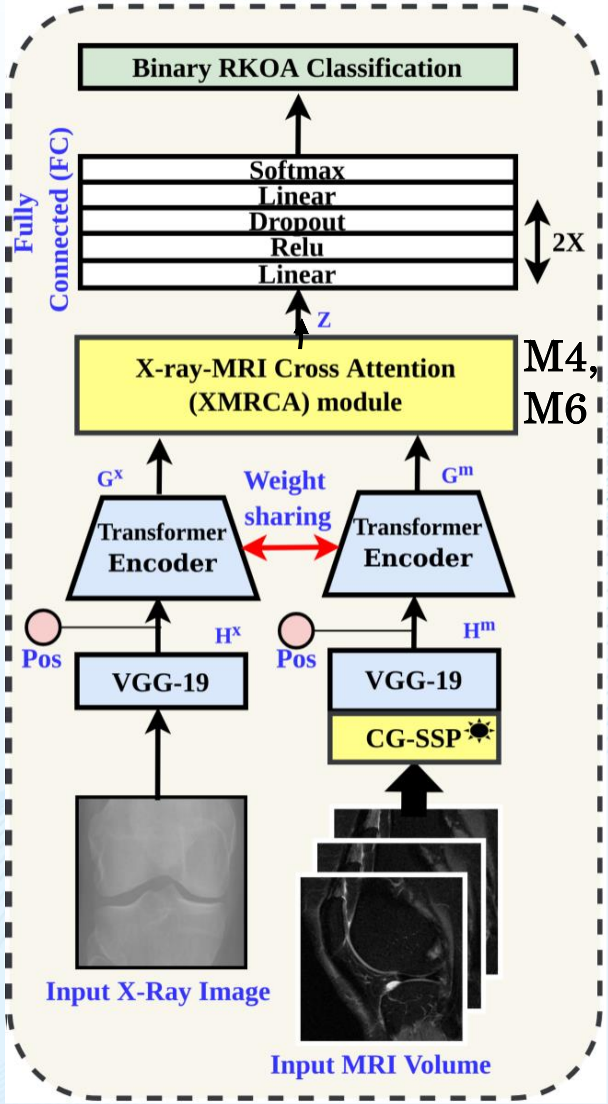
M5: $XMRCA^{H_m=Q}_{H_x=K,V}(H_x, H_m) \rightarrow$ TE, **M6**: $XMRCA^{G_m=Q}_{G_x=K,V}(G_x, G_m) \rightarrow$ FC

Table: Ablation study for XMRCA and CG-SSP modules

M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*	M6	M6*
Accuracy											
86.06	85.74	86.22	85.90	84.60	84.93	85.58	86.87	75.20	79.09	87.03	87.52
QWK											
0.71	0.7	0.71	0.69	0.67	0.72	0.69	0.71	0.47	0.57	0.72	0.73
GPU Load (GB)											
4.81	4.87	4.41	4.41	4.35	4.44	4.35	4.44	4.27	4.42	4.27	4.42

The * symbol beside the model name indicates the use of the CG-SSP module in that model.

MedCAM-OsteoCls Model



Ablation Study

Paper Title: MedCAM-OsteoCls: Medical Context Aware Multimodal Classification of Knee Osteoarthritis

2. Effect of CG-SSP module on MRNet and SKM-TEA datasets

MRnet: Meniscal Tear, and SKM-TEA: Femoral Cartilage Defects

Table: Ablation study for CG-SSP modules on diverse knee datasets

Sr.No.	Architecture	Acc. (%) ↑	Pre. ↑	Recall ↑	MAE ↓
MRNet Dataset (Bien et al., 2018)[Train:1018, Val:112, Test :120]					
Meniscal Tear					
1	M^m	61.67	0.61	0.62	0.38
2	M^m+CG-SSP	65.20	0.58	0.76	0.34
SKM-TEA Dataset (Desai et al., 2022)[Train:68, Val:27, Test :29]					
Cartilage Defect					
3	M^m	47.93	0.52	0.52	0.62
4	M^m+CG-SSP	58.62	0.59	0.58	0.41

- M^m = MedCAM-OsteoCls^m model (input modality = MRI) without CG-SSP module. MAE = Mean Absolute Error, Pre. = Precision.

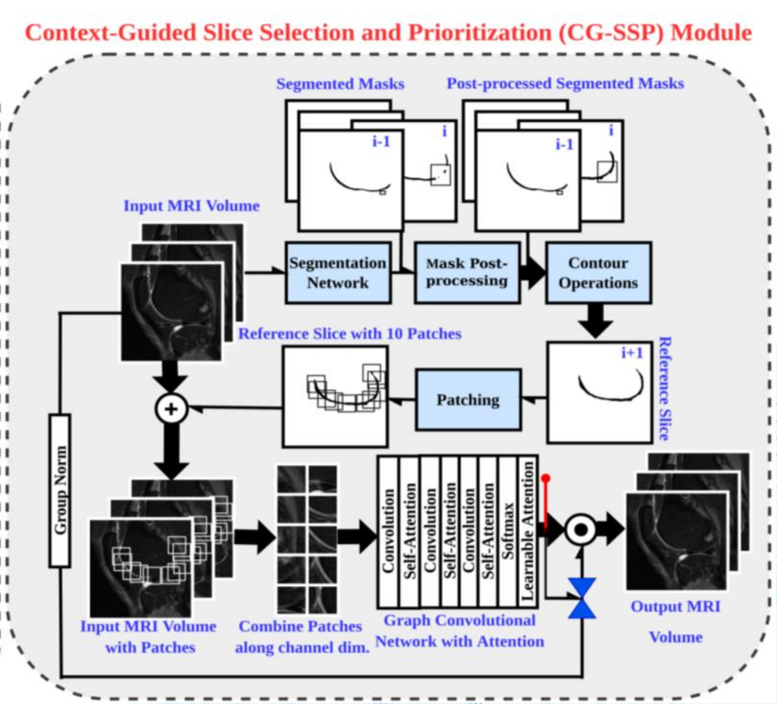


Figure: CG-SSP module

Key Findings

1. The proposed MedCAM-OsteoCls model is **computationally efficient** in terms of GPU utilization by **64.64%** compared to SOTA.
2. The proposed CG-SSP module is **first of its kind** in study of knee osteoarthritis (KOA) in locating **key diagnostic slices by learnable approach**.
3. The excellent performance is achieved by cross-attending critical **MRI slice features as query vector (Q)** with X-ray features serving as the Key (K) and Value (V) at an **advanced feature extraction stage.**
4. Integrating critical MRI slice features with X-ray features **enhanced recall,** thus reducing the **risk of missing treatment for RKOA subjects.**