

Multi-Task Framework for Efficient Knee Osteoarthritis Segmentation and Classification

Yun Xin Teoh

Sunway University

Marouane TLIBA

`tlibamaroine3@gmail.com`

University of Orléans

Mohamed Berrimi

University of Orléans

Khin Wee Lai

University of Malaya

Siew Li Goh

University of Malaya

Rachid Jennane

University of Orléans

Research Article

Keywords: Multitask Learning, 3D CNN, Knee Osteoarthritis, MRI

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8330779/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Multi-Task Framework for Efficient Knee Osteoarthritis Segmentation and Classification

Yun Xin Teoh¹, Marouane TLIBA², Mohamed Berrimi², Khin Wee Lai³,
Siew Li Goh^{3,4}, Rachid Jennane²

¹Sunway University, Kuala Lumpur, Malaysia .

²IDP, Université d’Orleans, Orleans, France.

³Department of Biomedical Engineering, Universiti Malaya, Kuala Lumpur , Malaysia.

⁴Faculty of Medicine , Universiti Malaya, Kuala Lumpur , Malaysia.

Contributing authors: yunxint@sunway.edu.my; tlibamaroine3@gmail.com;
mohamed.berrimi@univ-orleans.fr; lai.khinwee@um.edu.my; gsiewli@um.edu.my;
rachid.jennane@univ-orleans.fr;

Abstract

Knee osteoarthritis (KOA) is a leading cause of chronic disability worldwide, and magnetic resonance imaging (MRI) remains the gold standard for assessing cartilage and meniscal integrity. While 3D convolutional neural networks (CNNs) can leverage volumetric context for KOA analysis, they often incur high computational cost, increased memory demands, and susceptibility to learning from anatomically irrelevant regions. In this work, we propose a unified and flexible multi-task learning framework that accommodates both 2D slice-based and 3D volume-based convolutional backbones for joint KOA classification and knee structure segmentation. The architecture integrates a shared encoder configurable as either 2D or 3D followed by two task-specific branches: a dual-headed segmentation decoder for meniscus and cartilage masks, and a classification head operating on globally pooled features for KOA detection. We train and validate the framework on a large-scale dataset of 105,919 MRI slices, including expert-annotated segmentation masks and radiographically verified KOA labels. The multi-task design enables concurrent learning of anatomical and diagnostic signals at either the slice or volume level, improving task synergy and anatomical focus. Comparative evaluations demonstrate that the 2D variant achieves more consistent segmentation quality, faster convergence, and lower computational overhead than a 3D VGG-UNet baseline while also delivering improved KOA classification accuracy.

Keywords: Multitask Learning, 3D CNN, Knee Osteoarthritis, MRI

1 Introduction

Knee OsteoArthritis (KOA) is a prevalent and progressively debilitating musculoskeletal disease characterized by the deterioration of articular

cartilage and associated joint structures. It manifests clinically through chronic pain, stiffness, joint effusion, and loss of mobility [1], significantly impairing quality of life for millions worldwide. With aging populations and rising obesity rates, the global burden of KOA is expected to grow

substantially in the coming decades [2]. Early and accurate diagnosis is thus essential to guide intervention strategies that can delay progression and reduce disability.

Traditionally, KOA diagnosis has relied on clinical examination in combination with radiographic imaging, particularly X-rays assessed using the Kellgren-Lawrence (KL) grading scale [3]. However, X-rays primarily depict osseous structures and are limited in their ability to detect early degenerative changes in cartilage and soft tissues. In contrast, Magnetic Resonance Imaging (MRI) provides high-resolution visualization of joint components including cartilage, menisci, bone marrow, and ligaments, enabling more sensitive detection of early osteoarthritic features [4]. With the increasing availability of large-scale MRI datasets, deep learning has emerged as a powerful tool for automating KOA diagnosis from imaging data [5].

A substantial portion of recent literature has focused on deep learning methods using 3D convolutional neural networks (CNNs) applied to full MRI volumes. These models exploit the spatial continuity of the knee anatomy and can learn context-aware representations across adjacent slices [6]. While powerful in capturing volumetric features, 3D CNNs introduce several practical challenges: they are memory-intensive, require significant GPU resources and training time, and may inadvertently incorporate irrelevant anatomical context. Additionally, 3D annotation remains a bottleneck, as manual voxel-wise labeling is time-consuming and expensive, limiting the scalability of such approaches in real-world clinical settings [7].

To mitigate these constraints, 2D CNNs operating on individual MRI slices have gained traction as a computationally efficient alternative. These models require fewer resources, offer greater flexibility for data augmentation, and facilitate training on larger datasets. Moreover, with appropriately localized supervision, 2D models can achieve competitive or even superior performance compared to their 3D counterparts. However, the principal drawback of 2D models is their limited ability to capture inter-slice dependencies and global context. Multi-Task Learning (MTL) provides a promising framework to enhance representation learning in this context [8, 9]. By

jointly learning auxiliary tasks such as anatomical segmentation alongside primary classification objectives, MTL can guide the network toward clinically relevant structures and improve generalization [10]. In KOA assessment, segmenting cartilage and menisci key anatomical biomarkers can serve as an effective inductive bias that reinforces the diagnostic objective.

In this study, we propose a unified multitask learning framework that supports both 2D slice-level and 3D volume-level MRI analysis for KOA assessment. Our architecture comprises a shared encoder configurable as either a 2D or 3D backbone followed by two task-specific decoders: one for binary segmentation of cartilage and meniscus, and another for global KOA classification. This flexible design enables systematic comparison between 2D and 3D variants under consistent experimental conditions.

We conduct extensive experiments on a large MRI dataset of over 1,200 annotated knee volumes (approximately 120,000 DESS slices), with expert verified segmentation masks and KOA labels. Our evaluation highlights the trade-offs between 2D and 3D models in terms of classification accuracy, segmentation quality, training efficiency, and memory consumption. Notably, we demonstrate that our 2D variant achieves superior classification performance and segmentation consistency while requiring significantly fewer computational resources compared to a 3D VGG-UNet baseline.

The contributions of this paper are as follows:

- We introduce a modular multitask framework for KOA assessment that accommodates both 2D and 3D convolutional backbones, enabling flexible adaptation to various clinical and computational constraints.
- We provide a comprehensive empirical comparison between 2D and 3D models for joint KOA classification, analyzing performance, and efficiency.
- We show that slice-based 2D multitask models can outperform 3D volumetric approaches in KOA classification while offering better efficiency and robustness.
- We release our implementation and benchmarking pipeline to promote reproducibility and

facilitate future research in automated KOA diagnosis using multitask deep learning.

The remainder of this paper is organized as follows. Section 2 presents the proposed multitask framework, including problem formulation, architectural design, and training strategy. Section 3 details the experimental setup and evaluates the performance of 2D and 3D models across multiple configurations. Section 4 discusses the implications of our findings, and Section 5 concludes the paper with future research directions.

2 Proposed Method

This section presents our unified MTL framework for KOA assessment from MRI data. We first formalize the problem and describe the overall model architecture, which supports both 2D and 3D convolutional backbones. We then detail the MTL setup, including the classification and segmentation branches, and explain the training strategy and loss formulation.

2.1 Problem Formulation

Let $\mathcal{X} = \{(x_i, y_i, s_i)\}_{i=1}^N$ denote a dataset of N knee MRI samples, where each $x_i \in \mathbb{R}^{D \times H \times W}$ is a volumetric scan composed of D sagittal slices of spatial dimensions $H \times W$. The corresponding label $y_i \in \{0, 1\}$ indicates the presence ($y_i = 1$) or absence ($y_i = 0$) of KOA, while $s_i \in \{0, 1\}^{D \times H \times W}$ is a binary segmentation mask delineating key anatomical structures such as the cartilage and menisci.

The learning objective is to estimate a function $f_\theta : \mathbb{R}^{D \times H \times W} \rightarrow \{0, 1\} \times \{0, 1\}^{D \times H \times W}$, parameterized by $\theta \in \Theta$, that simultaneously predicts a classification output $\hat{y}_i = f_\theta^{\text{cls}}(x_i)$ and a segmentation output $\hat{s}_i = f_\theta^{\text{seg}}(x_i)$. This constitutes a MTL problem over two output spaces:

$$f_\theta(x_i) = (f_\theta^{\text{cls}}(x_i), f_\theta^{\text{seg}}(x_i)).$$

To train f_θ , we minimize a multitask loss over the dataset:

$$\mathcal{L}_{\text{total}}(\theta) = \frac{1}{N} \sum_{i=1}^N \left[\mathcal{L}_{\text{cls}}(f_\theta^{\text{cls}}(x_i), y_i) + \alpha \cdot \mathcal{L}_{\text{seg}}(f_\theta^{\text{seg}}(x_i), s_i) \right]$$

where $\mathcal{L}_{\text{cls}}$ is the classification loss, $\mathcal{L}_{\text{seg}}$ is the segmentation loss, and $\alpha \in \mathbb{R}^+$ is a weighting coefficient that balances the two tasks.

To accommodate varying computational constraints and clinical setups, we consider two modes of operation:

- **Volume-based (3D) mode:** The input x_i is processed as a 3D volume, and both predictions are derived from global features learned across the entire spatial context.
- **Slice-based (2D) mode:** The volume x_i is decomposed into slices $\{x_i^{(d)}\}_{d=1}^D$, with $x_i^{(d)} \in \mathbb{R}^{H \times W}$. The network f_θ processes each slice independently, and outputs are aggregated across d to produce volume-level predictions:

$$\hat{y}_i = \text{Agg}_d \left(f_\theta^{\text{cls}}(x_i^{(d)}) \right), \quad \hat{s}_i^{(d)} = f_\theta^{\text{seg}}(x_i^{(d)}),$$

$$\hat{s}_i = \bigcup_{d=1}^D \hat{s}_i^{(d)}.$$

This general formulation provides the basis for our proposed multitask framework, which is designed to flexibly integrate both 2D and 3D convolutional backbones under a unified loss structure. It enables joint anatomical and diagnostic learning from MRI while adapting to clinical and computational constraints.

2.2 Proposed Architecture

To address the dual objectives of KOA classification and anatomical segmentation from MRI, we propose a unified multitask architecture that flexibly supports both 2D slice-based and 3D volume-based convolutional backbones. The model is composed of three core components: a shared encoder for feature extraction, a classification head for KOA classification, and a segmentation decoder for anatomical delineation. This design enables the joint learning of complementary tasks, allowing structural priors from segmentation to inform classification, and vice versa. Figure 1 provides schematic overviews of the full multitask variants.

2.2.1 Shared Encoder

At the heart of the architecture lies a shared convolutional encoder responsible for extracting hierarchical visual representations from the input

MRI scans. The encoder is modularly designed to support both 2D and 3D variants, allowing the framework to adapt to different computational constraints and data representations.

In the 2D configuration, each sagittal slice is treated independently as a two-dimensional input image, processed through a deep stack of convolutional layers. We adopt a modified 2D VGG-style architecture, which includes multiple convolutional blocks interleaved with batch normalization, ReLU activations, and spatial max pooling. Each slice yields a feature map of reduced spatial resolution but enriched semantic content. To enable volumetric reasoning despite slice-wise processing, we aggregate per-slice embeddings via global average pooling. This approach ensures that the model attends to whole regions.

In contrast, the 3D encoder processes the full MRI volume as a unified input tensor. We implement a 3D VGG-UNet encoder consisting of volumetric convolutions that span across depth, height, and width dimensions. The 3D architecture captures spatial continuity and inter-slice anatomical coherence, allowing the model to learn contextualized volumetric features. Down-sampling is performed via 3D max pooling, and each block is followed by non-linear activation and normalization. While computationally more expensive, this configuration enables richer feature modeling of knee anatomy across all three axes.

2.2.2 Classification Branch

The classification head operates on the encoder’s deepest-level features to predict KOA presence. For 3D inputs, global average pooling is applied across all spatial dimensions to produce a compact feature vector, which is passed through one or more fully connected layers, culminating in a sigmoid activation for binary classification. For 2D models, slice-level features are first pooled and then aggregated across slices using max pooling. This strategy allows the 2D framework to approximate volumetric inference without incurring full 3D computation.

2.2.3 Segmentation Decoder

To guide the encoder toward learning anatomically grounded features and to support structural analysis, we incorporate a segmentation decoder parallel to the classification branch. The decoder

follows a U-Net-style design, consisting of successive upsampling layers interleaved with convolutional refinement blocks. Skip connections from the encoder layers are used to fuse low-level spatial detail with high-level semantic information, a crucial property for accurate delineation of cartilage and meniscal structures.

For 3D inputs, the decoder employs transposed 3D convolutions to progressively restore the spatial dimensions of the input volume. A final $1 \times 1 \times 1$ convolution with sigmoid activation produces the predicted segmentation mask $\hat{s}_i \in \{0, 1\}^{D \times H \times W}$. In the 2D variant, each slice is decoded independently using transposed 2D convolutions and a 1×1 output layer, with predictions stacked across the depth dimension to reconstruct the full volume mask.

Figure 1 shows the complete multitask architecture, integrating both classification and segmentation branches. The encoder is shared, allowing the two heads to regularize each other: the segmentation task encourages model feedback from fine-grained anatomical boundaries, while the classification task imposes global disease-level supervision.

2.3 Multitask Learning and Training Strategy

MTL offers a principled framework for simultaneously optimizing multiple related objectives. In our case, the primary task is binary classification of KOA presence, while the auxiliary task is semantic segmentation of key anatomical structures, specifically the femoral cartilage, tibial cartilage, and menisci. These tasks are synergistic: accurate segmentation of structurally relevant regions provides strong spatial priors for KOA detection, and diagnostic classification enforces global consistency that regularizes the segmentation process.

Our architecture is explicitly designed to support this multitask setting via a shared encoder, from which both the classification and segmentation branches derive their inputs. During training, the network learns a joint representation that captures both anatomical structure and disease-specific patterns.

Let $\{x_i, y_i, s_i\}_{i=1}^N$ be the training set, where x_i denotes an MRI volume or a set of slices, $y_i \in \{0, 1\}$ is the binary KOA label, and $s_i \in$

$\{0, 1\}^{D \times H \times W}$ is the ground-truth segmentation mask. The network outputs a classification score $\hat{y}_i \in [0, 1]$ and a predicted segmentation map $\hat{s}_i$. The training objective is defined as a weighted sum of task-specific loss functions:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}}(\hat{y}_i, y_i) + \alpha \cdot \mathcal{L}_{\text{seg}}(\hat{s}_i, s_i),$$

where $\alpha > 0$ is a scalar hyperparameter that controls the relative importance of the segmentation task.

For the classification loss $\mathcal{L}_{\text{cls}}$, we adopt the binary cross-entropy loss:

$$\mathcal{L}_{\text{cls}} = -y_i \log \hat{y}_i - (1 - y_i) \log(1 - \hat{y}_i),$$

which encourages the model to produce high confidence predictions for KOA presence or absence.

For the segmentation loss $\mathcal{L}_{\text{seg}}$, we employ the Dice loss, which is well-suited for class-imbalanced medical masks:

$$\mathcal{L}_{\text{seg}} = 1 - \frac{2 \sum_j \hat{s}_{ij} s_{ij} + \epsilon}{\sum_j \hat{s}_{ij} + \sum_j s_{ij} + \epsilon},$$

where j indexes over all voxels (or pixels in the 2D case), and ϵ is a small constant to ensure numerical stability.

In the 2D variant of the model, segmentation and classification losses are computed per slice and aggregated across the volume. For classification, we apply average or attention-based pooling over slice-level logits to produce a volume-level prediction. For segmentation, slice-wise Dice losses are averaged across the depth dimension. This formulation enables training on individual slices, which reduces memory consumption and increases training efficiency.

In contrast, the 3D variant computes both losses over the entire volume at once, leveraging spatial continuity across slices. While this approach offers richer spatial reasoning, it requires substantially more GPU memory and is more sensitive to inter-volume anatomical variability. We adopt an end-to-end training strategy using the Adam optimizer with a fixed learning rate schedule. Early stopping and validation set monitoring are employed to mitigate overfitting, especially in the 3D case. The multitask design not only

improves performance on both tasks but also provides inherent regularization by constraining the model to explain the data through complementary structural and diagnostic signals.

3 Experimental Analysis

In this section, we present the experimental setup and quantitative evaluation of the proposed multitask learning framework. We describe the dataset, implementation details, and performance metrics, followed by a series of experiments comparing different input representations, convolutional architectures (2D vs. 3D), and task configurations (single-task vs. multitask). Results are analyzed to assess classification accuracy, efficiency, and the benefits of joint learning.

3.1 Dataset

This study leveraged the Osteoarthritis Initiative (OAI) database [11], a publicly available resource comprising multimodal data from 4,796 patients. The OAI dataset’s comprehensiveness makes it particularly valuable for investigating KOA, enabling a nuanced understanding of the condition and associated injuries.

We specifically focused on 3D DESS MRI data from the OAI, utilizing femoral and tibial bone and cartilage segmentations provided by the Zuse Institute Berlin (ZIB) [12]. While the OAI provides volumetric 3D MRI data, this study focuses on 2D analysis. To this end, we extracted 2D slices from the 3D volumes and aligned each slice with its corresponding mask from the ZIB dataset. Empty slices, those lacking knee joint information and corresponding to fully black masks, were removed.

This process yielded a final dataset of 105,919 MRI slices and corresponding masks, totaling 211,838 knee images. The dataset exhibits a balanced gender distribution and comprises normal cases (49.5%) and KOA cases (50.5%) classified using the MOAKS grading system [13].

3.2 Experiments

To evaluate the effectiveness of our proposed multitask framework and determine the optimal data representation for KOA classification, we conducted two sequential experiments. The first

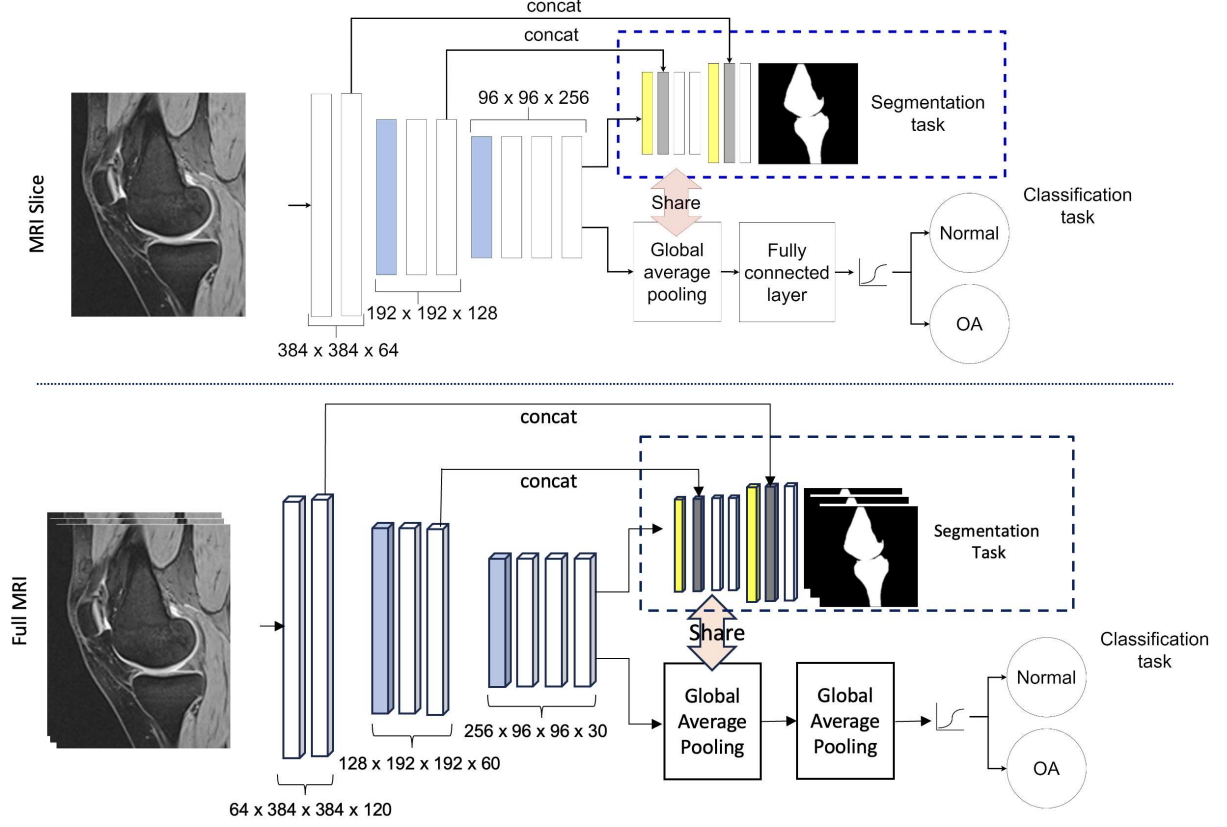


Fig. 1 Architecture of the proposed Multi-task-OA model integrating anatomical mask segmentation as an auxiliary task. The top panel illustrates the 2D slice-based variant, while the bottom panel illustrates the 3D volumetric approach.

focused on identifying the most informative input representation, while the second compared the classification performance of single-task and multitask models using the selected input.

3.2.1 Experiment 1: Evaluation of Input Representations

The goal of the first experiment was to assess how different input formats influence KOA classification performance, and find out the optimal input representation. We compared three types of inputs: (a) MRI, (b) binary MRI segmentation masks of anatomical structures, and (c) segmented MRI obtained by masking raw inputs using cartilage and meniscus annotations. A baseline single-task classification model (Single-task-OA) was trained separately on each input type while maintaining consistent architecture and training settings across experiments.

As illustrated in Figure 2, these modalities offer varying levels of anatomical abstraction. MRI alone contain rich structural detail but also include irrelevant background regions. Binary masks focus solely on the spatial extent of annotated structures, while segmented MRI preserve both texture and spatial localization. Performance comparisons were based on classification accuracy and area under the ROC curve (AUC) on a held-out test set.

3.2.2 Experiment 2: KOA Classification

The second experiment was designed to evaluate the effectiveness of KOA classification across both single-task and multitask learning setups, using consistent 2D and 3D convolutional architectures. Each model was trained to predict KOA presence from **MRI data** using the optimal input representation identified in Experiment 1.

To isolate the impact of multitask supervision and input dimensionality, we implemented both the Single-task-OA and Multi-task-OA models with 2D and 3D convolutional backbones. In total, four model variants were trained in this study: a **2D single-task** model performing slice-wise classification, a **3D single-task** model operating on full MRI volumes, a **2D multitask** model that jointly learns slice-wise classification and segmentation, and a **3D multitask** model that performs joint learning on full volumes.

All models shared a consistent encoder-decoder architecture, differing only in the spatial dimensionality of their convolutional operations. The number of layers, down-sampling, and placement of batch normalization and activation functions were matched between 2D and 3D versions to ensure architectural parity. Table 1 details model parameters and memory requirements for each case. In the multi-task setup, the classification head was trained jointly with a segmentation decoder for anatomical structures. In contrast, the single-task models were only optimised for KOA classification.

Training procedures—including data preprocessing, learning rate, optimizer, regularization parameters, and early stopping were held constant across all variants. This controlled design ensures that observed differences in performance can be attributed to the task formulation (single vs. multitask) and input dimensionality (2D vs. 3D), rather than extraneous implementation differences.

3.3 Results Analysis

3.3.1 Impact of Input Representation

Table 3 presents the classification performance of 2D and 3D convolutional models trained on three types of input representations: raw MRI data, binary mask images, and segmented slices. The results show that raw MRI consistently produced the highest classification accuracy for both architectures, with the 2D model reaching 90.84% and the 3D model 70.90%. These results highlight the importance of preserving full anatomical and intensity information, as raw slices provide rich visual cues essential for reliable KOA detection.

In contrast, binary mask inputs, which provide only coarse spatial information, resulted in

the lowest performance, indicating that critical diagnostic features are lost when raw intensity patterns are discarded. Segmented MRI slices, which retain partial structural context, performed slightly better than masks but remained inferior to raw inputs. Across all input types, the 2D model consistently outperformed the 3D model, demonstrating the efficiency of 2D architectures in extracting discriminative features under limited training data and computational resources.

3.3.2 Comparison of Learning Paradigms and Architectures

Table 2 presents a broader comparison across single-task and multitask learning paradigms, dimensionalities (2D vs. 3D), and published baselines. Our proposed 2D multitask model achieved a classification accuracy of 96.70%, outperforming all other variants, including the 3D multitask model (86.57%) and the two single-task models (90.84% for 2D and 70.90% for 3D). These results demonstrate that multitask learning, when paired with 2D convolutional backbones, can significantly enhance performance by leveraging auxiliary anatomical segmentation as a strong inductive bias. Furthermore, our 2D multitask model

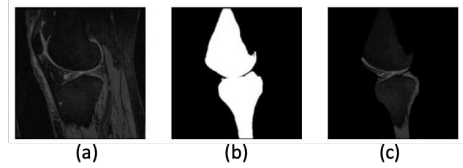


Fig. 2 Different input representations evaluated in Experiment 1: (a) raw MRI slice, (b) MRI binary segmentation mask, and (c) segmented MRI slice (masked anatomical regions)

is more parameter-efficient, with a model size of only 9.56 MB, compared to 64.30 MB for OA-MTL and 139.95 MB for RES-MTL [3]. While prior 3D methods such as 3D-ResCNN-GAP [7] and 3D CNN [14] demonstrate competitive performance (93.20% and 83.00%, respectively), they lack multitask supervision and incur substantially higher computational demands.

The comparison reveals three key insights: (i) multitask learning consistently outperforms single-task setups in both 2D and 3D configurations, (ii) 2D architectures are not only more

Table 1 Model variants with corresponding parameter counts and memory requirements. *Note:* Classification refers to the single-task setup, while Segmentation + Classification corresponds to the multitask setup.

Model	Total parameters	Memory size (MB)
Single-task-OA-2D	1,768,513	6.75
Multi-task-OA-2D	2,506,242	9.56
Single-task-OA-3D	3,513,634	13.40
Multi-task-OA-3D	5,600,227	21.36

Table 2 Comparison of KOA classification accuracy (Acc., %) across our models and baselines. Keys: **Inputs** – I1 = 2D MRI slices, I2 = 3D MRI volumes. **Learning setup** – L1 = Single-task, L2 = Multitask. Baselines: 3D CNN [14], 3D-ResCNN-GAP [7], MMRI-SE-ResNet [15], OA-MTL [3], RES-MTL [3]. Model size is reported in MB.

Model	Size (MB)	Input (I)	Setup (L)	Acc. (%)
Single-task-OA (2D)	6.75	I1	L1	90.84
Multi-task-OA (ours 2D)	9.56	I1	L2	96.70
Single-task-OA (ours 3D)	13.40	I2	L1	70.90
Multi-task-OA (ours 3D)	21.36	I2	L2	86.57
3D CNN	–	I2	L1	83.00
3D-ResCNN-GAP	–	I2	L1	93.20
MMRI-SE-ResNet	–	I1	L1	78.39
OA-MTL	64.30	I2	L2	82.50
RES-MTL	139.95	I2	L2	77.50

accurate but also more memory-efficient, and (iii) segmentation as an auxiliary task helps regularize feature learning and improves diagnostic precision. These results affirm the value of multitask design and the practicality of slice-based 2D modeling for scalable, high-performance KOA assessment from MRI data.

4 Discussion

This study presents a comparative analysis of 2D and 3D convolutional architectures within a unified multitask learning framework for KOA assessment from MRI. While 3D models are theoretically better suited for volumetric data due to their ability to capture inter-slice continuity and anatomical context, our results consistently show that 2D multitask models outperform their 3D counterparts in both classification accuracy and computational efficiency.

4.1 Why 2D Outperformed 3D in Practice

The underperformance of 3D models, despite their theoretical advantages, can be attributed to several practical challenges. Firstly, 3D CNNs contain significantly more parameters than their 2D

counterparts, making them prone to overfitting, particularly when there is limited data. Although our dataset comprises over 105,919 MRI slices, the effective number of unique 3D samples after pre-processing remains comparatively small due to the high dimensionality and redundancy across slices. In contrast, 2D models can leverage a much larger training set at the slice level, promoting better generalization. Secondly, 3D models process entire volumes, increasing the likelihood of incorporating irrelevant or noisy anatomical regions. Without sophisticated attention mechanisms or the ability to localise regions of interest, this broader field of view may dilute the essential discriminative signals for KOA classification. Conversely, 2D models trained with multitask supervision implicitly learn to focus on critical structures—such as cartilage and menisci—by jointly optimizing segmentation and classification objectives.

Finally, the high computational and memory demands of 3D models limit batch size and training iterations, constraining their optimization dynamics. This bottleneck becomes particularly problematic in multitask setups where joint loss functions require stable gradients and coordinated learning across tasks.

4.2 Utility of Multitask Learning

Our findings also emphasize the utility of MTL in MRI-based KOA diagnosis. By incorporating anatomical segmentation as an auxiliary task, MTL not only improves classification performance but also enhances model interpretability and robustness. The segmentation task serves as an inductive bias, guiding the shared encoder to prioritize anatomically relevant regions, thus mitigating the impact of background noise and inter-patient variability. This is particularly useful for slice-based 2D models, where spatial context is limited and task synergy compensates for the lack of 3D continuity.

4.3 Clinical and Computational Perspectives

From a clinical standpoint, the proposed 2D multitask framework offers a compelling balance between accuracy, interpretability, and resource efficiency. Its lightweight architecture facilitates real-time deployment in clinical workflows, particularly in low-resource settings or where 3D pipelines are impractical. Moreover, the dual-task setup offers richer outputs—diagnostic classification and anatomical segmentation—making it more informative for clinicians and more adaptable to downstream tasks such as prognosis or treatment planning. From a computational perspective, the results advocate for a paradigm shift: rather than defaulting to volumetric models for volumetric data, future systems should consider hybrid or slice-wise strategies that optimize performance and scalability. Leveraging the strengths of 2D models, particularly when enhanced through multitask learning, appears to be a pragmatic and effective alternative.

Table 3 Classification accuracy (%) of 2D and 3D models using different input representations.

Input	2D model (MRI slices)	3D model (full volumes)
MRI (raw)	90.84	70.90
MRI binary mask	76.60	54.48
Segmented MRI	81.61	50.00

4.4 Future Directions

While our study demonstrates the advantages of 2D multitask architectures, several avenues remain open for exploration. Future work could integrate temporal modeling across slices using sequence models (recent attention mechanism or state-space models) to recover some of the contextual continuity lost in 2D setups. Additionally, anatomical priors could be incorporated into 3D models to mitigate noise and improve interpretability. Another promising direction is semi-supervised or self-supervised multitask learning, which can alleviate the need for large-scale manual annotations and further enhance generalization.

5 Conclusion

We presented a unified multitask learning framework for KOA assessment from knee MRI, supporting both 2D and 3D convolutional architectures. The model jointly performs KOA classification and anatomical segmentation, encouraging anatomically guided representation learning. Our results show that 2D multitask models outperform their 3D counterparts in both accuracy and efficiency, despite the volumetric nature of MRI. This highlights the benefits of slice-wise modeling with shared supervision. The framework offers a practical and interpretable solution for automated KOA diagnosis and is well-suited for clinical deployment.

6 Declarations

Funding: This work was supported by the France–Malaysia Collaboration Program (MyTIGER 2023, Grant SCAC-MyT-2023-08) and the French National Research Agency (ANR-20-CE45-0013-01).

7 Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported.

8 Data Availability

This study uses publicly available data from the Osteoarthritis Initiative (<https://nda.nih.gov/oai>) and the ZIB knee segmentation dataset (<https://doi.org/10.5281/zenodo.1327317>). Processed

data generated during this study are available from the corresponding author upon reasonable request.

References

- [1] Törnblom, M. *et al.* Development of radiographic knee osteoarthritis and the associations to radiographic changes and baseline variables in individuals with knee pain: a 2-year longitudinal study. *BMJ Open* **14**, e081999 (2024).
- [2] Kang, Y. *et al.* The burden of knee osteoarthritis worldwide, regionally, and nationally from 1990 to 2019, along with an analysis of cross-national inequalities. *Archives of Orthopaedic and Trauma Surgery* 1–13 (2024).
- [3] Yeoh, P. S. Q., Goh, S. L., Hasikin, K., Wu, X. & Lai, K. W. 3d efficient multi-task neural network for knee osteoarthritis diagnosis using mri scans: Data from the osteoarthritis initiative. *IEEE Access* **11**, 135323–135333 (2023).
- [4] Hu, J. *et al.* Deepkoa: a deep-learning model for predicting progression in knee osteoarthritis using multimodal magnetic resonance images from the osteoarthritis initiative. *Quantitative Imaging in Medicine and Surgery* **13**, 4852 (2023).
- [5] Daneshmand, M., Panfilov, E., Bayramoglu, N., Korhonen, R. K. & Saarakkala, S. Deep learning based detection of osteophytes in radiographs and magnetic resonance images of the knee using 2d and 3d morphology. *Journal of Orthopaedic Research* (2024).
- [6] Tack, A., Shestakov, A., Ludke, D. & Zachow, S. A multi-task deep learning method for detection of meniscal tears in mri data from the osteoarthritis initiative database. *Frontiers in Bioengineering and Biotechnology* **9** (2021).
- [7] Berrimi, M., Hans, D. & Jennane, R. A semi-supervised multiview-mri network for the detection of knee osteoarthritis. *Computerized Medical Imaging and Graphics* **114**, 102371 (2024).
- [8] Tliba, M., Chetouani, A., Valenzise, G. & Dufaux, F. Representation learning optimization for 3d point cloud quality assessment without reference (2022). In: Proc. IEEE International Conference on Image Processing (ICIP), pp. 3702–3706.
- [9] Tliba, M., Chetouani, A., Valenzise, G. & Dufaux, F. Balancing representation abstractions and local details preservation for 3d point cloud quality assessment (2024). In: Proc. IEEE ICASSP, pp. 4335–4339.
- [10] Zhao, Y., Wang, X., Che, T., Bao, G. & Li, S. Multi-task deep learning for medical image computing and analysis: A review. *Computers in Biology and Medicine* **153**, 106496 (2023).
- [11] Peterfy, C. G., Schneider, E. & Nevitt, M. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthritis and Cartilage* **16**, 1433–1441 (2008).
- [12] Ambellan, F., Tack, A., Ehlke, M. & Zachow, S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative. *Medical Image Analysis* **52**, 109–118 (2019).
- [13] Hunter, D. J., Guermazi, A., Lo, G. H. *et al.* Evolution of semi-quantitative whole joint assessment of knee oa: Moaks (mri osteoarthritis knee score). *Osteoarthritis and Cartilage* **19**, 990–1002 (2011).
- [14] Guida, C., Zhang, M. & Shan, J. Knee osteoarthritis classification using 3d cnn and mri. *Applied Sciences* **11**, 5196 (2021).
- [15] Manfredonia, D., Harata, S., Sakuma, T., Trovò, F. & Kato, S. Prediction of kellgren-lawrence grade of knee osteoarthritis by deep residual networks using mr image with segmented image and slice position (2024). In: Proc. 16th International Conference on Agents and Artificial Intelligence (ICAART), Vol. 3, pp. 1300–1307.