People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
University of Echahid Hamma Lakhdar - El Oued
Faculty of Exact Sciences
Department of Informatics

ACADEMIC MASTER DISSERTATION

Domain: Mathematics and Informatics
Division: Informatics
Specialty: Artificial Intelligence and Data Science

Entitled:

Advanced Deep Learning and Imaging Techniques for Early Detection and Progression Analysis of Knee Osteoarthritis

Dissertation Submitted in Partial Fulfillment of the Requirements for the Master Degree in Artificial Intelligence and Data Science

Presented by	Kahla Chiraz
Supervisor	Dr. Belila Khaoula
Co-Supervisor	Dr. Meriem HAMOUD
President	Dr. Abbas Messaoud
Examiner	Dr. Maamra Oum El Hana

Academic Year: 2025/2026

Acknowledgements

All praise is due to Allah, the Most Gracious and Most Merciful, for granting me the strength, patience, and ability to complete this work.

I would like to express my sincere gratitude to **Prof. Abdelwahab Mansour**, Dean of the Faculty of Exact Sciences, for providing the academic environment and institutional support that made this research possible.

I would also like to express my gratitude to **Dr. Lala El Arabi**, radiologist, for his valuable medical insights and the clinical information he generously provided throughout this work.

My deepest thanks go to my supervisors, **Dr. Belila Khaoula** and **Dr. Meriem HAMOUD**, whose expertise, dedication, and continuous guidance were the cornerstone of this dissertation. Their commitment to excellence and their tireless efforts in shaping this work are deeply appreciated.

I would also like to extend my heartfelt gratitude to **my beloved family (my mother, my father, and my siblings: Tarek, Akram, Rimasse, Wassim, Mayar, and Chahd)**, whose unwavering support, patience, and encouragement throughout five years of study have been my greatest source of strength. Their sacrifices and belief in me made this journey possible.

I am also profoundly grateful to my dear friends, **Imad Kir** and **Khadidja Dlal**, for their constant support, motivation, and presence during the most challenging moments of this journey. Their friendship is a treasure I deeply value.

Finally, I extend my sincere thanks to all my **friends and colleagues** who offered their help and moral support along the way.

This work is dedicated to all of you.

Kahla Chiraz

El Oued, June 2026

Abstract

KNEE is one of the most complex joint structures in the human body, which is particularly prone to various injuries. Osteoarthritis is a progressive degenerative joint disease and one of the leading causes of disability worldwide, making early and accurate severity assessment crucial for OA progression monitoring, prediction of disease progression, and timely interventions to prevent irreversible damage. Magnetic Resonance Imaging has been widely used in the diagnosis of knee diseases due to its comprehensive imaging effect on bone, cartilage and ligament. However, automated OA analysis from knee MRI remains challenging due to anatomical complexity, inter-patient variability, and the presence of heterogeneous structural changes. In this work, we propose a cascaded deep learning framework integrating hierarchical segmentation and severity classification. Initially, a coarse binary segmentation model was employed to identify and isolate the knee region of interest from volumetric MRI scans, reducing irrelevant background information and improving anatomical localization. These cropped volumes were then used as inputs for a second fine-grained segmentation stage to extract clinically relevant knee anatomical structures. To investigate segmentation efficiency, both 3D U-Net and 3D Swin-UNETR architectures were explored, where the transformer-based 3D Swin-UNETR demonstrated superior performance due to its enhanced ability to capture long-range spatial dependencies and multiscale contextual information. The refined segmented regions were then used as the foundation for downstream classification tasks, including binary OA severity classification and multi-class classification of different osteoarthritis progression stages. Experimental results showed that the cascaded segmentation strategy significantly improved region localization and enhanced classification reliability by focusing on structurally relevant quantitative biomarkers.

Keywords: Knee, Osteoarthritis, MRI, Cascade-segmentation, Severity-classification, Deep learning, Machine learning.

Résumé

L'arthrose du genou est une maladie articulaire dégénérative progressive et l'une des principales causes d'incapacité dans le monde. Une évaluation précoce et précise de sa sévérité est essentielle pour le suivi de la progression de la maladie, la prédiction de son évolution et la mise en œuvre d'interventions thérapeutiques adaptées. L'Imagerie par Résonance Magnétique (IRM) est largement utilisée pour le diagnostic des pathologies du genou grâce à sa capacité à visualiser simultanément les structures osseuses, cartilagineuses et ligamentaires. Cependant, l'analyse automatique de l'arthrose à partir des images IRM reste difficile en raison de la complexité anatomique du genou, de la variabilité interindividuelle et de l'hétérogénéité des modifications structurelles associées à la maladie.

Dans ce travail, nous proposons un cadre d'apprentissage profond en cascade intégrant la segmentation hiérarchique et la classification de la sévérité de l'arthrose. Le cadre proposé démontre l'efficacité de la combinaison d'une segmentation 3D coarse-to-fine et d'une classification guidée par la segmentation pour l'évaluation automatique de la sévérité de l'arthrose du genou à partir des images IRM.

Mots-clés : Genou, Arthrose, IRM, Segmentation en cascade, Classification de la sévérité, Apprentissage profond, Apprentissage automatique.

الملخص

تُعدّ الركبة من أكثر التراكيب المفصلية تعقيدًا في جسم الإنسان، وهي معرّضة بشكل خاص لمختلف الإصابات. ويُعدّ الفُصال العظمي مرضًا مفصليًا تنكسيًا تدريجيًا، وأحد الأسباب الرئيسية للإعاقة على مستوى العالم، مما يجعل التقييم المبكر والدقيق لشدته أمرًا بالغ الأهمية لرصد تطور المرض، والتنبؤ بمساره، والتدخل في الوقت المناسب للحدّ من الأضرار غير القابلة للعلاج. ويُستخدم التصوير بالرنين المغناطيسي على نطاق واسع في تشخيص أمراض الركبة نظرًا لقدرته الشاملة على تصوير العظام والغضاريف والأربطة. غير أن التحليل الآلي للفُصال العظمي من صور الرنين المغناطيسي للركبة لا يزال يمثّل تحديًا بسبب التعقيد التشريحي، والتباين بين المرضى، والتغيرات البنيوية غير المتجانسة.

في هذا العمل، نقترح إطار عمل تعلّم عميق متتاليًا يدمج التجزئة الهرمية وتصنيف الشدة. في البداية، تم توظيف نموذج تجزئة ثنائي خشن لتحديد منطقة الركبة وعزلها من مجلدات الرنين المغناطيسي الحجمية، بهدف تقليل المعلومات الخلفية غير ذات الصلة وتحسين التموضع التشريحي. ثم استُخدمت هذه المجلدات المقتصّة مدخلاتٍ لمرحلة تجزئة دقيقة ثانية لاستخراج التراكيب التشريحية ذات الصلة سريريًا في الركبة. ولفحص كفاءة التجزئة، تم استكشاف بنيتَي U-Net 3D وSwin-UNETR 3D، حيث أظهر نموذج Swin-UNETR المستند إلى المحوّلات أداءً متفوقًا بفضل قدرته المحسّنة على التقاط الاعتماديات المكانية بعيدة المدى والمعلومات السياقية متعددة المقاييس.

ثم استُخدمت المناطق المجزّأة والمُحسَّنة أساسًا لمهام التصنيف اللاحقة، بما فيها التصنيف الثنائي لشدة الفُصال العظمي، والتصنيف متعدد الفئات لمراحل تطور المرض. وقد أظهرت النتائج التجريبية أن استراتيجية التجزئة المتتالية حسّنت بشكل ملحوظ من دقة تحديد المناطق وزادت من موثوقية التصنيف، من خلال التركيز على مؤشرات حيوية كمية ذات صلة تركيبية كحجم الغضروف الهلالي، والتغطية الظنبوبية، وسُماكة الغضروف وانبثاقه، والمميزات الشدية والنسيجية وترددية النطاق والسريرية.

الكلمات المفتاحية: الركبة، الفُصال العظمي، التصوير بالرنين المغناطيسي، التجزئة المتتالية، تصنيف الشدة، التعلم العميق، التعلم الآلي.

TABLE OF CONTENTS

List of Figures 4

List of Tables 6

List of Abbreviations 7

General Introduction 9

I Literature Review on MRI-Based Osteoarthritis Analysis and Research Gaps 14

1.1 Introduction . . . 14

1.2 Osteoarthritis and MRI Biomarkers . . . 14

1.2.1 Cartilage Degeneration in Osteoarthritis . . . 15

1.2.2 Structural MRI Biomarkers . . . 16

1.3 Computer-Aided Diagnosis (CAD) Systems . . . 17

1.4 Image Preprocessing in Medical Imaging . . . 18

1.4.1 Noise in MRI . . . 20

1.4.2 Wavelet-Based Denoising . . . 20

1.4.3 Transformer-Based Image Restoration . . . 22

1.4.4 Intensity Normalization and Gamma Correction . . . 23

1.4.5 Contrast Enhancement using CLAHE . . . 24

1.4.6 Integrated Preprocessing Pipeline . . . 25

1.5 Deep Learning for Medical Image Segmentation . . . 26

1.5.1 Convolutional Neural Networks . . . 26

1.5.2 U-Net Architecture . . . 27

1.5.3 Limitations of CNN-Based Models . . . 28

1.6 Transformer-Based medical image Segmentation Models . . . 29

1.6.1 Vision Transformers . . . 30

1.6.2 Self-Attention Mechanisms in Medical Imaging . . . 30

1.6.3 UNETR . . . 31

1.6.4 Swin-UNETR . . . 32
1.7 Feature Extraction for Disease Prediction . . . 33
1.7.1 Intensity-Based Features . . . 34
1.7.2 Texture-Based Features (GLCM, Haralick, and LBP) . . . 34
1.7.3 Shape and morphological features . . . 35
1.7.4 Deep Feature Embeddings . . . 35
1.8 State of the Art Review and Research Gap Analysis in MRI-Based Osteoarthritis Assessment . . . 37
1.8.1 Research Gap Analysis . . . 39
1.9 Research Problem Formulation and Open Challenges in MRI-Based Osteoarthritis Analysis . . . 40
1.9.1 MRI Complexity and Clinical Interpretation Challenges . . . 40
1.9.2 MRI Image Quality and Anatomical Boundary Challenges . . . 41
1.9.3 Computational and Segmentation Challenges . . . 41
1.9.4 Feature Extraction and Classification Challenges . . . 42
1.9.5 Need for an Integrated Framework . . . 43
1.10 motivation of the proposed framework . . . 43
1.11 Conclusion . . . 44

II Proposed Framework for 3D MRI Preprocessing, Segmentation, and OA Prediction 46

2.1 Introduction . . . 46
2.2 Overview of the Proposed Framework . . . 46
2.3 MRI Dataset Description . . . 49
2.4 Knee Anatomical Complexity and Segmentation Challenges . . . 51
2.5 Two-Stage Transformer-Based Knee MRI Segmentation . . . 52
2.5.1 Stage 1: Binary Knee Localization . . . 54
2.5.2 Stage 2: Fine Multi-Class Anatomical Segmentation . . . 55
2.5.3 Training Strategy . . . 57
2.5.4 Loss Functions . . . 60
2.5.5 Evaluation Metrics . . . 61

2.6 Region-Wise Feature Extraction Strategy . . . 62
2.6.1 Intensity Features . . . 63
2.6.2 Texture Features . . . 64
2.6.3 Shape and Morphological Features . . . 65
2.6.4 Thickness Features . . . 65
2.6.5 Frequency-Based Features . . . 66
2.6.6 Clinical Features . . . 67
2.6.7 Feature Fusion . . . 67
2.7 Osteoarthritis Classification Strategy . . . 69
2.7.1 Binary Classification Formulation . . . 69
2.7.2 Three-Class Classification Formulation . . . 70
2.7.3 Five-Class Classification Formulation . . . 70
2.8 Conclusion . . . 71
III Performance Assessment of the Proposed Framework 74
3.1 Experimental Setup . . . 74
3.1.1 Hardware and Software Environment . . . 74
3.1.2 Binary Segmentation Hyperparameter Configuration . . . 75
3.1.3 Multi-Class Segmentation Hyperparameter Configuration . . . 76
3.1.4 Online Data Augmentation Strategy . . . 78
3.2 Evaluation of Preprocessing process . . . 81
3.2.1 Visual Assessment . . . 81
3.2.2 Impact on Segmentation Performance . . . 82
3.3 Segmentation Results . . . 83
3.3.1 Model Selection Strategy . . . 83
3.3.2 Binary Knee Localization Results . . . 84
3.3.3 ROI Extraction Results . . . 85
3.3.4 Post-Processing Results . . . 86
3.3.5 Multi-Class Segmentation Results . . . 88
3.3.5.1 Qualitative Segmentation Results . . . 90
3.4 Evaluation of Osteoarthritis Severity Classification Performance . . . 91
3.4.1 Dataset Distribution . . . 92

3.4.2 Preliminary Data Augmentation Analysis for Classification . . . 93
3.4.3 Feature Extraction Verification . . . 94
3.4.4 Feature Importance Analysis . . . 95
3.4.5 Binary Classification Results . . . 100
3.4.6 KL2 Detection Results . . . 102
3.4.7 Five-Class Classification Results . . . 104
3.5 Evaluation on Independent Test Set . . . 107
3.5.1 Qualitative Evaluation . . . 107
3.5.2 Quantitative Evaluation . . . 108
3.5.3 Classification Performance Comparison . . . 109
3.5.4 Scientific Contribution Analysis . . . 110
3.6 Streamlit-Based Application . . . 110
3.7 Discussion . . . 112
3.7.1 Interpretation of Findings . . . 112
3.7.2 Clinical Relevance . . . 113
3.7.3 Robustness Analysis . . . 114
3.8 Conclusion . . . 114
General Conclusion and Future Perspectives **117**
Bibliography **120**

LIST OF FIGURES

1.1 Example of noise and low contrast effects in knee MRI images. 20
1.2 Wavelet-based denoising applied to knee MRI images. 21
1.3 Transformer-based image restoration using Restormer. 23
1.4 Gamma correction applied to MRI images for intensity stabilization. 24
1.5 CLAHE-based contrast enhancement applied to knee MRI images. 25
1.6 Basic CNN workflow for MRI medical image analysis showing convolution, pooling, and classification stages reproduced from [1]. 27
1.7 3D U-Net architecture reproduced from the GitHub repository [2, 3]. . . . 28
1.8 Scaled dot-product attention and multi-head attention mechanisms reproduced from [4]. 31
1.9 Swin-UNETR architecture reproduced from the MONAI GitHub repository [5, 6]. 33
2.1 The proposed multi-stage framework for knee MRI analysis and osteoarthritis classification. 48
2.2 Visualization of a 3D knee MRI volume through axial, coronal, and sagittal anatomical planes extracted from the OAIZIB-CM dataset. 50
2.3 Example of multi-planar knee MRI segmentation illustrating major anatomical structures. The red regions represent femoral structures, the blue regions correspond to tibial structures, while the green regions indicate cartilage and meniscal anatomical components. 52
2.4 Example of direct multi-class segmentation failure on full MRI volumes before applying the proposed two-stage framework. The model occasionally confused background regions with anatomical structures, resulting in noisy predictions and unstable segmentation outputs. 53
2.5 Overview of the proposed two-stage segmentation framework for knee MRI analysis. The first stage performs binary knee localization and ROI extraction, while the second stage performs detailed multi-class segmentation of anatomical structures. 54
2.6 Conversion of the original multi-class segmentation mask into a binary mask used for knee localization. 55

2.7 Example of mask showing different anatomical structures in Sagittal, Coronal, and Axial views. 56

2.8 Example of multi-class anatomical segmentation masks showing different anatomical structures in Sagittal, Coronal, and Axial views. 56

2.9 Three-dimensional visualization of the anatomical structures segmented in this study, including the femur, tibia, femoral cartilage, tibial cartilage, medial meniscus, and lateral meniscus. Adapted from [7]. 57

2.10 Example of online data augmentation applied during training. Different spatial and intensity transformations are dynamically generated to improve data diversity and reduce overfitting while preserving anatomical consistency. 59

2.11 Complete framework of the proposed osteoarthritis classification system, including multiclass segmentation, region-wise feature extraction, feature fusion, and classification tasks (binary, three-class, and five-class prediction). 68

3.1 Examples of online data augmentation operations applied during multi-class segmentation training. Spatial transformations include flipping and rotation, whereas intensity-based transformations include intensity scaling, intensity shifting, and contrast adjustment. The augmentations were dynamically generated during training without permanently increasing the dataset size. 80

3.2 Illustration of the proposed MRI preprocessing pipeline. The framework sequentially applies wavelet-based denoising, Restormer restoration, gamma correction, and CLAHE enhancement in order to improve image quality while preserving anatomical structures. 81

3.3 Visual comparison between representative original MRI slices and their corresponding preprocessed versions obtained using the proposed preprocessing pipeline. The enhanced images exhibit improved contrast, reduced noise, and clearer visualization of anatomical structures. 82

3.4 Comparison between Swin-UNETR and 3D U-Net during binary knee localization training. 83

3.5 Training loss and validation Dice score obtained during binary knee localization using the Swin-UNETR architecture. 84

3.6 Examples of automatic ROI extraction using binary masks predicted by the Swin-UNETR localization model. The predicted masks were utilized to generate knee-centered MRI volumes without relying on manual annotations. 86

3.7 Effect of post-processing on binary knee localization results. The refinement stage improved mask quality and reduced segmentation artifacts. . . 87
3.8 Training loss, validation Dice score, and validation IoU obtained during multi-class segmentation training using the Swin-UNETR architecture. . . 88
3.9 Per-class Dice and IoU scores achieved by the proposed Swin-UNETR model at the best validation epoch. 89
3.10 Representative multi-class segmentation results obtained using the proposed Swin-UNETR framework. Ground-truth masks and predicted segmentations are shown in axial, coronal, and sagittal views together with overlay visualizations. 90
3.11 Additional example of multi-class segmentation results obtained on a different MRI examination. Ground-truth masks and predicted segmentations are shown in axial, coronal, and sagittal views together with overlay visualizations. 91
3.12 Distribution of Kellgren–Lawrence grades in the classification dataset after excluding subjects with missing labels. 92
3.13 Representative examples of MRI slices, multi-class anatomical segmentations, and region-wise measurements obtained from subjects belonging to different Kellgren–Lawrence grades. 95
3.14 Top 20 most important handcrafted features selected by the CatBoost classifier for osteoarthritis prediction. 96
3.15 Distribution of representative discriminative features across different osteoarthritis severity groups. 97
3.16 Confusion matrix obtained for the three-class osteoarthritis classification task. 98
3.17 PCA projection of the extracted anatomical features for the three osteoarthritis severity groups. 99
3.18 Confusion matrix obtained for the No OA versus OA classification task. . . 101
3.19 Confusion matrix obtained for the KL0–1 versus KL3–4 classification task. 102
3.20 ROC curve obtained for the KL2 detection task. 103
3.21 Confusion matrix obtained for the KL2 detection task. 104
3.22 Confusion matrix obtained for the five-class osteoarthritis classification task. 106
3.23 Complete inference pipeline for Subject 405 from the independent test set. 108
3.24 Complete inference pipeline for Subject 418 from the independent test set. 108

3.25 Streamlit interface showing patient clinical information, MRI visualization, and multi-class anatomical segmentation results generated by the proposed framework. 111

3.26 Classification results generated by the Streamlit application, including prediction confidence scores and probability distributions for the different classification models. 112

LIST OF TABLES

1.1 Summary of Recent Deep Learning Approaches for Osteoarthritis Analysis 39

2.1 Main training parameters used for segmentation experiments. 58

2.2 Online data augmentation techniques and parameters used during training. 59

2.3 Summary of the extracted feature categories, anatomical regions, and descriptors. 68

2.4 Binary osteoarthritis classification strategy. 70

2.5 Three-class osteoarthritis classification strategy. 70

2.6 Five-class osteoarthritis classification strategy. 71

3.1 Hardware and software environment used in the experiments. 75

3.2 Final hyperparameter configuration used for binary segmentation models. . 75

3.3 Final hyperparameter configuration used for multi-class segmentation. . . . 77

3.4 Online data augmentation strategy used during multi-class segmentation training. 79

3.5 Comparison between Swin-UNETR and 3D U-Net during binary knee localization. 83

3.6 Binary knee localization performance using Swin-UNETR. 85

3.7 Effect of post-processing on binary localization performance. 87

3.8 Overall performance of the multi-class Swin-UNETR segmentation model. 88

3.9 Per-class segmentation performance at the best validation epoch. 89

3.10 Preliminary analysis of augmentation impact on osteoarthritis classification performance. 93

3.11 Classification report for the three-class osteoarthritis prediction task(Accuracy = 0.6919) 98

3.12 Classification report for the No OA vs OA task (Accuracy = 0.7728, AUC = 0.8396).100

3.13 Classification report for the KL0–1 vs KL3–4 task (Accuracy = 0.8384, AUC = 0.9107).100

3.14 Classification report for the KL2 detection task, One-vs-Rest (Accuracy = 0.7076, AUC = 0.690

3.15 Five-fold cross-validation results for the five-class classification task. 105

3.16 Overall performance of the five-class osteoarthritis classification task. . . . 105

3.17 Per-class performance for the five-class classification task. 105

3.18 Binary classification performance on the independent test set. 109
3.19 Comparison with representative studies from the literature. 109

LIST OF ABBREVIATIONS

Abbreviation	**Meaning**
OA	Osteoarthritis
MRI	Magnetic Resonance Imaging
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
DL	Deep Learning
ViT	Vision Transformer
UNETR	U-Net Transformer
ROI	Region of Interest
DWT	Discrete Wavelet Transform
CLAHE	Contrast Limited Adaptive Histogram Equalization
GLCM	Gray Level Co-occurrence Matrix
LBP	Local Binary Pattern
KL	Kellgren–Lawrence
IoU	Intersection over Union
DSC	Dice Similarity Coefficient
AUC	Area Under the Curve
FCN	Fully Convolutional Network
AMP	Automatic Mixed Precision
FFT	Fast Fourier Transform
BMI	Body Mass Index
GPU	Graphics Processing Unit
DESS	Dual-Echo Steady-State
OAI	Osteoarthritis Initiative
PCA	Principal Component Analysis
ML	Machine Learning
U-Net	U-shaped Network
Swin-UNETR	Shifted Window U-Net Transformer
NIfTI	Neuroimaging Informatics Technology Initiative
KL0–KL4	Kellgren–Lawrence Grades 0 to 4
3D	Three-Dimensional

GENERAL INTRODUCTION

Osteoarthritis (OA) is considered one of the most common degenerative joint diseases worldwide and represents an important health problem, particularly among older adults. The disease mainly affects synovial joints, especially the knee joint, and is commonly associated with pain, stiffness, and progressive reduction in mobility. These symptoms may significantly affect daily activities and overall quality of life. Although OA is more frequently observed with aging, additional factors such as obesity, repetitive mechanical stress, joint injuries, and lifestyle-related conditions may also contribute to disease progression [8, 9].

One of the major difficulties related to osteoarthritis is that the disease usually develops progressively over time. In many situations, structural changes inside the joint begin before severe symptoms become clearly noticeable. As degeneration advances, cartilage tissues gradually lose part of their elasticity and mechanical strength, reducing their ability to absorb pressure and support normal movement. Consequently, patients may experience long-term structural damage and movement limitations if the disease is not detected during its early stages [8, 10].

Early detection of osteoarthritis is therefore highly important in clinical practice. Identifying pathological changes during the initial stages of degeneration may help clinicians apply preventive interventions before severe structural damage occurs. In contrast, delayed diagnosis may considerably limit therapeutic options and eventually lead to irreversible joint deterioration or total knee replacement procedures [9].

The knee joint is particularly vulnerable to osteoarthritis because it continuously supports body weight and plays an essential role in movement and stability. Several important anatomical structures may be affected, including cartilage, menisci, ligaments, femoral regions, and tibial regions. Degeneration affecting these structures can alter joint biomechanics and negatively influence movement efficiency and stability [8].

In routine clinical practice, X-ray imaging remains one of the most commonly used techniques for osteoarthritis diagnosis because of its simplicity and accessibility. However, conventional radiographic imaging mainly provides information about bone structures and remains limited when evaluating soft tissues such as cartilage and menisci. This limitation becomes more critical during the early stages of osteoarthritis, where structural abnormalities may still be subtle and difficult to observe using standard radiographic methods [8, 10].

For this reason, Magnetic Resonance Imaging (MRI) has become increasingly important in musculoskeletal imaging and knee osteoarthritis analysis. Unlike conventional radiography, MRI provides detailed visualization of both hard and soft tissues. It allows clinicians to evaluate cartilage condition, meniscal integrity, ligament structures, and bone abnormalities with much higher anatomical detail. Such information is particularly valu-

able for detecting early pathological changes before severe degeneration becomes visible in radiographic images [8, 10, 9].

Despite these advantages, knee MRI analysis remains relatively complex in practice. MRI volumes usually contain large amounts of anatomical information distributed across multiple slices, which can make manual interpretation difficult and time-consuming. In some situations, clinicians may focus on a particular anatomical region while subtle abnormalities in surrounding tissues become more difficult to identify because of the complexity of the MRI data itself. This complexity may increase interpretation time and contribute to inter-observer variability or missed subtle abnormalities during routine clinical assessment [9].

Additional challenges arise from the variability between MRI scans and patient anatomy. Differences related to age, gender, body structure, disease severity, acquisition protocols, and scanner settings may produce considerable variations in image appearance and quality. MRI volumes may also contain noise, contrast inconsistencies, and intensity variations, which can complicate both visual interpretation and automated image analysis systems [11].

MRI data may also provide quantitative imaging biomarkers useful for disease characterization and progression monitoring. Parameters such as cartilage thickness, tissue morphology, texture variations, and structural asymmetry can provide valuable information about the condition of knee anatomical structures [12]. Another important challenge is related to MRI image quality itself. Medical images may contain noise, low contrast, intensity variations, and acquisition-related artifacts. In some situations, noise patterns may appear similar to anatomical structures, which can complicate both manual interpretation and automated segmentation tasks. Since even small structural changes inside cartilage or bone regions may have clinical importance, poor image quality can negatively affect the reliability of diagnosis and analysis [13, 14, 11].

Because of these challenges, artificial intelligence techniques have attracted increasing attention in medical imaging research. In recent years, computer-aided diagnosis (CAD) systems based on deep learning approaches have shown promising performance in several medical applications. These systems are capable of processing large imaging datasets and extracting meaningful patterns that may not always be easily observable during manual examination [15, 9].

Deep learning models are now widely investigated for tasks such as medical image segmentation, feature extraction, and disease classification. In osteoarthritis analysis, AI-based systems can help improve diagnostic consistency, reduce inter-observer variability, and accelerate MRI interpretation. Consequently, combining MRI imaging with artificial intelligence represents a promising direction for improving osteoarthritis assessment and

supporting earlier clinical decision-making [15, 16, 6].

Based on the above discussion, there is a clear need for an integrated and robust framework capable of improving MRI quality, stabilizing anatomical segmentation, and extracting discriminative multimodal representations from knee MRI volumes. Such a framework may contribute to more reliable osteoarthritis analysis, improve anatomical interpretation, and enhance early disease prediction performance [6, 16, 17, 9]. This work investigates an integrated framework for knee MRI analysis combining advanced preprocessing techniques, coarse-to-fine segmentation strategies, multimodal feature extraction, and osteoarthritis classification approaches. The proposed framework aims to improve MRI interpretation, enhance segmentation robustness, and support more reliable osteoarthritis assessment [6, 16, 17].

We noticed that Within artificial intelligence systems, segmentation models are particularly sensitive to image quality variations. When MRI volumes are unclear or poorly enhanced, models may struggle to detect accurate anatomical boundaries, especially in small and complex regions such as meniscal tissues. In some cases, background regions may also be confused with real anatomical structures, leading to unstable segmentation outputs and inaccurate feature extraction results [15, 3]. Furthermore, in direct multi-class segmentation approaches the large background of regions surrounding the knee joint influence the segmentation process. Processing the entire MRI volume may reduce segmentation stability and increase the difficulty of accurately localizing fine anatomical structures. Therefore, coarse-to-fine segmentation strategy based on preliminary knee localization and ROI extraction has been applied for improving segmentation robustness and reducing irrelevant background interference [6, 17]

A preprocessing has been applied to knee MRI. The objective of preprocessing is not limited to improving visual quality only, but also includes reducing noise, enhancing contrast, and clarifying anatomical boundaries in order to facilitate more stable and reliable learning. Indeed, appropriate preprocessing techniques can also reduce analysis difficulty and improve the efficiency of AI-based medical systems [13, 14, 11].

Traditional handcrafted features alone may fail to fully represent the complex structural information contained in three-dimensional MRI volumes, particularly within small anatomical regions. On the other hand, deep features learned automatically by neural networks, despite their strong representation capability, may not always capture all structural variations related to age, gender, or patient-specific anatomical characteristics [12, 15]. In this sens, the proposed framework relies on a single imaging modality (MRI) combining multiple complementary feature representations, including intensity-based, texture-based, morphological, and deep features, providing a richer and more discriminative description of osteoarthritic degeneration patterns. This thesis is organized into three main chapters

in addition to a general introduction and a general conclusion.

Chapter I: presents the theoretical background and literature review related to osteoarthritis, MRI biomarkers, medical image preprocessing techniques, deep learning-based segmentation models, transformer architectures, and multimodal feature extraction methods. The chapter also discusses recent advances in artificial intelligence for knee MRI analysis and highlights the main research gaps addressed in this work.

Chapter II: describes the proposed methodology adopted for knee MRI analysis. It presents the preprocessing pipeline, the proposed coarse-to-fine transformer-based segmentation framework, multimodal feature extraction strategies, feature fusion, and osteoarthritis classification models.

Chapter III: presents the experimental results and performance evaluation of the proposed framework. It includes preprocessing analysis, segmentation evaluation using Dice and IoU metrics, classification results, comparative experiments, and discussion of the obtained findings.

Finally, the thesis concludes with a general conclusion summarizing the main contributions and results of this work, while also presenting several limitations and future research directions for improving automated osteoarthritis analysis systems.

CHAPTER I

Literature Review on MRI-Based Osteoarthritis Analysis and Research Gaps

1.1 Introduction

This chapter presents the theoretical background and literature review related to knee osteoarthritis analysis using Magnetic Resonance Imaging (MRI) and artificial intelligence techniques. We first introduce osteoarthritis pathology, cartilage degeneration, and MRI biomarkers used for evaluating structural joint abnormalities and disease progression.

Subsequently, Computer-Aided Diagnosis (CAD) systems and their role in automated medical image analysis are discussed, particularly in the context of knee MRI interpretation and osteoarthritis assessment.

We then review medical image preprocessing techniques employed to improve MRI quality before deep learning analysis, including wavelet-based denoising, transformer-based image restoration, intensity normalization, gamma correction, and CLAHE-based contrast enhancement.

In addition, different deep learning approaches for medical image segmentation are presented, starting from convolutional neural networks and U-Net architectures to more advanced transformer-based segmentation models such as Vision Transformers, UNETR, and Swin-UNETR.

We also discuss feature extraction strategies used for osteoarthritis prediction, including intensity-based features, texture descriptors, morphological biomarkers, handcrafted radiomics features, and deep feature embeddings.

Furthermore, recent state-of-the-art approaches are analyzed through a research gap study in order to highlight the limitations of existing osteoarthritis analysis systems and motivate the development of the proposed framework.

Finally, we expose the main research problems and open challenges associated with knee MRI analysis, segmentation stability, multimodal feature extraction, and osteoarthritis classification.

1.2 Osteoarthritis and MRI Biomarkers

Osteoarthritis (OA) is one of the most prevalent degenerative musculoskeletal diseases worldwide and represents a major cause of pain, disability, and reduced mobility, particularly among elderly populations [9]. Recent epidemiological studies estimate that more than **500 million** individuals are affected by osteoarthritis globally, making it a significant public health and socioeconomic challenge.

Among individuals older than **60 years**, symptomatic knee osteoarthritis affects ap-

proximately **10%** of men and **13%** of women worldwide [9]. In addition, osteoarthritis is considered one of the leading causes of long-term disability because of chronic pain, reduced mobility, and progressive joint dysfunction.

The knee joint is particularly vulnerable to osteoarthritis progression because it continuously supports body weight and undergoes repetitive biomechanical loading during daily activities. The disease progressively affects several anatomical structures including articular cartilage, subchondral bone, menisci, ligaments, and synovial tissues [8, 10].

1.2.1 Cartilage Degeneration in Osteoarthritis

Articular cartilage degeneration represents one of the principal pathological characteristics of osteoarthritis. Under normal physiological conditions, cartilage acts as a smooth elastic tissue responsible for reducing friction and absorbing mechanical stress during joint movement. However, progressive biochemical and structural degeneration gradually reduces cartilage integrity and mechanical resistance, leading to tissue deterioration and joint instability.

As osteoarthritis progresses, several structural abnormalities may appear, including cartilage thinning, osteophyte formation, synovial inflammation, subchondral bone remodeling, meniscal degeneration, and joint space narrowing [9, 8]. These degenerative alterations progressively impair joint biomechanics and negatively affect patient mobility and quality of life.

In addition to osteoarthritis-related degeneration, several other knee abnormalities may affect joint structures and contribute to pain, instability, and functional limitation. Common abnormalities include meniscal tears, cartilage defects, anterior cruciate ligament (ACL) injuries, posterior cruciate ligament (PCL) lesions, medial collateral ligament (MCL) injuries, lateral collateral ligament (LCL) injuries, infrapatellar fat pad (IFP) abnormalities, synovial plica, and cyst formation [18].

These pathological conditions frequently coexist because traumatic and degenerative processes within the knee joint often overlap during disease progression. In many situations, pre-existing structural abnormalities may accelerate cartilage degeneration and increase the risk of developing osteoarthritis, thereby generating substantial clinical and socioeconomic burden [9].

Recent clinical studies estimated that knee abnormalities and osteoarthritis may generate annual medical costs ranging approximately between **700** and **15,600 USD** per patient depending on disease severity and therapeutic management strategies. Furthermore, within the first five years following osteoarthritis diagnosis, each patient may require an average excess of approximately **16.8** medical visits, **0.7** hospitalization days, and nearly

420 prescription doses [18].

Osteoarthritis progression is also highly heterogeneous and compartment-specific. Degeneration patterns may differ considerably between medial and lateral compartments as well as between femoral, tibial, and meniscal regions. Consequently, localized structural degeneration may not always be accurately represented using global knee analysis approaches [12, 9].

Furthermore, osteoarthritis frequently remains asymptomatic during its early stages, which complicates early diagnosis and therapeutic intervention. Delayed detection may considerably increase the risk of irreversible joint degeneration and severe structural damage requiring surgical procedures such as total knee arthroplasty (TKA) [10].

Consequently, there is an urgent need to improve early osteoarthritis diagnosis and progression prediction through reliable imaging biomarkers and automated medical image analysis systems. For this reason, large-scale clinical initiatives such as the Osteoarthritis Initiative (OAI) consortium were established to collect multimodal datasets including Magnetic Resonance Imaging (MRI), Computed Tomography (CT), clinical information, and genetic data for osteoarthritis research and biomarker discovery [8].

1.2.2 Structural MRI Biomarkers

Magnetic Resonance Imaging (MRI) is considered one of the most important imaging modalities for osteoarthritis assessment because of its ability to provide detailed visualization of both hard and soft knee anatomical structures. Unlike conventional radiography, which mainly evaluates visible bone alterations and joint space narrowing, MRI enables detailed analysis of cartilage, menisci, ligaments, synovial tissues, and subchondral bone structures [8, 10].

MRI has become the preferred non-invasive imaging modality for evaluating knee abnormalities because of its superior soft-tissue contrast and high anatomical visualization capability. In clinical practice, Proton Density-weighted Fat Suppression (PDwFS) MRI sequences are commonly employed for knee analysis because they provide improved visualization of cartilage, menisci, ligaments, and surrounding tissues [8].

This capability is particularly important because many early osteoarthritic abnormalities remain difficult or impossible to visualize using standard X-ray imaging. MRI allows earlier detection of cartilage degeneration, meniscal lesions, bone marrow abnormalities, inflammatory changes, and structural tissue alterations before severe degeneration becomes radiographically visible [8].

Traditional two-dimensional MRI acquisitions may suffer from limited spatial resolution, partial volume artifacts, and restricted anatomical visualization. Consequently,

modern three-dimensional MRI sequences have become increasingly important for osteoarthritis analysis because they provide improved spatial resolution and enable arbitrary plane reconstruction for more accurate anatomical assessment [10, 8].

A single knee MRI examination may contain more than **100** slices, generating large volumes of anatomical information that require detailed interpretation by radiologists and clinicians. This complexity considerably increases clinical workload and interpretation difficulty, particularly during the analysis of subtle osteoarthritic abnormalities [3].

In addition to qualitative structural abnormalities, MRI also provides quantitative biomarkers that can support automated osteoarthritis analysis systems. Parameters such as cartilage thickness, joint space narrowing, osteophyte formation, bone marrow lesions, meniscal degeneration, subchondral bone remodeling, tissue morphology, texture heterogeneity, volumetric measurements, and structural asymmetry may provide detailed information regarding disease progression and anatomical degeneration [12] [9, 8].

To evaluate osteoarthritis severity, the Kellgren–Lawrence (KL) grading system remains one of the most widely used clinical grading approaches. The KL system classifies osteoarthritis progression from Grade 0, representing a healthy knee joint, to Grade 4, corresponding to severe osteoarthritis associated with major structural degeneration, osteophyte formation, and substantial loss of joint functionality [8, 9].

1.3 Computer-Aided Diagnosis (CAD) Systems

The increasing complexity of medical imaging data has accelerated the development of Computer-Aided Diagnosis (CAD) systems for automated medical image analysis. These systems are designed to support clinicians during diagnostic interpretation and improve the reliability, efficiency, and consistency of medical decision-making [15].

CAD systems are particularly important in knee MRI analysis because MRI volumes contain large amounts of fine anatomical information distributed across multiple three-dimensional slices. Manual interpretation of these datasets is often time-consuming and cognitively demanding, especially in osteoarthritis analysis where early pathological changes may appear subtle and difficult to identify.

In some clinical situations, manual segmentation of a complete three-dimensional knee MRI series may require up to **6 h** for a trained clinical expert depending on anatomical complexity and segmentation precision requirements [3].

Several radiological studies additionally reported lesion miss rates ranging approximately between **10%** and **45%** depending on workload conditions, image complexity, and observer experience [3]. Consequently, automated CAD systems may contribute to

reducing interpretation variability and improving early disease detection.

Another important limitation concerns MRI image quality itself. Medical images may contain noise, low contrast, intensity inhomogeneity, and acquisition-related artifacts that complicate both manual interpretation and automated segmentation tasks. In some cases, noise patterns may appear similar to real anatomical structures, increasing the difficulty of accurately identifying pathological regions [13, 14, 11].

Modern CAD systems are capable of performing multiple medical image analysis tasks including anatomical segmentation, ROI localization, feature extraction, and disease classification. These systems help reduce analysis time, minimize observer-dependent variability, and facilitate earlier detection of osteoarthritic degeneration patterns [16, 15].

Early CAD systems for knee analysis mainly relied on conventional image processing and machine learning techniques for segmentation and abnormality detection tasks. However, recent advances in deep learning have significantly improved automated segmentation, feature extraction, and disease classification performance in complex knee MRI analysis applications [17].

Convolutional Neural Networks (CNNs) demonstrated strong capability for local feature extraction and anatomical segmentation, whereas Transformer-based architectures introduced self-attention mechanisms capable of modeling long-range spatial relationships within complex MRI volumes [6, 16].

Transformer-based frameworks such as Swin-UNETR have recently shown promising performance in medical image segmentation tasks because of their ability to simultaneously capture local anatomical details and global contextual information. However, these architectures frequently require high computational resources and large annotated datasets for stable training [6, 17].

Despite the remarkable progress achieved by modern CAD systems, several challenges still remain unresolved, including segmentation instability, anatomical variability, class imbalance, computational complexity, and the difficulty of accurately capturing subtle osteoarthritic degeneration patterns across heterogeneous knee anatomical compartments.

1.4 Image Preprocessing in Medical Imaging

Image preprocessing is considered one of the most important stages in medical image analysis, especially when working with Magnetic Resonance Imaging (MRI) data. Although MRI provides highly detailed anatomical information, these images often suffer from several limitations that affect the visibility of anatomical structures and reduce the reliability of both clinical interpretation and automated analysis systems [15, 10].

One of the most common issues in MRI images is the presence of noise. This noise may appear because of acquisition conditions, scanner limitations, or environmental interference during the imaging process. In many cases, noise appears as random intensity variations that may resemble real anatomical details, making tissue differentiation more difficult.

This issue becomes particularly critical in osteoarthritis analysis because subtle pathological changes may occur in very small anatomical regions such as cartilage surfaces and meniscal boundaries. Consequently, even minor image degradation may negatively affect segmentation precision and feature extraction reliability. Moreover, structures such as cartilage, menisci, and ligaments often have very similar intensity distributions. As a result, the boundaries between these tissues are not always clearly visible. Indeed, the three-dimensional nature of MRI data increases the complexity of analysis because each slice may contain subtle anatomical details that require careful interpretation.

MRI images may also suffer from low contrast and non-uniform intensity distribution, commonly referred to as intensity inhomogeneity. In such cases, the same tissue may appear with different intensity values in different regions of the image, which can negatively affect both segmentation accuracy and model stability during training.

Another challenge comes from differences between imaging devices and acquisition protocols. Variations in scanner settings and patient positioning may lead to inconsistencies in image appearance and intensity scaling across datasets, making it difficult to build models that generalize effectively.

MRI images may additionally suffer from partial volume effects and through-plane blurring, particularly near thin anatomical structures such as cartilage surfaces and meniscal boundaries. These phenomena reduce boundary precision and complicate both manual interpretation and automated segmentation tasks [3].

For these reasons, preprocessing has become an essential step before applying deep learning models to MRI data. The objective of preprocessing is not only to improve image appearance, but also to preserve important anatomical structures, reduce unwanted noise, and simplify the learning process for segmentation and classification models [13, 14].

In deep learning applications, preprocessing also contributes to stabilizing data distribution, reducing irrelevant variations, and facilitating more reliable model convergence during training.

In this work, a multi-stage preprocessing pipeline is adopted by combining several complementary techniques, including wavelet-based denoising, transformer-based image restoration using Restormer, gamma correction, and CLAHE-based contrast enhancement. This integrated strategy aims to improve MRI image quality while preserving

meaningful anatomical information instead of focusing only on visual enhancement [19, 13, 14, 20, 11].

1.4.1 Noise in MRI

MRI images are frequently affected by different forms of noise that may reduce image clarity and make anatomical structures more difficult to identify. In medical imaging, noise does not only affect visual quality, but may also interfere with tissue boundaries and hide fine structural details required for accurate diagnosis.

Unlike natural images, MRI data contains complex anatomical information with subtle intensity differences between tissues. Consequently, even a small amount of noise may significantly affect the visibility of structures such as cartilage, menisci, and ligaments. This issue becomes more critical in osteoarthritis analysis, where many pathological changes appear in very small and sensitive regions.

Figure 1.1 illustrates an example of MRI degradation caused by noise and low contrast.

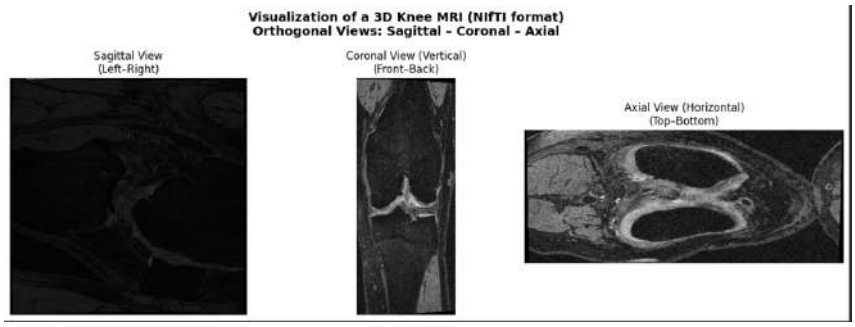


Figure 1.1: Example of noise and low contrast effects in knee MRI images.

1.4.2 Wavelet-Based Denoising

Wavelet-based denoising is used as the first stage of the preprocessing pipeline because of its ability to reduce noise while preserving important anatomical details. Unlike traditional smoothing filters that may blur edges and fine structures, wavelet-based methods analyze the image in the frequency domain [13].

Wavelet methods are particularly suitable for MRI preprocessing because they preserve fine anatomical structures while suppressing high-frequency noise components that may interfere with tissue boundary detection.

The approach relies on the Discrete Wavelet Transform (DWT), which decomposes the image into several frequency components:

$$I = LL + LH + HL + HH \tag{1.1}$$

The LL component contains the main low-frequency structural information, while the LH, HL, and HH components represent high-frequency details that usually contain edges and noise.

Since MRI noise is mainly concentrated in the high-frequency components, noise reduction is performed by attenuating these coefficients:

$$C' = \alpha C \tag{1.2}$$

where α is a scaling factor smaller than 1 used to suppress noise while preserving meaningful edge information.

This method helps clean MRI images without removing important structures such as cartilage boundaries and ligament edges, making it more suitable for medical imaging applications than conventional smoothing approaches.

In knee MRI analysis, preserving anatomical edges is particularly important because osteoarthritic degeneration frequently affects thin cartilage surfaces and meniscal boundaries that may become difficult to identify after excessive smoothing.

Wavelet-based denoising additionally contributes to improving segmentation stability and reducing noise-related artifacts during subsequent feature extraction and osteoarthritis analysis stages.

Figure 1.2 presents the effect of wavelet denoising on MRI image quality.

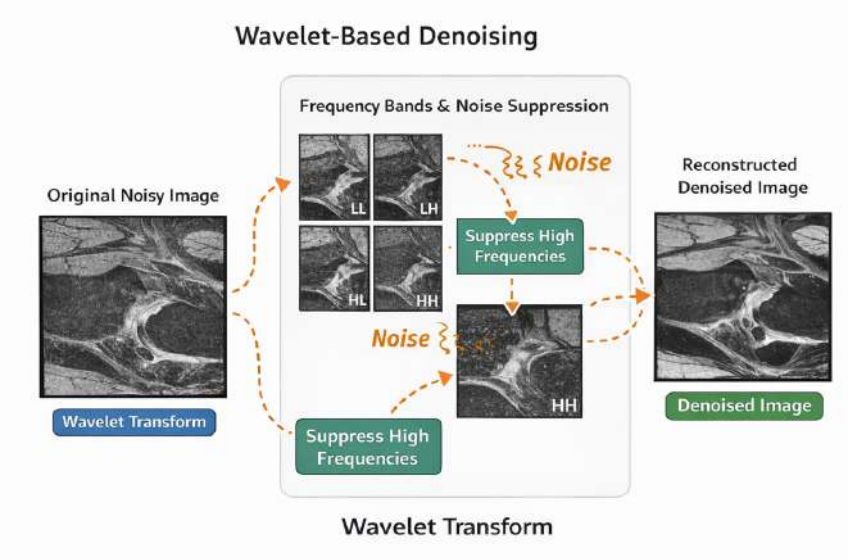


Figure 1.2: Wavelet-based denoising applied to knee MRI images.

1.4.3 Transformer-Based Image Restoration

After reducing noise, the Restormer model is applied to further improve image quality and recover fine anatomical details that may not be clearly visible in MRI images. The model is based on transformer architecture and uses self-attention mechanisms to understand relationships between distant image regions instead of focusing only on local neighborhoods like conventional convolutional neural networks (CNNs) [19, 16].

The self-attention mechanism can be expressed as:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1.3}$$

where Q, K, and V represent the query, key, and value matrices, while d represents the normalization dimension.

This mechanism allows the model to capture long-range dependencies and better understand complex anatomical structures within MRI images.

This global contextual understanding is particularly important in MRI analysis because anatomically related structures may appear across distant image regions within the three-dimensional volume.

Restormer improves structural clarity and helps recover fine details that may become less visible after denoising. This is particularly important in knee MRI analysis, where subtle structures such as cartilage surfaces and meniscus boundaries play a major role in segmentation quality.

Compared to CNN-based restoration methods, transformer-based approaches provide a more global understanding of the image and are more effective in handling complex structural relationships.

Figure 1.3 illustrates the Restormer architecture used for MRI restoration.

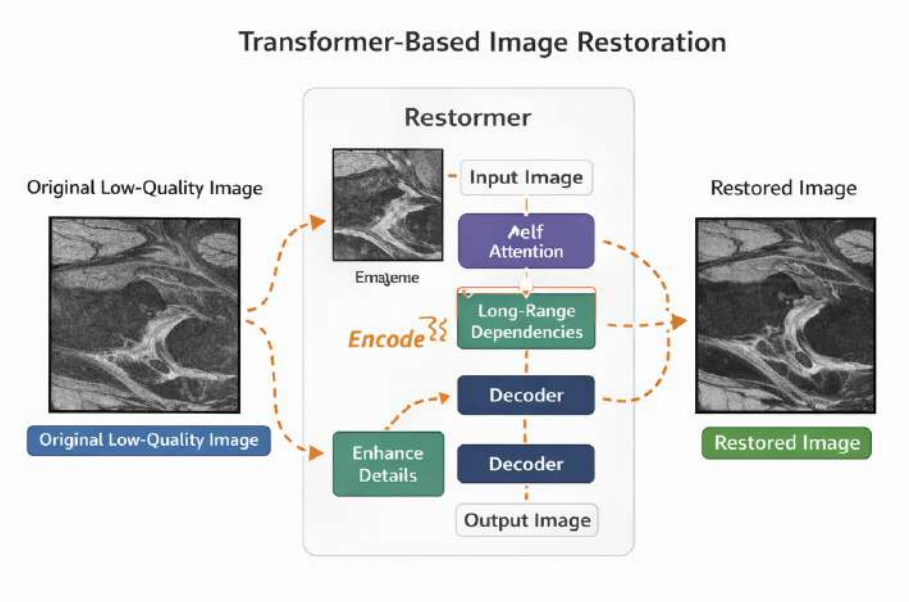


Figure 1.3: Transformer-based image restoration using Restormer.

1.4.4 Intensity Normalization and Gamma Correction

After restoration, intensity normalization is applied to ensure that pixel values remain within a stable range. This step helps reduce intensity inconsistencies and improves the stability of subsequent processing stages.

Gamma correction is then applied as a nonlinear intensity transformation defined by:

$$I_{out} = I_{in}^{\gamma} \tag{1.4}$$

where γ controls intensity transformation [20].

Although the visual impact may sometimes appear limited, gamma correction contributes to stabilizing intensity distribution and preparing the image for contrast enhancement.

This step becomes useful in osteoarthritis analysis because some pathological regions may appear with weak intensity variations that are difficult to distinguish without intensity stabilization. Figure 1.4 demonstrates un example of Gamma correction applied to MRI images.

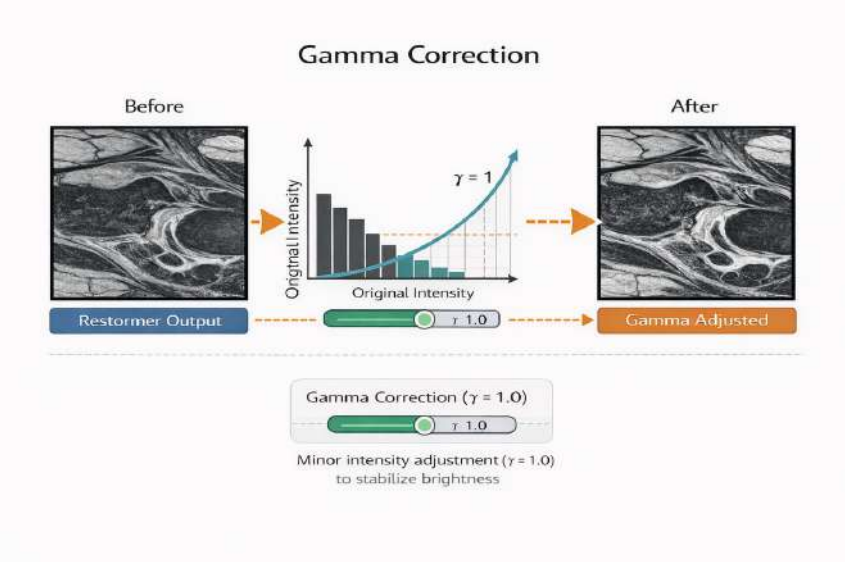


Figure 1.4: Gamma correction applied to MRI images for intensity stabilization.

1.4.5 Contrast Enhancement using CLAHE

The final stage of the preprocessing pipeline involves contrast enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE). This step plays an important role in improving the visibility of anatomical structures, particularly in MRI images where the contrast between tissues is often limited [14].

Unlike traditional histogram equalization methods that process the image globally, CLAHE divides the image into smaller local regions and enhances each region independently.

Local contrast enhancement is particularly important in knee MRI analysis because different anatomical regions may exhibit different intensity distributions and visibility conditions inside the same image.

This technique improves the visibility of subtle anatomical structures that are difficult to observe in low-contrast MRI images, particularly cartilage, ligaments, and menisci. In addition, CLAHE reduces excessive noise amplification that may occur with standard contrast enhancement approaches.

Figure 1.5 shows the effect of CLAHE on MRI contrast enhancement.

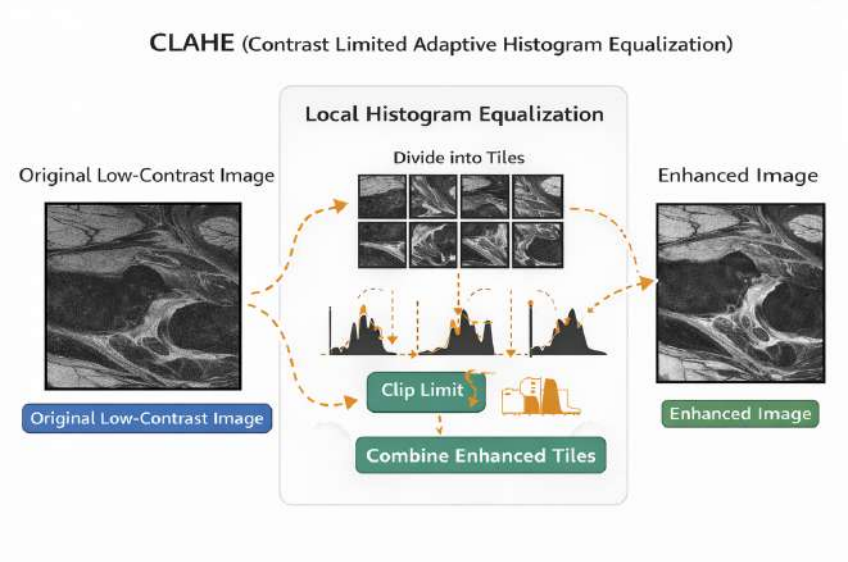


Figure 1.5: CLAHE-based contrast enhancement applied to knee MRI images.

1.4.6 Integrated Preprocessing Pipeline

The proposed preprocessing pipeline integrates all previously described techniques into a unified and sequential framework. Each stage is designed to address a specific limitation of MRI data, including noise reduction, structural enhancement, intensity stabilization, and contrast improvement.

Wavelet-based denoising is initially applied to suppress high-frequency MRI noise while preserving important anatomical edges. Restormer-based restoration is subsequently employed to recover fine structural details and improve global contextual consistency across anatomical regions. Gamma correction is then used to stabilize intensity distribution and reduce intensity variability before applying CLAHE-based local contrast enhancement. Finally, CLAHE improves tissue visibility and local contrast without excessively amplifying residual noise. This sequential organization helps preserve structural integrity while progressively improving MRI quality throughout the preprocessing pipeline.

Improving MRI quality before segmentation is particularly important because unclear anatomical boundaries and intensity inconsistencies may negatively affect coarse-to-fine segmentation stability, ROI localization accuracy, and multimodal feature extraction reliability.

Overall, this integrated preprocessing strategy produces MRI images with improved clarity and more consistent structural information, which directly supports more stable coarse-to-fine segmentation, improved anatomical localization, and more reliable multimodal feature extraction performance during subsequent analysis stages [15, 3, 6, 17].

1.5 Deep Learning for Medical Image Segmentation

Deep learning has significantly transformed the field of medical image analysis by improving the accuracy and efficiency of segmentation and classification systems compared to conventional image processing approaches. Unlike traditional techniques that rely on manually designed features, deep learning models are capable of automatically learning patterns and representations directly from data, allowing a more comprehensive understanding of anatomical structures in medical images [15, 3, 9].

Traditional image processing and handcrafted feature engineering approaches frequently struggle to handle the complexity and variability of three-dimensional knee MRI data. Consequently, deep learning models became increasingly important because of their capability to automatically learn hierarchical and discriminative representations directly from raw medical images.

This capability becomes particularly important in osteoarthritis analysis because knee MRI volumes contain highly complex anatomical information distributed across multiple slices and compartments, making manual feature engineering increasingly insufficient for robust analysis.

Medical image segmentation aims to identify and separate anatomical structures from surrounding tissues. This task is particularly important in Magnetic Resonance Imaging (MRI), where accurate delineation of cartilage, ligaments, and menisci is essential for disease analysis and prediction. However, segmentation remains challenging because MRI images often contain noise, low contrast, and highly complex anatomical details.

The difficulty becomes more pronounced when analyzing small anatomical structures such as cartilage surfaces and meniscal tissues, where even minor segmentation inaccuracies may significantly affect subsequent feature extraction and disease prediction stages.

1.5.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are among the most widely used models in medical image analysis because they are specifically designed for image processing tasks. CNNs automatically learn image patterns and extract meaningful features from raw data, which makes them highly effective for MRI analysis.

These networks rely on convolution operations to extract local features such as edges, textures, and structural patterns inside the image. After each convolution operation, the model generates feature maps that contain the important information learned from the input image.

However, convolution operations mainly focus on neighboring regions and local receptive fields. Consequently, CNNs are generally more effective at capturing local spatial patterns than modeling long-range contextual relationships distributed across distant anatomical regions.

CNNs also utilize pooling layers to reduce spatial dimensions and improve computational efficiency. This process enables the model to build hierarchical feature representations, starting from low-level details and progressively learning more complex semantic information.

CNN-based models achieved strong performance in medical imaging because they preserve important image structures while extracting meaningful local information, particularly in MRI images where tissue boundaries and textures play an important role [3].

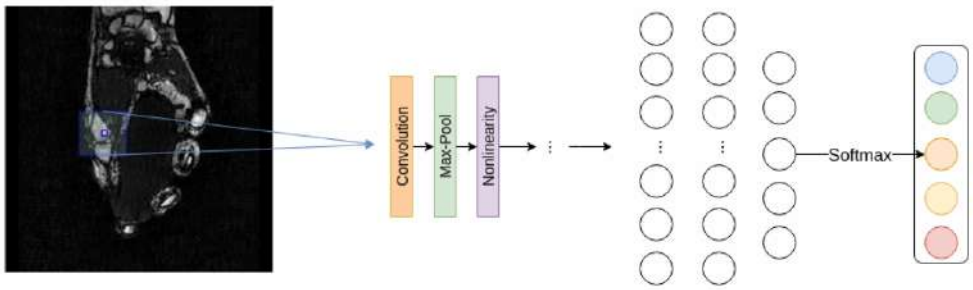


Figure 1.6: Basic CNN workflow for MRI medical image analysis showing convolution, pooling, and classification stages reproduced from [1].

1.5.2 U-Net Architecture

Although CNNs achieved promising results in image analysis, medical segmentation tasks require higher precision in detecting boundaries and small anatomical structures. For this reason, the U-Net architecture was specifically developed for biomedical image segmentation [3].

U-Net follows an encoder–decoder architecture. The encoder progressively extracts contextual information while reducing the spatial resolution of the image. In contrast, the decoder reconstructs the segmentation map and restores fine anatomical details.

Although encoder–decoder architectures achieved remarkable success in medical image segmentation, determining the optimal network depth remains highly task-dependent and varies according to anatomical complexity, dataset size, and segmentation objectives. In addition, the design of skip connections remains largely empirical and lacks strong the-

oretical foundations regarding the optimal integration of low-level and high-level feature representations [3].

This structure allows the model to capture global image information while preserving precise local structures such as cartilage, ligaments, and menisci.

This characteristic makes U-Net particularly suitable for medical image segmentation tasks involving complex anatomical boundaries and fine tissue structures commonly observed in MRI data.

One of the key components of U-Net is the use of skip connections. These connections transfer feature information directly from the encoder to the decoder, helping preserve spatial details that may otherwise be lost during pooling operations.

As a result, U-Net became one of the most successful architectures in medical image segmentation due to its ability to balance contextual understanding and accurate localization.

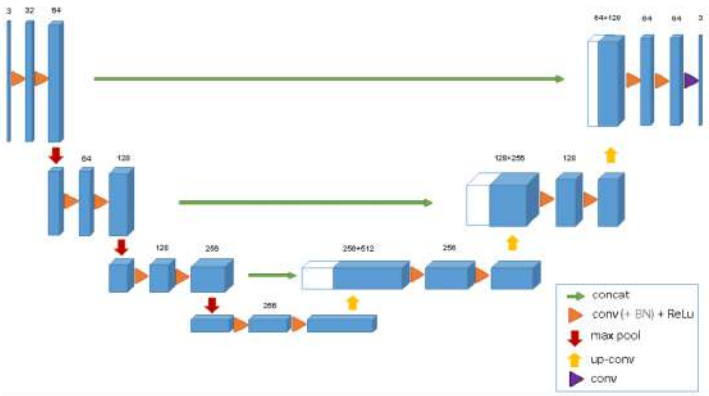


Figure 1.7: 3D U-Net architecture reproduced from the GitHub repository [2, 3].

1.5.3 Limitations of CNN-Based Models

Despite the strong performance of CNN-based models such as U-Net, they still suffer from several limitations, particularly when dealing with three-dimensional MRI data.

CNNs mainly rely on local understanding of the image, where the receptive field remains relatively limited. This makes capturing long-range relationships between distant anatomical regions more difficult.

In addition, pooling operations may lead to the loss of fine structural details, negatively affecting segmentation quality, especially in small and sensitive anatomical regions inside the knee joint.

These limitations become more significant in MRI analysis because medical images contain highly complex anatomical structures distributed across multiple slices. Conse-

quently, conventional CNN models may struggle to fully understand global contextual relationships compared to more advanced transformer-based architectures [21, 16, 6, 17].

In deep learning-based cartilage segmentation, segmentation performance frequently decreases near peripheral cartilage regions because of partial volume effects and through-plane blurring. These phenomena reduce boundary precision and make maintaining consistent segmentation performance across the entire three-dimensional MRI volume considerably more difficult [3].

Furthermore, processing complete three-dimensional MRI volumes often requires high computational resources and GPU memory consumption, particularly when dealing with high-resolution medical images and multi-region segmentation tasks.

Most existing CNN-based approaches mainly focus on isolated single-task objectives such as segmentation or classification independently. However, knee MRI analysis involves highly interconnected anatomical structures and pathological relationships, making integrated multi-stage analysis frameworks increasingly important for clinical applications.

Because of these limitations, recent research has increasingly focused on transformer-based architectures capable of modeling long-range dependencies and capturing global contextual information more effectively than conventional CNN-based approaches. These models have demonstrated promising performance in complex medical image segmentation tasks, particularly in three-dimensional MRI analysis [16, 6, 17].

1.6 Transformer-Based medical image Segmentation Models

Transformer-based architectures have recently attracted significant attention in medical image analysis because of their strong capability to model complex spatial relationships and capture global contextual information. These models were introduced to overcome several limitations associated with conventional convolutional neural networks (CNNs), particularly in medical imaging tasks involving highly detailed anatomical structures and three-dimensional MRI data [16, 6, 17].

Unlike CNN-based approaches that mainly focus on local image neighborhoods, transformer architectures provide a broader understanding of image content by modeling relationships between distant regions. This capability is particularly important in knee MRI analysis because anatomical structures such as cartilage, menisci, and surrounding tissues are spatially interconnected across multiple slices and compartments.

Transformer-based architectures additionally demonstrated improved capability in handling volumetric MRI analysis tasks where contextual information distributed across

the complete three-dimensional volume plays an important role in accurate anatomical segmentation and structural understanding.

1.6.1 Vision Transformers

Vision Transformers (ViT) represent one of the earliest transformer-based approaches adapted for computer vision tasks. Instead of processing the image as a continuous spatial grid, the image is divided into smaller patches. Each patch is transformed into a numerical representation called a token, which is then processed by the transformer architecture [16].

By learning relationships between tokens, the model can analyze interactions between different image regions and build a more global representation of anatomical structures. This differs from conventional CNNs, which mainly focus on extracting local spatial patterns through convolution operations.

The use of Vision Transformers becomes particularly useful in MRI analysis because pathological changes associated with osteoarthritis may appear across multiple regions inside the knee joint. Consequently, capturing long-range dependencies between anatomical structures can improve segmentation robustness and structural understanding.

However, standard Vision Transformers frequently require very large training datasets and high computational resources in order to achieve stable performance. These limitations become more challenging in medical imaging applications where annotated MRI datasets are often limited and expensive to obtain.

1.6.2 Self-Attention Mechanisms in Medical Imaging

The main component responsible for the effectiveness of transformer architectures is the self-attention mechanism. Self-attention enables the model to evaluate relationships between different image regions and determine which features are more relevant during the analysis process.

In medical imaging, this mechanism is highly important because anatomical structures inside MRI volumes are strongly interconnected. Structural changes occurring in one region may influence neighboring tissues or reflect broader pathological patterns associated with osteoarthritis progression.

The self-attention mechanism is mathematically expressed as:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1.5}$$

where Q, K, and V represent the query, key, and value matrices, while d represents the normalization dimension [4].

This mechanism allows transformer models to preserve important contextual information across multiple layers and improves their ability to capture fine anatomical structures and long-range spatial dependencies within MRI images.

Compared with conventional convolution operations, self-attention mechanisms provide a more flexible strategy for modeling global contextual relationships distributed across complex three-dimensional anatomical regions.

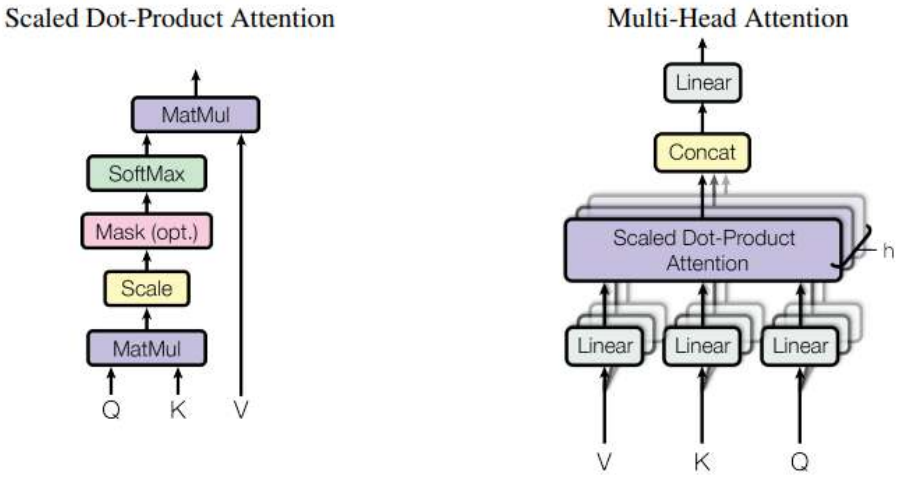


Figure 1.8: Scaled dot-product attention and multi-head attention mechanisms reproduced from [4].

1.6.3 UNETR

UNETR is among the first medical image segmentation architectures that successfully integrated transformer-based encoding with the U-Net framework. In this architecture, the transformer encoder is responsible for learning global contextual representations from the input image, while the decoder reconstructs fine anatomical details required for accurate segmentation [16].

This hybrid architecture combines the advantages of transformers and CNN-based decoders. The transformer component improves global contextual understanding, whereas the decoder preserves precise local anatomical structures. Consequently, UNETR became particularly suitable for MRI segmentation tasks involving complex anatomical boundaries and detailed tissue structures.

Unlike conventional CNN-based architectures, UNETR processes volumetric MRI data using transformer-based token representations capable of modeling long-range anatomical

dependencies across the complete three-dimensional volume.

However, UNETR architectures may still require high computational resources and GPU memory consumption, particularly when processing high-resolution MRI volumes and large patch embeddings.

1.6.4 Swin-UNETR

Although Vision Transformers and UNETR demonstrated promising segmentation performance, they may require high computational resources and large memory consumption, especially when processing three-dimensional medical images. To address these limitations, the Swin Transformer architecture was proposed as a more computationally efficient transformer-based approach [6].

Instead of processing the entire image globally, Swin Transformer analyzes the image using smaller local windows. This strategy significantly reduces computational complexity while still preserving the capability to capture long-range contextual information. In addition, shifted window mechanisms facilitate information exchange between neighboring image regions and improve contextual continuity.

By integrating Swin Transformer with the U-Net architecture, the Swin-UNETR model was developed for advanced medical image segmentation tasks. This architecture combines efficient contextual understanding with accurate local feature preservation, making it highly suitable for complex MRI anatomical structures and three-dimensional segmentation problems.

Compared with conventional CNN-based models, Swin-UNETR demonstrates improved capability in modeling long-range spatial dependencies and preserving fine anatomical structures, particularly in challenging MRI segmentation tasks involving cartilage and meniscal tissues [6, 17].

In addition, Swin-UNETR provides improved balance between segmentation accuracy and computational efficiency compared with conventional transformer architectures that process complete global image representations simultaneously.

Because osteoarthritis progression frequently affects heterogeneous anatomical compartments with subtle structural variations, the capability of Swin-UNETR to capture both local anatomical details and global contextual relationships becomes particularly advantageous for knee MRI analysis.

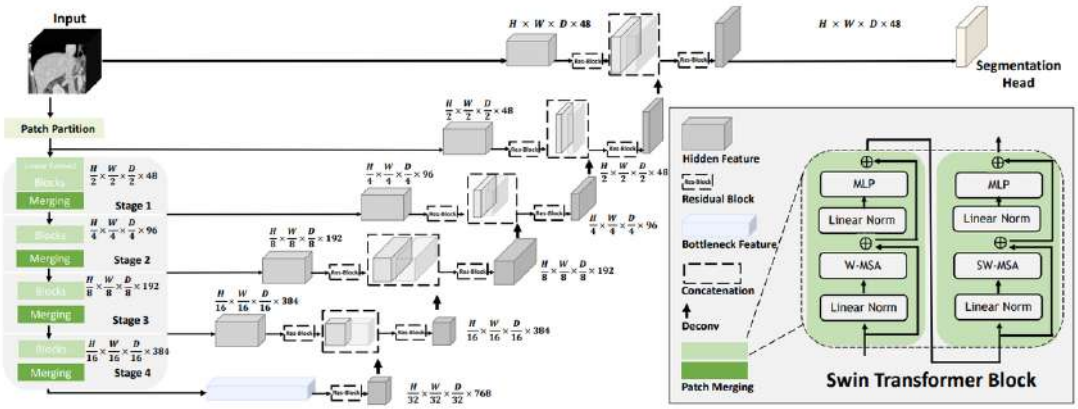


Figure 1.9: Swin-UNETR architecture reproduced from the MONAI GitHub repository [5, 6].

Overall, transformer-based segmentation architectures provide stronger contextual understanding and more robust representation of complex anatomical structures compared with conventional CNN-based approaches. These characteristics make transformer models increasingly suitable for challenging MRI segmentation tasks and osteoarthritis analysis applications involving highly detailed three-dimensional medical data [16, 6, 17].

1.7 Feature Extraction for Disease Prediction

Feature extraction is considered one of the fundamental stages in medical image analysis systems. Its objective is to transform MRI images into numerical representations that can be used to understand anatomical structures and predict disease severity [12, 15].

In osteoarthritis analysis, feature extraction becomes more challenging because degeneration patterns do not affect all anatomical regions uniformly. Structural changes may appear differently across cartilage, meniscus, femoral, and tibial regions, motivating the need for region-wise feature extraction strategies.

Osteoarthritis is considered a heterogeneous and compartment-specific disease because degeneration patterns may differ considerably between medial and lateral compartments as well as between femoral, tibial, and meniscal structures. Consequently, extracting global features from the complete knee joint may reduce sensitivity to localized pathological alterations associated with disease progression [9, 12]. Furthermore, osteoarthritis progression frequently exhibits structural asymmetry between knee compartments. Medial degeneration may differ significantly from lateral degeneration depending on biomechanical loading conditions and tissue-specific pathological evolution. Therefore, independent analysis of each anatomical region becomes important for preserving compartment-specific degeneration patterns.

MRI images contain a large amount of information related to bones, cartilage, and surrounding tissues. Therefore, relying only on raw image data is often insufficient for disease prediction, making feature extraction essential for identifying meaningful patterns associated with osteoarthritis progression.

MRI volumes contain highly complex multidimensional anatomical information distributed across multiple slices and tissues. Consequently, direct analysis of raw MRI images without quantitative feature representation may limit the capability of automated systems to identify subtle structural abnormalities associated with early osteoarthritis progression.

Small anatomical structures such as thin cartilage regions and meniscal tissues remain particularly challenging because subtle pathological changes may occur within very limited spatial regions.

1.7.1 Intensity-Based Features

One important category of features is intensity-based features. These features analyze pixel intensity values and study variations in brightness distribution across different tissues. Such information is important because MRI images vary between patients depending on age, gender, and anatomical structure. In addition, pathological changes may alter intensity distribution inside the joint.

These variations become particularly important in MRI analysis because osteoarthritic degeneration may produce subtle intensity differences that are difficult to detect through visual inspection alone.

Intensity variations may additionally reflect tissue degeneration, cartilage hydration changes, inflammatory processes, and alterations in subchondral bone structure associated with osteoarthritis progression.

1.7.2 Texture-Based Features (GLCM, Haralick, and LBP)

Texture-based features are also widely used to analyze internal tissue patterns. These features focus on relationships between neighboring pixels and study characteristics such as smoothness, roughness, and intensity transitions inside anatomical structures.

Methods such as Gray Level Co-occurrence Matrix (GLCM) and Haralick features are commonly used for texture analysis because they allow statistical evaluation of relationships between pixels and provide detailed information about internal tissue organization [12].

Texture analysis is particularly important in osteoarthritis because the disease pro-

gressively modifies cartilage structure and bone tissue organization, resulting in significant texture variations during disease progression.

Texture descriptors are particularly valuable because they capture microstructural tissue organization and spatial heterogeneity that may not always be visually observable during conventional MRI interpretation.

Local Binary Pattern (LBP) descriptors are additionally employed to characterize local micro-texture variations and structural irregularities within cartilage and meniscal tissues. LBP features are particularly effective for representing local spatial texture patterns and subtle tissue heterogeneity associated with osteoarthritic degeneration.

1.7.3 Shape and morphological features

Shape-based features are employed to analyze structural changes occurring inside the knee joint. Osteoarthritis may cause cartilage thinning, joint space narrowing, and deformation in bone structures. These morphological changes can be represented numerically and used for disease prediction.

Morphological analysis is particularly important because each anatomical compartment possesses specific volumetric and structural characteristics that may evolve differently during osteoarthritis progression.

Morphological features may also serve as imaging biomarkers reflecting osteoarthritis severity and structural joint degeneration.

Quantitative biomarkers such as cartilage thickness, meniscal volume, joint space measurements, and structural deformation indices may provide clinically relevant information regarding osteoarthritis severity and disease progression monitoring.

1.7.4 Deep Feature Embeddings

In addition to handcrafted features, deep feature embeddings have recently become important in medical image analysis. Unlike handcrafted approaches based on predefined rules, deep learning models automatically learn complex feature representations during training [15, 16].

Deep features are capable of capturing hidden relationships and complex patterns that may not be easily observed manually. However, relying exclusively on deep features may not always be sufficient in complex medical imaging tasks because some anatomical details remain difficult to interpret or may not be adequately represented.

In addition, deep learning models may sometimes learn highly abstract representations

that are difficult to interpret clinically. Consequently, relying exclusively on deep embeddings may reduce explainability and limit the interpretability of osteoarthritis biomarkers.

Deep embeddings may additionally become sensitive to dataset bias and acquisition variability, which may reduce generalization capability across heterogeneous MRI datasets and clinical environments.

Transformer-based deep learning models are particularly effective in this context because they can model long-range spatial dependencies and capture contextual anatomical relationships across MRI volumes.

Handcrafted descriptors provide interpretable quantitative biomarkers describing tissue morphology, texture organization, and structural degeneration, whereas deep embeddings capture hidden contextual and semantic MRI representations learned automatically from data.

For this reason, combining handcrafted features with deep features has become an important strategy in medical image analysis. Handcrafted features provide interpretable information related to intensity, texture, and shape, while deep features contribute by learning hidden and complex representations directly from MRI data.

Due to the complexity and richness of MRI images, this multimodal feature fusion strategy provides a more comprehensive representation of anatomical structures, which can improve disease prediction and classification performance in osteoarthritis analysis systems.

Although all extracted information originates from a single imaging modality (MRI), combining complementary feature representations including intensity, texture, morphology, and deep embeddings creates a multimodal analytical representation of knee anatomical structures.

Region-wise analysis additionally facilitates the extraction of inter-compartmental relationships and structural ratios between anatomical regions, which may provide important information regarding osteoarthritis progression and compartment-specific degeneration patterns.

These inter-regional descriptors may additionally quantify medial-lateral imbalance and compartment-specific degeneration asymmetry associated with osteoarthritis progression.

Overall, integrating handcrafted descriptors with deep feature embeddings provides complementary and heterogeneous representations of knee MRI data. This multimodal fusion strategy improves osteoarthritis prediction robustness, enhances sensitivity to localized pathological alterations, preserves compartment-specific degeneration patterns, and supports more reliable classification performance across different osteoarthritis severity

levels [12, 15, 9].

The extracted multimodal biomarkers may additionally support longitudinal osteoarthritis progression monitoring and facilitate more reliable disease severity assessment.

1.8 State of the Art Review and Research Gap Analysis in MRI-Based Osteoarthritis Assessment

Recent studies demonstrated significant progress in applying deep learning techniques for osteoarthritis analysis, particularly in segmentation, classification, and prediction tasks. These approaches improved medical image analysis performance and reduced human intervention compared with conventional image processing methods [9, 15].

The article entitled *Attention U-Net: Learning Where to Look for the Pancreas* introduced the Attention U-Net (AU-Net) architecture, which integrated attention mechanisms into the classical U-Net framework in order to improve segmentation accuracy and focus on relevant anatomical regions inside medical images [22]. However, the model still relied mainly on convolutional neural networks, which limited its ability to capture long-range contextual relationships in complex imaging data.

In addition, encoder–decoder architectures still suffer from unresolved challenges related to optimal network depth selection and skip connection design, which remain largely empirical and task-dependent [3].

The study *Automatic Knee Osteoarthritis Grading Using Deep Neural Networks* proposed the KOA-DNN framework based on ResNet convolutional neural networks for automatic five-class osteoarthritis grading using the Kellgren–Lawrence (KL) system [23]. Despite achieving improved classification accuracy, the framework remained sensitive to class imbalance problems and inherited the limitations of CNN-based local feature extraction.

Transformer-based architectures such as the Swin-UNETR model proposed in *Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images* attempted to overcome CNN limitations through shifted window mechanisms and global contextual learning [6, 6]. Nevertheless, the architecture required high computational resources, increased memory consumption, and computationally expensive training procedures.

The SDMT-Net framework introduced a multi-task transformer-based architecture for simultaneous segmentation and anatomical localization in three-dimensional MRI data [17]. Although the framework improved contextual feature learning and localization capability, the model remained computationally complex and required sophisticated opti-

mization strategies during training.

The paper *Deep Siamese Convolutional Neural Network for Osteoarthritis Classification* introduced the DS-CNN framework for osteoarthritis analysis under a binary classification setting [24]. Although the model produced high AUC classification performance, the framework required large annotated datasets and computationally demanding training procedures to achieve stable generalization performance.

Many earlier osteoarthritis analysis systems mainly relied on two-dimensional image representations, which may fail to preserve complete volumetric anatomical relationships distributed across three-dimensional MRI data.

Other studies explored multimodal learning strategies such as the MM-OA framework, which combined MRI and X-ray images in order to exploit complementary anatomical information from different imaging modalities [25]. Although multimodal learning improved prediction capability, it also introduced challenges related to modality alignment, preprocessing complexity, imaging variability, and feature fusion difficulties.

Radiomics-based approaches such as the RF-LASSO framework relied on handcrafted feature extraction methods including texture, intensity, and shape descriptors combined with LASSO feature selection [12]. However, handcrafted features remained highly sensitive to segmentation quality, image variability, and anatomical differences, which reduced feature stability and generalization capability.

Segmentation performance may additionally decrease near peripheral cartilage regions because of partial volume effects and through-plane blurring, which complicate accurate boundary delineation across MRI slices [3].

Recent foundation models such as the Segment Anything Model (SAM) introduced strong zero-shot generalization capability for segmentation tasks across various visual domains [26]. Despite their effectiveness, these models still require medical fine-tuning and domain-specific adaptation before being reliably applied to MRI-based osteoarthritis analysis systems.

Recent semi-supervised learning approaches such as SEN-MTKD further improved cartilage segmentation performance by combining mean teacher learning and knowledge distillation strategies, allowing better utilization of limited annotated MRI datasets [27]. The framework achieved improved Dice segmentation performance while reducing dependency on fully annotated training data. However, the training process remained computationally complex and sensitive to optimization stability.

More recently, multi-task deep learning systems such as the MTDL-System demonstrated promising performance for simultaneous classification of multiple knee abnormalities from MRI data [28]. The framework achieved strong diagnostic performance through

multi-task representation learning and improved automated abnormality detection capability. Nevertheless, the architecture increased model complexity and required extensive computational resources and large-scale annotated datasets.

Table 1.1 summarizes recent deep learning approaches proposed for osteoarthritis analysis, including segmentation, classification, multimodal learning, and radiomics-based frameworks.

Table 1.1: Summary of Recent Deep Learning Approaches for Osteoarthritis Analysis

Method	Year	Approach	Modality	Task	Performance	Limitation
AU-Net [22]	2018	Att. U-Net	MRI	Segmentation	Dice $\approx$ 0.84	Weak global context
Swin-UNETR [6]	2022	Transformer	3D MRI	3D segmentation	Dice $\approx$ 0.91	High GPU cost
SDMT-Net [17]	2023	MT Transformer	3D MRI	Seg. + localization	Dice = 0.839	Complex training
KOA-DNN [23]	2021	ResNet CNN	X-ray	5-class KL grading	Acc = 69.7%	Class imbalance
DS-CNN [24]	2018	Siamese CNN	X-ray	Binary classification	AUC = 0.98	Large dataset required
MM-OA [25]	2023	Multimodal DL	MRI + X-ray	Multi-class prediction	F1 = 0.84	Fusion complexity
RF-LASSO [12]	2012	Radiomics	MRI	Binary classification	AUC = 0.82	Feature instability
SAM [26]	2023	Foundation model	General images	Segmentation	Strong zero-shot	Needs fine-tuning
SEN-MTKD [27]	2025	Semi-supervised DL	MRI	Cartilage segmentation	Dice = 83.51%	Training complexity
MTDL-System [28]	2025	Multi-task DL	MRI	Knee abnormality classification(binary)	AUC = 0.902	Complex architecture

1.8.1 Research Gap Analysis

Overall, previous studies mainly focused on isolated tasks such as segmentation or classification independently [9, 27, 28]. Therefore, there remains a need for integrated MRI-based frameworks that combine preprocessing, segmentation, handcrafted feature extraction, deep feature learning, and multimodal feature fusion within a unified osteoarthritis analysis pipeline.

Preserving inter-compartmental relationships and medial-lateral anatomical asymmetry additionally represents an important challenge in osteoarthritis analysis because degeneration patterns frequently evolve differently across knee compartments [9, 27].

Furthermore, deep learning models may sometimes generate highly abstract feature representations that are difficult to interpret clinically, limiting explainability and direct biomarker interpretation [28, 16].

In addition, most previous studies focused either on binary classification or limited grading strategies, which may not fully capture the progressive nature of osteoarthritis

severity. Therefore, the proposed work investigates both three-class and five-class classification settings in order to better analyze disease progression and evaluate the capability of deep learning models to distinguish subtle KL grade variations [9, 28].

Another important challenge concerns balancing segmentation accuracy and computational efficiency, particularly for transformer-based architectures processing high-resolution three-dimensional MRI volumes [6, 16, 27].

Differences between MRI scanners, acquisition protocols, and patient populations may additionally reduce model generalization capability across heterogeneous clinical datasets [9, 28].

1.9 Research Problem Formulation and Open Challenges in MRI-Based Osteoarthritis Analysis

An urgent need remains for developing robust and efficient deep learning frameworks capable of improving knee MRI analysis, anatomical segmentation, and osteoarthritis prediction while maintaining both high accuracy and computational efficiency [15, 9].

Although Magnetic Resonance Imaging (MRI) provides highly detailed visualization of knee anatomical structures and surrounding tissues, analyzing three-dimensional MRI volumes remains a challenging task because of image complexity, anatomical variability, segmentation instability, and the heterogeneous nature of osteoarthritis progression [8, 10].

1.9.1 MRI Complexity and Clinical Interpretation Challenges

MRI volumes contain large amounts of anatomical information distributed across multiple slices and complex three-dimensional structures. This complexity increases the difficulty of image interpretation, particularly in osteoarthritis analysis where pathological changes may appear subtle during the early stages of disease progression [8, 10].

In addition, osteoarthritis does not affect all patients uniformly. Structural degeneration patterns may vary considerably between individuals depending on anatomical characteristics, disease severity, and biomechanical loading conditions. Consequently, manual MRI interpretation often requires extensive concentration and long examination time, especially when clinicians must analyze large numbers of patient scans [9].

Conventional diagnostic approaches still rely heavily on visual interpretation performed by specialists according to their clinical experience. However, this process may lead to variability between observers, particularly when anatomical boundaries are unclear or MRI structures become highly complex [8].

1.9.2 MRI Image Quality and Anatomical Boundary Challenges

Another important challenge is related to MRI image quality itself. Medical images may contain noise, low contrast, intensity variations, and acquisition-related artifacts. In some cases, noise patterns may appear similar to anatomical structures, which complicates both manual interpretation and automated segmentation tasks [13, 14, 11].

Anatomical boundary ambiguity also represents a major difficulty in knee MRI analysis. Cartilage and meniscal tissues frequently present blurry and poorly defined boundaries, particularly in regions affected by osteoarthritic degeneration. This ambiguity increases segmentation difficulty and may reduce anatomical localization accuracy [3, 16].

MRI volumes may additionally suffer from partial volume effects and through-plane blurring, particularly near thin cartilage surfaces and peripheral meniscal regions. These phenomena reduce boundary precision and complicate both manual interpretation and automated segmentation consistency across MRI slices [3].

The thin and irregular three-dimensional structure of cartilage, combined with patient-specific anatomical variability and osteoarthritis severity differences, further increases the complexity of accurate MRI segmentation and anatomical analysis [9, 8].

MRI data may additionally present significant variability between acquisition protocols, scanners, hospitals, and patient populations. Such variability introduces domain-shift problems that may reduce the robustness and generalization capability of deep learning models across different clinical environments [16, 6].

1.9.3 Computational and Segmentation Challenges

Another important challenge in knee MRI analysis is the high computational complexity associated with processing large three-dimensional MRI volumes. Medical image segmentation and feature extraction tasks often require extensive computational resources and long training times, particularly when using transformer-based deep learning architectures [6, 17].

Another limitation concerns the availability of large annotated medical datasets. Accurate MRI segmentation requires expert manual annotations, which is a time-consuming and expensive process. Consequently, many medical imaging datasets remain relatively limited compared to natural image datasets, which may affect model generalization capability [3, 15].

Within artificial intelligence systems, segmentation models are particularly sensitive to image quality variations. When MRI volumes are unclear or poorly enhanced, models may struggle to detect accurate anatomical boundaries, especially in small and complex

regions such as meniscal tissues. In some cases, background regions may also be confused with real anatomical structures, leading to unstable segmentation outputs and inaccurate feature extraction results [15, 3]

Direct multi-class segmentation approaches may also suffer from the influence of large background regions surrounding the knee joint. Processing the entire MRI volume may reduce segmentation stability and increase the difficulty of accurately localizing fine anatomical structures. Therefore, coarse-to-fine segmentation strategies based on preliminary knee localization and ROI extraction have become increasingly important for improving segmentation robustness and reducing irrelevant background interference [6, 17, 16].

1.9.4 Feature Extraction and Classification Challenges

Feature extraction represents another major challenge in MRI-based osteoarthritis analysis. Traditional handcrafted features alone may fail to fully represent the complex structural information contained in three-dimensional MRI volumes, particularly within small anatomical regions. On the other hand, deep features learned automatically by neural networks, despite their strong representation capability, may not always capture all structural variations related to age, gender, or patient-specific anatomical characteristics [12, 15].

Furthermore, deep feature representations may sometimes lack direct clinical interpretability because learned embeddings are often highly abstract and difficult to associate with explicit anatomical biomarkers.

In addition, osteoarthritis progression does not affect all knee anatomical structures uniformly. Degeneration patterns may differ between medial and lateral compartments as well as between cartilage and meniscus tissues. Consequently, extracting global features from the entire knee region may reduce sensitivity to localized pathological changes and anatomical asymmetries associated with disease progression [9, 12].

In addition, inter-compartmental relationships and medial-lateral anatomical asymmetry may provide important biomarkers associated with osteoarthritis progression and disease severity assessment.

Therefore, **combining handcrafted features, deep embeddings, intensity descriptors, texture features, and morphological characteristics** has become an important strategy for obtaining more comprehensive and discriminative MRI representations [12, 16].

Texture descriptors including GLCM, Haralick features, and Local Binary Pattern (LBP) representations are frequently employed to characterize tissue heterogeneity and

local microstructural degeneration patterns.

Another open challenge concerns osteoarthritis classification itself. Most previous studies focused either on binary classification or simplified grading strategies, whereas the full five-class Kellgren–Lawrence grading system remains considerably more difficult because adjacent OA grades frequently exhibit highly similar structural characteristics. Distinguishing subtle anatomical differences between neighboring grades therefore remains a challenging problem for automated prediction systems [9, 15].

1.9.5 Need for an Integrated Framework

Many previous studies focused on isolated tasks such as preprocessing, segmentation, or classification independently without fully integrating all analysis stages within a unified MRI analysis framework. Consequently, there remains a need for robust integrated frameworks capable of combining preprocessing, coarse-to-fine segmentation, region-wise feature extraction, multimodal feature fusion, and osteoarthritis prediction within a single interconnected pipeline [16, 6, 17].

Developing clinically applicable osteoarthritis analysis systems additionally requires balancing segmentation accuracy, computational efficiency, robustness, and interpretability in order to facilitate reliable deployment in real-world medical environments.

Based on these challenges, this work investigates the development of an integrated and robust framework capable of improving MRI quality, stabilizing anatomical segmentation, and extracting discriminative multimodal representations from knee MRI volumes in order to improve osteoarthritis analysis and disease prediction performance [9, 15].

1.10 motivation of the proposed framework

To overcome the limitations of existing state-of-the-art approaches, this work proposes an integrated MRI-based osteoarthritis analysis framework that combines preprocessing, segmentation, handcrafted feature extraction, deep feature learning, and classification within a unified pipeline. The proposed framework also investigates coarse-to-fine segmentation strategies in order to reduce background interference, improve anatomical localization stability, and facilitate more reliable feature extraction from localized knee regions.

The proposed approach also explores cascade-based analysis strategies in order to improve anatomical localization and disease prediction performance while handling the complexity of knee MRI data [6, 17, 16, 27].

1.11 Conclusion

We have presented the theoretical background and literature review related to knee osteoarthritis analysis using MRI images. The discussion covered the main concepts associated with osteoarthritis progression, MRI biomarkers, medical image preprocessing techniques, deep learning-based segmentation methods, transformer architectures, feature extraction strategies, and osteoarthritis classification approaches.

Particular attention was given to the challenges associated with MRI analysis, including image noise, low contrast, anatomical complexity, and the difficulty of accurately segmenting small anatomical structures such as cartilage and meniscal tissues. Different preprocessing techniques including wavelet denoising, transformer-based image restoration, gamma correction, and CLAHE enhancement were reviewed in order to highlight their importance in improving MRI image quality and facilitating more reliable deep learning analysis.

We have also reviewed the evolution of deep learning models for medical image segmentation, starting from convolutional neural networks and U-Net architectures to more advanced transformer-based models such as UNETR and Swin-UNETR. The limitations of CNN-based local feature learning and the advantages of transformer architectures in capturing long-range spatial dependencies were also discussed.

In addition, we have explored different feature extraction techniques used for osteoarthritis prediction, including handcrafted radiomics features, texture analysis, morphology-based descriptors, and deep feature embeddings. The importance of combining handcrafted and deep features within a multimodal analytical framework was also emphasized.

Several recent related works were analyzed in order to identify the strengths and limitations of existing osteoarthritis analysis systems. The literature review revealed that most previous studies focused mainly on isolated tasks such as segmentation or classification independently, while limited attention was given to integrated MRI-based frameworks capable of combining preprocessing, region-wise segmentation, multimodal feature extraction, feature fusion, and classification within a unified analysis pipeline.

Based on these observations, the next chapter introduces the proposed methodology through presenting the complete framework adopted for knee MRI analysis, including preprocessing techniques, the proposed coarse-to-fine two-stage segmentation strategy, ROI extraction, multimodal feature extraction, feature fusion, and osteoarthritis classification under both three-class and five-class grading settings.

CHAPTER II

Proposed Framework for 3D MRI Preprocessing, Segmentation, and OA Prediction

2.1 Introduction

After presenting the theoretical background and reviewing the main related works in the previous chapter, this chapter introduces the proposed methodology adopted for knee MRI analysis and osteoarthritis prediction.

The proposed framework was designed to address several challenges associated with MRI data, including image quality degradation, anatomical complexity, segmentation instability, and the difficulty of extracting reliable representations from three-dimensional medical images.

To overcome these limitations, this work proposes an integrated pipeline that combines preprocessing, two-stage segmentation, multimodal feature extraction, feature fusion, and classification within a unified framework.

In this chapter, we describe the dataset, preprocessing pipeline, segmentation models, feature extraction strategy, classification approach, evaluation metrics, and implementation details adopted during the experiments.

2.2 Overview of the Proposed Framework

Based on the limitations and research gaps discussed in Chapter 1, we propose an integrated MRI-based framework for knee osteoarthritis analysis. The proposed methodology combines preprocessing, two-stage segmentation, handcrafted and deep feature extraction, feature fusion, and classification within a unified pipeline.

The framework was specifically designed to address several practical limitations observed in previous studies, including low MRI visibility, unstable segmentation caused by large background regions, computational limitations of 3D MRI analysis, and the difficulty of distinguishing subtle KL grade variations.

The ultimate aim is to support osteoarthritis (OA) classification and progression assessment. Since knee MRI volumes are complex and often suffer from low contrast, noise, and illumination variations, several processing stages were designed to improve image quality and facilitate the analysis process.

The proposed methodology combines preprocessing, two-stage segmentation, multimodal handcrafted feature extraction, feature fusion, and classification within a unified pipeline.

Initially, MRI volumes undergo preprocessing operations including slice extraction, normalization, and region preparation. The preprocessed images are then passed through

a two-stage segmentation framework. In the first stage, a binary segmentation model identifies cartilage and meniscus regions from surrounding tissues. In the second stage, a multiclass segmentation model separates the detected structures into five anatomical compartments: femoral cartilage (Femoral), medial tibial cartilage (TibMed), lateral tibial cartilage (TibLat), medial meniscus (MedMen), and lateral meniscus (LatMen).

Following segmentation, multimodal handcrafted feature extraction is performed independently for each segmented anatomical region. The extracted descriptors include intensity-based statistics, GLCM texture features, shape and morphological characteristics, thickness measurements, frequency-based features obtained from FFT and wavelet analysis, as well as clinical variables [12, 8].

Region-wise feature computation helps preserve local anatomical information and improves the representation of osteoarthritic changes occurring across different knee compartments.

Because osteoarthritis progression involves structural, textural, morphological, and clinical variations, multiple complementary feature categories were combined in order to obtain a comprehensive representation of knee MRI characteristics before classification.

The extracted feature vectors consisted of intensity, texture, morphological, thickness, FFT, wavelet, and clinical descriptors obtained from the segmented anatomical regions. These features were subsequently fused and used as input for machine learning classifiers to perform osteoarthritis severity prediction according to the Kellgren–Lawrence grading system.

A schematic overview of the complete framework is illustrated in Figure 2.1.

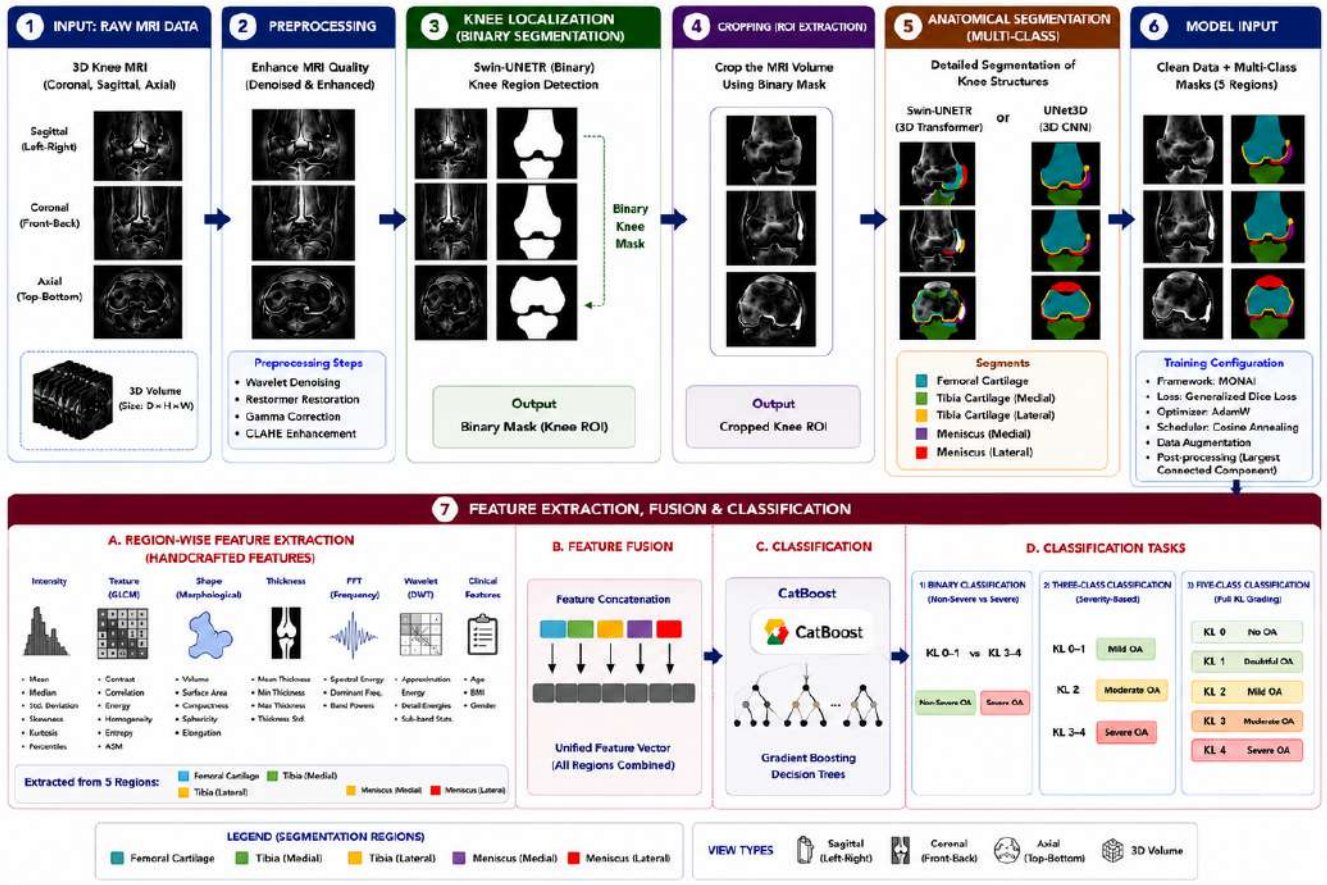


Figure 2.1: The proposed multi-stage framework for knee MRI analysis and osteoarthritis classification.

The workflow begins with a preprocessing stage designed to improve MRI image quality before model training. Multiple enhancement techniques are applied, including wavelet-based denoising, Restormer-based image restoration, gamma correction, and CLAHE contrast enhancement [13, 19, 20, 14]. These methods help reduce noise, improve structural visibility, and normalize image intensity.

After preprocessing, a binary segmentation model is first employed to identify the approximate knee region within the MRI volume. This step was introduced because preliminary experiments showed that large background regions negatively affected the training process and reduced segmentation efficiency.

Based on the detected knee location, a cropping stage is performed to isolate the knee joint and remove unnecessary background information [29]. The cropped MRI volumes are then used as input for the second segmentation stage. Focusing only on the knee region allows the model to better capture anatomical structures and improves region segmentation.

For the segmentation stage, two deep learning architectures are considered within the proposed framework: UNet3D and Swin-UNETR [3, 6]. Both models follow the same

preprocessing and training pipeline, allowing a consistent comparison between CNN-based and transformer-based segmentation approaches.

UNet3D is included due to its effectiveness in volumetric medical image segmentation and its ability to capture local spatial information. In contrast, the Swin-UNETR model incorporates transformer-based self-attention mechanisms to capture both local structural details and global contextual information within MRI volumes [16, 6].

Following segmentation, multimodal feature extraction is performed independently for each segmented region. The extracted features include intensity-based statistics, texture descriptors such as GLCM and LBP, morphological characteristics, and deep feature embeddings derived from the trained model [12, 15]. Region-wise feature computation helps preserve spatial variations between anatomical structures and improves representation of osteoarthritic changes.

Relying on a single feature representation was considered insufficient because osteoarthritis progression involves both structural and textural variations distributed across different anatomical regions. Therefore, the proposed framework combines handcrafted descriptors and deep feature embeddings in order to obtain a more comprehensive representation of knee MRI characteristics.

Finally, the extracted features are combined through a feature fusion stage before being used for osteoarthritis classification.

In addition to conventional binary analysis, the proposed framework investigates both three-class and five-class osteoarthritis classification scenarios based on the KL grading system. This strategy was adopted to better analyze disease progression and evaluate the capability of deep learning models to distinguish subtle differences between neighboring OA severity levels.

The three-class configuration provides a simplified representation that may improve classification stability and support rapid clinical assessment, whereas the five-class setting represents a more challenging and clinically detailed evaluation scenario.

The proposed framework integrates all processing stages into a unified pipeline for knee MRI analysis.

2.3 MRI Dataset Description

The proposed framework was developed using the OAIZIB-CM dataset obtained from the HuggingFace platform [30, 31]. The dataset is derived from the Osteoarthritis Initiative (OAI) database and consists of 507 three-dimensional knee MRI volumes corresponding to approximately 81,120 MRI slices acquired using the Dual-Echo Steady-State (DESS)

MRI sequence.

The MRI data cover different osteoarthritis severity levels according to the Kellgren–Lawrence (KL) grading system, with a strong tendency toward moderate and severe osteoarthritis cases. In addition, the dataset contains manually refined anatomical segmentation masks generated by experienced annotators, making it particularly suitable for segmentation and osteoarthritis prediction tasks [30].

Each MRI volume is stored in NIfTI (.nii) format and represents a complete volumetric knee scan. Since the dataset is three-dimensional, the MRI volumes can be visualized using different orthogonal anatomical planes, including axial, coronal, and sagittal views. These complementary views provide detailed spatial information about cartilage, bones, and surrounding anatomical structures. Most MRI volumes have approximate dimensions of $160 \times 384 \times 384$ [30].
Figure 2.2 illustrates the three orthogonal anatomical planes extracted from a 3D MRI volume. These complementary views provide detailed visualization of knee anatomical structures and highlight the volumetric complexity of MRI data used for segmentation and osteoarthritis analysis.

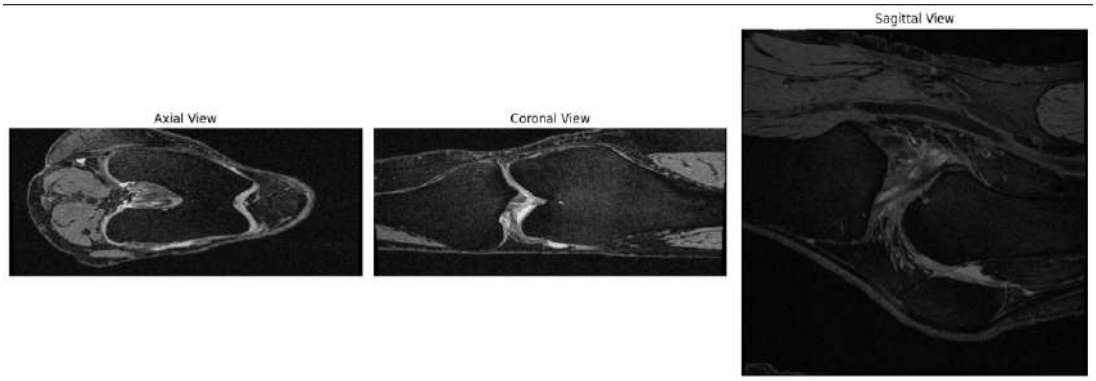


Figure 2.2: Visualization of a 3D knee MRI volume through axial, coronal, and sagittal anatomical planes extracted from the OAIZIB-CM dataset.

The dataset is organized into multiple directories containing MRI volumes, segmentation masks, and metadata files:

- **imagesTr**: training MRI volumes.
- **imagesTs**: testing MRI volumes.
- **labelsTr**: segmentation masks corresponding to training images.
- **labelsTs**: segmentation masks corresponding to testing images.
- **info**: metadata files and clinical annotations including KL grades.

The initial training set contained 404 MRI volumes, while the testing set contained 103 MRI volumes. However, several samples were excluded because corresponding KL grade annotations were missing from the metadata files. After the cleaning stage, 22 training samples and 5 testing samples were removed.

Each MRI volume is associated with a corresponding segmentation mask having identical spatial dimensions. The segmentation labels include several anatomical regions such as femoral cartilage, tibial cartilage, and meniscal structures used throughout the segmentation and feature extraction stages.

The remaining MRI volumes and segmentation masks were subsequently used for preprocessing, segmentation, feature extraction, and osteoarthritis classification experiments within the proposed framework.

The OAIZIB-CM dataset was selected because it provides volumetric MRI information, detailed anatomical segmentation masks, and clinically relevant osteoarthritis grading labels. These characteristics make the dataset suitable for region-wise segmentation, multimodal feature extraction, and osteoarthritis prediction tasks within the proposed framework.

2.4 Knee Anatomical Complexity and Segmentation Challenges

The knee joint is composed of several anatomical structures that appear differently in MRI images and perform different biomechanical functions inside the joint. In this work, the analyzed regions include the femoral cartilage (Femoral), medial tibial region (TibMed), lateral tibial region (TibLat), medial meniscus (MedMen), and lateral meniscus (LatMen) [8, 3].

These structures do not share identical anatomical characteristics. Some regions occupy relatively large areas inside the MRI volume, whereas others are smaller and contain fine structural details. Consequently, segmentation difficulty varies considerably between anatomical regions.

The femoral and tibial regions are generally easier to identify because they occupy larger spatial areas and their anatomical boundaries are more visible across neighboring MRI slices. Their structures also remain relatively stable through adjacent slices, allowing segmentation models to learn their representations more effectively.

In contrast, meniscal regions, particularly MedMen and LatMen, are considerably more difficult to segment. These structures are thinner, smaller, and often exhibit weak or partially unclear anatomical boundaries. As a result, separating meniscal tissues from

surrounding anatomical structures becomes more challenging.
Figure 2.3 presents an example of multi-planar knee MRI segmentation illustrating the major anatomical structures.

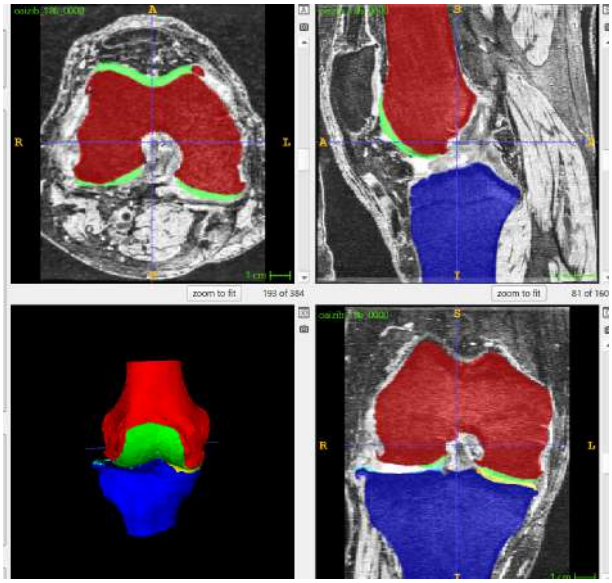


Figure 2.3: Example of multi-planar knee MRI segmentation illustrating major anatomical structures. The red regions represent femoral structures, the blue regions correspond to tibial structures, while the green regions indicate cartilage and meniscal anatomical components.

Another challenge arises from natural anatomical variability between patients. Knee structures may vary in shape, size, thickness, and visibility depending on factors such as age, gender, body morphology, and osteoarthritis severity. Therefore, the same anatomical region may present significantly different appearances across MRI volumes.

Osteoarthritis progression itself further increases segmentation complexity because cartilage degeneration and meniscal deterioration may progressively alter tissue appearance and reduce anatomical boundary visibility.

MRI quality also plays an important role in segmentation difficulty. Some MRI slices contain noise, low contrast, illumination inconsistencies, or intensity variations that reduce the visibility of small anatomical details, particularly within meniscal tissues [3].

2.5 Two-Stage Transformer-Based Knee MRI Segmentation

Due to the high complexity of knee MRI volumes, direct multi-class segmentation was not sufficiently stable during the early stages of experimentation. Initially, a multi-class

segmentation model was trained to directly segment the six anatomical regions from the complete MRI volume. Although the model achieved relatively acceptable Dice scores, several important issues were observed during visual inspection of the predictions. The model occasionally confused background regions with anatomical structures as exposed in figure 2.4 , and smaller regions such as the menisci were not segmented consistently, particularly near thin boundaries and peripheral areas.

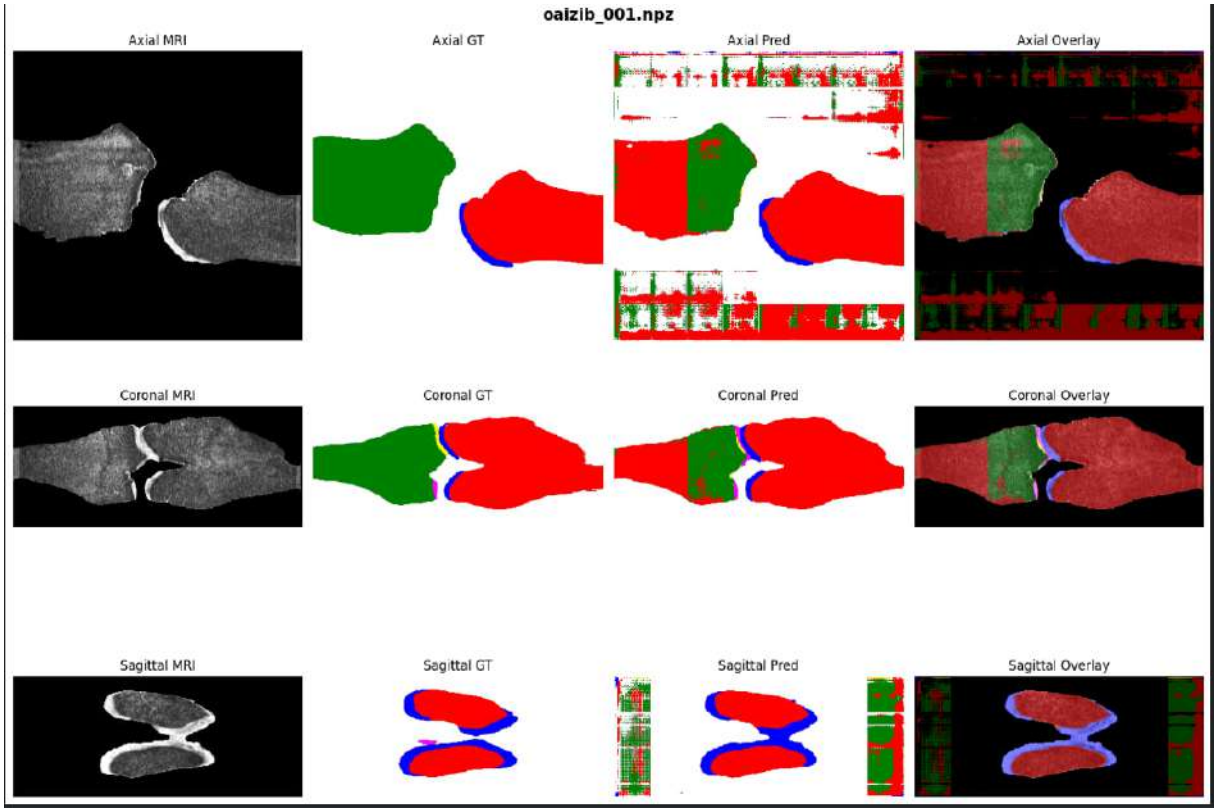


Figure 2.4: Example of direct multi-class segmentation failure on full MRI volumes before applying the proposed two-stage framework. The model occasionally confused background regions with anatomical structures, resulting in noisy predictions and unstable segmentation outputs.

Another issue appeared during the cropping stage. Since the predicted masks occasionally included irrelevant background regions, the generated cropped volumes were not always anatomically stable. In some cases, parts of the background were interpreted as part of the knee region, which negatively affected the quality of the cropped ROI. As a consequence, the second-stage model sometimes learned from noisy or partially corrupted regions.

To address these limitations, the segmentation pipeline was redesigned into a two-stage framework as illustrated in figure 2.5. Instead of directly performing detailed anatomical segmentation on the full MRI volume, the first stage was dedicated to knee localization using binary segmentation, while the second stage focused on detailed multi-class

anatomical segmentation after removing irrelevant background information.

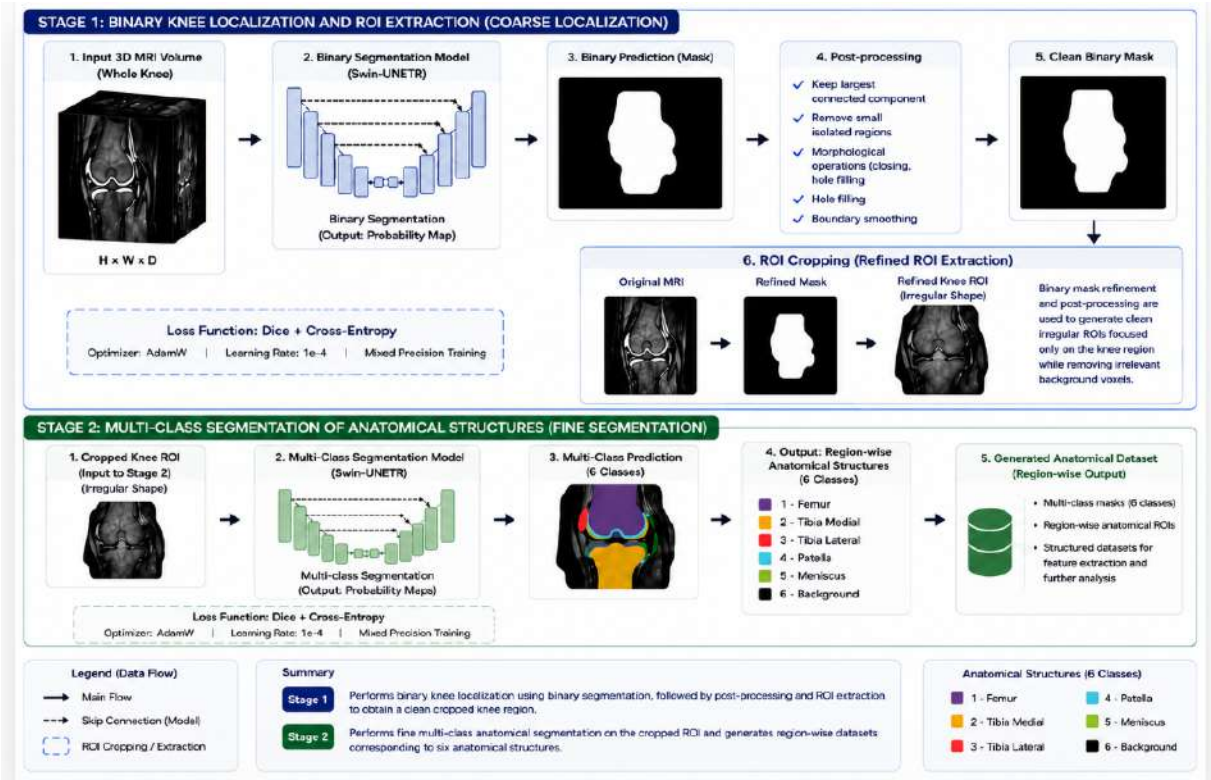


Figure 2.5: Overview of the proposed two-stage segmentation framework for knee MRI analysis. The first stage performs binary knee localization and ROI extraction, while the second stage performs detailed multi-class segmentation of anatomical structures.

2.5.1 Stage 1: Binary Knee Localization

In the first stage, a Swin-UNETR architecture was employed for binary segmentation. The model was trained to distinguish only between the knee region and the surrounding background. This formulation simplified the segmentation problem and allowed the network to learn the global structure of the knee more efficiently compared to direct six-class segmentation.

Swin-UNETR was selected because Transformer-based self-attention mechanisms are capable of modeling long-range spatial relationships inside 3D MRI volumes [6, 16]. This capability is particularly important in knee MRI analysis, where anatomical structures may vary in shape, orientation, intensity distribution, and spatial appearance across subjects. In addition, self-attention allows the model to simultaneously capture both global contextual information and fine local details, which is beneficial for small and difficult structures such as the menisci.

Pretrained weights were also utilized during some training stages in order to improve convergence stability and reduce the time required for optimization.

Instead of directly performing multi-class segmentation on the complete MRI volume, all anatomical labels were first merged into a single foreground class representing the knee region. The binary mask generation process can be formulated as:

$$M_{binary}(x) = \begin{cases} 1, & \text{if } M(x) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

where $M(x)$ represents the original segmentation mask and $M_{binary}(x)$ denotes the generated binary mask.
Figure 2.6 shows the conversion process of an original multi-class segmentation mask into a binary mask.

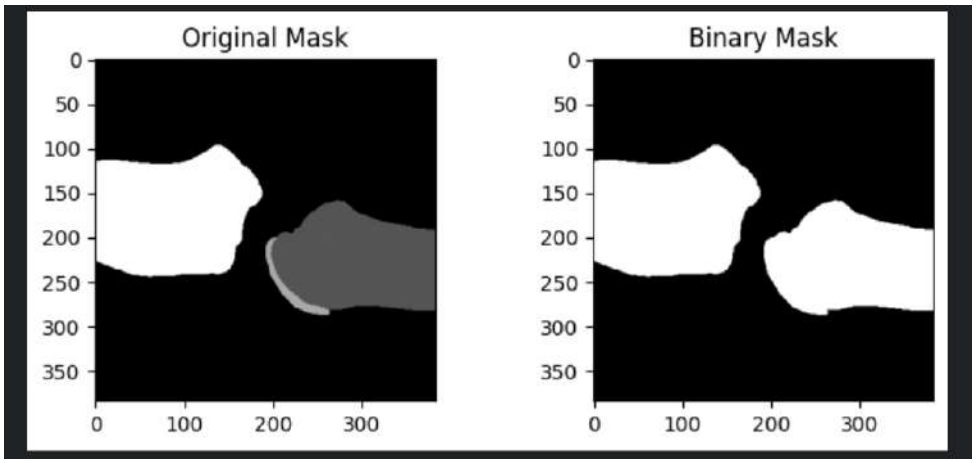


Figure 2.6: Conversion of the original multi-class segmentation mask into a binary mask used for knee localization.

After binary prediction, the generated masks occasionally contained isolated noisy regions and internal holes caused by prediction uncertainty. Since the masks were generated automatically by the model rather than obtained directly from the ground truth labels, a post-processing stage was introduced before ROI extraction.

2.5.2 Stage 2: Fine Multi-Class Anatomical Segmentation

In the second stage, a multi-class segmentation model was trained on the cleaned and cropped knee region. The segmentation task focused on the following anatomical classes:

- Background
- Femoral
- Tibia Med

- Tibia Lat
- Med Meniscus
- Lat Meniscus

Figure 2.7 illustrates the segmentation masks (ground truth) of the anatomical regions, used for training the model of multi-class anatomical segmentation.

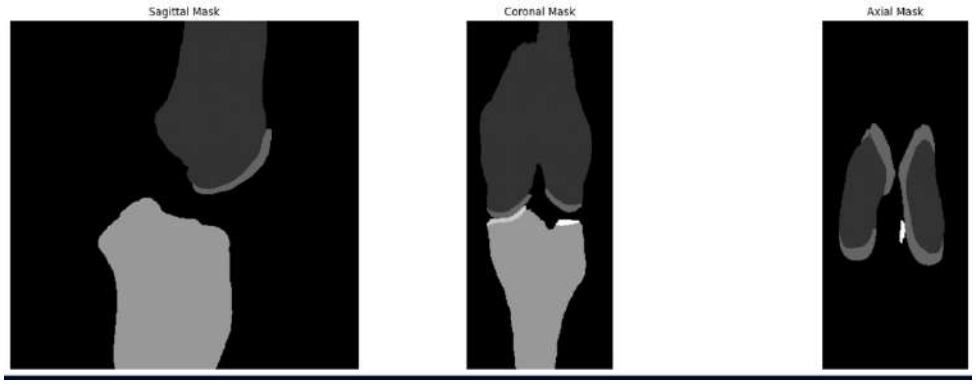


Figure 2.7: Example of mask showing different anatomical structures in Sagittal, Coronal, and Axial views.

Figure 2.8 shows the same masks with a highlight of the anatomical structures.

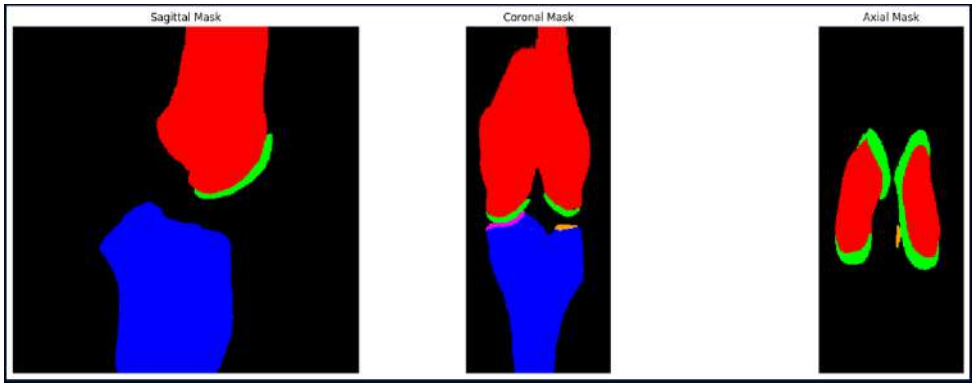


Figure 2.8: Example of multi-class anatomical segmentation masks showing different anatomical structures in Sagittal, Coronal, and Axial views.

The proposed framework focuses on the segmentation and analysis of the principal anatomical structures involved in knee osteoarthritis. These structures include the femur, tibia, femoral cartilage, tibial cartilage, medial meniscus, and lateral meniscus. Figure 2.9 presents a three-dimensional (3D) anatomical reconstruction of these structures, illustrating the regions considered throughout the segmentation and feature extraction stages.

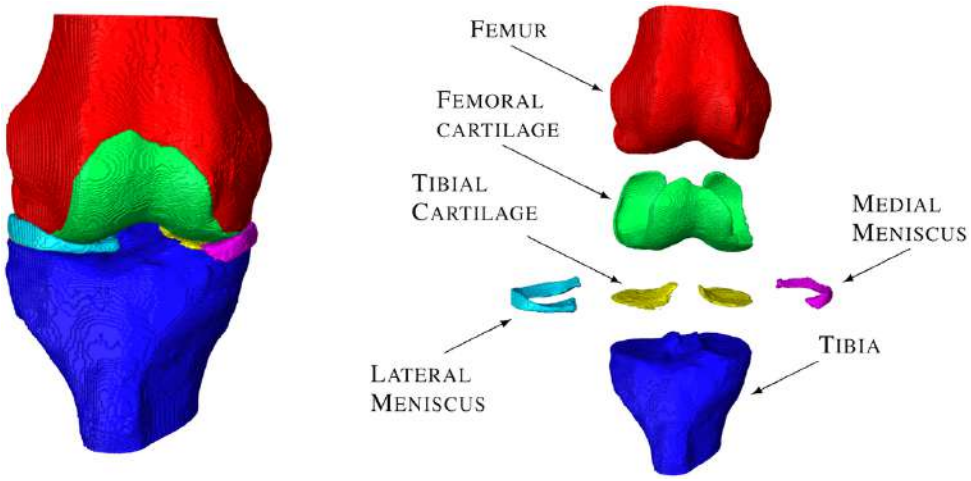


Figure 2.9: Three-dimensional visualization of the anatomical structures segmented in this study, including the femur, tibia, femoral cartilage, tibial cartilage, medial meniscus, and lateral meniscus. Adapted from [7].

Training on localized knee regions improved anatomical consistency and allowed the model to better learn fine structural boundaries compared to the initial direct segmentation experiments.

The binary mask was also reused during the second stage itself in order to suppress residual noisy regions before training. Only regions located inside the detected knee area were preserved, which improved training stability and reduced the influence of irrelevant background voxels on small anatomical structures.

In addition to Swin-UNETR, the same pipeline was later applied to UNet3D in order to compare Transformer-based segmentation with convolution-based segmentation while preserving identical preprocessing, cropping, and training procedures.

2.5.3 Training Strategy

Because of the large spatial dimensions of 3D knee MRI volumes, full-volume training was computationally expensive and exceeded the available GPU memory limits in Kaggle and Google Colab environments, especially for Transformer-based architectures.

Initially, larger patch sizes such as $128 \times 128 \times 128$ were tested in order to preserve larger anatomical context. However, memory consumption became excessively high, training speed decreased significantly, and training sessions became unstable because of GPU memory limitations. Therefore, a patch size of $96 \times 96 \times 96$ was selected as a compromise between preserving anatomical information and maintaining stable GPU memory utilization.

A patch-based training strategy was therefore adopted instead of processing the com-

plete MRI volume at once. This approach enabled stable training of 3D segmentation models with lower computational cost.

Rather than extracting patches completely randomly, the training process preferentially sampled patches containing foreground anatomical regions in order to reduce the proportion of empty background regions and increase the probability of learning meaningful anatomical structures.

A dedicated smart sampling strategy was also introduced for meniscus regions. Compared to larger bone structures, menisci were considerably more difficult to segment because of their smaller size, weak boundaries, and similarity to surrounding tissues. Consequently, part of the extracted patches was intentionally centered around meniscus regions in order to expose the model to more difficult anatomical examples during training.

Mixed precision training was employed to reduce GPU memory usage and accelerate optimization, particularly for large Transformer-based 3D models.

The training process used the AdamW optimizer together with the CosineAnnealingLR scheduler in order to improve convergence stability and reduce optimization oscillations during training [32].
Table 2.1 illustrates the main training parameters used for segmentation experiments.

Table 2.1: Main training parameters used for segmentation experiments.

Parameter	**Value**
Patch Size	$96 \times 96 \times 96$
Batch Size	1
Optimizer	AdamW
Scheduler	CosineAnnealingLR
Training Precision	Mixed Precision
Inference Strategy	Sliding Window Inference
Loss Functions	DiceCELoss / DiceFocalLoss
Augmentation	Online Augmentation (32,300 virtual augmented samples)

Several augmentation techniques were dynamically applied during training as illustrated in the table 2.2

Table 2.2: Online data augmentation techniques and parameters used during training.

Augmentation Technique	**Probability**	**Parameters / Configuration**
Random Flip	0.5	Applied independently along sagittal, coronal, and axial axes
Random Rotation 90°	0.5	Random 90-degree rotation across spatial dimensions
Random Rotation (3D)	0.3	Rotation range: $x, y, z \in [-0.15,\ 0.15]$ radians
Random Zoom	0.2	Zoom factor between 0.95 and 1.05
Gaussian Noise	0.2	Standard deviation = 0.01
Gaussian Smoothing	0.15	$\sigma_x \in [0.3,\ 0.7]$
Intensity Scaling	0.3	Scaling factor = $\pm 15\%$
Intensity Shifting	0.3	Offset = ± 0.1
Contrast Adjustment	0.3	Gamma range = $[0.85,\ 1.15]$

Figure 2.10 shows an example of the applied process of online data augmentation.

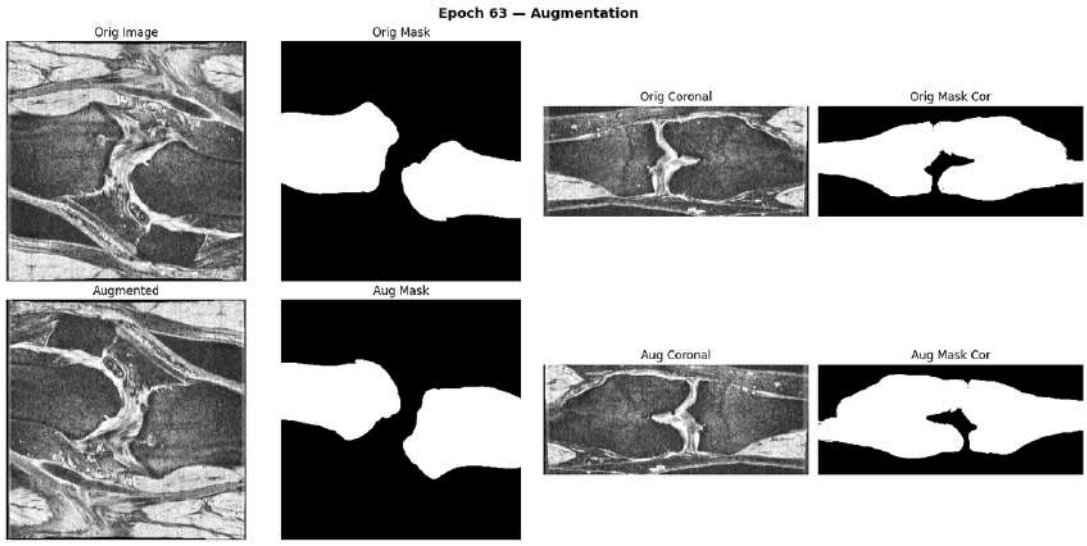


Figure 2.10: Example of online data augmentation applied during training. Different spatial and intensity transformations are dynamically generated to improve data diversity and reduce overfitting while preserving anatomical consistency.

Online augmentation was preferred over offline augmentation because 3D MRI volumes require large storage capacity. Generating and storing additional augmented datasets offline would considerably increase storage requirements and computational overhead. Therefore, augmentations were generated dynamically during each training iteration, al-

lowing the model to observe different variations of the same MRI volume across epochs without increasing disk usage [33].

During early experiments, stronger augmentation settings were tested. However, excessive transformations occasionally distorted anatomical structures and negatively affected the learning of fine segmentation boundaries. Consequently, lighter augmentations were selected in order to preserve realistic anatomical characteristics while still improving data diversity.

After obtaining stable initial results, fine-tuning strategies were introduced instead of restarting training entirely from scratch. Previously saved checkpoints were loaded and training continued using modified settings such as adjusted learning rates or different loss configurations. This strategy accelerated optimization and preserved previously learned anatomical representations.

Checkpoint saving was also integrated because Kaggle and Google Colab sessions are time-limited. This allowed interrupted training sessions to resume from the last saved epoch without losing previous optimization progress.

During inference, sliding-window inference was employed to process complete MRI volumes while maintaining stable GPU memory consumption and preserving spatial consistency across the 3D volume.

2.5.4 Loss Functions

Several loss functions were investigated during the project in order to improve segmentation stability and enhance the learning of difficult anatomical structures.

Initially, DiceCELoss was employed because it provides a balance between overlap optimization and voxel-wise classification stability [34].

Later, DiceFocalLoss was evaluated in an attempt to improve the segmentation of small structures such as the menisci. Although DiceFocalLoss occasionally improved some small anatomical regions, training stability was not always superior to DiceCELoss. Consequently, multiple configurations continued to be explored in order to identify the best compromise between segmentation quality and optimization stability.

Class weights were also introduced to assign higher importance to small anatomical structures compared to background regions and large bones. This strategy helped reduce the impact of class imbalance during training.

During experimentation, it was observed that the loss value occasionally remained relatively high even when Dice and IoU scores improved considerably. This behavior was mainly related to the nature of medical multi-class segmentation, particularly for small

structures such as the menisci.

Small anatomical regions are highly sensitive to minor boundary errors, and the use of class weights further increased the contribution of these regions to the overall loss computation. Consequently, even small segmentation inaccuracies produced noticeable increases in loss values while Dice scores remained relatively high because of improved global overlap.

In addition, Dice-based loss functions such as DiceCELoss and DiceFocalLoss do not necessarily decrease in the same manner as conventional classification losses. Therefore, model evaluation did not rely exclusively on loss values, and Dice as well as IoU metrics were continuously monitored to better reflect segmentation quality.

2.5.5 Evaluation Metrics

The segmentation models were evaluated using Dice Score and Intersection over Union (IoU), which are among the most widely used metrics in medical image segmentation tasks.

Dice-based metrics are particularly important in medical image segmentation because they directly evaluate overlap quality between predicted and ground-truth anatomical regions, especially in highly imbalanced segmentation problems.

The Dice coefficient is defined as:

$$Dice = \frac{2|P \cap G|}{|P| + |G|} \tag{2.2}$$

The Intersection over Union (IoU) metric is defined as:

$$IoU = \frac{|P \cap G|}{|P \cup G|} \tag{2.3}$$

The Precision metric is defined as:

$$Precision = \frac{TP}{TP + FP} \tag{2.4}$$

The Recall metric is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{2.5}$$

where P represents the predicted segmentation mask, G represents the ground-truth segmentation mask, TP represents true positives, FP represents false positives, and FN

represents false negatives.

Dice and IoU scores were computed independently for each anatomical structure because large regions such as Femoral and Tibia are generally easier to segment compared to small meniscus regions. This per-class evaluation provided a more detailed understanding of model behavior instead of relying only on global average scores.

In addition to Dice and IoU, Precision and Recall metrics were also monitored during experimentation in order to evaluate false positive and false negative segmentation behavior across different anatomical regions.

Loss values were continuously monitored throughout training in order to evaluate optimization stability and progressively refine the training configuration during experimentation.

After obtaining anatomically consistent segmentation masks, the segmented anatomical regions were subsequently used for multimodal feature extraction and osteoarthritis analysis. The extracted regions preserve localized structural information and provide the basis for quantitative characterization of osteoarthritic degeneration patterns across different knee compartments.

2.6 Region-Wise Feature Extraction Strategy

Following the segmentation stage, handcrafted features were extracted independently from each anatomical region in order to characterize tissue properties associated with knee osteoarthritis. Feature extraction was performed separately for the femoral cartilage (Femoral), medial tibial cartilage (TibMed), lateral tibial cartilage (TibLat), medial meniscus (MedMen), and lateral meniscus (LatMen).

Feature extraction represents an important stage in MRI analysis because raw image intensities alone may not adequately describe structural alterations associated with osteoarthritis. Therefore, segmented MRI regions were transformed into quantitative descriptors that facilitate machine learning-based analysis and improve the characterization of anatomical changes associated with disease progression [12, 15].

The proposed framework adopts a region-wise feature extraction strategy in which imaging descriptors are computed independently for each segmented anatomical compartment. This approach preserves local anatomical information and allows each region to contribute individually to the final classification process.

Multiple complementary feature categories were extracted from each anatomical region, including intensity, texture, shape, thickness, frequency-based, and clinical descriptors. These heterogeneous representations provide information regarding tissue composi-

tion, structural organization, anatomical morphology, and patient-specific characteristics.

The extracted descriptors were subsequently combined into a unified feature representation used for osteoarthritis classification [8, 12].

2.6.1 Intensity Features

Intensity-based descriptors were extracted from each segmented anatomical region in order to characterize the distribution of MRI signal values within the region of interest. These descriptors provide quantitative information regarding tissue brightness, signal variability, and intensity heterogeneity.

Several first-order statistical measurements were computed independently for each anatomical compartment, including mean intensity, standard deviation, minimum value, maximum value, median, percentile values, skewness, kurtosis, and interquartile range (IQR). These features summarize the intensity characteristics of the analyzed tissues and provide complementary information regarding local signal distribution [12].

The mean intensity is defined as:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2.6}$$

where x_i represents the intensity value of voxel i and N denotes the total number of voxels inside the analyzed region.

The standard deviation is computed as:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2} \tag{2.7}$$

where μ denotes the mean intensity value.

Skewness was additionally computed to evaluate the asymmetry of the intensity distribution:

$$\text{Skewness} = \frac{1}{N\sigma^3} \sum_{i=1}^{N} (x_i - \mu)^3 \tag{2.8}$$

Kurtosis was extracted to quantify the peakedness of the intensity distribution:

$$\text{Kurtosis} = \frac{1}{N\sigma^4} \sum_{i=1}^{N} (x_i - \mu)^4 \tag{2.9}$$

In addition, percentile-based measurements were extracted in order to provide robust descriptors of the intensity distribution and reduce sensitivity to extreme intensity values. Together, these first-order statistical descriptors provide a quantitative representation of tissue appearance and signal characteristics within each anatomical structure.

2.6.2 Texture Features

Texture descriptors were extracted using the Gray-Level Co-occurrence Matrix (GLCM) method. Unlike first-order intensity statistics, GLCM features characterize spatial relationships between neighboring intensity values and provide information regarding local tissue organization and structural heterogeneity.

Texture analysis is particularly useful in medical image analysis because tissue degeneration may alter internal anatomical organization before substantial intensity changes become visible. Therefore, texture descriptors provide complementary information that cannot be captured by intensity statistics alone [35].

For each anatomical region, six GLCM descriptors were computed: contrast, homogeneity, energy, correlation, dissimilarity, and angular second moment (ASM).

The Gray-Level Co-occurrence Matrix is represented as:

$$P(i, j, d, \theta) \tag{2.10}$$

where i and j denote gray-level intensities, d represents the spatial distance between neighboring pixels, and θ corresponds to the orientation angle.

Among the extracted descriptors, contrast measures local intensity variations and is defined as:

$$\text{Contrast} = \sum_{i,j} (i - j)^2 P(i, j) \tag{2.11}$$

where $P(i, j)$ denotes the normalized co-occurrence matrix.

The extracted GLCM features provide quantitative information regarding local image organization, structural complexity, and tissue heterogeneity within the analyzed anatomical compartments.

2.6.3 Shape and Morphological Features

Shape-related descriptors were extracted from each segmented anatomical structure in order to characterize anatomical geometry and spatial organization.

The extracted measurements included voxel count, voxel ratio, bounding-box dimensions, dimension ratios, centroid coordinates, fill ratio, and compactness. These descriptors provide quantitative information regarding the size, morphology, and spatial distribution of the segmented structures.

Voxel count was used as a morphological descriptor representing the size of the segmented anatomical structure:

$$N_v = \sum_{i=1}^{N} m_i \tag{2.12}$$

where m_i represents the binary mask value of voxel i and N_v denotes the total number of voxels belonging to the segmented structure.

Bounding-box dimensions were additionally extracted to characterize the spatial extent of each anatomical compartment. Dimension ratios were computed to describe relative geometric proportions, while centroid coordinates provided information regarding spatial localization.

Compactness and fill ratio were also included to quantify structural organization and occupancy characteristics of the segmented regions.

Morphological descriptors complement intensity and texture information by describing geometric characteristics directly derived from the segmentation masks [12].

2.6.4 Thickness Features

Thickness-based descriptors were extracted in order to quantify structural characteristics of the segmented anatomical regions.

For each anatomical structure, local thickness measurements were computed and summarized using several statistical descriptors, including mean thickness, minimum thickness, maximum thickness, standard deviation, and percentile-based thickness measurements.

Thickness analysis is particularly relevant in osteoarthritis studies because cartilage degeneration is frequently associated with progressive tissue thinning and structural deterioration. Consequently, thickness descriptors provide quantitative biomarkers related to tissue preservation and disease progression [36].

The mean thickness is defined as:

$$T_{mean} = \frac{1}{N} \sum_{i=1}^{N} t_i \tag{2.13}$$

where t_i denotes the local thickness value and N represents the total number of sampled thickness measurements.

The extracted thickness descriptors complement intensity and morphological features by providing direct measurements of anatomical structural integrity.

2.6.5 Frequency-Based Features

FFT Features

Frequency-domain descriptors were extracted using the Fast Fourier Transform (FFT) in order to characterize structural patterns that may not be directly observable in the spatial domain.

The Fourier transform decomposes image information into multiple frequency components and provides a representation of spatial variations and structural complexity within anatomical tissues. For each anatomical region, the FFT magnitude spectrum was computed and several statistical descriptors were extracted, including spectral mean, spectral standard deviation, maximum magnitude, and spectral energy.

These descriptors provide complementary information regarding tissue organization and image heterogeneity and may capture structural characteristics that are not fully represented by conventional intensity measurements [37].

The Fourier transform is defined as:

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)} \tag{2.14}$$

where $f(x,y)$ denotes the image intensity at spatial coordinates (x,y) and $F(u,v)$ represents the corresponding frequency-domain representation.

Wavelet Features

Wavelet-based descriptors were extracted using a single-level Haar wavelet decomposition. Unlike conventional frequency analysis, wavelet transforms preserve both spatial localization and frequency information, making them particularly suitable for medical image characterization.

For each analyzed anatomical region, four wavelet subbands (LL, LH, HL, and HH) were generated. These subbands represent approximation and detail information at different frequency levels.

For every subband, four statistical measurements were computed: mean, standard deviation, energy, and entropy. These descriptors provide multi-scale information regarding local image structures, tissue heterogeneity, and spatial intensity variations.

Wavelet descriptors complement FFT-based measurements by simultaneously capturing frequency-related and spatial information, thereby providing a richer representation of MRI tissue characteristics [38].

2.6.6 Clinical Features

In addition to imaging biomarkers extracted from segmented anatomical regions, several clinical variables were incorporated into the classification framework. These variables included patient age, body mass index (BMI), and gender.

Clinical factors are known to be associated with the onset and progression of knee osteoarthritis. Age is considered one of the strongest risk factors for cartilage degeneration, while increased BMI contributes to excessive mechanical loading of the knee joint. Gender-related differences have also been reported in the prevalence and severity of osteoarthritis.

The integration of clinical variables with imaging-derived biomarkers provides complementary patient-level information and enables a more comprehensive characterization of disease severity [39].

2.6.7 Feature Fusion

The extracted feature categories provide complementary information regarding tissue characteristics and osteoarthritic degeneration patterns. To exploit the complementary nature of these descriptors, all extracted features were concatenated into a unified feature representation:

$$F_{\text{fusion}} = [F_{\text{intensity}}, F_{\text{texture}}, F_{\text{shape}}, F_{\text{thickness}}, F_{\text{FFT}}, F_{\text{wavelet}}, F_{\text{clinical}}] \tag{2.15}$$

where $F_{\text{intensity}}$, F_{texture}, F_{shape}, $F_{\text{thickness}}$, F_{FFT}, F_{wavelet}, and F_{clinical} represent the extracted intensity, texture, shape, thickness, frequency-based, wavelet, and clinical feature vectors, respectively.

The resulting feature vector combines complementary imaging and clinical information

extracted from the femoral cartilage, medial tibial cartilage, lateral tibial cartilage, medial meniscus, and lateral meniscus. This unified representation was subsequently used as input for the osteoarthritis classification models [12].

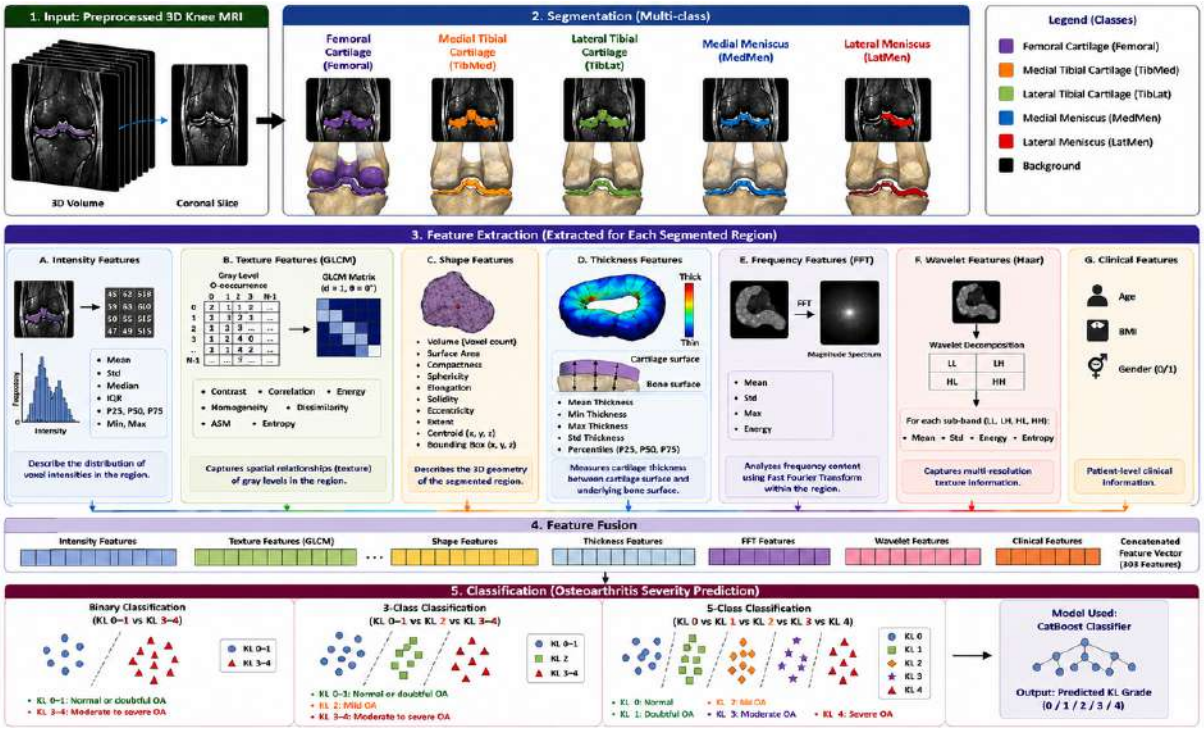


Figure 2.11: Complete framework of the proposed osteoarthritis classification system, including multiclass segmentation, region-wise feature extraction, feature fusion, and classification tasks (binary, three-class, and five-class prediction).

This unified representation was subsequently used as input for the osteoarthritis classification models. Table 2.3 summarizes the extracted feature categories and their corresponding descriptors.

Table 2.3: Summary of the extracted feature categories, anatomical regions, and descriptors.

Feature Category	**Anatomical Regions and Extracted Descriptors**
Intensity	Femoral cartilage, tibial cartilage, and menisci: mean, standard deviation, minimum, maximum, median, percentiles, skewness, kurtosis, and IQR.
Texture (GLCM)	Femoral cartilage, tibial cartilage, and menisci: contrast, homogeneity, energy, correlation, dissimilarity, and ASM.
Shape (Imaging Biomarkers)	All segmented anatomical structures: voxel count (cartilage volume), voxel ratio (tibial cartilage coverage), bounding-box dimensions (dx, dy, dz — cartilage spatial extent), dimension ratios, centroid coordinates, fill ratio, compactness, meniscal volume, mean cartilage thickness as a tissue preservation indicator, and joint space approximation derived from the relative positions of segmented femoral and tibial structures.
Thickness	Cartilage regions: mean thickness, minimum thickness, maximum thickness, standard deviation, and thickness percentiles.
FFT	All segmented regions: spectral mean, spectral standard deviation, maximum magnitude, and spectral energy.
Wavelet	All segmented regions: mean, standard deviation, energy, and entropy extracted from LL, LH, HL, and HH subbands.
Clinical	Patient-level information: age, BMI, and gender.

2.7 Osteoarthritis Classification Strategy

After feature extraction and feature fusion, the resulting feature vectors were used for osteoarthritis classification. The objective of this stage was to predict the severity of knee osteoarthritis according to the Kellgren–Lawrence (KL) grading system.

The classification framework was designed to investigate different formulations of the osteoarthritis grading problem. Instead of relying on a single prediction scenario, several classification strategies were evaluated in order to analyze disease severity from different clinical perspectives.

The extracted feature vectors consisted of intensity, texture, morphological, thickness, FFT, wavelet, and clinical descriptors computed independently from the femoral cartilage, medial tibial cartilage, lateral tibial cartilage, medial meniscus, and lateral meniscus regions. These region-wise descriptors were subsequently concatenated into a unified feature representation used as input for the classifier.

A CatBoost classifier was adopted as the final machine learning model [40]. CatBoost is a gradient boosting algorithm based on decision trees and is particularly effective for heterogeneous tabular data containing complex nonlinear relationships between features. The algorithm iteratively constructs an ensemble of decision trees, where each new tree attempts to correct the errors produced by previously learned trees.

The prediction function of CatBoost can be represented as:

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x) \tag{2.16}$$

where $h_m(x)$ denotes the prediction of the m-th decision tree, γ_m represents its corresponding weight, and M is the total number of trees in the ensemble.

CatBoost was selected because it can effectively model nonlinear interactions between imaging biomarkers and clinical variables while maintaining strong generalization performance on structured medical datasets.

2.7.1 Binary Classification Formulation

In addition to multi-class analysis, a binary classification strategy was investigated in order to distinguish between non-severe and severe osteoarthritis cases.

The original KL grades were grouped into two categories as follows:

Table 2.4: Binary osteoarthritis classification strategy.

KL Grade	Classification Class
0–1	Non-Severe OA
3–4	Severe OA

In this formulation, KL grades 0 and 1 represent knees with absent or minimal osteoarthritic signs, whereas KL grades 3 and 4 correspond to advanced osteoarthritis characterized by substantial structural degeneration. Grade KL2 was excluded from this binary setting because it represents an intermediate stage that may introduce ambiguity between healthy and severely affected cases.

This binary formulation allows the classifier to focus on clearly distinguishable disease states and provides a simplified assessment of osteoarthritis severity.

2.7.2 Three-Class Classification Formulation

To investigate intermediate disease severity, a three-class classification strategy was adopted. The original KL grades were grouped according to osteoarthritis severity as follows:

Table 2.5: Three-class osteoarthritis classification strategy.

KL Grade	Classification Class
0–1	Mild OA
2	Moderate OA
3–4	Severe OA

This grouping strategy reduces the difficulty of distinguishing highly similar neighboring KL grades while preserving clinically meaningful severity categories. Grades KL0 and KL1 were combined because both generally correspond to absent or minimal radiographic signs of osteoarthritis, whereas grades KL3 and KL4 represent advanced disease stages characterized by substantial structural degeneration and joint space narrowing. Grade KL2 was maintained as an independent category because it represents the transition toward definite osteoarthritic changes. Similar grouping strategies have been adopted in previous osteoarthritis classification studies to reduce inter-class ambiguity and improve classification stability [7].

2.7.3 Five-Class Classification Formulation

To evaluate the capability of the proposed framework to perform fine-grained osteoarthritis assessment, a five-class classification formulation was also considered.

In this setting, each KL grade was treated as an independent class:

Table 2.6: Five-class osteoarthritis classification strategy.

KL Grade	Classification Class
0	No OA
1	Doubtful OA
2	Mild OA
3	Moderate OA
4	Severe OA

The five-class formulation directly follows the original Kellgren–Lawrence grading system, where each disease severity level is treated as an independent class [41].

Unlike binary and three-class formulations, the five-class strategy preserves the complete KL grading system and therefore represents the most challenging classification scenario. The classifier must distinguish subtle structural differences between neighboring disease stages, particularly between KL1, KL2, and KL3, where radiographic and anatomical changes may overlap considerably.

The three classification formulations provide complementary perspectives for evaluating osteoarthritis severity and allow comprehensive analysis of the proposed framework under different clinical classification settings.

2.8 Conclusion

This chapter presented the proposed framework developed for automated knee MRI analysis and osteoarthritis prediction. The proposed methodology integrates preprocessing, two-stage segmentation, multimodal feature extraction, feature fusion, and classification within a unified MRI analysis pipeline.

We first introduced the OAIZIB-CM dataset and described the preprocessing pipeline adopted to improve MRI image quality. Multiple enhancement techniques including wavelet-based denoising, Restormer restoration, gamma correction, and CLAHE enhancement were sequentially combined in order to improve anatomical visibility while preserving structural consistency.

We also presented the proposed two-stage segmentation framework. A binary Transformer-based Swin-UNETR model was first employed for knee localization and ROI extraction, followed by detailed multi-class anatomical segmentation using both Swin-UNETR and UNet3D architectures.

In addition, the training strategy, augmentation procedures, loss functions, and evaluation metrics adopted during the experiments were described in detail. Particular atten-

tion was given to computational limitations associated with 3D MRI processing and the segmentation difficulty of small anatomical structures such as the menisci.

Finally, the chapter introduced the proposed region-wise multimodal feature extraction strategy, which combines intensity, texture, morphological, thickness, frequency-domain, wavelet, and clinical descriptors in order to obtain comprehensive MRI representations for osteoarthritis analysis and prediction.

The next chapter presents the experimental results and performance evaluation of the proposed framework, including preprocessing analysis, segmentation performance, feature analysis, and osteoarthritis classification results.

CHAPTER III

Performance Assessment of the Proposed Framework

3.1 Experimental Setup

In this chapter, we present the experimental results obtained using the proposed framework for automated knee MRI analysis and osteoarthritis prediction. The evaluation encompasses binary knee localization, ROI extraction, multi-class anatomical segmentation, feature extraction, and osteoarthritis classification. In addition, the obtained results are compared with previously reported studies in order to assess the effectiveness of the proposed methodology.

All experiments were conducted using the preprocessing pipeline, segmentation framework, feature extraction strategy, and classification methodology described in Chapter 2. Both quantitative and qualitative analyses were performed to evaluate the performance of each stage of the proposed framework.

3.1.1 Hardware and Software Environment

This section describes the experimental environment and configuration used to train and evaluate the proposed deep learning models. It includes details regarding the hardware resources, software frameworks, and implementation settings adopted throughout this work.

The experiments were conducted using Google Colab and Kaggle cloud platforms with GPU acceleration in order to ensure efficient training and inference of the proposed deep learning models. NVIDIA Tesla T4 GPUs with CUDA acceleration were utilized for all experiments involving 3D MRI preprocessing, segmentation, feature extraction, and classification tasks.

The proposed framework was implemented using Python programming language with PyTorch and MONAI libraries [42, 43]. Additional libraries including NumPy, OpenCV, Scikit-learn, Matplotlib, and SimpleITK were also employed for preprocessing, visualization, feature extraction, and evaluation purposes.

Table 3.1 presents the main hardware resources, software frameworks, and computational tools adopted throughout the experimental study.

Table 3.1: Hardware and software environment used in the experiments.

Component	Specification
Programming Language	Python 3.12
Deep Learning Framework	PyTorch
Medical Imaging Framework	MONAI
GPU	NVIDIA Tesla T4
Platform	Google Colab / Kaggle
CUDA Support	Enabled

3.1.2 Binary Segmentation Hyperparameter Configuration

Several experiments and parameter tuning procedures were conducted in order to determine the most suitable configuration for training the proposed binary segmentation models. The final hyperparameter settings were selected according to validation performance, convergence stability, and computational efficiency.

In addition to the transformer-based Swin-UNETR architecture [44], a conventional 3D U-Net model [45] was also implemented in order to perform comparative evaluation under identical experimental conditions.

Table 3.2 summarizes the final hyperparameter configuration adopted for training the proposed binary segmentation models.

Table 3.2: Final hyperparameter configuration used for binary segmentation models.

Parameter	Swin-UNETR	3D U-Net
Input Size	$96 \times 96 \times 96$	$96 \times 96 \times 96$
Input Channels	1	1
Output Channels	1	1
Initial Feature Dimension	24	N/A
Batch Size	1	1
Learning Rate	1×10^{-4}	1×10^{-4}
Minimum Learning Rate	1×10^{-6}	1×10^{-6}
Optimizer	AdamW	AdamW
Weight Decay	1×10^{-5}	1×10^{-5}
Scheduler	Cosine Annealing LR	Cosine Annealing LR
Loss Function	Dice + Cross-Entropy	Dice + Cross-Entropy
Epochs	60	100
Validation Split	80% / 20%	80% / 20%
Mixed Precision Training	Enabled	Enabled
GPU	NVIDIA Tesla T4	NVIDIA Tesla T4

Although the original MRI volumes had dimensions of approximately $160 \times 384 \times 384$, patch-based training was employed during optimization. Consequently, cropped

volumetric regions of size $96 \times 96 \times 96$ were extracted and provided as input to both segmentation models in order to reduce GPU memory consumption and focus training on anatomically relevant regions.

A cosine annealing learning rate scheduler was employed during training in order to progressively decrease the learning rate from 1×10^{-4} to 1×10^{-6}. This strategy improves convergence stability and enables finer optimization during later training stages.

Both models were trained using identical preprocessing, augmentation, optimization, and validation settings in order to ensure a fair comparative evaluation. The primary difference between the experiments was the network architecture itself.

Although the 3D U-Net model required a larger number of training epochs to achieve stable convergence, the transformer-based Swin-UNETR architecture demonstrated faster convergence and improved segmentation stability during training.

A batch size of one was adopted due to the high GPU memory requirements associated with 3D MRI volumes and transformer-based architectures. In addition, mixed precision training was enabled to reduce memory consumption and accelerate training.

3.1.3 Multi-Class Segmentation Hyperparameter Configuration

Several experiments and optimization strategies were conducted in order to improve the performance of the proposed multi-class anatomical segmentation framework. The final configuration was selected according to validation Dice score, convergence stability, and anatomical segmentation quality.

Based on the comparative evaluation performed during the binary knee localization stage, the transformer-based Swin-UNETR architecture was selected for the multi-class segmentation experiments. The selection was motivated by its superior segmentation accuracy, improved convergence behavior, and better training stability compared with the conventional 3D U-Net architecture.

Table 3.3 summarizes the final hyperparameter configuration adopted for the multi-class segmentation experiments.

Table 3.3: Final hyperparameter configuration used for multi-class segmentation.

Parameter	Value
Training Strategy	Two-Stage ROI Segmentation
Input Data	Cropped Knee MRI Volumes
Segmentation Type	Multi-Class Segmentation
Number of Classes	6
Input Patch Size	$96 \times 96 \times 96$
Input Channels	1
Output Channels	6
Initial Feature Dimension	24
Batch Size	1
Learning Rate	1×10^{-4}
Minimum Learning Rate	1×10^{-7}
Optimizer	AdamW
Weight Decay	1×10^{-5}
Loss Function	Weighted DiceCE Loss
Scheduler	Cosine Annealing LR
Mixed Precision Training	Enabled
Validation Split	80% / 20%
Epochs	80
GPU	NVIDIA Tesla T4

Although the original MRI volumes had dimensions of approximately $160 \times 384 \times 384$, a coarse-to-fine ROI extraction strategy was employed before multi-class segmentation. First, the Stage 1 binary segmentation model automatically generated predicted knee localization masks from the MRI volumes. These predicted masks were subsequently refined using post-processing operations including connected component analysis and hole filling in order to remove small artifacts and improve anatomical consistency.

The refined binary masks were then utilized to automatically crop anatomically relevant knee regions from MRI volumes. Consequently, a new cropped ROI dataset focused exclusively on knee anatomical structures was generated and used as input for Stage 2 multi-class segmentation training.

This model-guided ROI extraction strategy reduced irrelevant background information, improved anatomical focus during training, and simulated more realistic inference conditions without relying directly on manually cropped ground-truth regions.

In addition, patch-based training with volumetric regions of size $96 \times 96 \times 96$ was adopted in order to reduce GPU memory consumption and improve training efficiency.

To address severe class imbalance between anatomical structures, weighted Dice Cross-Entropy loss was employed during training. Higher weights were assigned to meniscal regions because of their smaller anatomical size and increased segmentation difficulty

compared with femoral and tibial cartilage structures.

A cosine annealing learning rate scheduler was employed during training in order to progressively decrease the learning rate from 1×10^{-4} to 1×10^{-7}. This strategy improves convergence stability and enables finer optimization during later training stages.

The multi-class segmentation experiments were therefore conducted exclusively using the Swin-UNETR architecture. This decision reduced computational complexity and allowed subsequent analyses to focus on the best-performing segmentation framework identified during the preliminary evaluation stage.

A batch size of one was adopted due to the high GPU memory requirements associated with 3D MRI volumes and transformer-based architectures. In addition, mixed precision training was enabled to reduce memory consumption and accelerate training.

3.1.4 Online Data Augmentation Strategy

Online data augmentation was employed during training in order to increase data diversity, improve model generalization, and reduce overfitting. Since the available MRI dataset was relatively limited and exhibited substantial anatomical variability between subjects, several spatial and intensity-based transformations were dynamically applied during training [46].

Unlike offline augmentation approaches that permanently generate additional samples before training, the adopted strategy performed augmentation online during data loading. Consequently, new image variations were generated at every training epoch without increasing the physical size of the dataset.

The augmentation operations included random flipping along the three anatomical axes, random 90-degree rotations, intensity scaling, intensity shifting, and contrast adjustment. Spatial transformations were applied simultaneously to MRI volumes and segmentation masks in order to preserve anatomical correspondence, whereas intensity-based augmentations were applied only to MRI images.

Table 3.4 summarizes the augmentation techniques and parameter values used during multi-class segmentation training.

Table 3.4: Online data augmentation strategy used during multi-class segmentation training.

Transformation	Parameter	Probability
Random Flip	Axis = 0 (Depth)	0.4
Random Flip	Axis = 1 (Height)	0.4
Random Flip	Axis = 2 (Width)	0.4
Random Rotation (90°)	Random Orientation	0.4
Random Intensity Scaling	Factor = 0.1	0.2
Random Intensity Shift	Offset = 0.08	0.2
Random Contrast Adjustment	Gamma = (0.9, 1.1)	0.2

Because the available MRI dataset remained relatively limited for training deep learning models, the online augmentation strategy exposed the network to thousands of different image variations throughout the learning process. This substantially improved robustness against anatomical variability, intensity inconsistencies, and acquisition-related differences across MRI examinations.

Figure 3.1 presents examples of online data augmentation operations applied during multi-class segmentation training.

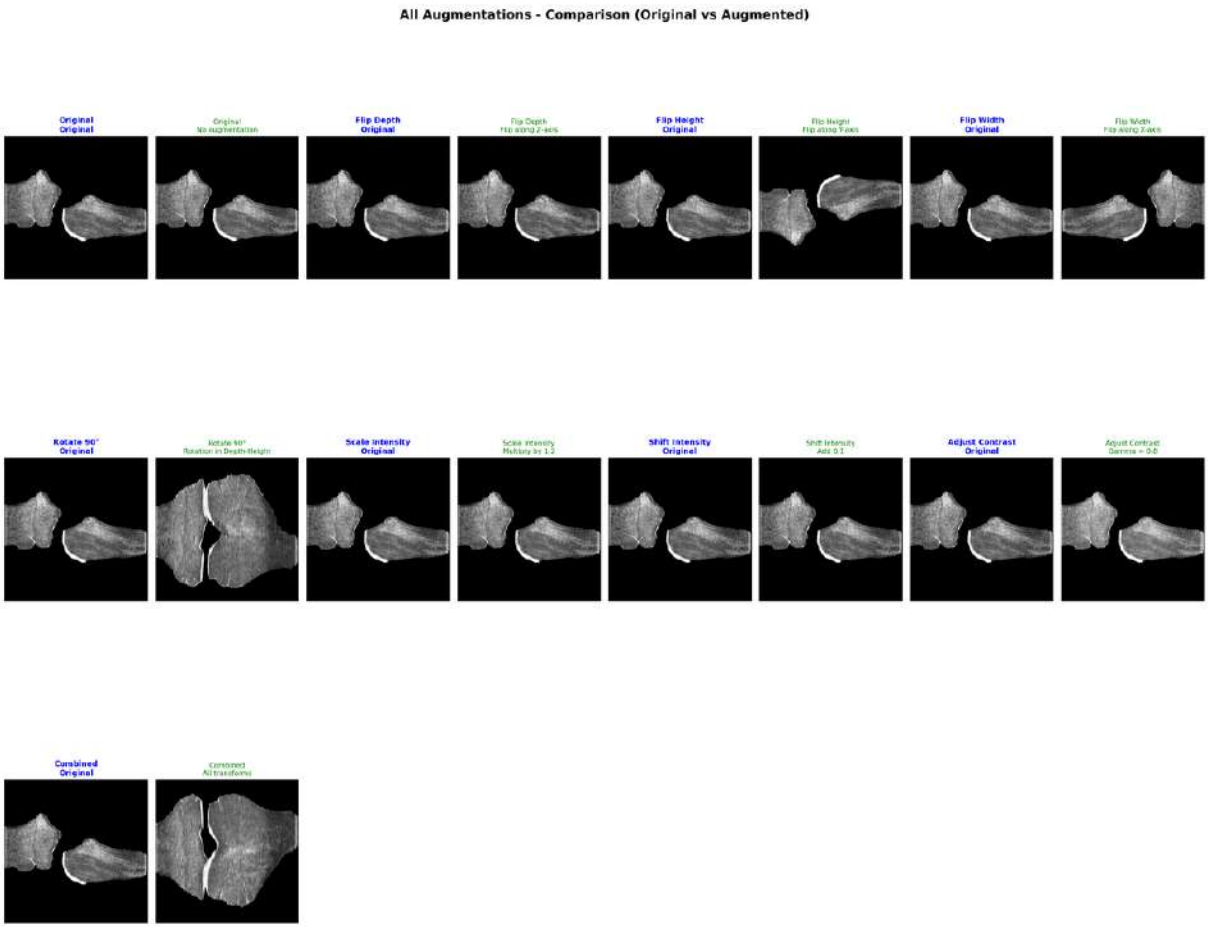


Figure 3.1: Examples of online data augmentation operations applied during multi-class segmentation training. Spatial transformations include flipping and rotation, whereas intensity-based transformations include intensity scaling, intensity shifting, and contrast adjustment. The augmentations were dynamically generated during training without permanently increasing the dataset size.

As illustrated in Figure 3.1, the proposed augmentation strategy generated diverse anatomical orientations and intensity distributions while preserving the underlying anatomical structures. These transformations enhanced model robustness and reduced the risk of overfitting during training.

Furthermore, because the augmentations were applied dynamically during each training epoch, the model was exposed to continuously varying image appearances, enabling improved generalization capability and greater resilience to inter-subject anatomical differences. This strategy was expected to be particularly beneficial for small anatomical structures such as the meniscal regions, which are generally more challenging to segment accurately than larger anatomical components.

3.2 Evaluation of Preprocessing process

3.2.1 Visual Assessment

The effectiveness of the proposed preprocessing pipeline was initially evaluated through qualitative visual inspection of representative MRI slices.

The proposed preprocessing framework is illustrated in Figure 3.2. The pipeline combines wavelet-based denoising, Restormer restoration, gamma correction, and CLAHE enhancement in order to improve MRI quality while preserving anatomical structures.

In contrast, the preprocessed images demonstrate improved contrast, reduced noise levels, and enhanced visibility of cartilage and surrounding knee tissues. Anatomical boundaries become more distinguishable, facilitating subsequent localization and segmentation tasks.

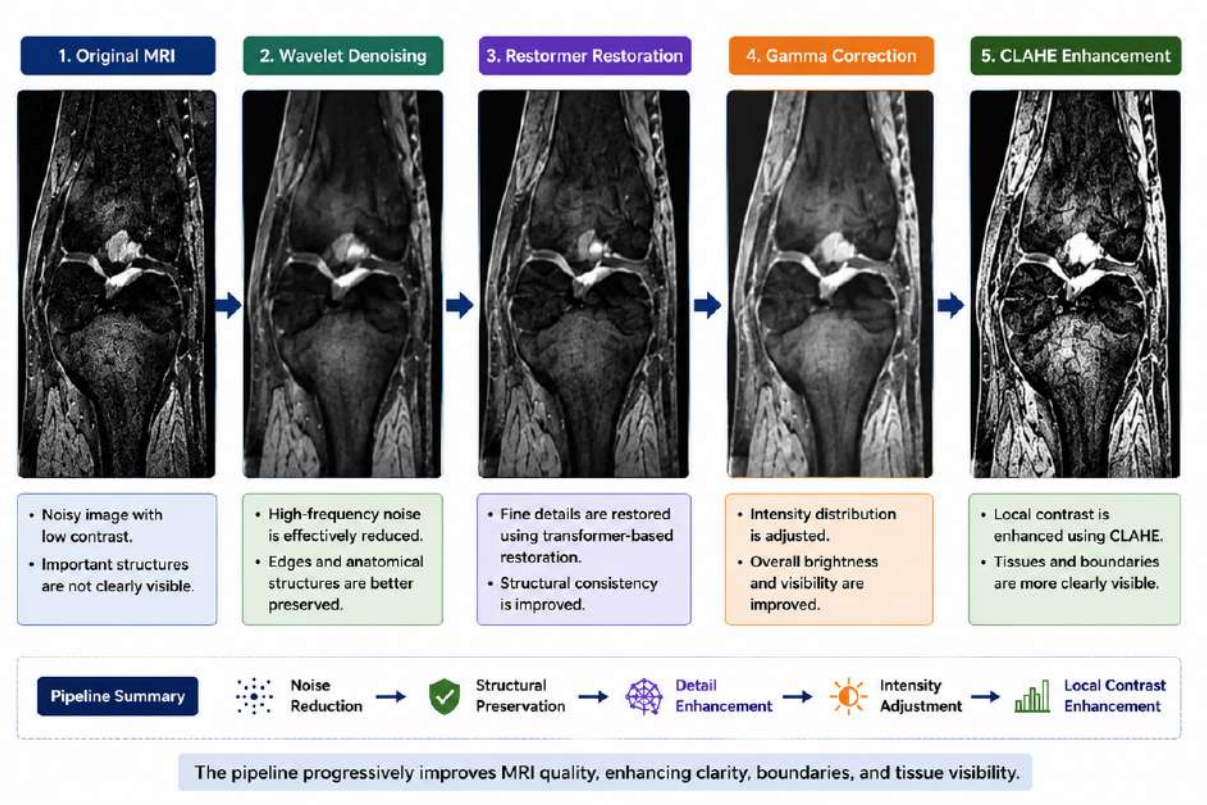


Figure 3.2: Illustration of the proposed MRI preprocessing pipeline. The framework sequentially applies wavelet-based denoising, Restormer restoration, gamma correction, and CLAHE enhancement in order to improve image quality while preserving anatomical structures.

Visual inspection as depicted in the figure 3.3 indicates that the proposed preprocessing pipeline successfully enhanced image quality while preserving relevant anatomical

information. These improvements provide higher-quality inputs for the subsequent segmentation and osteoarthritis analysis stages.

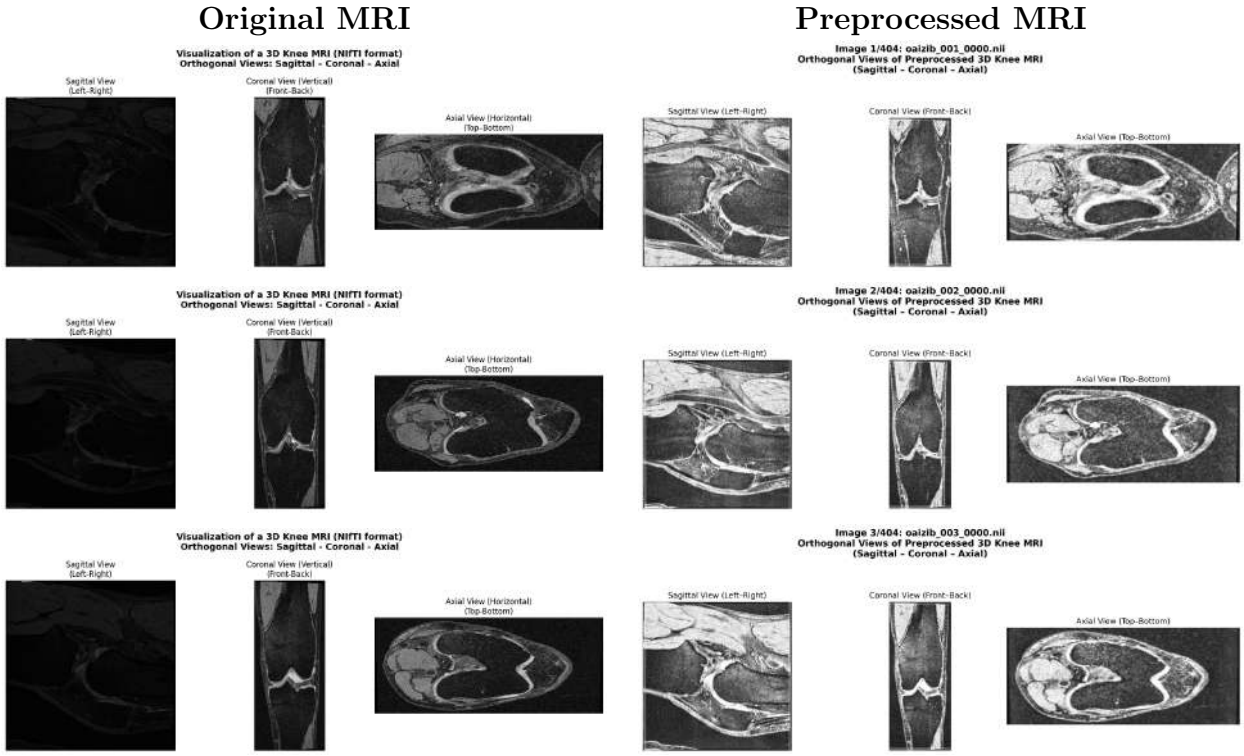


Figure 3.3: Visual comparison between representative original MRI slices and their corresponding preprocessed versions obtained using the proposed preprocessing pipeline. The enhanced images exhibit improved contrast, reduced noise, and clearer visualization of anatomical structures.

3.2.2 Impact on Segmentation Performance

The preprocessing stage played an important role in improving the quality of the MRI volumes used for subsequent segmentation tasks. Noise reduction and contrast enhancement improved the visibility of anatomical structures and facilitated the learning process of the deep learning models.

By reducing image artifacts and increasing tissue contrast, the preprocessing pipeline enabled more accurate knee localization during the binary segmentation stage and improved the delineation of anatomical structures during multi-class segmentation. In addition, the enhanced image quality contributed to more stable model convergence and reduced sensitivity to inter-subject intensity variations.

Overall, the preprocessing framework provided higher-quality input data for the proposed segmentation pipeline and contributed to the robustness of the subsequent feature extraction and osteoarthritis classification stages.

3.3 Segmentation Results

3.3.1 Model Selection Strategy

Before conducting the final segmentation experiments, a comparative evaluation was performed between the transformer-based Swin-UNETR architecture and the conventional 3D U-Net architecture. Both models were trained under identical preprocessing, augmentation, optimization, and validation settings in order to ensure a fair comparison.

The objective of this preliminary experiment was to identify the most suitable segmentation architecture for the subsequent stages of the proposed framework, including ROI extraction, multi-class anatomical segmentation, features extraction, and osteoarthritis classification.

Table 3.5 summarizes the quantitative results obtained during the binary knee localization stage.

Table 3.5: Comparison between Swin-UNETR and 3D U-Net during binary knee localization.

Metric	**Swin-UNETR**	**3D U-Net**
Best Validation Dice	0.9752	0.9532
Best Validation Loss	0.1783	0.2329
Training Epochs	100	100
Selected for Subsequent Experiments	Yes	No

Figure 3.4 illustrates the comparative performance of both architectures during training.

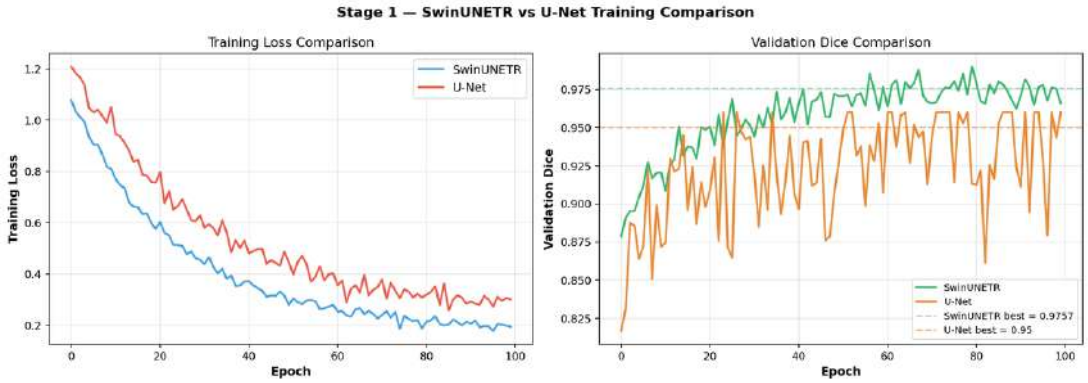


Figure 3.4: Comparison between Swin-UNETR and 3D U-Net during binary knee localization training.

The quantitative results presented in Table 3.5 and Figure 3.4 demonstrate that the Swin-UNETR architecture outperformed the conventional 3D U-Net model during binary

knee localization.The best validation Dice score reached 0.9752, confirming the effectiveness of the proposed binary localization stage and its suitability for automatic ROI extraction.

Furthermore, Swin-UNETR reached stable convergence using fewer training epochs, whereas the 3D U-Net model required additional optimization while still producing inferior performance. These observations suggest that the transformer-based architecture is more effective at capturing complex anatomical relationships within knee MRI volumes.

The ability of Swin-UNETR to model both local anatomical details and global contextual information through self-attention mechanisms contributes to improved localization accuracy and more stable optimization behavior. Similar observations have been reported in previous studies demonstrating the effectiveness of transformer-based architectures for medical image segmentation [44].

Consequently, Swin-UNETR was selected as the primary segmentation architecture for all subsequent experiments. Therefore, the remaining segmentation analyses presented in this chapter are based exclusively on the Swin-UNETR framework.

3.3.2 Binary Knee Localization Results

The first stage of the proposed framework focused on binary knee localization, where the objective was to distinguish the knee joint region from the surrounding background tissues. Accurate localization is essential because it enables subsequent ROI extraction and reduces the amount of irrelevant information processed during anatomical segmentation.

The selected Swin-UNETR model was trained using the hyperparameter configuration described previously. Figure 3.5 illustrates the evolution of the training loss and validation Dice score during optimization.



Figure 3.5: Training loss and validation Dice score obtained during binary knee localization using the Swin-UNETR architecture.

The training curves demonstrate stable convergence throughout the optimization process. The training loss decreased progressively from approximately 0.90 during the initial epochs to approximately 0.18 at the end of training. Simultaneously, the validation Dice score increased rapidly during the early training stages and gradually stabilized around 0.97.

The obtained results indicate that the model successfully learned the spatial characteristics of the knee joint and achieved highly accurate localization performance. Furthermore, the relatively small fluctuations observed during the later epochs suggest good generalization capability and stable optimization behavior.

The best validation Dice score reached 0.9752, while the lowest validation loss reached 0.1783, confirming the effectiveness of the proposed binary localization stage and its suitability for automatic ROI extraction. Table 3.6 summarizes the obtained results.

Table 3.6: Binary knee localization performance using Swin-UNETR.

Metric	Value
Best Validation Dice	0.9752
Best Validation Loss	0.1783
Training Epochs	100
Architecture	Swin-UNETR

3.3.3 ROI Extraction Results

Following binary knee localization, the predicted masks were utilized to automatically extract anatomically relevant knee regions from the original MRI volumes. The objective of this stage was to remove irrelevant background information and generate compact regions of interest (ROIs) suitable for subsequent multi-class anatomical segmentation.

Unlike conventional approaches that rely on manually generated masks, the proposed framework performed ROI extraction using the binary masks predicted by the Swin-UNETR localization model. Consequently, the cropping procedure reflects realistic inference conditions and enables a fully automated segmentation pipeline.

Figure 3.6 illustrates representative examples of the automatic ROI extraction process.

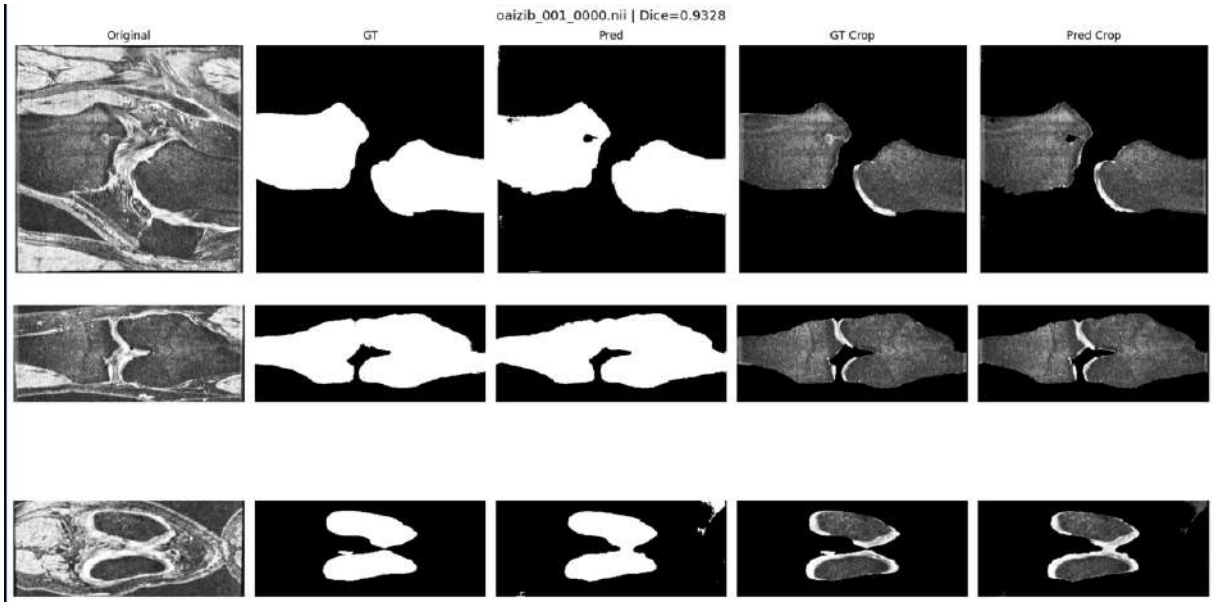


Figure 3.6: Examples of automatic ROI extraction using binary masks predicted by the Swin-UNETR localization model. The predicted masks were utilized to generate knee-centered MRI volumes without relying on manual annotations.

Visual inspection confirms that the predicted binary masks accurately identified the knee joint region and enabled successful extraction of anatomically relevant MRI volumes. Despite minor differences between the predicted and ground-truth masks, the resulting cropped regions preserved the essential anatomical structures required for subsequent analysis.

The generated ROI volumes substantially reduced the amount of irrelevant background information while maintaining the femoral, tibial, cartilaginous, and meniscal regions of interest. This automatic cropping strategy improved anatomical focus and reduced computational complexity during the subsequent multi-class segmentation stage.

Furthermore, because ROI extraction was performed using masks predicted by the Swin-UNETR localization model rather than manually generated annotations, the proposed framework maintained a fully automated workflow and simulated realistic deployment conditions.

3.3.4 Post-Processing Results

A post-processing stage was applied to the predicted binary masks in order to improve anatomical consistency and remove small segmentation artifacts. The adopted refinement strategy was described previously in Chapter 2.

Figure 3.7 illustrates the effect of the post-processing stage on representative binary segmentation outputs.

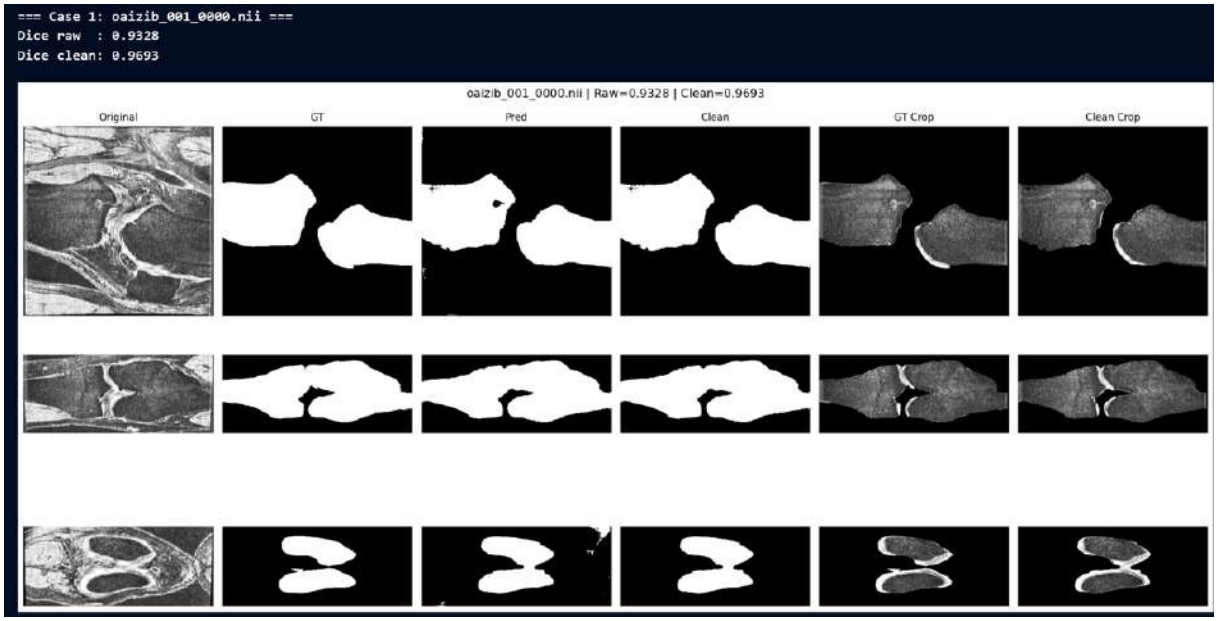


Figure 3.7: Effect of post-processing on binary knee localization results. The refinement stage improved mask quality and reduced segmentation artifacts.

As illustrated in Figure 3.7, the post-processing stage successfully removed isolated artifacts and corrected minor irregularities in the predicted masks. For the representative example shown, the Dice score increased from 0.9328 before refinement to 0.9693 after refinement, corresponding to an improvement of approximately 3.65 percentage points.

The refined masks exhibited improved anatomical consistency and more accurate delineation of the knee region. These cleaner masks were subsequently utilized to generate the final ROI dataset used for training the multi-class segmentation model.

Table 3.7: Effect of post-processing on binary localization performance.

Stage	Dice Score
Raw Prediction	0.9328
Post-Processed Prediction	0.9693
Improvement	+0.0365

The results demonstrate that the post-processing stage contributed to improving mask quality and anatomical consistency by eliminating small isolated artifacts and correcting minor segmentation irregularities. Although the binary localization model already achieved high accuracy, the refinement stage further enhanced the agreement between the predicted masks and the ground-truth annotations.

The resulting clean masks provided more reliable knee-centered regions of interest and served as the basis for constructing the final ROI dataset used during multi-class anatomical segmentation. Consequently, the post-processing stage contributed to improving the

quality of the inputs provided to the subsequent segmentation model.

3.3.5 Multi-Class Segmentation Results

The second stage of the proposed framework focused on the segmentation of multiple anatomical structures within the knee joint. The objective was to accurately delineate the femoral bone, tibial regions, and meniscal structures from the automatically generated ROI volumes obtained during the previous stage.

The Swin-UNETR architecture was trained using the hyperparameter configuration described in Chapter 2. The best segmentation performance was achieved at Epoch 64, yielding a mean Dice score of 0.7936 and a mean IoU of 0.7057.

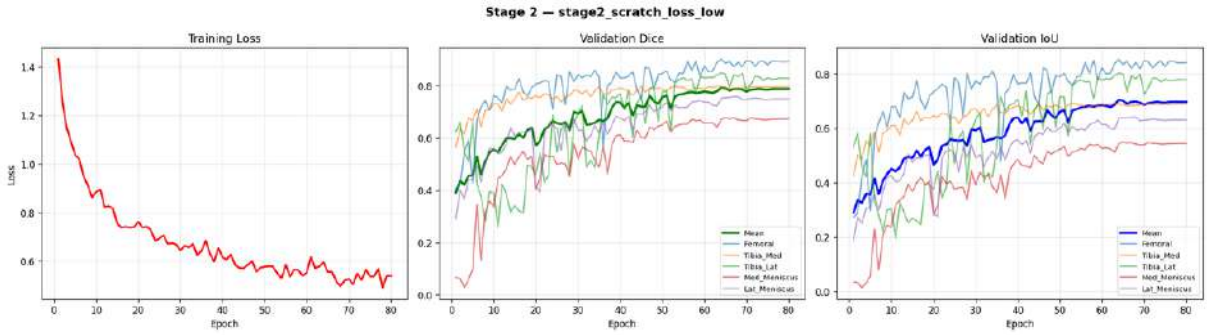


Figure 3.8: Training loss, validation Dice score, and validation IoU obtained during multi-class segmentation training using the Swin-UNETR architecture.

The training curves demonstrate stable convergence throughout the optimization process. The training loss progressively decreased from approximately 1.45 at the beginning of training to nearly 0.55 during the final epochs. Simultaneously, both Dice and IoU scores consistently improved before stabilizing as training progressed. The best segmentation performance was achieved at Epoch 64, indicating effective optimization and good generalization capability. Table 3.8 illustrates the overall performance of the multi-class Swin-UNETR segmentation model.

Table 3.8: Overall performance of the multi-class Swin-UNETR segmentation model.

Metric	Value
Best Dice Score	0.7936
Best IoU Score	0.7057
Best Epoch	64
Loss at Best Epoch	0.5948

Table 3.9 shows the segmentation performance per-class.

Table 3.9: Per-class segmentation performance at the best validation epoch.

Anatomical Structure	Dice Score	IoU Score
Femoral Bone	0.9012	0.8542
Medial Tibia	0.7951	0.6886
Lateral Tibia	0.8487	0.8039
Medial Meniscus	0.6709	0.5441
Lateral Meniscus	0.7522	0.6377

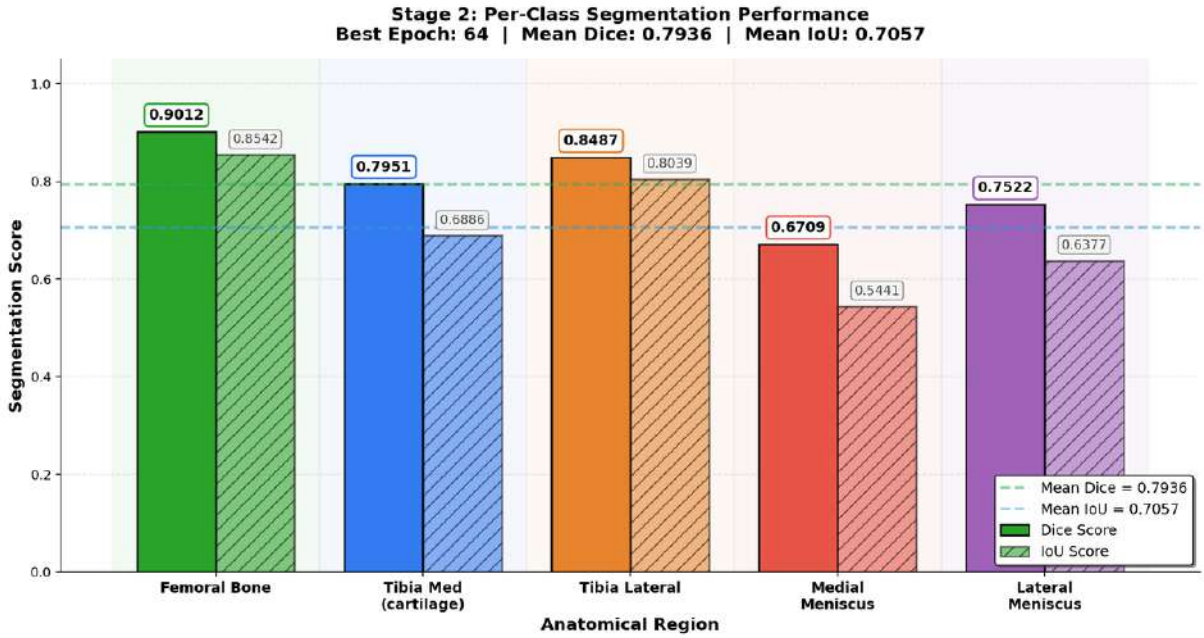


Figure 3.9: Per-class Dice and IoU scores achieved by the proposed Swin-UNETR model at the best validation epoch.

The quantitative results presented in Tables 3.8 and 3.9 demonstrate the effectiveness of the proposed Swin-UNETR framework for multi-class anatomical segmentation.

The highest segmentation accuracy was achieved for the femoral bone, with a Dice score of 0.9012 and an IoU score of 0.8542. This result can be attributed to the relatively large size and well-defined anatomical boundaries of the femoral structure.

The tibial regions also achieved good segmentation performance, with Dice scores of 0.7951 and 0.8487 for the medial and lateral tibial compartments, respectively. These results indicate that the model successfully captured the anatomical characteristics of the major bone structures within the knee joint.

The meniscal structures exhibited lower segmentation performance compared with the bone regions. The medial meniscus achieved a Dice score of 0.6709, whereas the lateral meniscus achieved a Dice score of 0.7522. This reduction in performance is expected because menisci are considerably smaller anatomical structures and present more complex

shapes and less distinct boundaries in MRI volumes.

Overall, the obtained results demonstrate that the proposed framework successfully segmented multiple knee anatomical structures while maintaining good performance across both large and small regions of interest.

3.3.5.1 Qualitative Segmentation Results

Qualitative evaluation was performed by visually comparing the predicted segmentation masks with the corresponding ground-truth annotations in the three anatomical views (axial, coronal, and sagittal).

Figures 3.10 and 3.11 present representative examples of the multi-class segmentation results obtained using the proposed Swin-UNETR model. The figures illustrate the original MRI slices, ground-truth annotations, predicted masks, and overlay visualizations highlighting the segmented anatomical structures.

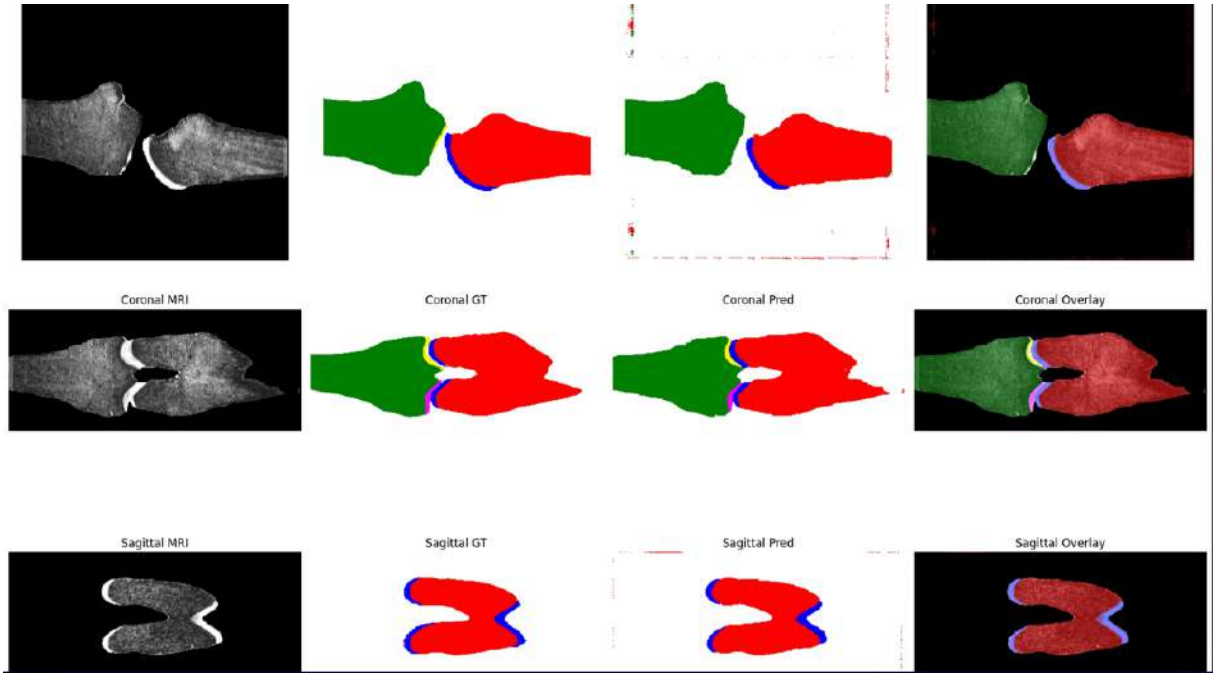


Figure 3.10: Representative multi-class segmentation results obtained using the proposed Swin-UNETR framework. Ground-truth masks and predicted segmentations are shown in axial, coronal, and sagittal views together with overlay visualizations.

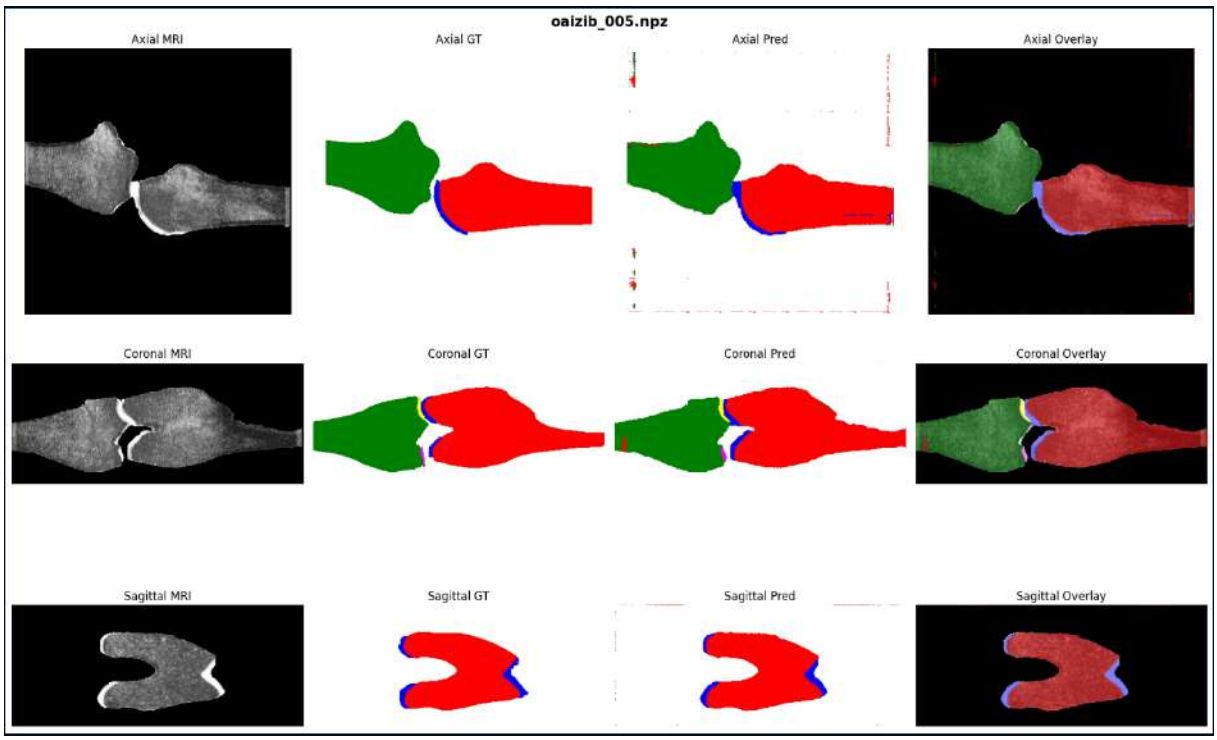


Figure 3.11: Additional example of multi-class segmentation results obtained on a different MRI examination. Ground-truth masks and predicted segmentations are shown in axial, coronal, and sagittal views together with overlay visualizations.

Visual inspection of both examples demonstrates a strong agreement between the predicted segmentations and the corresponding ground-truth annotations across the three anatomical views. The model successfully delineated the femoral region, tibial compartments, and meniscal structures with anatomically consistent boundaries.

The overlay visualizations further confirm that the predicted masks closely follow the target anatomical structures. Minor discrepancies are mainly observed around thin meniscal boundaries and transition regions between adjacent anatomical structures, which is consistent with the quantitative results reported previously.

Overall, the qualitative results support the quantitative findings and demonstrate the capability of the proposed framework to generate reliable multi-class anatomical segmentations suitable for subsequent feature extraction and osteoarthritis classification.

3.4 Evaluation of Osteoarthritis Severity Classification Performance

Following multi-class anatomical segmentation, handcrafted features were extracted from the segmented knee regions and combined with clinical information to perform osteoarthri-

tis prediction. The objective of this stage was to evaluate the ability of the proposed framework to distinguish between different osteoarthritis severity levels using region-based anatomical descriptors.

The original dataset contained 404 MRI examinations. However, 21 subjects were excluded from the classification experiments because their corresponding Kellgren–Lawrence (KL) grades were unavailable in the clinical metadata. Consequently, a total of 383 subjects were retained for feature extraction and osteoarthritis prediction.

For the independent test set, 103 MRI examinations were initially available. Nevertheless, five subjects were excluded due to missing KL annotations, resulting in a final test cohort of 98 subjects used for evaluation.

3.4.1 Dataset Distribution

Figure 3.12 presents the distribution of KL grades within the final classification dataset.

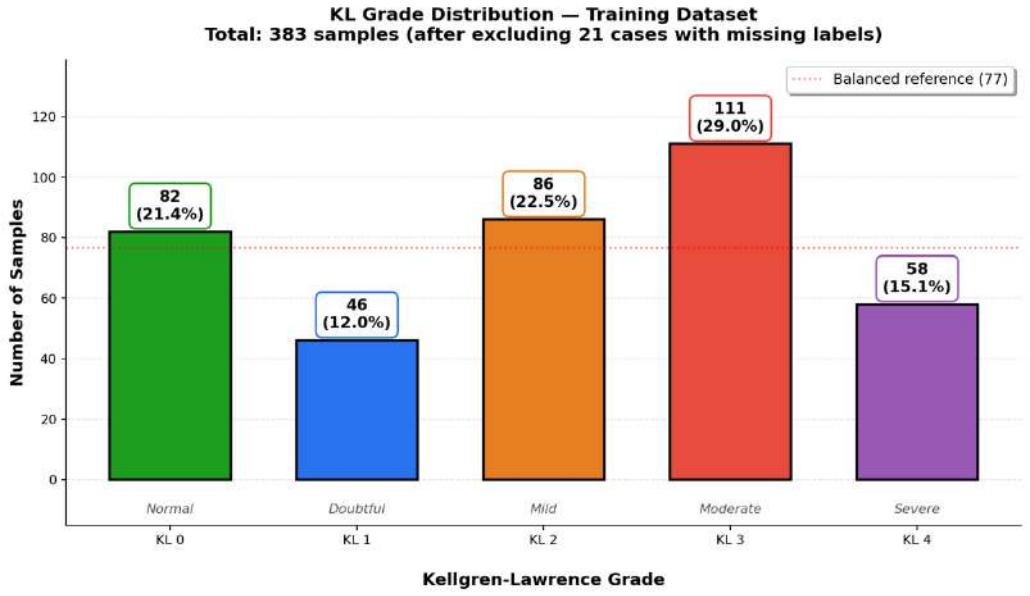


Figure 3.12: Distribution of Kellgren–Lawrence grades in the classification dataset after excluding subjects with missing labels.

The final classification dataset consisted of 383 MRI examinations distributed across five KL grades. Specifically, the dataset contained 82 KL0 cases, 46 KL1 cases, 86 KL2 cases, 111 KL3 cases, and 58 KL4 cases. The class distribution was moderately imbalanced, with KL3 representing the largest category (29.0%) and KL1 representing the smallest category (12.0%).

Such class imbalance increases the complexity of osteoarthritis classification because minority classes are represented by fewer training examples. Consequently, additional

evaluation metrics beyond overall accuracy were considered during performance assessment in order to provide a more reliable assessment of classification performance.

3.4.2 Preliminary Data Augmentation Analysis for Classification

Preliminary data augmentation experiments were initially conducted during the early stages of the osteoarthritis classification pipeline in order to investigate whether increasing training data diversity could improve classification performance and reduce overfitting.

Several augmentation operations including spatial flipping, rotation, and intensity modification were applied to the MRI regions before feature extraction and classification. The augmented dataset increased the number of available training samples from 383 original subjects to 1532 augmented samples.

At this preliminary stage, the classification framework relied on an initial feature extraction configuration composed mainly of statistical and region-based descriptors extracted from the segmented anatomical structures. A hierarchical classification strategy was adopted to progressively distinguish healthy subjects, mild osteoarthritis cases, and severe osteoarthritis cases.

Table 3.10 summarizes the obtained results for both the original and augmented datasets.

Table 3.10: Preliminary analysis of augmentation impact on osteoarthritis classification performance.

Experiment	Dataset Size	Accuracy
Original Dataset Only	383 subjects	61.88%
With Data Augmentation	1532 samples	59.29%

The obtained results indicated that augmentation did not improve classification performance. In contrast, a slight decrease in overall accuracy was observed after introducing augmented samples into the training process.

This behavior suggests that the primary limitation of the classification task was not solely related to insufficient training data quantity. Instead, the experimental analysis indicated that the major challenge originated from the intrinsic overlap between osteoarthritis severity stages, particularly for intermediate KL grades such as KL2.

Furthermore, the preliminary feature extraction configuration used during these experiments relied mainly on handcrafted statistical and anatomical descriptors. Consequently, the augmentation process introduced additional variability without significantly improving inter-class separability.

Based on these observations, augmentation strategies were not further integrated into the final classification framework. Instead, the subsequent experiments focused on improving feature representation quality through the extraction of more advanced handcrafted descriptors, including wavelet-based features, FFT descriptors, texture features derived from GLCM matrices, first-order statistical features, and clinical variables such as age, BMI, and gender.

These advanced descriptors provided substantially richer anatomical and structural information for osteoarthritis severity prediction and were therefore adopted in the final classification pipeline.

3.4.3 Feature Extraction Verification

Before performing osteoarthritis classification, a visual verification step was conducted to ensure the reliability of the extracted anatomical regions. Representative subjects from different Kellgren–Lawrence (KL) grades were selected and inspected after the multi-class segmentation stage.

Figure 3.13 presents representative MRI slices together with their corresponding anatomical segmentations and region-wise measurements. The figure illustrates examples from KL0 to KL4 and demonstrates the ability of the proposed framework to consistently identify the major knee anatomical structures across different osteoarthritis severity levels.

The visual results confirm that the segmentation framework successfully delineated the femoral region, tibial compartments, and meniscal structures for subjects exhibiting different disease severities. The extracted region-wise measurements further demonstrate that the proposed pipeline was capable of generating anatomically meaningful quantitative descriptors from the segmented structures.

This verification step provides additional confidence regarding the quality of the extracted anatomical regions before feature computation and osteoarthritis classification. Consequently, the generated feature set can be considered sufficiently reliable for subsequent machine learning analysis.

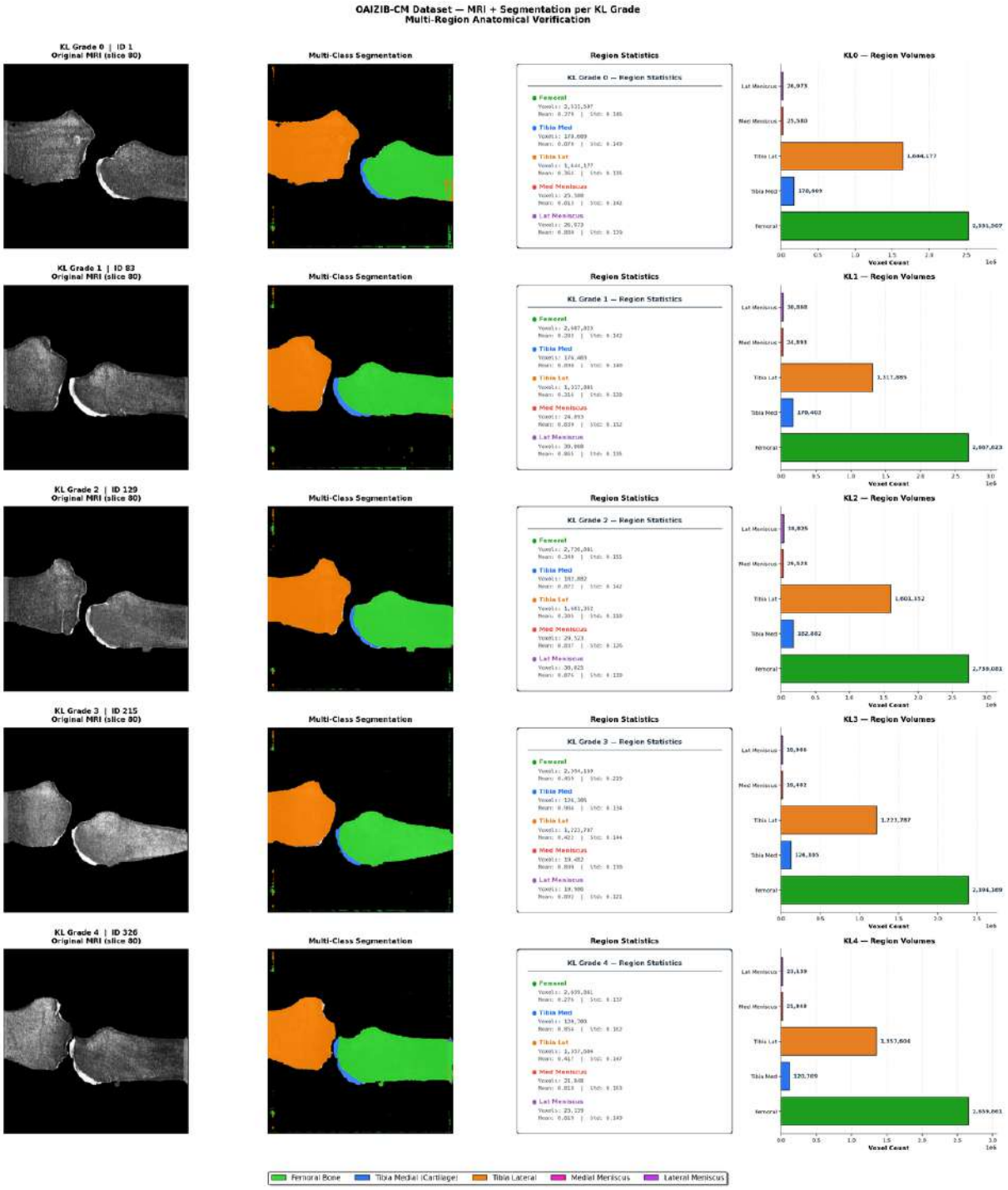


Figure 3.13: Representative examples of MRI slices, multi-class anatomical segmentations, and region-wise measurements obtained from subjects belonging to different Kellgren–Lawrence grades.

3.4.4 Feature Importance Analysis

To better understand the contribution of individual descriptors to osteoarthritis prediction, feature importance analysis was performed using the CatBoost classifier. The objec-

tive was to identify the most informative features and determine which anatomical regions contributed most strongly to classification performance.

Figure 3.14 presents the twenty most important features selected by the classification model.

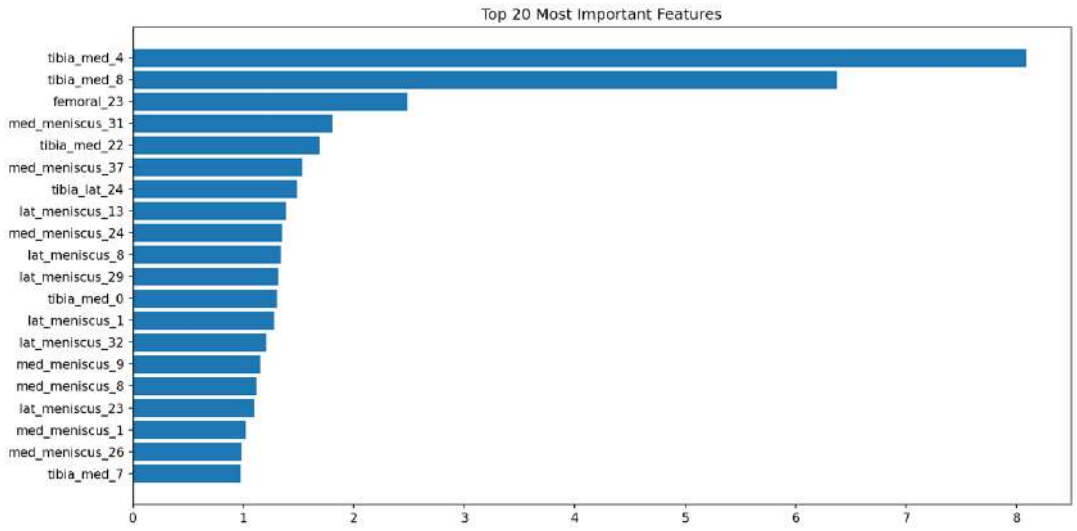


Figure 3.14: Top 20 most important handcrafted features selected by the CatBoost classifier for osteoarthritis prediction.

The feature importance ranking indicates that several highly discriminative descriptors originated from the tibial and meniscal regions. In particular, features extracted from the medial and lateral tibial compartments achieved the highest importance scores, suggesting that these structures contain valuable information related to osteoarthritis progression.

In addition, meniscal descriptors appeared frequently among the most influential features, highlighting the importance of meniscal alterations in characterizing disease severity. These findings are consistent with the known role of cartilage and meniscal degeneration in osteoarthritis development.

To further investigate feature discriminability, the distributions of representative descriptors were analyzed across the different osteoarthritis severity groups as depicted in the figure 3.15.

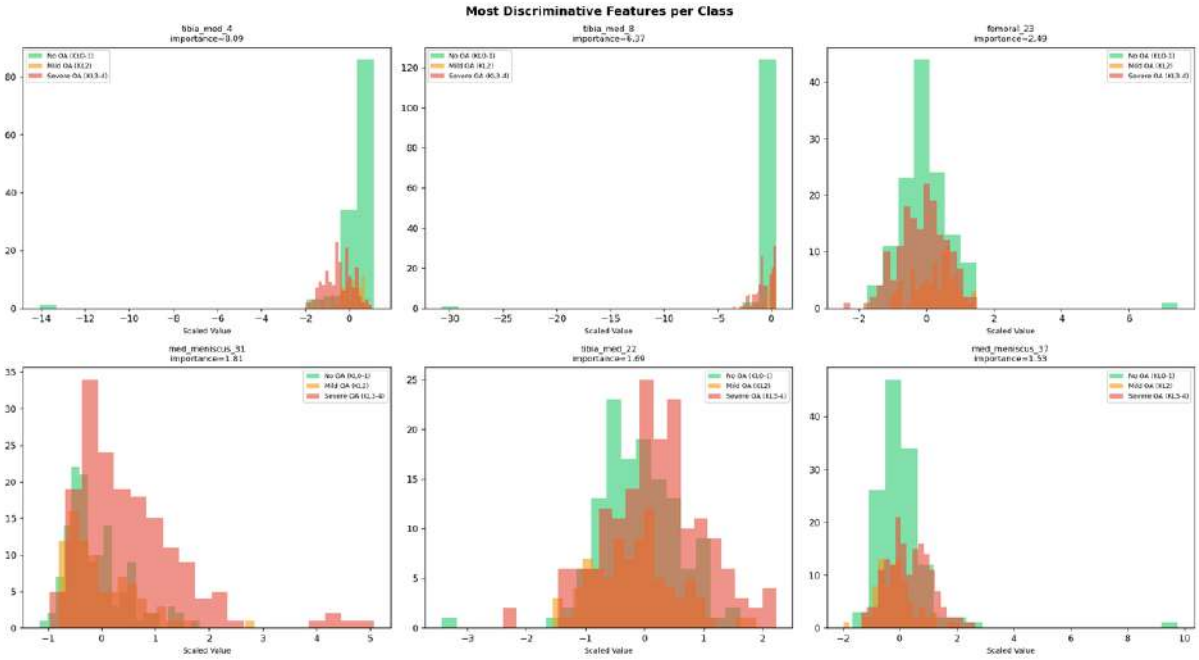


Figure 3.15: Distribution of representative discriminative features across different osteoarthritis severity groups.

Figure 3.15 illustrates the distributions of several representative features across the severity groups. Clear differences can be observed between classes for multiple descriptors, particularly those extracted from tibial and meniscal structures.

Although partial overlap exists between neighboring severity groups, several features demonstrate distinct distribution patterns that facilitate class separation. These observations confirm the discriminative capability of the extracted anatomical descriptors and support their use for osteoarthritis prediction.

Overall, the feature importance analysis demonstrates that the proposed handcrafted feature extraction strategy successfully captured clinically relevant anatomical characteristics associated with osteoarthritis severity.

The average classification accuracy obtained across the five folds was 69.18%, indicating that the extracted anatomical descriptors contain meaningful information related to osteoarthritis severity. However, a detailed analysis of the classification results revealed important differences in performance across the three severity groups.

The detailed classification report is presented in Table 3.11.

Table 3.11: Classification report for the three-class osteoarthritis prediction task(Accuracy = 0.6919)

Class	Precision	Recall	F1-score	Support
KL0–1 (No OA)	0.68	0.80	0.74	128
KL2 (Mild OA)	0.64	0.19	0.29	86
KL3–4 (Severe OA)	0.71	0.86	0.78	169

The obtained results indicate that the classifier successfully distinguished the No OA (KL0–1) and Severe OA (KL3–4) groups. In contrast, the KL2 category achieved substantially lower recall, suggesting that a large proportion of mild osteoarthritis cases were confused with neighboring severity groups.

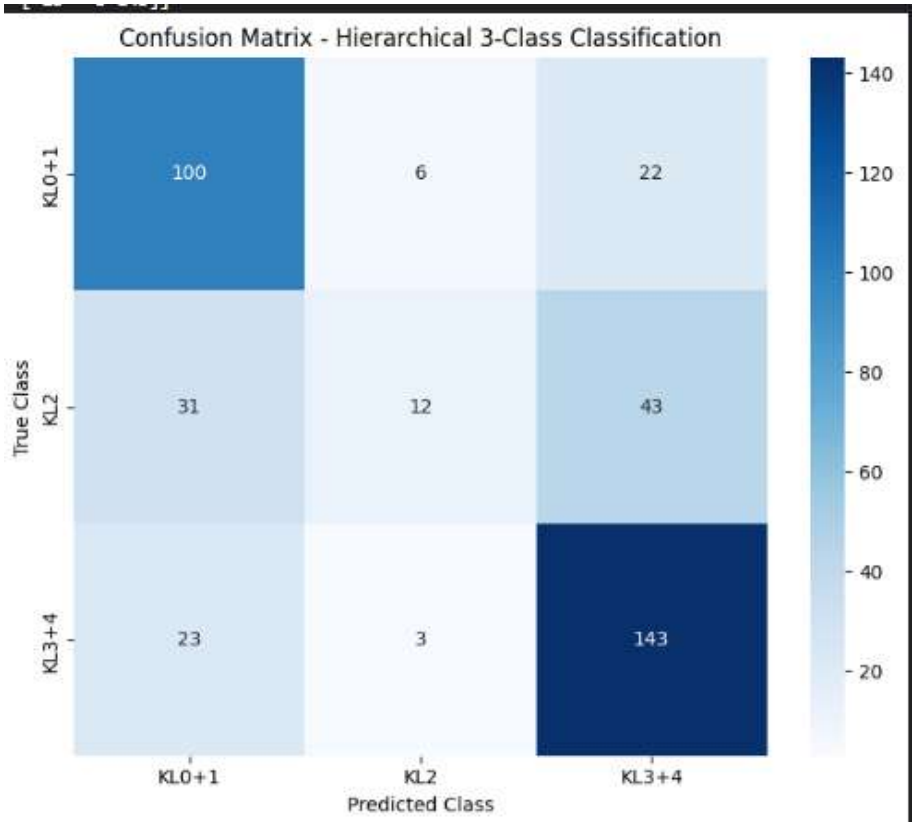


Figure 3.16: Confusion matrix obtained for the three-class osteoarthritis classification task.

Figure 3.16 further illustrates the classification behavior. The confusion matrix reveals that the majority of KL0–1 and KL3–4 subjects were correctly classified, whereas KL2 cases were frequently misclassified as either No OA or Severe OA. Only 12 out of 86 KL2 subjects were correctly identified, confirming the difficulty of separating the intermediate disease stage from adjacent severity categories.

To better understand this behavior, Principal Component Analysis (PCA) was applied to the extracted feature space as mentioned in the figure 3.17.

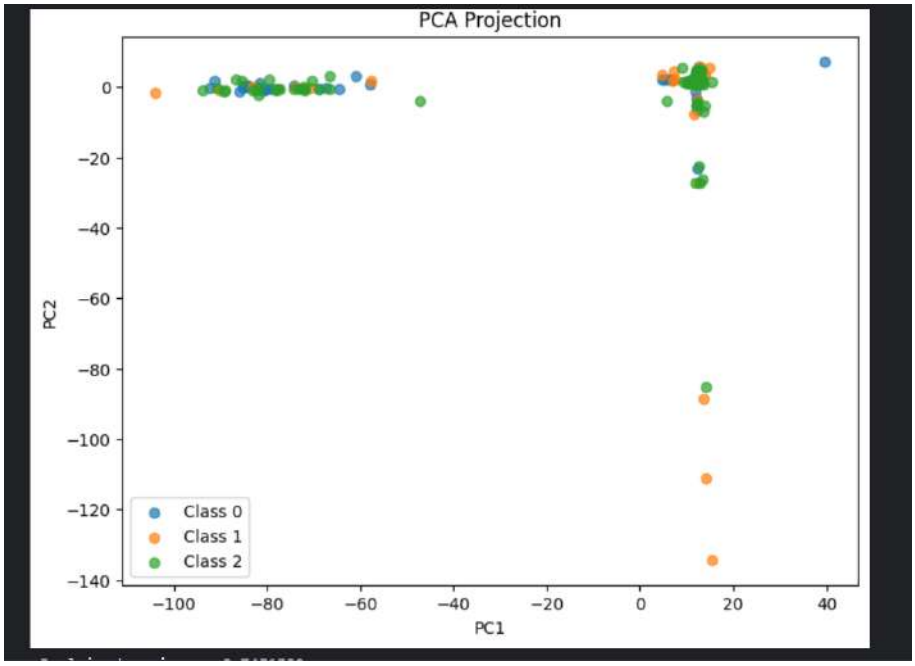


Figure 3.17: PCA projection of the extracted anatomical features for the three osteoarthritis severity groups.

As illustrated in Figure 3.17, a substantial overlap can be observed between the KL2 category and the neighboring severity groups. The PCA visualization confirms that mild osteoarthritis occupies an intermediate position between healthy and advanced disease stages. This overlap provides a visual explanation for the reduced classification performance observed for KL2 and highlights the intrinsic difficulty of identifying this transitional stage.

The findings obtained from both the confusion matrix and PCA analysis indicate that the main limitation of the three-class formulation lies in the discrimination of KL2 subjects. While the No OA and Severe OA groups exhibit relatively distinct anatomical characteristics, KL2 cases share features with both neighboring categories, making accurate classification considerably more challenging.

Furthermore, the PCA analysis suggests that the observed overlap is not solely related to dataset size but also reflects the biological continuity of osteoarthritis progression. Consequently, even with additional training samples, distinguishing KL2 from adjacent severity stages would remain a challenging task because the corresponding anatomical characteristics partially overlap.

Based on these observations, additional classification strategies were investigated. Binary classification experiments were conducted to evaluate the separability of healthy and advanced osteoarthritis cases, while a dedicated KL2 detection model was developed to specifically address the most challenging stage of disease progression. The results of these

experiments are presented in the following sections.

3.4.5 Binary Classification Results

To further investigate the discriminative capability of the extracted anatomical descriptors, two binary classification experiments were conducted. The first experiment focused on distinguishing subjects without osteoarthritis from subjects exhibiting osteoarthritic changes. The second experiment aimed to separate non-severe cases (KL0–1) from advanced osteoarthritis cases (KL3–4), excluding the intermediate KL2 category.

Tables 3.12 and 3.13 summarize the classification performance obtained for both binary scenarios.

Table 3.12: Classification report for the No OA vs OA task (Accuracy = 0.7728, AUC = 0.8396).

Class	Precision	Recall	F1-score	Support
No OA	0.68	0.59	0.64	128
OA	0.81	0.86	0.83	255

Table 3.13: Classification report for the KL0–1 vs KL3–4 task (Accuracy = 0.8384, AUC = 0.9107).

Class	Precision	Recall	F1-score	Support
KL0–1	0.83	0.79	0.81	128
KL3–4	0.85	0.88	0.86	169

The results demonstrate that binary classification substantially improved prediction performance compared with the three-class scenario. This improvement can be attributed to the clearer separation between the target groups and the reduced ambiguity associated with intermediate disease stages.

The first binary experiment, which distinguished subjects without osteoarthritis from osteoarthritic subjects, achieved an accuracy of 77.28% and an AUC of 0.8396. These results indicate that the extracted anatomical and clinical features provide reliable information for detecting the presence of osteoarthritis.

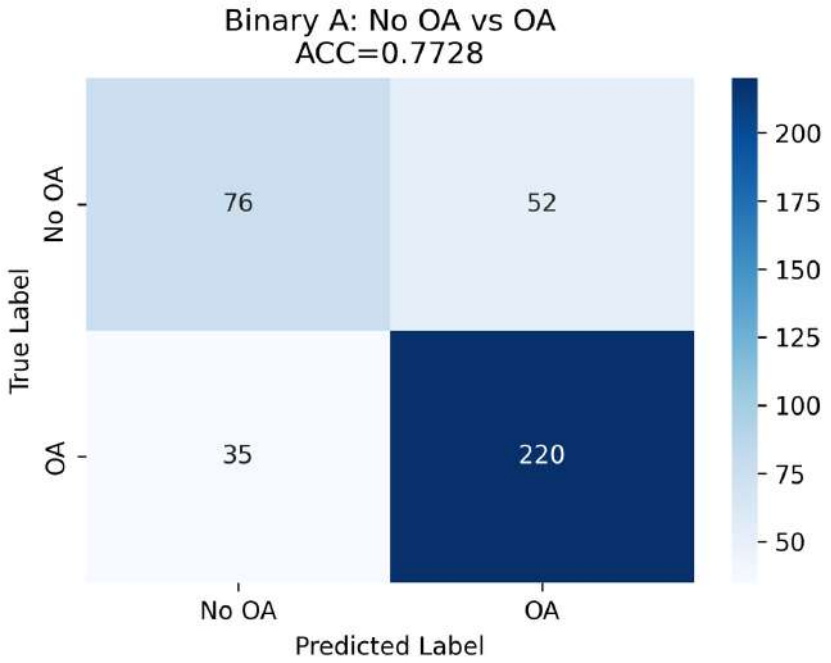


Figure 3.18: Confusion matrix obtained for the No OA versus OA classification task.

As illustrated in Figure 3.18, a substantial proportion of osteoarthritic subjects were correctly identified. The model achieved a precision of 0.81, a recall of 0.86, and an F1-score of 0.83 for the OA class, demonstrating strong sensitivity for detecting osteoarthritis cases.

The second binary experiment focused on distinguishing non-severe cases (KL0–1) from advanced osteoarthritis cases (KL3–4). To increase class separability, the intermediate KL2 category was excluded from the analysis, resulting in a dataset of 297 subjects.

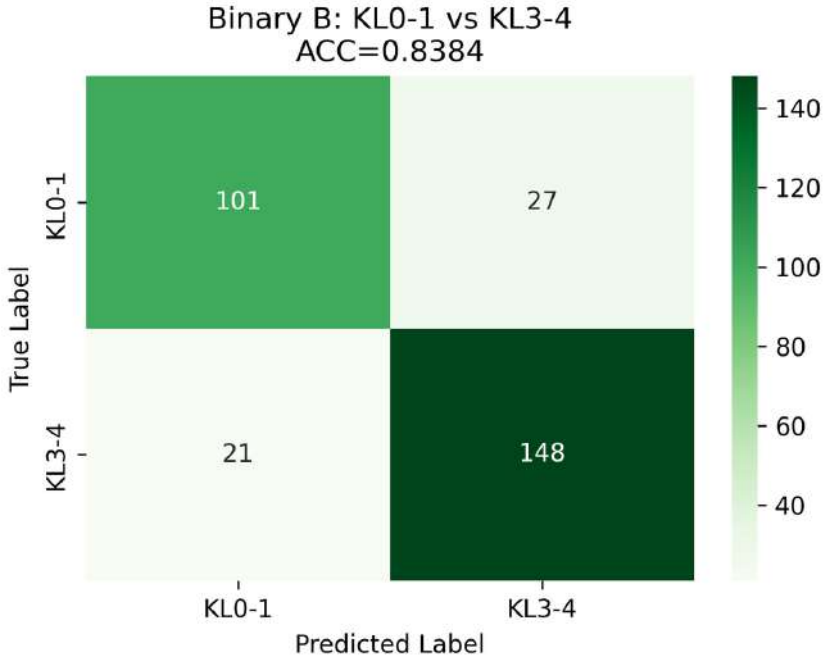


Figure 3.19: Confusion matrix obtained for the KL0–1 versus KL3–4 classification task.

This experiment achieved the highest performance among all classification tasks, with an accuracy of 83.84% and an AUC of 0.9107. The confusion matrix shown in Figure 3.19 indicates that most subjects were correctly assigned to their respective severity groups.

The classifier achieved a precision of 0.85, a recall of 0.88, and an F1-score of 0.86 for the KL3–4 category. The superior performance observed in this experiment can be explained by the more pronounced anatomical differences between early-stage and advanced osteoarthritis, which facilitate class discrimination.

Overall, the binary classification results confirm the effectiveness of the proposed feature extraction framework and demonstrate that the extracted anatomical descriptors contain substantial information related to osteoarthritis severity.

3.4.6 KL2 Detection Results

Because KL2 represents a transitional stage between early and advanced osteoarthritis, an additional experiment was conducted to specifically identify KL2 subjects. This classification task is clinically important because KL2 corresponds to the onset of definite radiographic osteoarthritis and often exhibits characteristics shared with both healthy and more advanced disease stages.

The classification model was trained using the extracted anatomical descriptors combined with clinical information. Five-fold cross-validation was employed to evaluate the

robustness of the proposed framework.

Table 3.14: Classification report for the KL2 detection task, One-vs-Rest (Accuracy = 0.7076, AUC = 0.6909).

Class	Precision	Recall	F1-score	Support
Not KL2	0.85	0.75	0.80	297
KL2	0.39	0.55	0.46	86

Figure 3.20 illustrates the ROC curve obtained for the KL2 detection task.

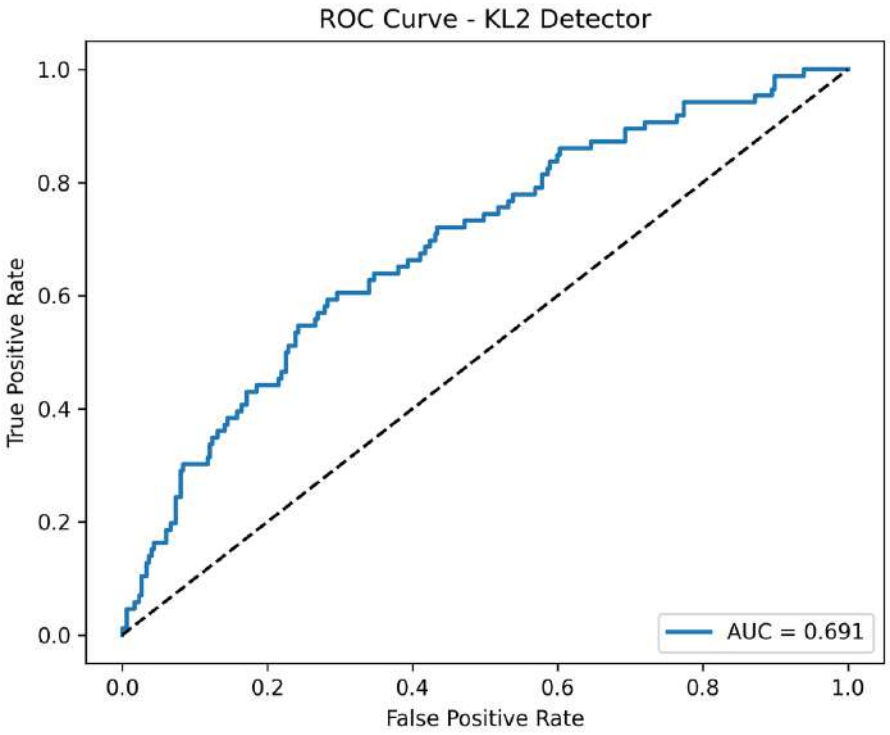


Figure 3.20: ROC curve obtained for the KL2 detection task.

Figure 3.21 depicts the confusion matrix obtained for the KL2 detection task.

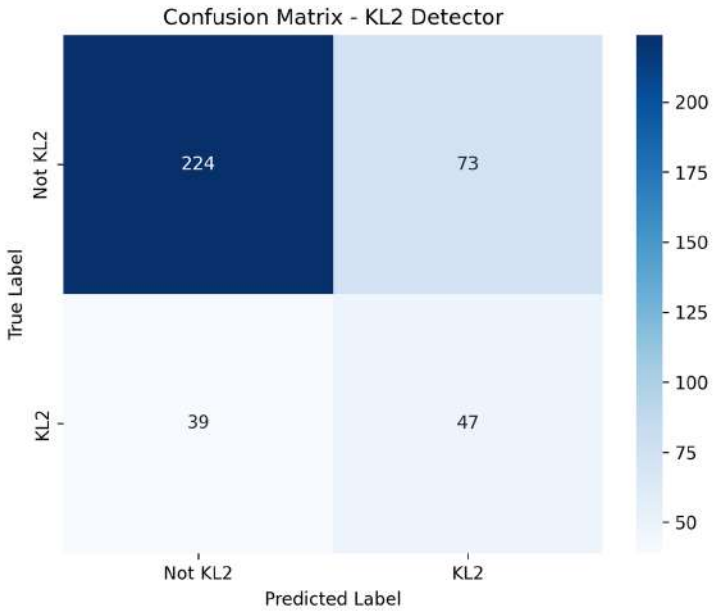


Figure 3.21: Confusion matrix obtained for the KL2 detection task.

The obtained results indicate that KL2 detection remains a challenging classification problem. The proposed framework achieved an accuracy of 70.76% and an AUC of 0.6909, reflecting the difficulty of distinguishing KL2 subjects from neighboring osteoarthritis stages.

The classifier achieved a recall of 55%, indicating that more than half of the KL2 subjects were correctly identified. Although the precision remained relatively modest, this behavior is expected because KL2 represents an intermediate disease stage that shares anatomical characteristics with both early and advanced osteoarthritis.

Compared with the binary classification experiments, KL2 detection produced lower performance. This observation further supports the hypothesis that transitional osteoarthritis stages exhibit less distinct anatomical boundaries, making classification substantially more challenging.

Nevertheless, the obtained results demonstrate that the extracted anatomical descriptors contain useful information for identifying KL2 subjects and provide a promising basis for early osteoarthritis assessment.

3.4.7 Five-Class Classification Results

The final and most challenging classification experiment consisted of predicting the complete Kellgren–Lawrence grading scale, where subjects were assigned to one of five categories (KL0, KL1, KL2, KL3, or KL4). Unlike the binary and three-class scenarios, this task requires the classifier to distinguish between adjacent disease stages that often

exhibit subtle anatomical differences.

The classification dataset consisted of 383 subjects distributed across the five KL grades, including 82 KL0 cases, 46 KL1 cases, 86 KL2 cases, 111 KL3 cases, and 58 KL4 cases.

Five-fold cross-validation was performed to evaluate the robustness of the proposed framework. The classification accuracies obtained for each fold are presented in the table 3.15.

Table 3.15: Five-fold cross-validation results for the five-class classification task.

Fold	Accuracy
1	0.4545
2	0.4805
3	0.3377
4	0.4342
5	0.4605
Mean	0.4335
Std	0.0501

The obtained results demonstrate moderate variability across the validation folds, with a mean accuracy of 43.35% and a standard deviation of 5.01%. The lower performance compared with the binary and three-class experiments reflects the increased difficulty associated with fine-grained osteoarthritis grading.

Table 3.16 illustrates overall performance of the five-class osteoarthritis classification task.

Table 3.16: Overall performance of the five-class osteoarthritis classification task.

Metric	Value
Accuracy	0.4334
Balanced Accuracy	0.3728
Weighted F1-score	0.4026

Table 3.17 depicts Per-class performance for the five-class classification task.

Table 3.17: Per-class performance for the five-class classification task.

Class	Precision	Recall	F1-score
KL0	0.47	0.60	0.52
KL1	0.25	0.07	0.10
KL2	0.39	0.37	0.38
KL3	0.44	0.64	0.52
KL4	0.55	0.19	0.28

Figure 3.22 shows the confusion matrix obtained for the five-class osteoarthritis classification task.

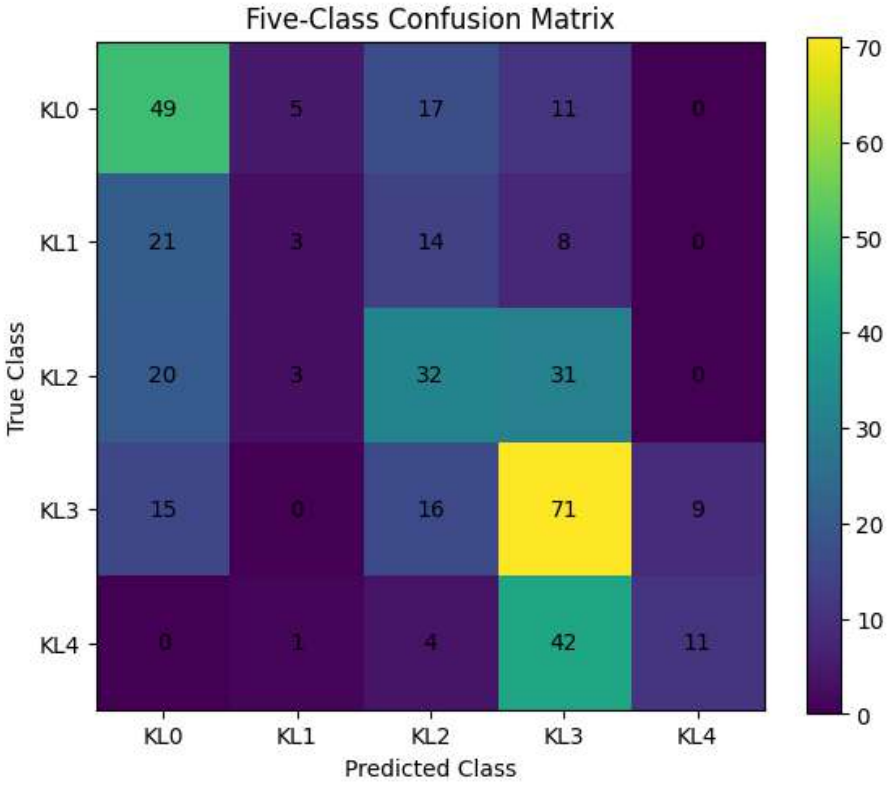


Figure 3.22: Confusion matrix obtained for the five-class osteoarthritis classification task.

The five-class classification task achieved an overall accuracy of 43.34% and a weighted F1-score of 0.403. As expected, performance decreased compared with the binary and three-class experiments because the classifier was required to distinguish between adjacent osteoarthritis grades exhibiting highly similar anatomical characteristics.

The highest recall values were obtained for KL0 and KL3, indicating that the model was relatively successful in identifying healthy subjects and subjects with advanced osteoarthritis. In contrast, KL1 and KL2 exhibited lower performance, reflecting the difficulty of distinguishing early and transitional disease stages.

Visual inspection of the confusion matrix reveals that most classification errors occurred between neighboring grades, particularly between KL1 and KL2 and between KL2 and KL3. This behavior is consistent with the progressive nature of osteoarthritis, where anatomical changes occur gradually rather than abruptly.

Although the five-class classification task remains considerably more challenging than binary prediction, the obtained results demonstrate that the extracted anatomical descriptors retain meaningful information related to disease severity and are capable of supporting fine-grained osteoarthritis assessment.

3.5 Evaluation on Independent Test Set

To further assess the generalization capability of the proposed framework, an additional evaluation was conducted using an independent test set that was not involved in model training, validation, or hyperparameter optimization. The objective of this experiment was to investigate the performance of the complete pipeline on previously unseen MRI examinations.

The independent test cohort consisted of 76 subjects, including 33 KL0–1 cases and 43 KL3–4 cases. For each subject, the complete framework was applied, including binary knee localization, ROI extraction, multi-class anatomical segmentation, feature extraction, and osteoarthritis classification.

3.5.1 Qualitative Evaluation

Figures 3.23 and 3.24 present representative examples of the complete inference pipeline applied to two unseen MRI examinations from the independent test set. In both cases, the binary localization model successfully identified the knee region and generated anatomically focused ROIs. The cropped volumes were subsequently processed by the multi-class segmentation model.

The predicted anatomical structures closely matched the corresponding ground-truth annotations, yielding a mean Dice score of 0.8487. The femoral and lateral tibial regions achieved particularly high segmentation accuracy, whereas the meniscal structures remained comparatively more challenging because of their thin morphology and complex anatomical boundaries.

These qualitative observations demonstrate the effectiveness of the proposed coarse-to-fine segmentation strategy and confirm its ability to generalize to previously unseen MRI examinations.

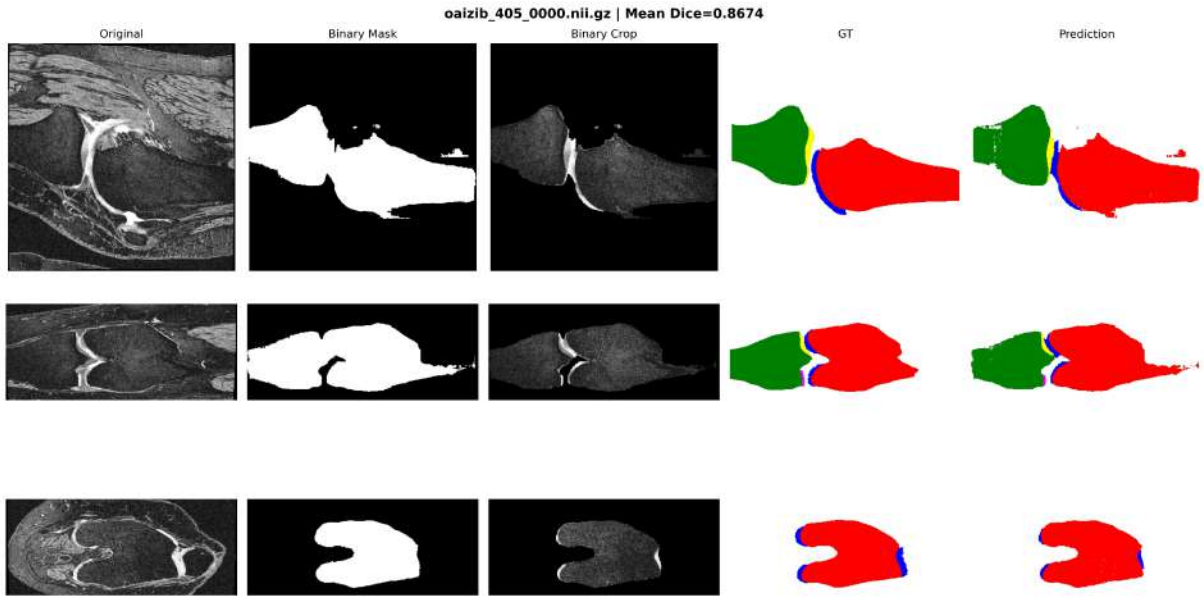


Figure 3.23: Complete inference pipeline for Subject 405 from the independent test set.

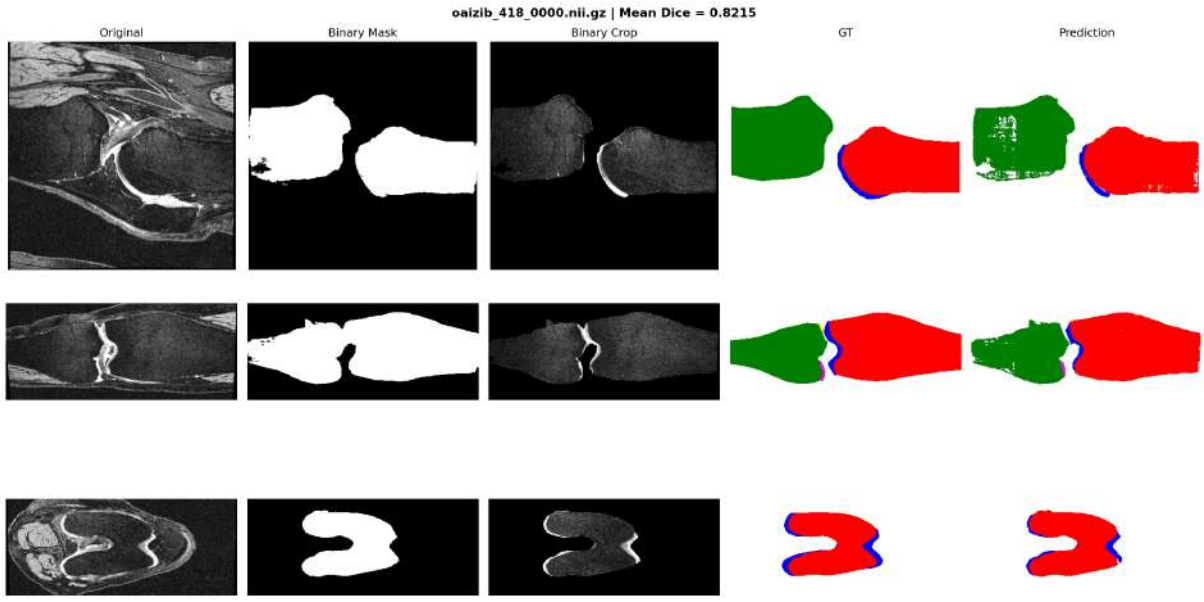


Figure 3.24: Complete inference pipeline for Subject 418 from the independent test set.

3.5.2 Quantitative Evaluation

The classification performance obtained on the independent test set is summarized in Table 3.18. The proposed binary classification model achieved an accuracy of 81.58

Table 3.18: Binary classification performance on the independent test set.

Metric	Value
Number of cases	76
KL0–1 subjects	33
KL3–4 subjects	43
Accuracy	81.58%
AUC	0.8478

The obtained results indicate that the proposed framework maintained stable predictive performance when applied to previously unseen MRI examinations. This observation suggests that the extracted anatomical biomarkers and classification strategy possess satisfactory generalization capability beyond the training cohort.

3.5.3 Classification Performance Comparison

To further evaluate the effectiveness of the proposed framework, the obtained classification results were compared with representative osteoarthritis prediction studies discussed in Chapter 1. Table 3.19 summarizes the comparison.

Table 3.19: Comparison with representative studies from the literature.

Method	Year	Task	Dataset Type	Annotation Type	Performance
RF-LASSO [47]	2012	Binary OA Prediction	OAI (MRI)	Radiomics	AUC = 0.82
KOA-DNN [48]	2021	Five-Class Classification	OAI (X-ray)	KL Grade	Acc = 69.7%
Multi-Task OA Framework [49]	2023	OA Severity Prediction	OAI (MRI)	KL Grade	AUC = 0.89
MTDL-System [50]	2025	Binary OA Classification	Private (MRI)	KL Grade	AUC = 0.902
Proposed Framework	**2026**	**Binary (KL0–1 vs KL3–4)**	**OAIZIB-CM (MRI)**	**KL + Mask**	**AUC = 0.9107**

The comparison presented in Table 3.19 demonstrates that the proposed framework achieved competitive classification performance relative to recently published approaches. In particular, the proposed binary classification model achieved an AUC of 0.9107 when distinguishing KL0–1 subjects from KL3–4 subjects.

This performance slightly exceeds the AUC reported by the MTDL-System study (AUC = 0.902), indicating that the extracted anatomical descriptors provide highly discriminative information for osteoarthritis severity assessment.

Furthermore, the obtained results were achieved using handcrafted anatomical features

extracted from automatically segmented knee structures, without relying exclusively on end-to-end image-based classification. This observation highlights the effectiveness of combining anatomical segmentation and region-based feature extraction for osteoarthritis prediction.

3.5.4 Scientific Contribution Analysis

Beyond the quantitative results, the proposed framework introduces several methodological contributions that distinguish it from previously reported studies.

First, the framework integrates multiple processing stages within a unified automated pipeline, including MRI preprocessing, binary knee localization, ROI extraction, multi-class anatomical segmentation, feature extraction, and osteoarthritis classification. Most previous studies focused on only one or two of these stages independently.

Second, a two-stage segmentation strategy was adopted. Binary localization was initially performed using a Swin-UNETR model to identify the knee region automatically. The resulting masks were subsequently utilized to generate anatomically focused ROIs for multi-class segmentation. This strategy reduced background interference and improved anatomical focus during training and inference.

Third, the proposed methodology combined anatomical descriptors extracted from multiple segmented structures with clinical information, including age and gender, to support osteoarthritis prediction.

Finally, a dedicated KL2 detection experiment was conducted in order to investigate the transitional osteoarthritis stage separately. Because KL2 represents an intermediate stage between healthy and advanced osteoarthritis, it is frequently reported as one of the most challenging categories to classify accurately. To the best of our knowledge, dedicated KL2 detection is rarely investigated as an independent classification task in the studies reviewed in Chapter 1.

Overall, the proposed framework provides a complete and automated solution for knee MRI analysis, combining segmentation and classification within a single unified pipeline while maintaining competitive performance relative to recent state-of-the-art approaches.

3.6 Streamlit-Based Application

To demonstrate the practical applicability of the proposed framework, an interactive graphical user interface (GUI) was developed using the Streamlit framework. The application integrates the different stages of the proposed pipeline within a unified environment

for automated knee osteoarthritis analysis from MRI data.

The developed interface allows users to provide patient clinical information, including age, body mass index (BMI), and gender, in addition to loading MRI examinations for automated processing. Once the MRI volume is uploaded, the application automatically executes the successive stages of the proposed framework, including knee localization, ROI extraction, multi-class anatomical segmentation, feature extraction, and osteoarthritis severity prediction.

Figure 3.25 illustrates the segmentation module of the developed application. The interface provides visualization of MRI slices in different anatomical planes (axial, coronal, and sagittal) together with the corresponding segmentation overlays generated by the proposed Swin-UNETR model. This visualization enables users to inspect the segmented anatomical structures and verify the quality of the generated predictions.

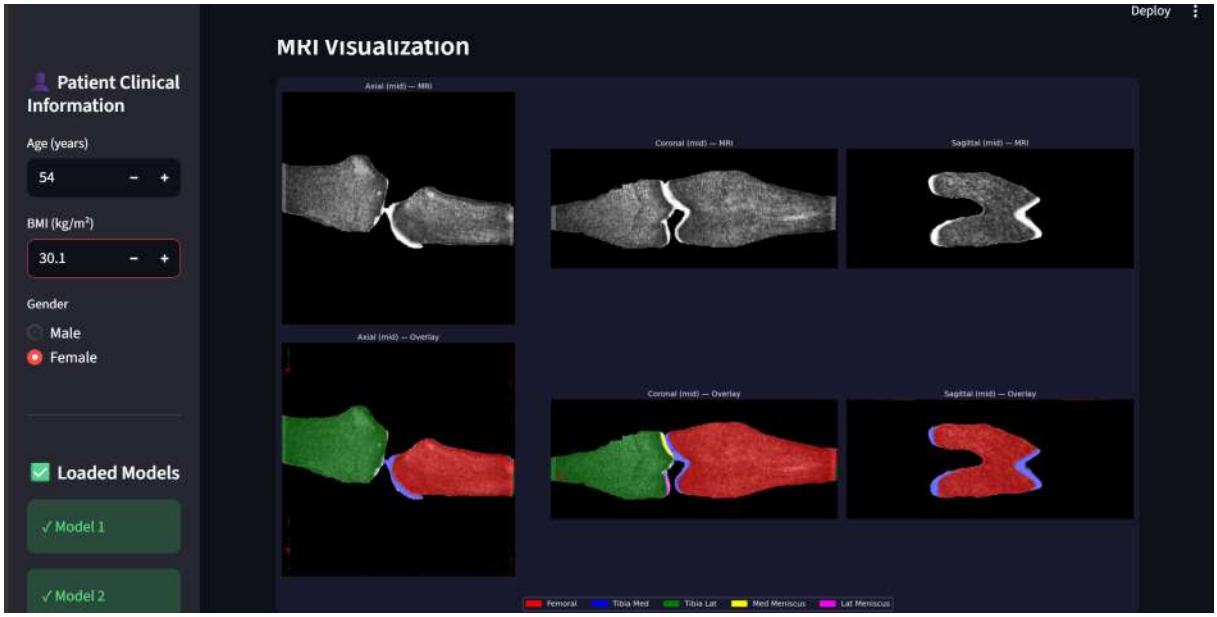


Figure 3.25: Streamlit interface showing patient clinical information, MRI visualization, and multi-class anatomical segmentation results generated by the proposed framework.

Figure 3.26 presents the classification module of the application. The interface summarizes the predictions generated by the different classification models and provides confidence scores together with probability distributions. These visual outputs improve the interpretability of the classification results and facilitate the assessment of osteoarthritis severity.

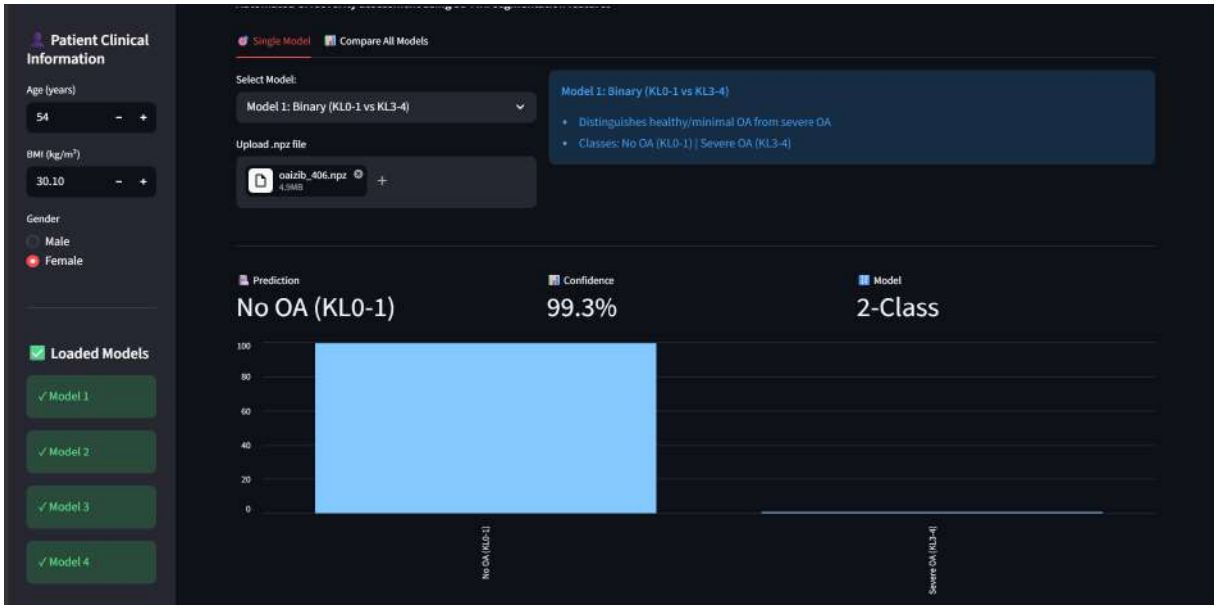


Figure 3.26: Classification results generated by the Streamlit application, including prediction confidence scores and probability distributions for the different classification models.

The developed Streamlit application demonstrates the feasibility of deploying the proposed methodology in an accessible and user-friendly environment. By integrating image visualization, anatomical segmentation, clinical information, feature extraction, and classification results within a single platform, the application provides a complete end-to-end solution for automated knee osteoarthritis assessment. Such an implementation may facilitate future integration of artificial intelligence-based diagnostic tools into clinical decision-support systems and medical research workflows.

3.7 Discussion

3.7.1 Interpretation of Findings

The experimental results obtained throughout this study demonstrate the effectiveness of the proposed multi-stage framework for automated knee osteoarthritis analysis. The high performance achieved during the binary localization stage confirms the ability of the Swin-UNETR architecture to accurately identify knee joint regions within MRI volumes, thereby enabling reliable ROI extraction for subsequent processing.

The multi-class segmentation results further indicate that the proposed framework successfully captured the anatomical characteristics of multiple knee structures. Higher segmentation accuracy was achieved for large anatomical regions such as the femoral bone and tibial compartments, whereas lower Dice scores were observed for the meniscal

structures. This behavior is consistent with the anatomical complexity and relatively small size of the menisci compared with bone structures.

The classification experiments provided additional insights into osteoarthritis severity prediction. The three-class formulation achieved an overall accuracy of 69.18%, demonstrating that the extracted anatomical descriptors contain meaningful information related to disease progression. However, detailed analysis revealed that the KL2 category represented the most challenging class.

Both the confusion matrix and PCA visualization showed substantial overlap between KL2 subjects and the neighboring severity groups. This observation confirms the transitional nature of mild osteoarthritis and explains the reduced recall obtained for KL2 during three-class classification. Consequently, additional binary classification experiments and a dedicated KL2 detection model were developed in order to further investigate disease progression and improve classification interpretability.

Overall, the obtained findings demonstrate that combining anatomical segmentation with feature-based classification provides a meaningful representation of structural changes associated with knee osteoarthritis.

3.7.2 Clinical Relevance

Knee osteoarthritis is one of the most prevalent musculoskeletal disorders worldwide and represents a major cause of pain, disability, and reduced quality of life. Early detection and accurate assessment of disease severity are therefore essential for improving patient management and treatment planning.

The proposed framework contributes to this objective by providing a fully automated pipeline capable of processing MRI data without manual intervention. Automatic localization, segmentation, and classification reduce the need for time-consuming manual analysis and may improve diagnostic consistency across clinical evaluations.

Particularly important is the investigation of KL2 cases, which correspond to the transitional stage between healthy and advanced osteoarthritis. Accurate identification of this stage may facilitate earlier clinical intervention and support patient monitoring before severe structural degeneration occurs.

Although the proposed framework is intended primarily for research purposes, the obtained results demonstrate its potential applicability as a computer-aided decision support tool for osteoarthritis assessment and severity prediction.

3.7.3 Robustness Analysis

Several observations indicate the robustness of the proposed framework. First, stable convergence behavior was observed during both localization and segmentation training, as reflected by the progressive decrease in loss values and the consistent improvement of validation metrics.

Second, the binary localization model achieved a Dice score exceeding 97%, demonstrating reliable knee joint localization despite anatomical variability among subjects. The resulting ROI extraction process enabled the segmentation model to focus on anatomically relevant regions while reducing the influence of irrelevant background information.

Third, the segmentation model maintained satisfactory performance across multiple anatomical structures with different sizes and shapes. Although meniscal segmentation remained more challenging than bone segmentation, acceptable accuracy was achieved across all target regions.

The classification experiments also demonstrated consistent predictive capability across different evaluation scenarios. Rather than relying on a single classification formulation, multiple strategies were investigated, including five-class classification, three-class severity prediction, binary severity assessment, and dedicated KL2 detection. This progressive experimental process provided a comprehensive evaluation of the extracted anatomical descriptors and confirmed their relevance for osteoarthritis assessment.

Nevertheless, several limitations should be acknowledged. The available dataset remains relatively small compared with large-scale clinical studies, and class imbalance was observed for certain KL grades. Furthermore, the substantial overlap between KL2 and neighboring categories suggests that disease progression is inherently continuous rather than strictly separable into discrete classes. Future studies involving larger multi-center datasets and additional clinical variables may further improve model generalization and robustness.

3.8 Conclusion

This chapter presented the experimental evaluation of the proposed framework for automated knee osteoarthritis analysis using MRI data. The effectiveness of the preprocessing stage was first assessed through qualitative visual inspection, demonstrating improved image quality while preserving relevant anatomical structures.

The binary localization model achieved a Dice score of 97.52%, enabling accurate automatic extraction of knee-centered regions of interest. The generated ROI volumes

reduced irrelevant background information and provided focused inputs for the subsequent segmentation stage.

The multi-class Swin-UNETR model successfully segmented multiple knee anatomical structures, achieving a mean Dice score of 79.36% and a mean IoU score of 70.57%. The best segmentation performance was obtained for the femoral bone, while the meniscal structures remained the most challenging regions because of their smaller size and more complex morphology.

The extracted anatomical features were subsequently utilized for osteoarthritis classification. Several classification strategies were investigated, including five-class classification, three-class severity prediction, binary classification, and dedicated KL2 detection. The obtained results demonstrated that the extracted anatomical descriptors contain meaningful information related to disease severity and progression.

Among all investigated classification scenarios, the best performance was achieved for the KL0–1 versus KL3–4 classification task, yielding an accuracy of 83.84% and an AUC of 0.9107.

Finally, the comparative analysis highlighted the competitiveness of the proposed framework and emphasized its main contribution: the integration of preprocessing, localization, ROI extraction, segmentation, feature extraction, and osteoarthritis classification within a unified automated pipeline.

The next chapter presents the general conclusion of this work, summarizes the main contributions, discusses the limitations of the proposed framework, and outlines potential directions for future research.

GENERAL CONCLUSION

AND FUTURE PERSPECTIVES

This work addressed the problem of automated knee osteoarthritis assessment using magnetic resonance imaging (MRI). Osteoarthritis remains one of the most common degenerative joint diseases worldwide, and accurate evaluation of disease severity is essential for diagnosis, treatment planning, and patient monitoring.

To address this challenge, a complete deep learning framework was developed for automatic knee MRI analysis. The proposed methodology combined image preprocessing, knee localization, ROI extraction, multi-class anatomical segmentation, feature extraction, and osteoarthritis classification within a unified processing pipeline.

The preprocessing stage improved image quality and enhanced anatomical visibility while preserving clinically relevant structures. Subsequently, a binary Swin-UNETR model was employed for automatic knee localization, achieving a Dice score of 97.52% and enabling reliable extraction of knee-centered regions of interest.

The extracted ROIs were then utilized for multi-class segmentation of major knee anatomical structures. The segmentation model achieved a mean Dice score of 79.36% and a mean IoU score of 70.57%, demonstrating its ability to accurately delineate multiple structures despite anatomical variability and class imbalance.

Following segmentation, anatomical descriptors were extracted and utilized for osteoarthritis severity prediction. Several classification formulations were investigated, including five-class classification, three-class severity prediction, binary classification, and dedicated KL2 detection. The experimental results demonstrated that the extracted anatomical features contain valuable information related to disease progression. The best classification performance was obtained for the KL0–1 versus KL3–4 experiment, achieving an accuracy of 83.84% and an AUC of 0.9107.

A detailed analysis of the three-class classification task revealed that KL2 represents the most challenging category because of its transitional nature between healthy and advanced osteoarthritis. Both the confusion matrix and PCA visualization confirmed the substantial overlap between KL2 and neighboring severity groups. These findings motivated the investigation of additional binary classification strategies and the development of a dedicated KL2 detection model.

Overall, the obtained results demonstrate the feasibility and effectiveness of combining localization, segmentation, feature extraction, and classification within a unified framework for automated knee osteoarthritis assessment.

Main Contributions

The principal contributions of this work can be summarized as follows:

- Development of a complete automated framework for knee osteoarthritis analysis using MRI data.
- Implementation of a binary Swin-UNETR localization model for automatic knee ROI extraction, achieving a Dice score of 97.52%.
- Development of a multi-class anatomical segmentation framework capable of segmenting multiple knee structures with a mean Dice score of 79.36%.
- Extraction of anatomical descriptors from segmented knee structures for osteoarthritis severity prediction.
- Investigation of multiple classification formulations, including five-class, three-class, and binary classification strategies.
- Identification of KL2 as the most challenging osteoarthritis category through confusion matrix analysis and PCA visualization.
- Development of a dedicated KL2 detection strategy to further investigate intermediate disease stages.
- Integration of all processing stages into a unified automated pipeline for knee MRI analysis.

Limitations

Despite the promising results obtained in this study, several limitations should be acknowledged.

First, the available dataset remains relatively limited compared with large-scale clinical studies. Increasing the number of subjects could further improve model generalization and robustness.

Second, class imbalance was observed across certain osteoarthritis severity grades, particularly for intermediate disease stages. This imbalance may influence classification performance and contribute to prediction uncertainty.

Third, the KL2 category exhibited substantial overlap with neighboring severity groups, reflecting the continuous nature of osteoarthritis progression and making accurate discrimination particularly challenging.

Finally, the proposed framework was evaluated using a single dataset. Additional validation on external multi-center datasets would provide a more comprehensive assessment of generalization capability and clinical applicability.

Future Perspectives

Several directions may be explored to further improve the proposed framework:

- Evaluation of the proposed methodology on larger and more diverse MRI datasets.
- Investigation of self-supervised and semi-supervised learning strategies to improve feature representation.
- Integration of additional clinical variables and patient metadata for enhanced osteoarthritis prediction.
- Development of explainable artificial intelligence techniques to improve model interpretability and clinical trust.
- Investigation of longitudinal MRI studies for monitoring osteoarthritis progression over time.
- Validation of the proposed framework in real clinical environments and decision-support systems.

Bibliography

[1] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2017.

[2] A. Aghdam, "3d u-net architecture repository." https://github.com/amir-aghdam/3D-UNet, 2021. Accessed: 2026-05-12.

[3] O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical image segmentation," *MICCAI*, 2015.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017.

[5] Project MONAI, "Swin-unetr research repository." https://github.com/Project-MONAI/research-contributions, 2022. Accessed: 2026-05-12.

[6] A. Hatamizadeh *et al.*, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," *arXiv preprint arXiv:2201.01266*, 2022.

[7] A. Tiulpin *et al.*, "Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach," *PLOS ONE*, vol. 16, no. 10, p. e0258855, 2021.

[8] D. J. Hunter and A. Guermazi, "Imaging techniques in osteoarthritis," *PM&R*, vol. 4, no. 5, pp. 68–74, 2012.

[9] P. Herrera *et al.*, "Artificial intelligence approaches for osteoarthritis analysis: A review," *Arthritis Research & Therapy*, vol. 27, p. 81, 2025.

[10] K. Mead, T. Cross, G. Roger, R. Sabharwal, S. Singh, and N. Giannotti, "MRI deep learning models for assisted diagnosis of knee pathologies: A systematic review," *European Radiology*, vol. 35, no. 5, pp. 2457–2469, 2024.

[11] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2D trained network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1522–1534, 2018.

[12] P. Lambin *et al.*, "Radiomics: extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012.

[13] J. Yang *et al.*, "Wavelet-based denoising methods for medical images: A review," *Biomedical Signal Processing and Control*, vol. 68, p. 102589, 2021.

[14] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics Gems IV*, pp. 474–485, 1994.

[15] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[16] A. Hatamizadeh *et al.*, "Unetr: Transformers for 3d medical image segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 574–584, 2022.

[17] X. Li, S. Lv, M. Li, J. Zhang, Y. Jiang, Y. Qin, H. Luo, and S. Yin, "Sdmt: Spatial dependence multi-task transformer network for 3d knee mri segmentation and landmark localization," *IEEE Transactions on Medical Imaging*, vol. 42, no. 8, 2023.

[18] Z. Xie, Z. Qiu, Y. Li, X. Chen, Q. Ye, M. Wang, X. Chen, Y. Shao, C. Liu, *et al.*, "Development of a multi-task deep learning system for classification of nine common knee abnormalities on mri: a large-scale, multicentre, stepwise validation study," *eClinicalMedicine*, vol. 89, p. 103534, 2025.

[19] S. W. Zamir *et al.*, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5728–5739, 2022.

[20] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Pearson, 4 ed., 2018.

[21] F. Isensee *et al.*, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[22] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[23] P. Chen, L. Gao, X. Shi, K. Allen, and L. Yang, "Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 84–92, 2019.

[24] A. Tiulpin *et al.*, "Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach," *Scientific Reports*, vol. 8, 2018.

[25] C. Guida, M. Zhang, and J. Shan, "Improving knee osteoarthritis classification using multimodal intermediate fusion of X-ray, MRI, and clinical information," *Neural Computing and Applications*, vol. 35, no. 13, pp. 9763–9772, 2023.

[26] A. Kirillov *et al.*, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.

[27] Z. Wang, Z. Li, P. Zhang, *et al.*, "Semi-supervised knee cartilage segmentation with multi-task knowledge distillation," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 3, pp. 2143–2154, 2025.

[28] Z. Xie, Z. Qiu, Y. Li, *et al.*, "Development of a multi-task deep learning system for classification of nine common knee abnormalities on MRI: A large-scale, multicentre, stepwise validation study," *eClinicalMedicine*, vol. 89, p. 103534, 2025.

[29] H. Roth *et al.*, "Automatic region of interest localization in medical images using deep learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2313–2323, 2018.

[30] F. Ambellan, A. Tack, M. Ehlke, and S. Zachow, "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks," *Medical Image Analysis*, vol. 52, pp. 109–118, 2019.

[31] Y. YAO, "Oaizib-cm dataset." https://huggingface.co/datasets/YongchengYAO/OAIZIB-CM, 2024. Accessed: 2026-05-08.

[32] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *ICLR*, 2019.

[33] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, 2019.

[34] F. Milletari *et al.*, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *3DV*, 2016.

[35] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[36] F. Eckstein *et al.*, "Quantitative mri of cartilage and bone: degenerative changes in osteoarthritis," *NMR in Biomedicine*, vol. 19, no. 7, pp. 822–854, 2006.

[37] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Pearson, 4th ed., 2018.

[38] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 2008.

[39] D. J. Hunter and S. Bierma-Zeinstra, "Osteoarthritis," *The Lancet*, vol. 393, no. 10182, pp. 1745–1759, 2019.

[40] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "Catboost: unbiased boosting with categorical features," in *Advances in Neural Information Processing Systems*, vol. 31, 2018.

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

[42] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[43] M. J. Cardoso *et al.*, "Monai: An open-source framework for deep learning in healthcare," *arXiv preprint arXiv:2211.02701*, 2022.

[44] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," *arXiv preprint arXiv:2201.01266*, 2022.

[45] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 424–432, 2016.

[46] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.

[47] C. Kokkotis, S. Moustakidis, T. Tsatalas, and D. Tsaopoulos, "Machine learning in knee osteoarthritis: A review," *OARSI Open*, vol. 2, no. 3, p. 100069, 2020.

[48] A. S. Mohammed, A. A. Hasanaath, G. Latif, and A. Bashar, "Knee osteoarthritis detection and severity classification using residual neural networks on preprocessed X-ray images," *Diagnostics*, vol. 13, no. 8, p. 1380, 2023.

[49] J.-B. Schiratti, R. Dubois, P. Herent, D. Caháne, J. Dachary, T. Clozel, G. Wainrib, *et al.*, "A deep learning method for predicting knee osteoarthritis radiographic progression from MRI," *Arthritis Research & Therapy*, vol. 23, no. 1, p. 262, 2021.

[50] Z. Xie, Z. Qiu, Y. Li, *et al.*, "Development of a multi-task deep learning system for classification of nine common knee abnormalities on MRI: A large-scale, multicentre, stepwise validation study," *eClinicalMedicine*, vol. 89, p. 103534, 2025.