CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Osteoarthritis Detection Through Precise Demarcation of Knee Anatomical Structure

by

Aleena Javed

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Computing
Department of Computer Science

2025

CERTIFICATE OF APPROVAL

Osteoarthritis Detection Through Precise Demarcation of Knee Anatomical Structure

by

Aleena Javed

(MCS233012)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Muhammad Majid	UET, Taxila
(b)	Internal Examiner	Dr. Aamer Nadeem	CUST, Islamabad
(c)	Supervisor	Dr. Rizwan Bin Faiz	CUST, Islamabad

Dr. Rizwan Bin Faiz

Thesis Supervisor

April, 2026

Dr. M. Masroor Ahmed

Head

Dept. of Computer Science

April, 2026

Dr. M. Abdul Qadir

Dean

Faculty of Computing

April, 2026

Author's Declaration

I, **Aleena Javed** hereby state that my MS thesis titled "**Osteoarthritis Detection through Precise Demarcation of Knee Anatomical Structure.**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Aleena Javed)
MCS233012

Plagiarism Undertaking

I solemnly **Aleena Javed** declare that research work presented in this thesis titled "**Osteoarthritis Detection Through Precise Demarcation of Knee Anatomical Structure.**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Aleena Javed)
MCS233012

Acknowledgement

First and foremost, I offer my sincere gratitude to Allah Almighty, the Most Merciful, for granting me the strength, patience, and guidance to complete this thesis successfully. Without His blessings, this achievement would not have been possible.

I am deeply grateful to my supervisor, **Dr. Rizwan Bin Faiz**, for his invaluable guidance, continuous support, and insightful feedback throughout my research. His encouragement and expertise have been fundamental in shaping both the quality and direction of this work.

I would also like to express my heartfelt appreciation to my **Parents**, whose love, prayers, and unwavering support have been my foundation. Their sacrifices and belief in me have been a constant source of motivation. Finally, I am thankful to my **Brother**, whose encouragement and understanding have inspired me to persevere through challenges and complete this journey

(Aleena Javed)

Abstract

Osteoarthritis (OA) is a common musculoskeletal disorder causing pain, disability, and reduced quality of life. It involves progressive degeneration of cartilage, subchondral bone remodeling, and joint-space narrowing. Early OA is often underdiagnosed, as symptoms and radiographic signs, like Kellgren Lawrence grading, appear only after substantial structural damage.

An automated framework is developed that integrates deep learning–based segmentation, cartilage thickness estimation, and machine learning–based classification for early osteoarthritis (OA) detection from knee MRI.

A 3D U-Net was trained on 507 OAI-ZIB CM MRIs to segment femur, femoral cartilage, tibia, medial and lateral tibial cartilage. Dice Loss guided training; performance was evaluated via Dice Similarity Coefficient (DSC). Cartilage thickness was computed voxel-wise and summarized into region-level morphometric features, combined with demographic data (age, BMI, gender) for KL 0 vs. KL 1 OA classification. Random Forest and XGBoost classifiers were compared, with SMOTE addressing class imbalance.

DSC exceeded 0.97 for bones and 0.78–0.85 for cartilage. Cartilage-thickness features effectively distinguished healthy vs. early OA. XGBoost with SMOTE and hyperparameter tuning achieved 90% accuracy, outperforming Random Forest.

Segmentation-derived cartilage thickness is a reliable biomarker for early OA. The proposed framework provides a reproducible, scalable pipeline integrating MRI segmentation, morphometry, and machine-learning classification, addressing a key gap in automated OA assessment.

Contents

List of Figures

List of Tables

Abbreviations

BMI	Body Mass Index
CNN	Convolutional Neural Network
cGAN	Conditional Generative Adversarial Network
DESS	Dual Echo Steady State (MRI sequence)
KL	Kellgren–Lawrence (grading scale)
KOA	Knee Osteoarthritis
MONAI	Medical Open Network for AI (Python-based medical imaging framework)
MRI	Magnetic Resonance Imaging
NIfTI	Neuroimaging Informatics Technology Initiative (file format: .nii / .nii.gz)
OAI	Osteoarthritis Initiative
OA	Osteoarthritis
ROI	Region of Interest
SMOTE	Synthetic Minority Oversampling Technique
SSM	Statistical Shape Model
VOE	Volume Overlap Error

Symbols

p_i	Predicted voxel value at index i
g_i	Ground-truth voxel value at index i
L_{Dice}	Dice Loss used for segmentation training
M	Binary cartilage mask
B	Structural element used in morphological operations
$\ominus$	Morphological erosion operator
$\oplus$	Morphological dilation operator
S	Extracted cartilage surface (boundary voxels)
(x, y)	Spatial voxel coordinates on cartilage surfaces
$\|x - y\|^2$	Euclidean distance between two voxels
$T(x)$	Local cartilage thickness at voxel x
$\overline{T}$	Mean cartilage thickness for a region
N	Number of valid thickness voxels
$\hat{y}$	Final predicted class label
$h_t(x)$	Output of the t^{th} decision tree in Random Forest
mode$(\cdot)$	Majority vote function in Random Forest
g_i	Gradient of loss for sample i (XGBoost)
h_i	Hessian (second derivative) for sample i
$f_k(x)$	Output of the k^{th} boosted tree in XGBoost
η	Learning rate (shrinkage factor)

Chapter 1

Introduction

1.1 Background

Osteoarthritis (OA) is one of the most predominant musculoskeletal disorders which is characterized by progressive degeneration of articular cartilage, subchondral bone remodelling and the joint space narrowing. It is a leading cause of disability among the adult's worldwide [1].

The knee joint is the most regularly affected area due to its weight-bearing purpose and predisposition to mechanical stress.

While early detection of this structural change is crucial for slowing the disease progression and improving the patient outcomes. Figure 1.1 below illustrate the anatomical differences between a healthy knee joint and an osteoarthritic knee which highlighting the cartilage degeneration and the bony changes frequently associated with Osteoarthritis (OA).

The Magnetic Resonance Imaging (MRI) has developed as the gold-standard imaging modality for the evaluation of the structural alteration in the knee joint which is unlike radiography enables detailed visualization of soft tissues which including femoral and tibial cartilage, menisci and ligaments [2].

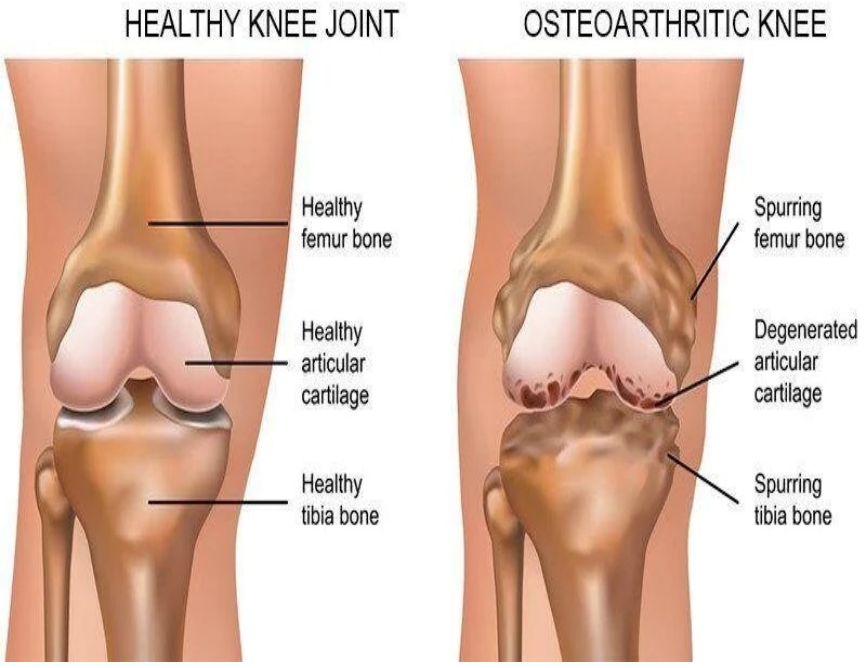

FIGURE 1.1: Healthy knee joint and osteoarthritis knee.

As cartilage thinning is among the earliest biomarkers of the OA the accurate identification of cartilage morphology is important.

Traditional segmentation methods such as the thresholding, atlas-based approaches and the region growing have certain limitation to reliably delineate the thin cartilage tissues particularly in the medial compartment where the degeneration starts prior[3][4].

On the other hand, manual segmentation is although accurate by it is time-consuming and also requires expert knowledge which makes it infeasible for large-scale studies.

Deep learning models specifically the Convolutional Neural Networks (CNNs) has revolutionized the medical image segmentation in their study which strengths includes strong classification accuracy.

U-net and its other variant have showed exceptional performance in various biomedical domains [5].

In the knee MRI research studies have shows the 3D CNNs capability of capturing volumetric context for accurate segmentation[6][7].

A recent study has introduced a democratized deep learning framework for the cartilage segmentation utilizing the MRIs form the Osteoarthritis Initiative (OAI).

Their model achieved a Dice score which is comparable to the state-of-art techniques which underlined reproducibility and the accessibility of the segmentation pipelines [8].

However, their study findings and other studies in literature highlighted a persistent challenge in the medial tibial cartilage segmentation due to its low contrast, thin geometry and partial volume effects. [9][10][11].

While most of the existing research remains limited to the anatomical delineation without integrating quantitative morphometrics such as cartilage thickness despite significant advancement in segmentation.

Additionally, only a limited number of studies have attempted to integrate imaging-based biomarkers with machine learning classifiers for predicting the severity of osteoarthritis (OA).

Most existing approaches either focus solely on feature extraction or apply classification techniques without fully leveraging the potential of combined methodologies.

Furthermore, none of the reviewed studies performed classification after a dedicated segmentation stage, despite substantial evidence suggesting that morphological features—such as joint space narrowing, cartilage degradation, and bone structure alterations—are strongly correlated with disease progression.

Incorporating segmentation prior to classification could therefore enhance the accuracy and interpretability of OA severity prediction models.[4][12].

1.2 Problem Satement

Although deep learning methods achieve high accuracy in knee structure segmentation, most studies treat segmentation and osteoarthritis (OA) classification as

separate tasks.

Classification is often performed directly on raw images without incorporating segmentation-derived biomarkers such as cartilage thickness, volume, and morphology. Although deep learning models can extract features from raw images, this approach may overlook subtle structural changes that are clinically important for early osteoarthritis (OA) detection.

Segmentation-derived biomarkers provide quantitative insights into cartilage degeneration, including thinning, surface irregularities, and regional variations, which are key indicators of early disease progression.

The lack of integration between segmentation and classification also reduces interpretability, as predictions are made without linking them to specific anatomical changes. Consequently, this underutilizes clinically meaningful quantitative features and may limit the accuracy and reliability of early OA assessment.

1.3 Research Questions

i. How can knee anatomical structures, including femur, tibia, and corresponding femoral and tibial cartilage regions, be accurately segmented from MRI data?

ii. What is the impact of integrating segmentation-derived features, particularly cartilage thickness, with clinical variables for improving osteoarthritis classification performance?

1.4 Research Objectives

i. To implement an accurate segmentation framework for extracting knee anatomical structures, including (Bones) femur, tibia, and their corresponding femoral and tibial cartilage regions, from MRI data.

ii. To quantitatively compute segmentation-derived cartilage thickness and integrate these morphometric features with clinical variables for osteoarthritis (OA) classification.

iii. To evaluate the impact of combining cartilage thickness biomarkers with clinical features on improving the performance and interpretability of OA classification models.

1.5 Purpose

The purpose of the study is to design and evaluate the automated, end-to-end processing pipeline for knee MRI data that integrate the deep learning-based segmentation of femoral and tibial structures, quantification of the cartilage thickness and OA classification based on machine learning.

This study goals are to address the limitations in the existing literature by building a framework that capable of not only segment cartilage precisely but can also interpret the extracted features for disease-stage prediction.

1.6 Scope of the Research

The scope of this research is limited to:

i. MRI scans obtained from the publicly available OAIZB-CM dataset.

ii. Segmentation based on five knee structures: Femur, Femoral cartilage, Tibia, Medial Tibial cartilage, and Lateral Tibial cartilage (coded 1 to 5).

iii. Estimation of cartilage thickness based on post-processed segmentation masks.

iv. Binary classification of OA severity into Healthy and OA categories.

v. Use of Random Forest and XGBoost machine learning classifiers.

vi. Evaluation metrics restricted to Dice coefficient for segmentation and accuracy for classification.

1.7 Significance of the Research

This research embraces a considerable implication within the field of medical image analysis and assessing osteoarthritis mainly based on the integration of deep-learning-based segmentation with machine learning classification.

While most of the existing studies treat cartilage segmentation and OA severity classification as a separate task but the present work combines the precise cartilage morphometry with predictive modelling to classify OA severity.

This integrated approach directly addresses a core research gap by linking anatomical structures with downstream clinical prediction in same pipeline.

Secondly, a major contribution of this study is the commitment to the automation and reproducibility by utilizing open-source robust framework specifically 3D U-Net architectures which implemented within the MONAI platform.

This research is ensuring transparent, reproducible experimentation which is consistent with modern standards in medial AI research. This research approach aligns with the broader movement healthcare by democratizing deep learning by making advance tools which can accessible, verifiable and adaptable.

Lastly, the study advances the literature on OA imaging by synthesizing prior work from literature along with foundational and contemporary works on cartilage segmentation, limitation of MRIs and morphometric biomarkers.

Collectively, these contributions place the research study as expressive progression in the automated assessment of OA and as a valuable reference to the future studies on medical image computations.

1.8 Research Contribution

This study proposes an integrated framework that combines knee MRI–based segmentation, cartilage thickness quantification, and osteoarthritis (OA) classification within a unified pipeline.

A 3D U-Net model was implemented for accurate segmentation of femur, tibia, and corresponding cartilage regions. Cartilage thickness was extracted as a quantitative biomarker and integrated with clinical features to improve OA classification performance.

The effectiveness of this approach was validated using machine learning models such as Random Forest, XGBoost, and SVM, demonstrating enhanced predictive capability and interpretability.

1.9 Organization of Thesis

Chapter 1

The first chapter provides an overview of the background, problem statement, research questions and objectives, purpose and scope, significance of the study, achieved objectives, and thesis organization.

Chapter 2

The second chapter, *Literature Review*, presents a synthesis of previous research on knee segmentation, cartilage morphometry, and classification of OA.

It critically examines both deep learning and traditional approaches, highlighting the research gaps addressed in this study.

Chapter 3

Third chapter *Methodology*, discusses the steps undertaken during experimentation, including dataset preprocessing, model specification and architecture, thick-

ness computation, and machine learning classification for OA prediction.

Chapter 4

Chapter 4, *Results*, presents the outcomes of experimentation and implementation, including experimental setup, segmentation performance, thickness analysis, and OA classification results.

Chapter 5

Chapter 5, *Conclusion*, provides the conclusion of the study, highlighting key findings and limitations.

Chapter 2

Literature Review

2.1 Introduction

Knee Osteoarthritis (KOA) is a chronic and degenerative joint disease that severely impacts the quality of life by reducing mobility, causing persistent pain, and even tually leading to joint stiffness and disability. As the global aging population increases, KOA has emerged as one of the most prevalent forms of arthritis, par ticularly among individuals over the age of 50.

Early diagnosis and continuous monitoring are crucial in managing its progression and in implementing timely therapeutic interventions Deep learning has gained widespread adopt ion across medical imaging due to its ability to learn complex spatial patterns directly from high-dimensional data.

In musculoskeletal imaging, deep learning particularly convolutional neural networks (CNNs) has demonstrated strong performance in tasks such as knee cartilage and bone segmentation, where traditional algorithms struggle with low-contrast boundaries and anatomical variability.

This chapter reviews the evolution of knee MRI analysis methods, including classical machine-learning approaches, modern deep-learning architectures, and hybrid pipelines that integrate imaging biomarkers with predictive models.

Understanding the strengths and limitations of these approaches provides essential context for selecting and designing an appropriate framework for automated cartilage segmentation, morphometric analysis, and osteoarthritis classification in this study.

2.2 Background

Osteoarthritis (OA) is a chronic degenerative joint disease characterized by progressive deterioration of articular cartilage, remodelling of subchondral bone, synovial inflammation, and structural changes within the joint [1].

The knee is the most frequently affected joint due to its complex biomechanical role in weight bearing, load distribution, and movement, rendering it highly susceptible to degenerative processes.

The degradation of hyaline cartilage is often gradual and initially microscopic, making early detection challenging.

Traditional radiographic approaches such as the Kellgren-Lawrence (KL) grading system are limited in their ability to visualize early biochemical and morphological cartilage changes because they detect only late manifestations like joint-space narrowing and osteophytes [13].

Magnetic resonance imaging (MRI) has become the preferred modality for investigating OA.

Due to its high soft-tissue contrast and ability to capture three-dimensional structure with excellent spatial resolution [14].

MRI enables direct assessment of cartilage morphology, bone marrow lesions (BMLs), meniscal pathology, and synovial tissue changes.

Quantitative MRI biomarkers, such as cartilage thickness and volume, serve as essential indicators for monitoring progression in both clinical and research settings [4].

However, accurate extraction of these biomarkers relies on reliable segmentation, especially because cartilage is thin, curved, and displays low contrast at its boundaries.

Deep learning, particularly convolutional neural networks (CNNs), has emerged as a transformative solution for automating anatomical segmentation tasks in MRI analysis [15].

CNN-based methods outperform classical image-processing approaches in accuracy and reproducibility.

Yet segmentation of knee cartilage remains challenging due to variations in MRI acquisition protocols, intensity inhomogeneities, and the inherently thin geometry of cartilage tissues [9].

These factors contribute to segmentation errors that can propagate into downstream tasks such as surface modeling and thickness estimation [11]. This theoretical foundation underscores the necessity of developing robust segmentation models capable of handling heterogeneous datasets and capturing fine anatomical detail.

2.3 Literature Review

MRI-derived biomarkers play a fundamental role in detecting early degenerative changes in OA. Among the most clinically relevant morphological biomarkers are cartilage thickness, cartilage volume, subchondral bone shape, meniscal extrusion, and the presence of BMLs [4].

Quantitative measurement of cartilage geometry is particularly important because cartilage thinning is one of the earliest detectable signs of OA progression.

Accurate segmentation is essential for deriving these biomarkers. Even small segmentation errors can lead to significant biases in thickness measurement due to the narrow width of cartilage, especially in tibial regions [8].

Errors in boundary delineation propagate directly into surface-normal based thickness estimation and volumetric calculations.

Norman et al. (2018) demonstrated that segmentation inaccuracies could substantially affect relaxometry and morphometry analyses, impacting clinical decision-making.

Additionally, structural pathologies frequently coexist in OA, complicating biomarker extraction. For example, osteophytes, synovial inflammation, and meniscal damage often distort cartilage shape [16]

Folkesson et al. (2007) and Moon et al., (2025). observed that partial volume effects caused by overlapping tissue boundaries lead to ambiguous signal intensities, making segmentation more difficult.

MRI sequences, such as T1 and T2 mapping, also introduce variability based on contrast mechanisms, acquisition parameters, and scanner models.

These complexities emphasize the importance of robust segmentation frameworks capable of handling diverse imaging conditions.[3][17].

2.4 Evolution of Cartilage Segmentation Methods

2.4.1 Classical Image Processing Approaches

Early cartilage segmentation techniques relied heavily on classical methods such as thresholding, region growing, edge detection, and active contour models.

While computationally efficient, these methods perform poorly in low-contrast environments characteristic of knee MRI [3].

Atlas-based segmentation introduced anatomical priors but suffered from poor generalizability across diverse populations [10].

Statistical shape models (SSMs) improved upon atlas methods by incorporating population-level variation; however, manual initialization and sensitivity to intensity variations limited their scalability [18].

Based on the pixels of nearby neighbors, we can estimate a region's or area's boundary as the difference between two regions. The segmentation and initializa tion processes begin with a small number of seed points.

The process then divides the regions according to the pixel intensity. It only takes one seed point and the homogeneous properties of nearby pixels to complete the process. In an effort to improve MRI-based diagnosis of meniscal injuries and osteoarthritis.

Mallikarju naswamy et al. explores knee joint menisci segmentation using the Seeded Region Growing (SRG) algorithm, aiming to enhance MRI-based diagnosis of osteoarthri tis and meniscal injuries.

The method initiates segmentation from manually se lected seed points, ensuring precise extraction of meniscal structures.

The study also integrates 3D visualization, allowing for better structural assessment, and quantifies meniscal dimensions, comparing results with manual expert evaluations and commercial software like MIMICS.

Validating findings through radiologists and existing tools reinforces the algorithm's reliability for clinical applications.

Despite its accuracy, the method requires manual seed selection, introducing vari ability [19].

Meanwhile, O ztuk and Albayrak propose a better voxel- classification-based region- growing method to segment cartilage of human knee using high field MRI scanners by incorporating vicinity-correlated subsampling to obtain a better qual ity and a cost-efficient approach. They effectively divide femural, tibial, patellar cartilages and this method provides high Dice Similarity Coefficients 82.672.6that vcinity-correlated sparse subsampling has the most effective segmentation. Al

though the method enhances the computation mode, it still uses voxel classification method, thereby lacking generalizability over varying MRI scans .

Anshad et al. (2019) introduce Chondroblasts-to mosaic Segmentation This is the slice of chondroblasts- toma By use of modified region growing algorithm meant to cut out from medical pictures. They propose an improvement to the conven tional region grow- ing technique through the use of an adaptive thresholding and spatial constraints that would greatly increase the accuracy of tumor delineation, minimize manual correction, as well as observer to observer variability.

The algorithm performance is proved by using such measures as Dice coefficients, Jaccard distance, spatial overlaps, and volume measurement errors which is much better than other segmentation methods.

The changes guarantee no-table clinical advantages, despite being based on con ventional segmentation techniques, since they allow tumor segmentation that is more precise and reliable during diagnosis and treatment planning [20].

Similarly, Norouzi et al. (2014) In a similar manner, a comprehensive review of edge detection and region- growing methods provided by Norouzi et al. lays out the main issues existing in the tools, which include a lack of robustness when they encounter noisy or low-contrast imaging [21].

Meanwhile, Williams et al. (2010) applied SSMs to assess knee cartilage morphology in MRI, enabling precise re gional analysis across diverse patient cohorts [25].

Despite their contributions, conventional methods are inherently labor-intensive and struggle with scalability when applied to large datasets.

Additionally, these methods often require manual adjustments, which not only increase user variability but also limit the reliability and efficiency of the results.

The edge detector Flow Laplacian of Gaussian (FLoG) takes edge information and transforms it into an area-based representation, accomplished through a flow- fill operation, making post processing less complex.

The algorithm to grow an onion also folds in estimates of bone regions and gives robust segmentation in dif- ficult imaging environments such as soft tissue adhesion and partial volume effects. Their method is more effective and relevant in comparison to standard methods, with regard to the analysis of the clinical knee [22].

In the meantime, Norouzi et al offer two-iterative step segmentation (two-iterative segmentation) in extracting knees bone on MRI images using combined Otsu thresholding and a Support Vector Machines (SVMs) to produce better accuracy. The starting point is an Otsu thresholding to recognize bone candidate areas, then the space of venue and shape features are refined by SVM. Ad-justment, segmentation is done iteratively, and Canny edge filtering, and morphological operations improve pre- cision.

The overall accuracy is measured by the Dice Similarity Index, which the approach receives at the level of 0.96. Although functional, it is conventional, in the sense that it operates with the frameworks of segmentation, with intensity as a parameter of substantial remarks, but not deep learning automation.

The study also provides a convenient source of comparison between the thresholding based approaches and CNN-driven models in knee osteoarthritis [23].

2.5 Graph Based Segmentation Method

The graph-cut method has gained significant interest in image segmentation due to its ability to incorporate both regional and boundary information, enhancing accuracy and robustness in delineating image structures [24].

Graph-based segmentation methods encompass a set of algorithms that model pixels or voxels in an image sequence, along with their interconnections, as a weighted undirected graph. Given a graph $G = (V, E)$, each pixel is represented as a node $v \in V$, while the relationship between neighboring nodes is expressed as an edge $e \in E \subseteq V \times V$ with assigned weights influencing segmentation accuracy.

Following the normalized cut approach, graph segmentation has gained increasing attention. In this context, the separation of a graph is referred to as a "cut."

A binary graph segmentation divides the graph into two subsets, G_m and G_n, such that $G_m \cup G_n = V$ and $G_m \cap G_n = \emptyset$, reducing the degree of difference between the two regions.

The difference is quantified as the total weight of the removed edges, expressed as:

$$\text{Cut}(G_m, G_n) = \sum_{u \in G_m} \sum_{v \in G_n} w(u, v) \tag{2.1}$$

Here, G_m and G_n represent the vertices of two separate subgraphs, and the sum of edge weights constitutes the cut.

Minimizing the cut differentiates the subgraphs, but achieving an optimal partitioning is challenging.

A possible solution is to apply optimization techniques to minimize the cut in equation 2.1, as discussed in the comprehensive review of graph-

based segmentation methods [25].

Wu and Leahy (1993) propose a graph-theoretic clustering approach for image segmentation, modeling pixels as nodes connected by weighted edges.

Their method minimizes the cut cost using normalized cuts and minimum spanning trees (MSTs) to optimize segmentation boundaries.

Leveraging spectral graph theory, they refine segmentation efficiency, demonstrating superior boundary preservation and object recognition.

Despite high computational complexity, their approach remains influential in graph-based learning and medical imaging applications.

Ababneh and Gurcan (2010) present an efficient graph-cut segmentation method for knee bone osteoarthritis medical images, aiming to enhance accuracy and computational efficiency in medical image analysis. Their approach models pixels as

nodes in a weighted graph, applying max-flow/min-cut optimization to achieve precise segmentation of knee bone structures.

The method integrates both regional and boundary cues, ensuring robust delineation while supporting automatic and interactive segmentation, allowing user input for refinement in complex cases.

Tested on knee osteoarthritis datasets, the algorithm demonstrates high accuracy, superior edge preservation, and reduced computational complexity compared to conventional methods. Despite requiring parameter tuning for optimal performance, the approach significantly improves segmentation reliability, making it a valuable tool for automated medical imaging applications [26].

Yin et al. (2010) introduce LOGISMOS, a layered optimal graph-based segmentation framework designed for multi-object and multi-surface extraction in medical imaging. The method utilizes a weighted directed graph, where each column represents potential surface locations, optimizing spatial relationships through max-flow graph algorithms.

Applied to knee cartilage segmentation using MRI datasets from the Osteoarthritis Initiative, LOGISMOS achieves high accuracy, with Dice Similarity Coefficients of 0.84, 0.80, and 0.80 for femoral, tibial, and patellar cartilage, respectively, and mean cartilage thickness errors ranging from -0.11 mm to -0.03 mm.

The framework incorporates shape-based, edge-based, and region-based cost functions, ensuring precise segmentation even in complex anatomical structures [27].

Kashyap et al. (2018) discuss the Just-Enough Interaction (JEI) approach as a post-processing step integrated into LOGISMOS, enhancing segmentation accuracy. They employ a hierarchical random forest (RF) classifier.

where the second RF model generates learning-based cost functions that refine segmentation boundaries.

The study reports a significant reduction in signed error for femur bone surface positioning (0.03 mm) compared to the tibia [28].

2.5.1 Atlas-based Segmentation Methods

Atlas-based segmentation methods utilize anatomical priors by aligning patient-specific imaging data with pre-constructed atlases to enhance segmentation accuracy and ensure anatomical consistency. These methods have proven especially useful in standardizing

analyses across diverse datasets.

Shan et al. (2014) introduced an atlas-based technique for cartilage segmentation, achieving high anatomical consistency across knee MRIs by leveraging multi-label segmentation frameworks [29].

Zarychta (2022) also presents an atlas-based segmentation approach for extracting femur, tibia, and patella from CT and MRI scans using automated image matching, normalization, and fuzzy segmentation techniques.

The method integrates fuzzy c-means and fuzzy connectedness for refinement, achieving Dice coefficients of 85.52 to 89.48. The automated workflow minimizes manual intervention, ensuring efficient segmentation [30].

Brehler et al. (2019) propose an atlas-based algorithm for automatic anatomical measurements in knee tomographic datasets.

The method utilizes a set of atlases containing knee images, segmented bone structures, and anatomical landmarks.

A multistage registration process transfers these landmarks to the target volume, eliminating the need for manual segmentation.

The algorithm was validated on 24 cone-beam CT scans, achieving intraclass correlation coefficients (ICC) above 0.95, demonstrating high reliability.

Compared to inter-reader variability (ICC ranging from 0.19 to 0.95), atlas-based segmentation has been widely explored for anatomical structure extraction, particularly in brain and knee imaging.

The study highlights its potential to enhance workflow efficiency and reduce operator dependency in anatomical assessments [31].

Similarly, Pedoia et al. (2016) conducted a comprehensive review of atlas-guided approaches in musculoskeletal imaging, highlighting their significant utility in arthritis research, where precise delineation of cartilage and other tissues is critical [32].

Shan et al. (2014) propose an automatic atlas-based segmentation method for femoral and tibial cartilage extraction from MR knee images. The approach integrates multi-atlas segmentation with non-local patch-based label fusion, ensuring robust cartilage identification. A key innovation is the three-label segmentation framework, which spatially separates femoral and tibial cartilage while preserving anatomical regularity through anisotropic spatial regularization. The method was validated on 706 images from the Pfizer Longitudinal Study, demonstrating reliable segmentation performance. Comparisons with other atlas-based strategies and local classifiers highlight its effectiveness in handling thin cartilage structures [33].

Heckemann et al. (2006) introduce a method that combines label propagation and decision fusion to enhance brain MRI segmentation accuracy, achieving a similarity index of 0.836 [34]. Rohlfing et al. (2004) evaluate various atlas selection strategies for image segmentation, finding that multi-classifier decision fusion yields the highest segmentation accuracy with a mean similarity index of 0.86 [35].

While multi-atlas segmentation has proven effective for brain image segmentation, it has rarely been utilized to represent the knee complex.

Nikolopoulos et al. (2020) present a multi-atlas segmentation and mesh refinement approach for personalized knee geometry modeling.

The study focuses on generating subject-specific finite element models to simulate knee joint mechanics, optimize treatments, and prevent injuries. The method integrates deformable registration and joint label fusion to segment knee structures, followed by anatomically-adapted mesh refinement to enhance model accuracy.

The segmentation performance was validated on 7 Open Knee project participants and 78 Osteoarthritis Initiative cases, demonstrating competitive accuracy. The study emphasizes the importance of structured meshes with well-shaped hexahedra for knee cartilages and menisci, ensuring reliable finite element simulations. The developed tools are available for public use, facilitating advancements in patient-specific diagnostics and interventions [36].

Despite their strengths, atlas-based methods face challenges such as sensitivity to registration errors, which can compromise accuracy, and high computational demands.

Validation on larger datasets and adaptability to varied imaging modalities remain challenges that limit their efficiency in large-scale applications.

Addressing these limitations requires the development of more robust and efficient workflows to optimize their clinical viability.

2.5.2 Manual Annotation

Manual annotation enables the availability of ground truth data necessary to use supervised learning models and can be used to ensure that the given outputs of segmentation comply with clinical norms.

Mukherjee et al. (2023) suggest a strategy that lends itself well to manual tracking of patellar osteophytes, as the authors observe morphological transformations that herald the onset of early osteoarthritis. Their approach enables accurate patella segmentation based on clinical images, allowing intricate investigation of structural changes over time [37].

Moreover, manual annotation is prone to inter-observer variability, which may cause discrepancies in segmentation accuracy across different datasets. These challenges necessitate the development of adequate automation methods and hybrid workflows that combine expert knowledge with segmentation powered by Artificial Intelligence.

2.6 Supervised and Unsupervised Methods

Where Supervised models rely on annotated data that delivered overall high accuracy in the anatomical structure identification, unsupervised approaches dive into the data patterns without prelabelled data.

Sikkandar et al. (2022) exemplified this strategy to train a supervised CNN to predict knee osteoarthritis severity with an impressive 93.2 Percent on the availability of giant labeled datasets, which can be a limitation in medical imaging [29].

Alternatively, Aldieri et al. (2024) proposed an unsupervised femur segmentation model that uses the anatom- ical priors to reduce the variance in the segmentation of a femur without the need of labeled data.[38] Ambellan et al. (2018) have discovered a conjunctive approach constituting statistical shape models (SSMs) and CNNs, pairing the old techniques of anatomical modelling and AI-powered learning [6].

Although supervised models may be very accurate because of trained data that is well structured, unsupervised models still suffer in precision when it comes to critical diagnostic tasks. The combined strategies should be used to deal with such issues by striking a compromise between automation, dataset availability, and anatomical consistency whih will enhance

segmentation results on medical images.

Tamez-Peña et al. (2012) introduced a fully automated, unsupervised segmentation method for analyzing anatomical knee features using 3D dual echo steady state (DESS) MRI data from the Osteoarthritis Initiative (OAI). Their approach leveraged a multi-atlas segmentation framework, where manually segmented MRI series served as reference atlases to guide the automated delineation of femoral and tibial cartilage.

The method demonstrated high accuracy, with Dice similarity coefficients of 0.88 and 0.84 for femoral and tibial cartilage respectively, when compared to manual segmentation.

Quantitative metrics such as cartilage volume, surface area, thickness, and subchondral bone plate curvature were extracted with strong test-retest precision thickness measurements showed variability as low as 0.014 mm (0.6%) at the femur.

Longitudinal analysis revealed a Pearson correlation coefficient of 0.94 for femoral cartilage thickness changes, underscoring the method's reproducibility and sensitivity to disease progression.

This study provides a robust foundation for noninvasive, quantitative assessment of osteoarthritis-related morphological changes, offering valuable insights for longitudinal monitoring and clinical trials [39].

Hwang et al. (2025) explored the potential of machine learning techniques for early detection and classification of knee osteoarthritis (OA) using single-leg standing (SLS) kinematics.

The study involved 43 manufacturing workers aged 40–70, whose frontal plane movements were captured via 2D video analysis during SLS tasks.

Horizontal displacements of six anatomical landmarks—including the pelvis and ankle—were analyzed to identify biomechanical deviations associated with early OA (EOA).

Using unsupervised K-means clustering, the researchers identified three distinct movement profiles, while supervised models such as Random Forest achieved perfect classification accuracy (AUC = 1.000) in distinguishing EOA from non-EOA cases.

Pelvic and ankle displacements emerged as the most influential predictors. This approach offers a cost-effective and accessible alternative to traditional imaging, highlighting the value of dynamic movement analysis in identifying subtle kinematic changes that precede structural joint damage.

The findings support the integration of machine learning with functional movement assessments for early OA screening and personalized rehabilitation strategies [34].

Seo et al. (2020) provided a comprehensive overview of machine learning techniques applied to biomedical image segmentation, emphasizing both classical and deep learning approaches. The review highlighted traditional methods such as Markov Random Fields, k-means clustering, and Random Forests, noting their sample efficiency and interpretability despite limitations in segmentation accuracy.

In contrast, deep learning architectures—including artificial neural networks (ANNs) , convolutional neural networks (CNNs), and recurrent neural networks (RNNs) were shown to outperform classical models by enabling automatic feature extraction and robust semantic segmentation.

The authors discussed the evolution of supervised, semi-supervised, and unsupervised learning paradigms, along with challenges like data scarcity, model generalization, and computational complexity. By presenting segmentation results from recent studies and outlining heuristics to improve training efficiency, the paper serves as a valuable resource for understanding the technical landscape and practical applications of machine learning in medical imaging [40].

Zhong et al. (2023) presented an unsupervised domain adaptation (UDA) framework to enhance automated classification of knee osteoarthritis (OA) phenotypes using MRI data. The study addressed the challenge of limited labeled data in clinical settings by leveraging a large, high-quality source dataset from the Osteoarthritis Initiative (OAI) and adapting it to a smaller, locally collected target dataset.

Their four-step pipeline included automatic segmentation, source classifier training, encoder adaptation, and performance validation.

The UDA model significantly outperformed both the source-only and target-only classifiers, achieving an AUROC of 0.90 for cartilage/meniscus phenotypes and 0.75 for subchondral bone phenotypes.

These results underscore the effectiveness of domain adaptation in improving classification accuracy and generalizability across heterogeneous datasets.

The approach offers a scalable solution for phenotype-based OA analysis, facilitating more precise diagnostics and personalized treatment planning [30].

Felfeliyan et al. (2021) developed a domain adaptation framework using CycleGAN to enable unsupervised segmentation of knee MRI data, addressing the challenge of domain shift across different imaging protocols.

Their method translated TSE Fat Suppressed MRI sequences into pseudo-DESS images, which were then segmented using an improved Mask R-CNN (I-MaskRCNN) trained on DESS data.

This pipeline achieved a Dice coefficient of 0.76 for cartilage and femoral head segmentation, closely matching manual annotations without retraining.

The CycleGAN model preserved anatomical consistency through cycle and identity losses, while the I-MaskRCNN enhanced boundary detection via skip connections and encoder layers.

Validated on Osteoarthritis Initiative (OAI) data, the approach demonstrated rapid processing times (14–16 ms for standard images), making it suitable for large-scale clinical deployment.

This technique offers a scalable solution for hardware-independent, automated analysis of knee MRI, facilitating efficient OA assessment across diverse imaging environments [41].

Yu et al. (2024) introduced a fully automated, unsupervised segmentation framework for detecting bone marrow edema-like lesions (BMELs) in knee MRI, a key biomarker linked to osteoarthritis progression and pain.

Their method leveraged conditional generative models, including diffusion-based anomaly detection, to identify BMELs without relying on manual annotations—addressing challenges like poor inter-rater reliability and annotation bias.

By training on healthy knee MRIs and detecting deviations, the model achieved robust segmentation across diverse lesion shapes and intensities.

The study also analyzed expert annotations to benchmark intra- and inter-rater variability, highlighting the limitations of traditional manual grading systems such as WORMS and MOAKS.

This approach offers a scalable, unbiased alternative for BMEL quantification, enhancing early diagnosis and longitudinal monitoring of knee osteoarthritis [42].

Kamper, Livescu, and Goldwater (2017) introduced the Embedded Segmental K-means (ES-KMeans) model, a novel approach for unsupervised segmentation and clustering of speech in zero-resource settings.

Unlike traditional Bayesian models, ES-KMeans employs hard clustering and segmentation, offering a clear objective function while significantly improving computational efficiency.

The model represents variable-length speech segments as fixed-dimensional acoustic word embeddings, enabling direct comparison in embedding space without alignment.

Evaluated on English and Xitsonga datasets, ES-KMeans outperformed heuristic methods in word segmentation and achieved comparable accuracy to Bayesian models.

while being five times faster and requiring fewer hyperparameters. However, the trade-off was lower cluster purity.

The model also scaled effectively to larger corpora, performing competitively in the Zero Resource Speech Challenge 2017 across five languages.

This work bridges the gap between probabilistic rigor and practical scalability, advancing unsupervised speech processing for low-resource languages [43].

Li et al. (2024) proposed a source-free unsupervised domain adaptation (UDA) framework for multi-tissue segmentation of knee joint MRI, addressing the challenge of inaccessible labeled source data due to privacy constraints. Their two-stage approach begins with matching batch normalization statistics and entropy

minimization to generate pseudo segmentation labels, which are then refined using a voting strategy.

In the second stage, an uncertainty-aware cross pseudo supervision mechanism further enhances segmentation accuracy.

The model demonstrated superior performance compared to existing source-free UDA methods, with an average Dice coefficient improvement of 11.8%.

Notably, the framework proved effective in both forward and reverse adaptation across different MRI pulse sequences and systems.

This study marks a significant advancement in scalable, privacy-preserving segmentation techniques for knee osteoarthritis assessment [44].

2.6.1 Machine Learning Approaches

Prior to deep learning, classical machine learning methods such as random forests, support vector machines (SVMs), and structured classification models improved segmentation accuracy through voxel-based classification [45].

These models utilized hand-crafted features designed to capture texture, gradients, or contextual cues.

Although they achieved modest improvements, manually engineered features could not represent complex cartilage morphology sufficiently, limiting performance.

2.7 Rise of Deep Learning Approaches

Deep learning revolutionized medical image segmentation with the introduction of fully convolutional neural networks (FCNs).

U-Net (Ronneberger et al., 2015) and V-Net (Milletari et al., 2016) architectures provided end-to-end learning of semantic segmentation.

Initially, 2D CNNs achieved strong performance on slice-based MRI but were limited by the lack of volumetric context [5][46][7].

As a result, 3D U-Nets and hybrid 2D-3D systems were developed to capture spatial dependence across slices [6].

These models delivered state-of-the-art results on public datasets such as OAI and ZIB. Rodríguez-Vila et al. (2022) expanded accessibility by democratizing deep learning pipelines, providing fully transparent segmentation workflows.

Their work underscores that reproducibility and open access are essential for accelerating progress.

TABLE 2.1: Summaries of Literature Review on Knee Segmentation Methods

Study	Dataset	Segmentation Method	Dice Score
Ambellan et al. (2019) [6]	OAI ZIB	3D U-Net + Shape Models	Femoral: 89.2 ± 2.4; Tibial: 88.1 ± 3.8
Norman et al. (2018) [16]	UCSF cohort	2D U-Net CNN	Femoral: 86.7 ± 3.2; Tibial: 78.8 ± 3.3
Rodríguez-Vila et al. (2022) [8]	OAI + ZIB	CNN (open-source pipeline)	Femoral: 98.6 ± 0.3; Tibial: 98.6 ± 0.4
Gan et al. (2021) [7]	OAI	3D CNN	Femoral: 0.84–0.90
Tiulpin et al. (2018) [47]	OAI	CNN (OA classification)	N/A (classification)
Desai et al. (2016) [48]	OAI challenge	Mixed CNNs (benchmark)	Femoral: 87 ± 3; Tibial: 85 ± 4
Fripp et al. (2010) [49]	OAI	Atlas-based Segmentation	Femoral: 0.75–0.82
Kessler et al. (2020) [50]	OAI	Conditional GAN (cGAN)	Femoral: 98.5 ± 0.2; Tibial: 98.5 ± 0.3
Prasoon et al. (2013) [51]	In-house	Three 2D-CNNs (triplanar)	Medial Tibial Cartilage (cMT): 82.49%
Yao et al. (2024) [52]	SKI10 (train 60, test 40)	2D SegNet + deformable model	VOE only (no Dice reported)
Panfilov et al. (2024) [53]	OAI	CNN	Patellar: 80%; Femoral: 84%; Tibial: 80%

2.8 Performance Challenges in Knee Cartilage Segmentation

Despite methodological advances, several persistent challenges continue to limit segmentation performance.

Variability in MRI image quality, differences in acquisition protocols, and the presence of noise and artifacts can significantly affect the accuracy of automated segmentation methods.

These challenges collectively impact segmentation reliability and highlight the need for more robust, generalizable, and data-efficient approaches.

2.8.1 Challenges in Knee Cartilage Segmentation

2.8.1.1 Thin Geometry and Poor Contrast

Cartilage, especially medial tibial cartilage, is extremely thin and exhibits poor contrast relative to subchondral bone. These anatomical properties increase susceptibility to noise, cause blurred boundaries, and result in systematic under-segmentation [9]. Norman et al. [16] show that segmentation inaccuracies directly influence downstream relaxometry calculations.

2.8.1.2 Anatomical Variability

Osteoarthritis introduces structural irregularities such as osteophytes, cartilage delamination, meniscal extrusion, and bone marrow lesions.

These abnormalities disrupt morphological patterns, making segmentation significantly more difficult[4].

2.8.1.3 Domain Shifts Across MRI Protocols

MRI acquisitions vary widely by scanner vendor, magnetic field strength, imaging sequence, and coil configuration. Models trained on a single dataset often fail to generalize. Rodríguez-Vila et al. emphasize that networks trained only on OAI data may underperform on other hospital datasets [8].

2.8.1.4 Compartment Specific Difficulties

Tibial cartilage consistently yields the lowest Dice scores due to curved geometry, low thickness, and close proximity to menisci [6].

Thus, tibial segmentation remains a bottleneck in automated workflows.

2.9 Reproducibility Transparency and Dataset Bias in Deep Learning

Reproducibility is a critical concern in deep learning for medical imaging. Many studies lack sufficient detail regarding hyperparameters, preprocessing routines, training schedules, or evaluation criteria, making independent replication challenging. (Pineau et al., 2021) and Renard et al. (2020) further emphasize that deep learning introduces additional sources of variability, including stochastic initialization, optimization dynamics, dataset composition, and cross-validation splits. Dataset bias also limits generalizability.

Public datasets such as OAI, while valuable, reflect specific imaging protocols and demographic distributions. Models trained solely on these datasets tend to overfit them [54][55].Norman et al. (2018) and Desai et al. (2016) demonstrate that segmentation performance varies widely when models are evaluated on unseen clinical data. Standardization of evaluation metrics remains another challenge.Dice similarity coefficient is most commonly used, but studies also report surface distance, volumetric error, and thickness error, making cross-study comparisons difficult [16][48]. (Litjens et al., 2017) and Rodríguez-Vila et al. (2022) advocate for open-source solutions to address reproducibility concerns and promote transparent benchmarking [15][48].

2.10 Advances in Cartilage Thickness Estimation

Cartilage thickness is one of the most clinically meaningful biomarkers in the assessment of OA. Early methods relied on Euclidean distance transforms, Laplacian

modeling, and geometric surface mapping [56].

These methods were highly sensitive to segmentation errors, especially near boundaries.

Modern thickness estimation approaches frequently utilize surface-normal projections between opposing cartilage surfaces, achieving improved anatomical realism [12]. However, since thickness is computed directly from segmentation masks, even minor boundary errors can yield substantial inaccuracies in thickness estimation [8].Deep learning has also been applied to predict thickness maps directly from MRI, bypassing explicit segmentation [57]. Still, most analytical pipelines continue to rely on segmentation-based measurements because of their inherent interpretability and reliability.

Segmentation provides a clear and direct representation of anatomical structures, allowing researchers and clinicians to visually verify and validate the accuracy of the extracted features.

Unlike purely voxel-based or black-box methods, segmentation-derived metrics offer transparent insights into regional cartilage morphology, thickness, and integrity, which are crucial for understanding disease progression.

Furthermore, these measurements are reproducible across different datasets and imaging modalities, making them a dependable foundation for downstream analyses such as osteoarthritis classification, biomarker development, and longitudinal studies of cartilage degeneration.

2.11 Machine Learning and Deep Learning for OA Classification

Traditional OA severity classification is based on the KL radiographic scale, but this method is subjective and insensitive to early structural changes [13]. Automated classification of knee osteoarthritis (OA) has been widely investigated

to address the subjectivity and limited sensitivity of conventional clinical grading systems such as the Kellgren–Lawrence (KL) scale. Radiographic assessment primarily relies on visual evaluation of joint space narrowing and osteophyte formation, which often fails to detect early OA characterized by subtle cartilage degeneration.

Consequently, data-driven classification methods based on imaging biomarkers have gained increasing attention for improving diagnostic accuracy and reproducibility [58][59].

Zhang et al. (2023) extracted morphometric features, including cartilage thickness and volume, along with quantitative susceptibility mapping (QSM) values, which provide information on cartilage tissue composition.

Statistical analysis indicated that these features varied significantly with osteoarthritis severity ($p < 0.05$),highlighting their potential as reliable biomarkers for early OA detection.

When employed in a support vector machine (SVM) classifier, these automatically quantified cartilage metrics enabled effective differentiation between OA and normal knees, achieving a high AUC of 0.94 [60].

Early computational approaches for OA classification predominantly utilized 2D convolutional neural networks (CNNs) applied to X-ray images. Antony et al. (2016) demonstrated that deep CNNs could be used to quantify OA severity from knee radiographs, achieving moderate accuracy in five-grade KL classification.

In subsequent work, improved performance by incorporating automated knee joint localization using fully convolutional networks [61].

Guida et al. (2021) developed a 3D convolutional neural network (3D-CNN) for knee osteoarthritis classification using MRI volumes.

Unlike traditional 2D approaches, their method leverages the full 3D structure of the knee to capture spatial and contextual information of cartilage and bone tissues.The model was trained on MRI data and evaluated for both binary OA/non

OA classification and multi-class Kellgren–Lawrence grading.[62] Pedoia et al. (2019) demonstrated that segmenting cartilage and meniscus structures prior to classification improves localization of disease-relevant regions and enhances OA severity staging.

Similarly, Wang et al. (2019) showed that features automatically extracted from segmented MRI volumes using 3D CNNs outperform models based solely on demographic and clinical variables when predicting long-term outcomes such as total knee replacement risk [63][32] Eckstein et al. (2014) highlighted the critical role of quantitative imaging biomarkers in osteoarthritis (OA) research, emphasizing that measures such as cartilage thickness and volume are strongly associated with OA progression.

The authors reviewed advances and limitations in imaging cartilage and bone for clinical trials, noting that morphometric biomarkers provide objective and reproducible metrics that can improve the assessment of disease severity and progression.[64].Tiulpin et al. (2018) proposed a deep learning–based framework for automatic knee osteoarthritis diagnosis from plain radiographs. The study employed convolutional neural networks (CNNs) to classify OA severity according to the Kellgren–Lawrence grading system. [47].

Rani et al. (2024) developed a custom 12-layer CNN for automated knee osteoarthritis detection and KL severity grading from OAI X-ray images, achieving 92.3% binary accuracy and 78.4% multiclass accuracy, outperforming several prior models and demonstrating the effectiveness of deep learning for objective KOA assessment[65] Abdelbasset Brahim et al. (2019) proposed a machine learning–based computer-aided diagnosis (CAD) system for early knee osteoarthritis detection using X-ray images from the Osteoarthritis Initiative (OAI) dataset.The method includes Fourier-domain preprocessing, predictive normalization using multivariate linear regression, and feature extraction via independent component analysis (ICA). Naïve Bayes and Random Forest classifiers were employed for binary OA/non-OA classification, achieving an accuracy of 82.98 %, sensitivity of 87.15 %, and statistically significant results ($p < 0.05$) [66]. Chen et

al. (2025) investigated the use of machine learning for knee osteoarthritis (KOA) diagnosis using clinical and demographic data.

The study evaluated multiple supervised learning algorithms, including Random Forest, XGBoost, Logistic Regression, Support Vector Machine, Decision Tree, and Naïve Bayes, to classify individuals as OA or non-OA.

Ensemble methods, particularly Random Forest and XGBoost,achieved the best performance,as measured by area under the ROC curve (AUC) and accuracy [67].

Shamir et al. (2009) demonstrated that early detection of radiographic knee osteoarthritis can be improved using computer-aided analysis combined with machine learning. The study employed classifiers on hand-crafted features, including cartilage thickness, shape descriptors, and clinical variables such as age, BMI, and gender, to predict OA severity[4].

TABLE 2.2: Comparison of OA Classification Methods

Author	**Dataset**	**Method**	**Classification**	**Accuracy**
Antony (2017)[61]	OAI/MOST (X-ray)	CNN+FCN	Multi-class	60.3%
Guida (2021)[62]	OAI (MRI)	3D CNN	Binary; Multi-class	83%; 54%
Tiulpin (2018)[47]	MOST/OAI (X-ray)	Siamese CNN	Multi-class + Binary	66.7%; AUC 0.93
Zhang (2023)[60]	MRI (65)	nnU-Net+SVM	Binary	AUC 0.94
Rani (2024)[65]	OAI (X-ray)	12L-CNN	Binary; Multi-class	92.3%; 78.4%

In summary, Various machine learning and deep learning methods have been applied for knee osteoarthritis (OA) classification across different studies. However, these approaches differ significantly in terms of imaging modality, dataset size, preprocessing strategies, and model design. Some studies have used X-ray images.while others relied on MRI data; similarly, certain methods performed classification directly on whole images, whereas others incorporated prior segmentation and extracted morphometric features before classification.

In addition, the tasks varied between binary OA detection and multi-class Kellgren– Lawr- ence (KL) grading.

Due to these differences in datasets, experimental setups, and evaluation protocols, direct comparison of performance metrics across studies is challenging. These

variations highlight the diversity of methodologies explored in OA classification research

2.12 Research Gap

i. Most studies focus only on performing segmentation of cartilage or bone structures without extending the analysis toward disease classification.

ii. The outputs of segmentation are often not utilized for osteoarthritis (OA) severity prediction.

The lack of integration between segmentation and classification leads to several limitations:

i. Most existing studies perform segmentation as an isolated task without utilizing the segmented outputs for downstream OA classification.

ii. There is a lack of systematic measurement and quantitative evaluation of cartilage thickness derived directly from segmentation results.

iii. The absence of an integrated segmentation–feature extraction–classification pipeline limits both predictive performance and clinical interpretability.

Therefore, there is a need for an integrated pipeline that combines segmentation with feature-based OA classification.

Chapter 3

Proposed Research Methodology

Following the critical analysis of existing research presented in Chapter 2, several important gaps were identified. This chapter outlines the methodological framework developed to address these gaps and provides a detailed description of the proposed approach.

3.1 Proposed Methodology

This study proposes an integrated framework for early osteoarthritis (OA) assessment using knee MRI data. The methodology combines deep learning–based cartilage segmentation, quantitative cartilage thickness estimation, and machine-learning classification.After standardized MRI preprocessing, a 3D U-Net model is used to segment femoral, medial tibial, and lateral tibial cartilage regions.Voxel-wise cartilage thickness is then computed, and region-level statistical features are extracted. These morphometric features, along with clinical attributes, are used to classify subjects into healthy (KL = 0) and early OA (KL = 1) categories using Random Forest and XGBoost classifiers.

The framework is designed to be accurate, reproducible, and clinically interpretable for early OA detection.

3.2 Methodological Framework

This section presents the complete methodological framework adopted in this study for automated knee MRI analysis, consisting of three major components:

i. Deep learning-based segmentation of knee joint anatomical structures using a 3D U-Net model,

ii. Voxel-based estimation of cartilage thickness from segmentation masks, and

iii. Machine learning classification of osteoarthritis (OA) severity using morphometric (cartilage thickness) and clinical features.

The proposed workflow was designed to address the research objectives outlined in Chapter 1 and to ensure a transparent, reproducible, and state-of-the-art approach consistent with contemporary practices in medical image computing.

The complete system workflow is illustrated below in Figure 3.1

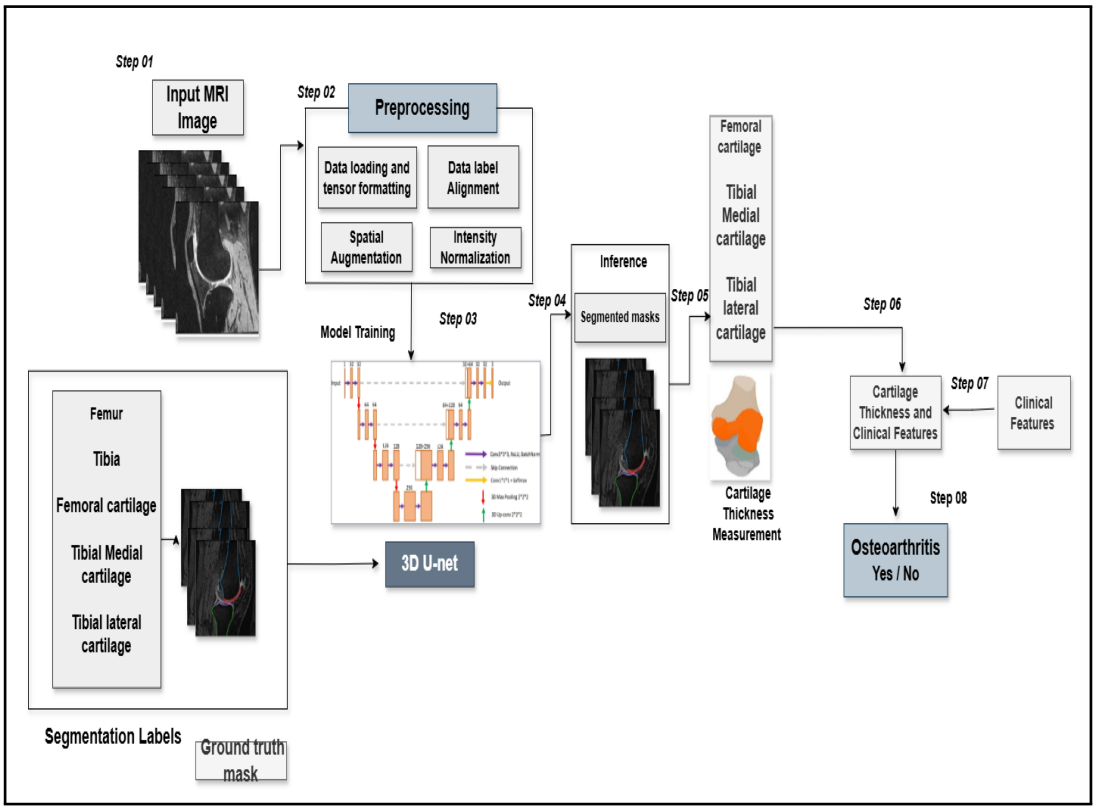

FIGURE 3.1: Proposed Methodology Diagram.

3.3 Dataset Description

This study utilizes the OAIZIB-CM dataset, a publicly available repository released under the CC-BY-NC-4.0 license and derived from the larger OAIZIB/OAI project.

The dataset comprises a total of 507 knee MRI volumes, of which 404 scans are allocated for training and 103 scans for testing.

All scans are stored in NIfTI (.nii.gz) format to preserve volumetric information and facilitate standardized processing.

The imaging modality employed in this study is high-resolution 3D sagittal Dual-Echo Steady-State (DESS) magnetic resonance imaging (MRI). This technique is particularly advantageous because it provides exceptionally detailed visualization of knee cartilage and surrounding anatomical structures, allowing for precise delineation of cartilage boundaries. The high spatial resolution and contrast of DESS MRI make it especially well suited for accurate cartilage segmentation and quantitative thickness analysis.

For each MRI volume, expert radiologists or trained annotators have provided segmentation masks that carefully label the regions of interest (ROIs). These annotated masks serve as the ground truth for subsequent computational analyses, ensuring that the extracted morphometric features accurately reflect the underlying anatomical structures. The ROIs typically include key components of the knee joints.

Table 3.1: Segmentation Labels and Corresponding ROIs

Label	**Region of Interest (ROI)**
1	Femur
2	Femoral Cartilage
3	Tibia
4	Medial Tibial Cartilage
5	Lateral Tibial Cartilage

The ground truth segmentation masks were generated using the CartiMorph framework [52], ensuring clinically reliable annotations.

3.4 Preprocessing Pipeline

Preprocessing was performed to ensure that all MRI volumes and segmentation masks were standardized before training the 3D U-Net model.

Variations in intensity ranges, orientations, and spatial dimensions can negatively influence segmentation quality; therefore, a structured pipeline was developed using the MONAI medical imaging framework.

The steps below describe how each transformation contributes to the overall performance and reliability of the proposed system.

The major steps included:

3.4.1 Data Loading and Tensor Formatting

MRI volumes and their corresponding segmentation masks were loaded using `LoadImaged` and converted into the required channel-first tensor format using `EnsureChannelFirstd`.

This step provides a uniform data structure for all samples, ensuring that the 3D U-Net receives consistent inputs, which improves training stability and reduces preprocessing-related errors.

3.4.2 Intensity Normalization

Pixel intensities were normalized to the range [0,1] using:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{3.1}$$

Normalizing all volumes to the same intensity scale allows the model to learn anatomical differences rather than scanner-dependent variations. As a result, the segmentation network generalizes better across subjects, leading to more reliable cartilage boundaries and subsequently more accurate thickness computation.

3.4.3 Spatial Augmentation

Random rotations, flips, affine transformations, and spatial cropping were applied during training. These augmentations expose the model to a wider variety of anatomical presentations and spatial configurations, improving robustness against positional and structural variability. This helps the network generate accurate segmentation masks even when encountering knees with different shapes, orientations, or mild pathological deviations—an essential requirement for obtaining consistent morphometric measurements.

3.4.4 Data and Label Alignment Verification

All volumes were checked to ensure identical shapes:

$$\dim(X) = \dim(Y) \tag{3.2}$$

Incorrectly aligned cases were discarded to avoid propagation of errors during segmentation.

3.5 3D U-Net Segmentation Model

Deep learning-based segmentation of knee anatomical structures was carried out using a 3D U-Net architecture in Figure 3.2.

which operates directly on volumetric MRI data.

The model extends the standard U-Net design into three dimensions, enabling it to capture spatial context across sagittal, coronal, and axial planes simultaneously.

This capability is essential for delineating thin cartilage regions and preserving anatomical continuity across slices, both of which are required for reliable morphometric analysis.

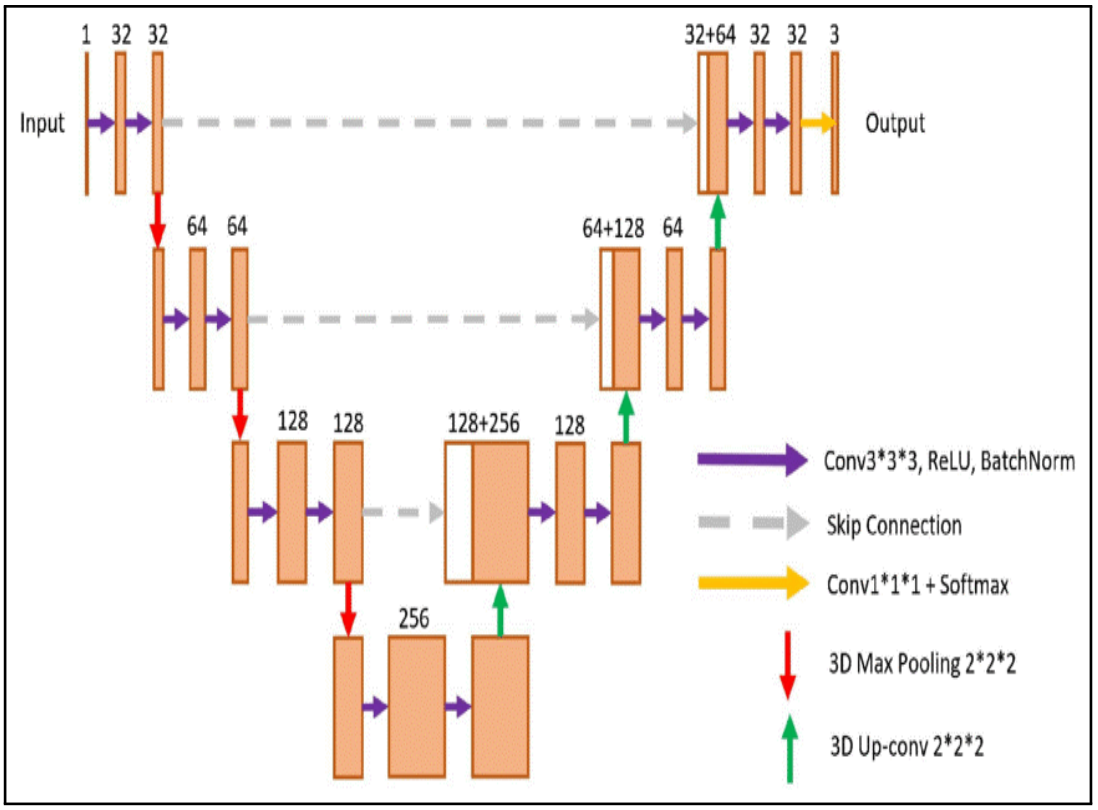

FIGURE 3.2: 3D U-NET Architecture

3.5.1 Architecture

The proposed segmentation network follows a 3D U-Net architecture consisting of five encoder–decoder levels to enable multi-scale feature extraction. The number of feature channels progressively increases from 16 to 32, 64, 128, and 256 across successive encoder stages, allowing the model to capture increasingly complex spatial representations. Residual blocks are incorporated within the skip connections to enhance gradient flow and preserve fine-grained spatial information during feature fusion. The network employs a SoftMax output layer with six channels, correspo-

nding to the background and five regions of interest (ROIs). Model optimization is performed using the Dice loss function, which is well suited for handling class imbalance and improving segmentation accuracy of cartilage structures.

3.5.2 Encoder Path

The encoder path extracts hierarchical feature representations from the input MRI volume. Each level applies a 3D convolutional operation followed by a non-linear activation function:

$$F_l = \sigma(\text{Conv3D}(F_{l-1}, k = 3) + b) \tag{3.3}$$

where F_{l-1} is the feature map from the previous level, k denotes the kernel size, b the bias term, and σ the activation function.

A 3D max-pooling layer is then used for spatial downsampling:

$$F_l^{\downarrow} = \text{MaxPool3D}(F_l) \tag{3.4}$$

This progressively reduces spatial dimensions while expanding the receptive field, enabling the network to learn coarse structural features essential for identifying cartilage, bone, and their boundaries.

3.5.3 Decoder Path

The decoder reconstructs high-resolution segmentation maps by gradually restoring spatial detail. Each decoder level begins with volumetric upsampling:

$$F_l^{\uparrow} = \text{UpConv3D}(F_{l+1}) \tag{3.5}$$

This upsampled feature map is concatenated with the corresponding encoder fea-

ture map through skip connections:

$$F_l^{\text{concat}} = [F_l^{\uparrow}, F_l] \tag{3.6}$$

The concatenated representation is then processed through additional 3D convolutional layers to refine structural boundaries and recover fine anatomical details across all regions of interest.

3.5.4 Dice Loss Function

The segmentation model was trained using the Dice Loss, which is well suited for handling the class imbalance between large bony structures and thin cartilage regions. The loss is defined as:

$$L_{\text{Dice}} = 1 - \frac{2\sum_i p_i g_i}{\sum_i p_i + \sum_i g_i} \tag{3.7}$$

where p_i is the predicted voxel and g_i is the ground-truth voxel. This formulation maximizes the spatial overlap between predicted and true segmentation masks, enabling the network to produce accurate delineations across all five anatomical structures.

3.5.4.1 Training Configuration

TABLE 3.2: Training Configuration for 3D U-Net

Hyperparameter	**Value**
Optimizer	Adam
Learning Rate	1×10^{-4}
Batch Size	1
Epochs	20
Loss	Dice Loss
Metric	DSC

3.6 Cartilage Thickness Computation

Cartilage thickness was quantified for three anatomical regions of interest: femoral cartilage, medial tibial cartilage, and lateral tibial cartilage. Thickness measurements were derived from the 103 post-processed segmentation masks generated by the 3D U-Net model.These measurements allowed for detailed morphometric analysis, capturing inter-subject variability and subtle regional differences in cartilage integrity. By analyzing thickness across multiple regions, we were able to better understand patterns of cartilage degeneration and their association with osteoarthritis progression.

Furthermore, these quantitative metrics provide a reliable basis for integrating imaging biomarkers with clinical variables, enhancing the accuracy and interpretability of downstream OA classification models.

3.6.1 Surface Extraction

For each binary cartilage mask M, the inner and outer cartilage surfaces were obtained using a morphological gradient operation:

$$S = M \ominus B - (M \oplus B) \tag{3.8}$$

where B denotes a three-dimensional structural element.

This operation isolates the boundary voxels of each cartilage region, required for subsequent calculation of local surface distance.

3.6.2 Distance Transform Thickness

Voxel-level thickness values were computed by measuring the Euclidean distance between each voxel x on one cartilage surface and the closest voxel y on the

opposing surface:

$$T(x) = \min_{y \in S_{\text{opposite}}} \|x - y\|_2 \tag{3.9}$$

This formulation provides a spatially resolved thickness map for each cartilage structure.

3.6.3 Mean Thickness per Region

For each cartilage type, the mean thickness value was computed by averaging all voxel-wise thickness measurements:

$$\bar{T} = \frac{1}{N} \sum_{i=1}^{N} T_i \tag{3.10}$$

where N represents the total number of valid thickness voxels within the region. These region-level measurements form part of the feature set used in OA severity classification.

3.7 Machine-Learning Classification

Following cartilage thickness computation, machine-learning models were trained to classify subjects into two categories based on Kellgren–Lawrence (KL) grading: Healthy (KL = 0) and Early OA (KL = 1)

The classification framework combines morphometric features with clinical attributes. Random Forest (RF) was implemented as a baseline, while XGBoost emerged as the optimal classifier after hyperparameter tuning.

3.7.1 Feature Set

The feature set used for osteoarthritis classification was designed to capture both structural cartilage changes and relevant clinical characteristics.

Mean cartilage thickness values were computed for the femoral, medial tibial, and lateral tibial regions, as these anatomical compartments are known to exhibit distinct degeneration patterns during early OA progression.In addition to morphometric features, demographic and clinical attributes including age, body mass index (BMI), and gender were incorporated, as these factors are strongly associated with OA risk and progression. To further characterize cartilage morphology, statistical descriptors of thickness, such as variance and standard deviation, were extracted to quantify regional heterogeneity and irregularity in cartilage thickness.

Together, these features provide a comprehensive representation of cartilage health and patient-specific risk factors, enabling robust discrimination between healthy and early OA subjects.

3.7.2 Random Forest Classifier

Random Forest is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to improve accuracy and reduce overfitting.

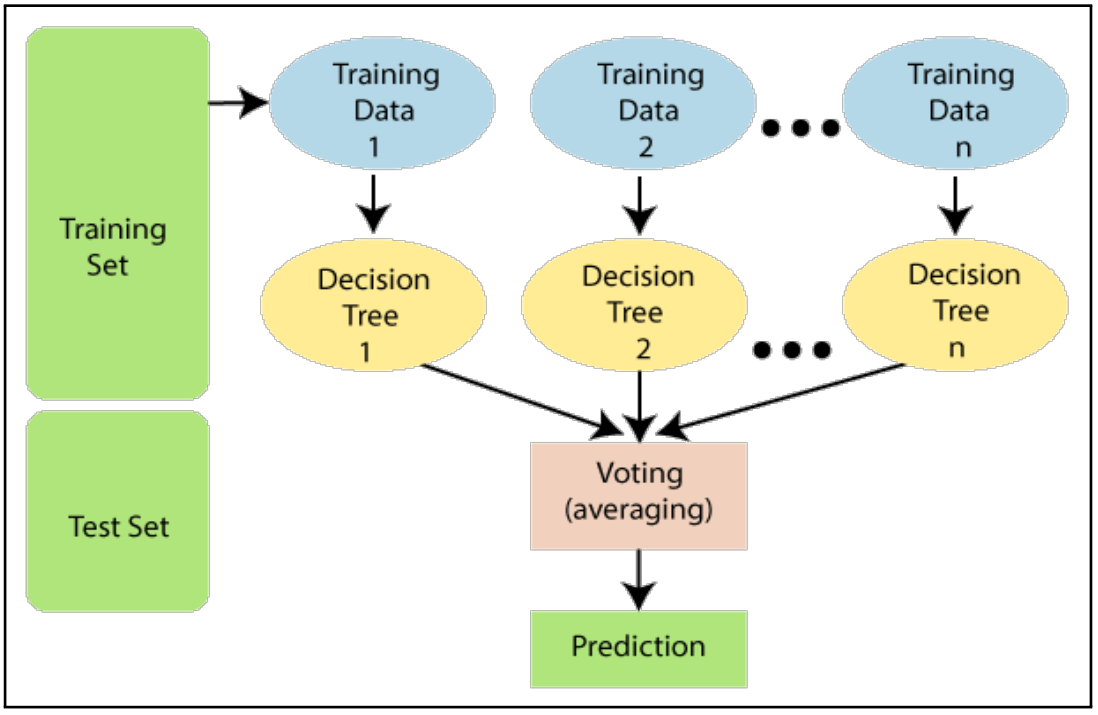

FIGURE 3.3: Random Forest Classifier

RF prediction is expressed as:

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \ldots, h_T(x)) \tag{3.11}$$

where:

x is the feature vector

$h_t(x)$ the prediction of the t^{th} tree, and

T the total number of trees

TABLE 3.3: Pseudo-code of Random Forest Classifier

Step	**Description**
Input	Training dataset D with features F and labels Y, number of trees T, number of features per split k
Output	Random Forest model RF
1	Initialize an empty forest $RF = \{\}$
2	For $i = 1$ to T: a. Draw a bootstrap sample D_i from dataset D b. Train a decision tree T_i on D_i c. For each node in T_i: i. Randomly select k features from F ii. Choose the best feature among k to split d. Grow the tree until stopping criteria e. Add T_i to forest RF
3	For each new sample x: a. Each tree $T_i \in RF$ predicts a class label y_i b. Collect predictions $\{y_1, y_2, \ldots, y_T\}$ c. Final output = majority_vote($\{y_1, y_2, \ldots, y_T\}$)

3.7.3 XGBoost Classifier

XGBoost is an optimized implementation of Gradient Boosting that builds trees sequentially, each correcting the errors of the previous one.

It uses regularization, parallelization, and efficient handling of sparse data to achieve high accuracy and speed

TABLE 3.4: Pseudo-code of XGBoost Classifier

Step	**Description**
Input	Training dataset D with features F and labels Y, number of boosting rounds K, learning rate η, loss function $L(y, \hat{y})$
Output	Boosted model with K trees
1	Initialize prediction $\hat{y} = 0$
2	For $k = 1$ to K: a. Compute gradients $g_i = \frac{\partial L(y_i, \hat{y}_i)}{\partial \hat{y}_i}$ for each sample b. Compute hessians $h_i = \frac{\partial^2 L(y_i, \hat{y}_i)}{\partial \hat{y}_i^2}$ for each sample c. Fit a regression tree $f_k(x)$ using $\{g_i, h_i\}$ as weights d. Update predictions: $\hat{y}_i \leftarrow \hat{y}_i + \eta \cdot f_k(x_i)$ e. Apply regularization to control complexity of f_k
3	Final prediction: $\hat{y}(x) = \sum_{k=1}^{K} \eta \cdot f_k(x)$

XGBoost prediction is expressed as:

$$\hat{y} = \sum_{k=1}^{K} f_k(x), \quad f_k \in F \tag{3.12}$$

Where:

i. F = space of regression trees

ii. $f_k(x)$ = the k^{th} boosted tree

iii. K = number of trees

iv. $\hat{y}$ = final prediction

3.8 Experimental Setup

All experiments were performed in a cloud-based environment (Google Colab) with the following configuration:

3.8.1 Hardware

All experiments were conducted in a cloud-based computational environment equipped with an NVIDIA T4 GPU featuring 16 GB of GDDR6 memory to support efficient deep learning model training and inference.

The system utilized an Intel Xeon-class virtual processor for general computation. while dynamic memory allocation provided between 12 and 25 GB of RAM depending on workload requirements.

This hardware configuration ensured adequate computational performance and scalability for both segmentation and machine-learning classification tasks.

3.8.2 Software

The software environment for this study was based on Python version 3.10, which served as the primary programming language for implementation and experimentation. Deep learning models were developed using PyTorch 2.0, while the MONAI 1.3 framework was employed for medical image preprocessing, augmentation, and standardized data handling.

Nibabel and NumPy were used for efficient loading, manipulation, and numerical processing of volumetric MRI data. Data visualization and result analysis were performed using Matplotlib and Seaborn. Machine-learning algorithms, including

Random Forest and XGBoost, along with SMOTE for class imbalance handling, were implemented using the Scikit-learn library.

3.8.3 Reproducibility

To ensure reproducibility of the experiments, random seeds were fixed across PyTorch, NumPy, and Python, guaranteeing consistent initialization and deterministic behavior of the models.

Model checkpoints were saved at fixed intervals during training, allowing recovery and evaluation of intermediate states.

Additionally, all results were systematically logged and version-controlled, providing a transparent and traceable record of the experimental workflow and enabling consistent comparison across different runs.

3.9 Evaluation Metrics

3.9.1 Segmentation Metric: Dice Similarity Coefficient

The performance of the segmentation model was evaluated using the Dice Similarity Coefficient (DSC), a widely used metric for measuring the overlap between predicted and ground-truth masks.

The DSC is defined as:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{3.13}$$

where

i. $|A|$ = number of predicted positive voxels

ii. $|B|$ = number of ground-truth positive voxels

iii. $|A \cap B|$ = number of correctly predicted voxels (overlap between prediction and ground truth)

3.9.2 Classification Metric: Accuracy

The performance of the classification models was evaluated using accuracy, which measures the proportion of correctly classified instances among all predictions. Accuracy is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.14}$$

where:

i. TP = True Positives

ii. TN = True Negatives

iii. FP = False Positives

iv. FN = False Negatives

3.10 Conclusion

This chapter presented a complete workflow for automated knee MRI analysis. It includes 3D U-Net segmentation of cartilage and bones, cartilage thickness estimation, and feature extraction for OA assessment. Extracted features were used with Random Forest and XGBoost classifiers to distinguish healthy from early OA cases. The methodology ensures accurate, reliable, and reproducible measurements, providing a strong foundation for the results in Chapter 4.

Chapter 4

Implementation and Experiments of the Proposed Methodology

This section provides a detailed overview of the experimental setup used to evaluate the proposed deep learning framework for automatic segmentation of knee anatomical structures, cartilage thickness quantification, and osteoarthritis (OA) severity classification.

The experiments were designed to address the research objectives outlined in Chapter 1, using the publicly available OAIZIB-CM dataset and implemented entirely in Google Colab with GPU acceleration.

4.1 Tools and Technologies Used

This research work utilized a combination of programming tools, deep-learning frameworks, medical image processing libraries, and machine-learning utilities to develop the complete segmentation and classification pipeline. These tools were carefully selected to ensure efficient data handling, robust model training, and accurate evaluation of imaging-derived features. The integration of multiple frameworks allowed for seamless preprocessing, augmentation, and post-processing of

MRI data, while also supporting advanced neural network architectures such as 3D U-Net for segmentation. Additionally, the use of specialized libraries facilitated feature extraction, statistical analysis, and visualization, enabling a comprehensive workflow that bridges imaging data with clinical variables. A detailed description of each tool is provided below:

4.1.1 Python

Python served as the primary programming language due to its extensive ecosystem for scientific computing, machine learning, and medical image analysis.

All stages of the pipeline—including preprocessing, model development, metric computation, and classification—were implemented in Python because of its readability, flexibility, and rich library support.

4.1.2 Google Colab

Google Colab provided the cloud-based computational environment required to train deep-learning models efficiently.

The platform's free access to NVIDIA T4 GPUs significantly reduced training time for the 3D U-Net segmentation model and enabled experimentation with different architectures and parameters.

Colab notebooks also ensured easy reproducibility and collaborative workflow management.

4.1.3 MONAI

MONAI played a central role in handling medical-image–specific requirements. It was used for MRI preprocessing, advanced 3D transformations (e.g., spacing, orientation, cropping), and implementation of the 3D U-Net architecture.

MONAI's built-in evaluators enabled accurate computation of segmentation metrics, most notably the Dice Similarity Coefficient, ensuring standardized and reproducible evaluation of the segmentation performance.

4.1.4 PyTorch

PyTorch functioned as the core deep-learning framework for constructing, training, and optimizing the segmentation model. It provided dynamic computation graphs, GPU acceleration, and essential modules for loss functions, backpropagation, and model checkpointing.

MONAI's components were seamlessly integrated with PyTorch, allowing smooth handling of 3D medical volumes.

4.1.5 Nibabel

Nibabel was used to load, read, and manipulate medical imaging data in NIfTI (.nii/.nii.gz) format. It enabled efficient extraction of volumetric MRI scans and their corresponding segmentation masks, facilitating direct integration into preprocessing pipelines.

Its support for affine information ensured accurate spatial alignment and orientation handling.

4.1.6 NumPy

NumPy supported all numerical and array-based operations required throughout the workflow.

This included pixel-intensity normalization, array reshaping, 3D volume manipulation, statistical calculations, and serving as the underlying data structure for tensors passed into deep-learning models.

4.1.7 Pandas

Pandas was used for managing structured, tabular data such as patient identifiers, clinical metadata, radiological variables, and extracted cartilage-thickness measurements. Its DataFrame structure facilitated easy merging, cleaning, transformation, and organization of both imaging-derived and non-imaging features for downstream classification tasks.

This structured format allowed for efficient handling of large datasets, ensuring that all features were correctly aligned with their corresponding subjects. Moreover, it enabled seamless integration of quantitative imaging biomarkers with clinical variables, supporting robust preprocessing.

4.1.8 Matplotlib

Matplotlib was used for visualizing segmentation results, plotting loss/accuracy curves, and displaying classifier performance metrics (e.g., confusion matrices, ROC curves). These visualizations played a crucial role in interpreting model behavior and presenting results clearly in the thesis.

4.1.9 Scikit-learn

Scikit-learn provided the machine-learning infrastructure for building and evaluating the classification models, specifically Random Forest and XGBoost.which were chosen for their ability to handle complex, high-dimensional data and provide interpretable feature importance scores.

It facilitated comprehensive dataset management, including partitioning into training, validation, and testing sets, as well as scaling and normalization of numerical features to ensure stable model convergence. Additionally, Scikit-learn supported the computation of a wide range of performance metrics, such as accuracy, precision, recall, F1-score, and area under the ROC curve, enabling a thorough assess-

ment of model performance across different classes.

To address the common issue of class imbalance in osteoarthritis datasets, SMOTE (Synthetic Minority Over-sampling Technique) was applied, generating synthetic samples for underrepresented classes and improving the model's ability to generalize.Collectively, these tools ensured a robust, fair, and reproducible evaluation of the classification stage, while also providing flexibility for hyperparameter tuning and cross-validation to optimize predictive performance..

4.2 Dataset Description

The experiments were conducted using the OAIZIB-CM dataset, a publicly available and CC-BY-NC-4.0–licensed medical imaging dataset for knee joint segmentation tasks. OAIZIB-CM is derived from the original OAIZIB dataset and integrates data from the Osteoarthritis Initiative (OAI) project as shown in Figure 4.1 and Figure 4.2. The dataset includes 507 MRI scans divided into a training set of 404 cases and a testing set of 103 cases, with corresponding ground-truth segmentation masks. Each scan is stored in NIfTI format and represents 3-D sagittal DESS MRI sequences of the knee joint.

FIGURE 4.1: Original MRI Image

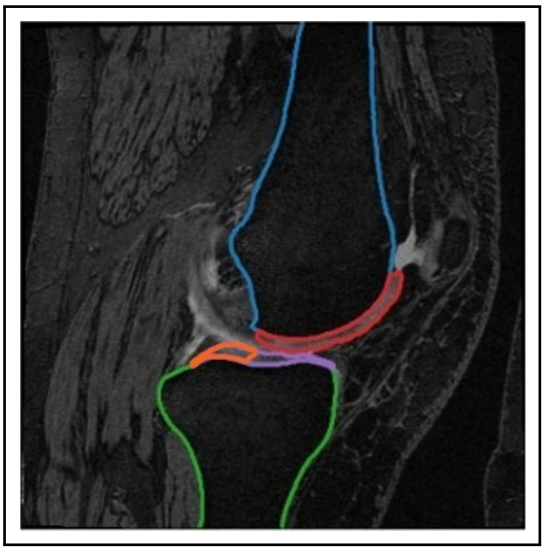

FIGURE 4.2: Ground Truth MRI Image

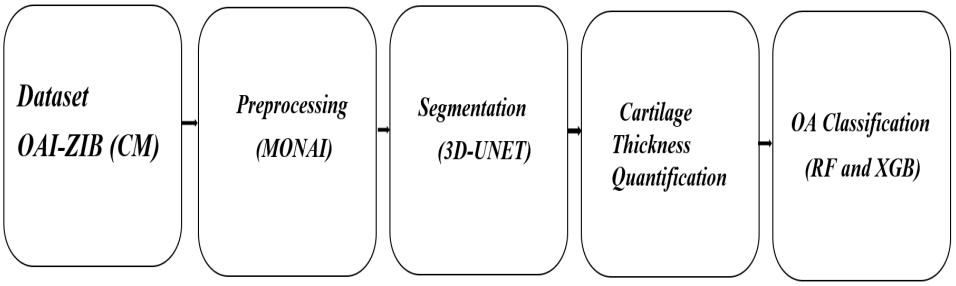

FIGURE 4.3: overview of the proposed Methodolgy

In this dataset, the tibial cartilage is subdivided into medial and lateral tibial cartilages to facilitate localized thickness and degeneration analysis.

TABLE 4.1: Anatomical structures annotated in the segmentation masks

Label ID	**Region of Interest (ROI)**
1	Femur
2	Femoral Cartilage
3	Tibia
4	Medial Tibial Cartilage
5	Lateral Tibial Cartilage

Ground-truth masks were provided by domain experts using the CartiMorph framework ensuring anatomically consistent and clinically validated annotations.

The dataset is distributed through both Hugging Face and Google Drive repositories, making it accessible for reproducible research.

4.2.1 MONAI Library

To prepare the MRI volumes for model training and inference, a series of preprocessing operations were performed using the MONAI library. The steps included:

i. Data loading: Reading MRI volumes and segmentation masks.

ii. Channel formatting: Aligning data tensors for 3D U-Net input.

iii. Intensity normalization: Scaling pixel values between 0 and 1.

iv. Data augmentation: Random flipping and 90-degree rotation to improve model robustness against spatial variations.

v. Type conversion: Ensuring compatibility with PyTorch tensors.

All scans were verified for image–label alignment to prevent mismatches during training and evaluation.

4.2.2 Google Colab

All experiments were executed in the Google Colab environment equipped with an NVIDIA T4 GPU (16 GB VRAM) and 16 GB of RAM. The software stack comprised Python 3.10, PyTorch 2.0, MONAI 1.3, and additional libraries such as NumPy, Matplotlib, and Nibabel for data handling and visualization.

The computational environment provided sufficient GPU memory and runtime to support full 3D volumetric training without significant memory bottlenecks.

4.2.3 Model Architecture and Training Setup

The segmentation network was implemented using a 3D U-Net architecture, which is a well-established and widely adopted model for volumetric medical image segmentation tasks.

This architecture is particularly suited for MRI and other 3D imaging modalities because it captures spatial context across multiple dimensions, enabling precise delineation of anatomical structures.

The 3D U-Net consists of an encoder-decoder structure with skip connections that allow fine-grained features from the contracting path to be effectively combined

with the upsampled features in the expanding path, improving segmentation accuracy, especially at boundaries.

The model was designed to process entire 3D volumes, capturing both local and global structural information, which is critical for accurate cartilage segmentation. The architecture included the following components:.

Table 4.2: Experimental Setup for 3D U-Net Segmentation

Component	**Description**
Model Architecture	3D U-Net for volumetric medical image segmentation
Layers	Five encoder–decoder levels
Channel Progression	16, 32, 64, 128, 256
Skip Connections	Residual units used to stabilize deeper layer training
Output Layer	SoftMax activation with six channels (background + 5 ROIs)
Optimizer	Adam
Learning Rate	1×10^{-4}
Loss Function	Dice Loss (to address imbalance between bone and cartilage regions)
Batch Size	1 (due to GPU memory constraints)
Epochs	20
Evaluation Metric	Dice Similarity Coefficient (DSC) computed per ROI
Evaluation Tool	MONAI Dice Metric module

The model was optimized using the Adam optimizer, with a learning rate of 1×10^4,

chosen for its efficiency in handling sparse gradients and fast convergence in deep neural networks. To address the inherent class imbalance between the large bone structures and the smaller, more delicate cartilage regions, Dice Loss was employed as the primary objective function. Dice Loss is particularly effective for segmentation tasks because it directly maximizes the overlap between predicted and ground-truth masks, ensuring accurate delineation of even the smallest anatomical structures.Each model was trained for 20 epochs, as illustrated in Figure 4.4, with a batch size of 1 due to GPU memory limitations associated with processing full 3D MRI volumes.Despite the small batch size, the training process was stabilized using gradient accumulation and learning rate scheduling to prevent overfitting and improve convergence.

The performance of the segmentation models was evaluated using the Dice Similarity Coefficient (DSC), which quantifies the overlap between predicted and reference masks on a per-region-of-interest (ROI) basis. This metric was computed for each cartilage and bone region using MONAI's DiceMetric module, providing a robust and reproducible measure of segmentation accuracy.

The use of DSC as the evaluation metric allowed for precise assessment of model performance across different anatomical structures and ensured that subtle variations in cartilage thickness and morphology were accurately captured.

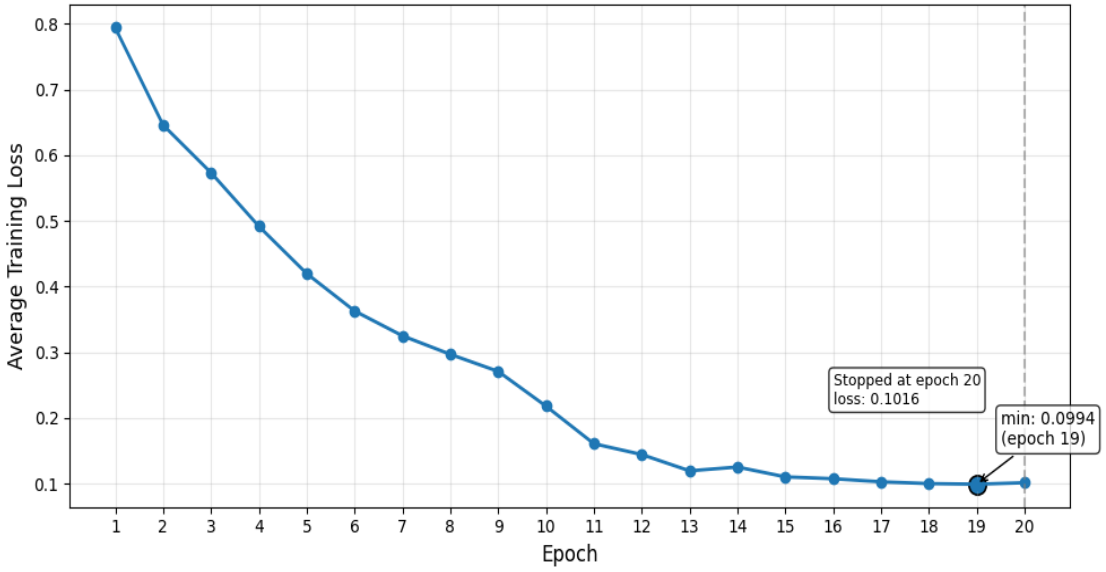

FIGURE 4.4: Average Loss per Epoch

4.2.4 Classification

After segmentation, cartilage thickness values were extracted for each region and used as input features for OA classification.

Two machine-learning algorithms were selected for this task: Random Forest and XGBoost.

i. Random Forest: Used as the baseline classifier due to its simplicity, stability, and strong performance on tabular clinical data. Initial experiments highlighted class imbalance (KL = 0 vs. KL = 1).

   To address this, the Synthetic Minority Oversampling Technique (SMOTE) was applied.

ii. XGBoost: Implemented as an advanced gradient-boosting classifier. After hyperparameter tuning, XGBoost achieved higher predictive performance, reaching up to 90% accuracy, and outperformed the baseline model.

Model evaluation was conducted using accuracy.

Integrating deep-learning segmentation with machine-learning classification allowed the development of a complete automated workflow for OA assessment.

4.2.5 Segmentation Results

This section presents the comprehensive quantitative and qualitative performance evaluation of the proposed 3D U-Net model for automated segmentation of knee anatomical structures from MRI data.

The evaluation focuses on assessing how accurately the model can delineate critical anatomical regions, providing insights into both its reliability and clinical applicability.The analysis specifically evaluates segmentation accuracy across five key anatomical regions: Femur, Femoral Cartilage, Tibia, Medial Tibial Cartilage, and

Lateral Tibial Cartilage.

These regions are essential for understanding joint morphology, cartilage health, and the progression of osteoarthritis, making their accurate segmentation a critical step for downstream morphometric analyses.The model's performance is primarily assessed using the Dice Similarity Coefficient (DSC), a widely adopted metric in medical image segmentation that quantifies the spatial overlap between the predicted segmentation masks and the ground-truth expert annotations. Higher DSC values indicate more accurate segmentation and better agreement with manual annotations.In addition to quantitative metrics, qualitative evaluations were performed by visual inspection of segmentation outputs overlaid on MRI slices.which helped to identify areas of under- or over-segmentation and provided additional context regarding the model's ability to capture complex cartilage morphology.

Together, these assessments offer a thorough understanding of the model's capabilities, highlighting its strengths in accurately segmenting both large bone structure and smaller, more intricate cartilage regions.

4.2.5.1 Quantitative Results

The model achieved consistent and reliable segmentation performance across all regions of interest. The average Dice coefficient (Mean ± Standard Deviation) for each region is presented in Table 4.3.

TABLE 4.3: Quantitative Segmentation Results Based on Dice Coefficient

Region of Interest (ROI)	**Mean Dice**	**Standard Deviation**
Femur	97.04	± 0.08
Femoral Cartilage	85.38	± 2.5
Tibia	97.57	± 0.54
Medial Tibial Cartilage	78.32	± 7.2
Lateral Tibial Cartilage	78.58	± 5.3

As shown in Table 4.3, the model performed exceptionally well in segmenting bony structures, particularly the Femur and Tibia, achieving Dice scores above 0.97.

The slightly lower performance observed in cartilage regions can be attributed to their thin geometry, smooth intensity transitions, and relatively low tissue contrast in MRI scans.

4.2.6 Segmentation Stability and Standard Deviation Thresholds

To determine an appropriate standard deviation (SD) threshold for segmentation stability, we referred to the validation statistics reported by *Ambellan et al. (2019)*, who proposed a hybrid statistical shape model (SSM) and CNN–based knee bone and cartilage segmentation framework using Osteoarthritis Initiative (OAI) data. Bone structures exhibit low variability, with DSC standard deviations of 0.3% for the femur and 0.3% for the tibia.whereas cartilage structures show higher variability, with SDs of 3.6% for femoral cartilage and 4.5% for tibial cartilage.

These values reflect state-of-the-art segmentation performance, influenced by anatomical complexity and segmentation volume, particularly in cartilage regions with smaller segmentation volumes.

Based on these results, SD thresholds of approximately 0.5%–1% for bone structures and 3%–5% for cartilage structures were adopted as reference values in this study.

Using these criteria, the proposed 3D U-Net–based CNN demonstrates highly stable segmentation, achieving DSC values of $97.04 \pm 0.08\%$ for the femur and $97.57 \pm 0.54\%$ for the tibia, indicating consistent bone segmentation well within the baseline variability. For cartilage regions, the model attains $85.38 \pm 2.5\%$ for femoral cartilage and $78.58 \pm 5.3\%$ for tibial cartilage, with standard deviations comparable to or slightly below the ranges reported by [6], supporting the robustness and reproducibility of the proposed method.

TABLE 4.4: Comparison of Segmentation Performance (DSC) for Bone and Cartilage Regions

Method	**Fem.Bone**	**Tib.Bone**	**Fem.Cart.**	**Tib.Cart.**
Hybrid (SSM and CNN)	98.6 ± 0.3	98.5 ± 0.3	89.9 ± 3.6	85.6 ± 4.5
GAN	98.5 ± 0.2	98.5 ± 0.3	89.7 ± 2.2	83.9 ± 4.0
CNN	98.6 ± 0.3	98.6 ± 0.4	88.2 ± 2.6	84.3 ± 4.5
Proposed CNN (3D-UNET)	97.04 ± 0.08	97.57 ± 0.54	85.38 ± 2.5	78.58 ± 5.3

4.2.6.1 Qualitative Results

Qualitative results were assessed visually by overlaying predicted segmentation masks onto corresponding MRI slices.3D visualization of ground truth and segmented MRI volume as shown in Figure 4.5 shows representative examples from the test set. The segmented contours demonstrate excellent anatomical alignment with ground-truth boundaries for both bone and cartilage regions.

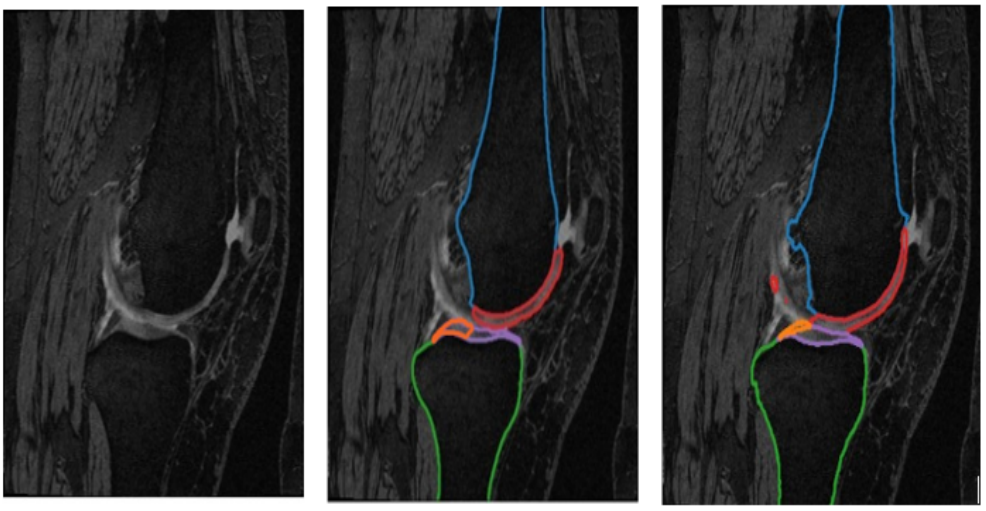

FIGURE 4.5: Original Image, Ground Truth Image and Segmented Image

4.2.7 Cartilage Thickness Evaluation

This section presents the results of quantitative cartilage thickness analysis derived from the segmented knee joint structures. The primary purpose of this evaluation is to quantify morphometric variations in key articular cartilage regions, specifically the Femoral Cartilage, Medial Tibial Cartilage, and Lateral Tibial Cartilage. By measuring thickness across these regions, the analysis aims to capture inter-subject variability as well as localized patterns of cartilage degeneration that are characteristic of osteoarthritis (OA) progression.These quantitative measurements provide insight into both the normal physiological differences in cartilage morphology and the pathological changes associated with disease advancement.Furthermore, comparing the observed trends with established anatomical and clinical knowledge allows for validation of the segmentation and measurement pipeline. This assessment not only highlights regions that are more susceptible to cartilage thinning but also informs potential clinical interventions and monitoring strategies for OA patients.

Overall, the results serve as a critical step in linking imaging-derived biomarkers with functional and clinical outcomes in knee osteoarthritis research.

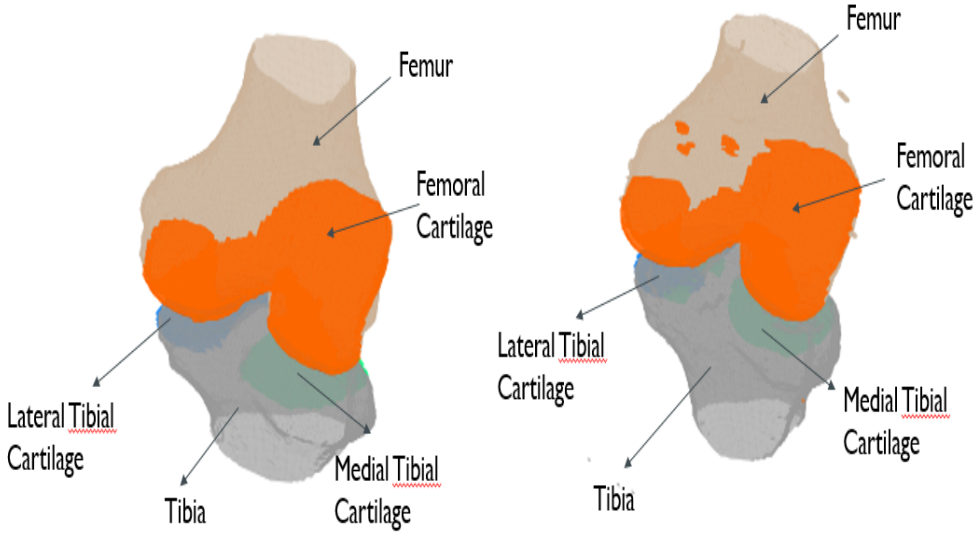

FIGURE 4.6: Ground Truth 3D MRI and Segmented 3D MRI

4.2.7.1 Descriptive Statistics of Cartilage Thickness

The mean, standard deviation, minimum, and maximum thickness values for each cartilage region were computed from the pos

t-processed segmentation masks of 103 test MRI cases. The descriptive statistics are presented in

TABLE 4.5: Descriptive statistics of cartilage thickness measurements

Cartilage Region	**Mean (mm)**	**Std. Dev. (mm)**	**Minimum (mm)**	**Maximum (mm)**
Femoral Cartilage	1.978	0.100	1.789	2.229
Medial Tibial Cartilage	1.608	0.134	1.383	2.023
Lateral Tibial Cartilage	1.805	0.135	1.511	2.144

The results indicate that the Femoral Cartilage possesses the highest mean thickness (1.978 mm), followed by the Lateral Tibial Cartilage (1.805 mm), and finally the Medial Tibial Cartilage (1.608 mm).

These findings are consistent with the normal physiological structure and functional biomechanics of the knee joint, where the femoral cartilage generally exhibits greater thickness compared to other regions.

This is primarily due to its critical load-bearing role and its anatomical curvature over the femoral condyles, which enables it to withstand higher mechanical stress and distribute joint forces effectively during movement. Additionally, the increased thickness of femoral cartilage contributes to joint stability, shock absorption, and smooth articulation, all of which are essential for maintaining proper knee function.

4.2.7.2 Anatomical Trends and Interpretation

Across all subjects, the same thickness pattern was observed: These findings are consistent with the normal physiological structure and functional biomechanics of the knee joint, where the femoral cartilage generally exhibits greater thickness compared to other regions. This is primarily due to its critical load-bearing role and

its anatomical curvature over the femoral condyles, which enables it to withstand higher mechanical stress and distribute joint forces effectively during movement. Additionally, the increased thickness of femoral cartilage contributes to joint stability, shock absorption, and smooth articulation, all of which are essential for maintaining proper knee function.

i. Femoral cartilage was the thickest,

ii. Lateral tibial cartilage was intermediate,

iii. Medial tibial cartilage was the thinnest.

This pattern agrees with clinical findings showing that the medial tibial cartilage often wears down earlier in osteoarthritis because it carries more body weight.

The small but consistent difference in thickness between the medial and lateral tibial regions (about 0.2 mm) supports the idea that thinning in the medial compartment can be an early sign of OA.

The low standard deviation values (< 0.15 mm) also indicate that the proposed segmentation method produced stable and consistent thickness measurements.

4.2.7.3 Visualization of Thickness Distributions

To visualize the variability in cartilage thickness across different regions, boxplots were generated for all three cartilage types, as shown in Figure 4.7. The plots illustrate that femoral cartilage maintains the narrowest interquartile range, indicating minimal variation among subjects.In contrast, the medial tibial cartilage demonstrates a greater spread in measurements, indicating more pronounced inter-subject variability.

This variability can be attributed to localized cartilage degeneration, uneven load distribution, and progressive wear patterns that commonly affect the medial compartment of the knee.Additionally, structural heterogeneity in this region may

arise from differences in biomechanical stress. Such variability is often associated with the early onset and progression of degenerative conditions like osteoarthritis.

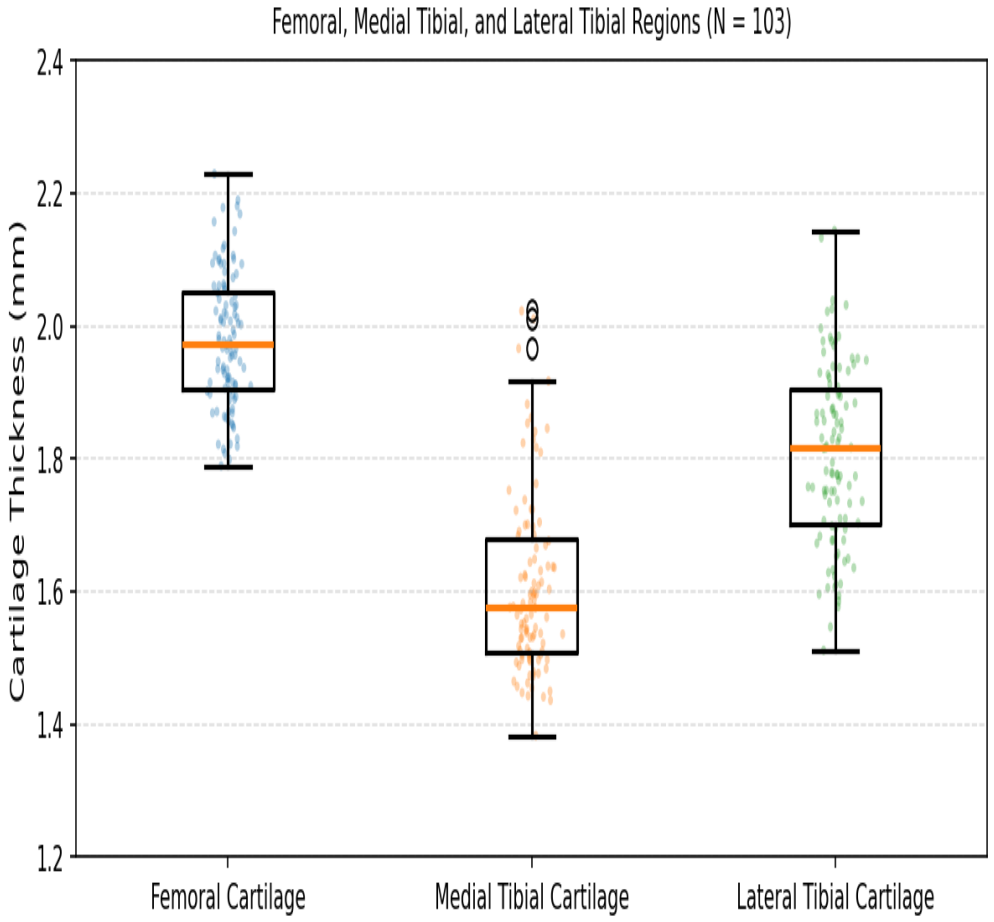

FIGURE 4.7: Boxplots showing cartilage thickness distribution for Femoral, Medial Tibial, and Lateral Tibial regions

4.2.8 Classification Results

This section presents the outcomes of osteoarthritis (OA) severity classification based on the morphometric features extracted from the segmented cartilage regions.Two machine learning models, Random Forest (RF) and Extreme Gradient Boosting (XGBoost), were trained using both imbalanced and balanced datasets to assess their ability to classify OA severity levels effectively. The goal of this experiment was to determine whether cartilage thickness and clinical parameters could accurately discriminate :healthy (KL = 0) and OA-affected (KL = 1) subj-

ects.

4.2.8.1 Data Balancing and Feature Preparation

Initially, the dataset exhibited an imbalanced class distribution, with 77 samples labeled as KL = 1 and 21 samples as KL = 0.

Such imbalance can lead to biased model predictions favoring the majority class.

To mitigate this, the Synthetic Minority Oversampling Technique (SMOTE) was applied, which synthetically generates new samples for the minority class by interpolating existing ones in the feature space.

After applying SMOTE, both classes were balanced to 61 samples each, resulting in a total of 122 data points for training and testing.

The features used for classification included:

i. Mean cartilage thickness values (Femoral, Medial Tibial, and Lateral Tibial)

ii. Demographic features (Age, BMI, Gender)

iii. Statistical measures (variance, standard deviation of thickness)

The balanced dataset improved model fairness and reduced bias toward the OA-dominant class.

4.2.8.2 Random Forest

The Random Forest classifier demonstrated robust performance with high sensitivity for OA detection. Table 4.6 presents the evaluation metrics.

Table 4.6: Random Forest Classifier Performance

Metric	**Value**
Accuracy	0.80

The confusion matrix for Random Forest as shown in Figure 4.9 indicates that the model achieved perfect recall for OA-positive cases (Class 1), correctly identifying all diseased instances However, it occasionally misclassified a few healthy cases as OA due to the oversampling process.

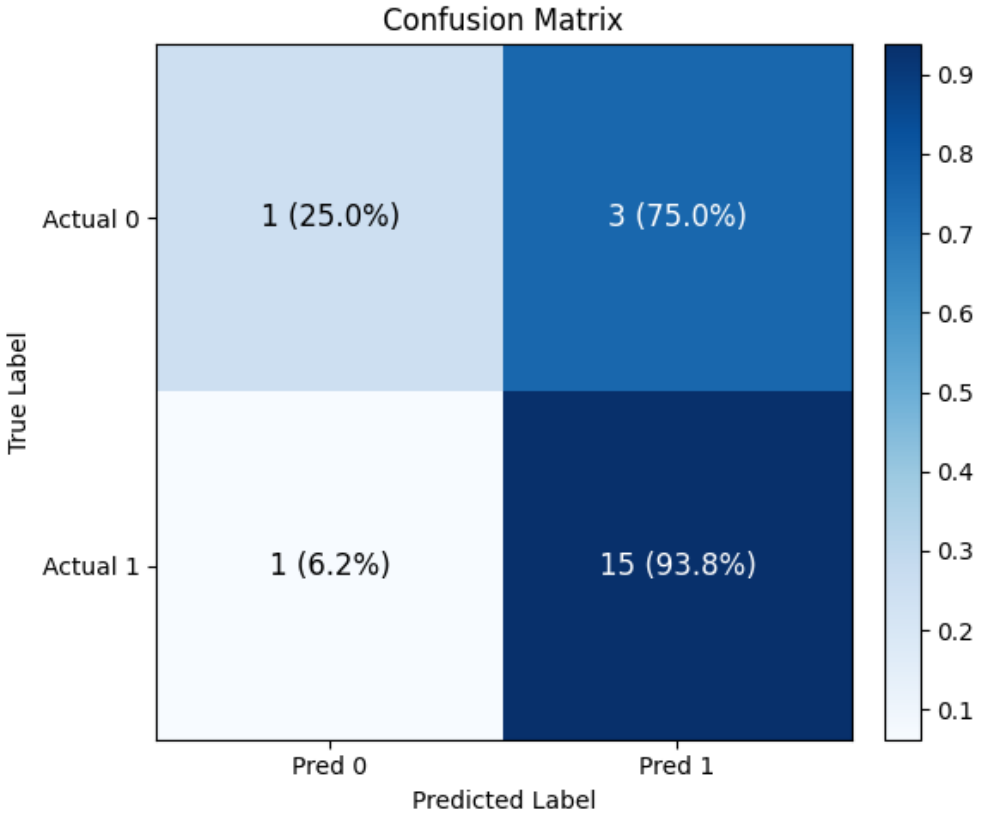

FIGURE 4.8: Confusion Matrix of Random Forest

4.2.8.3 XGBoost Results

The XGBoost classifier achieved comparable accuracy and recall values to the Random Forest model but exhibited slightly higher accuracy performance, as summarized in Table 4.8.

TABLE 4.7: XGBoost Classifier Performance

Metric	**Value**
Accuracy	0.90

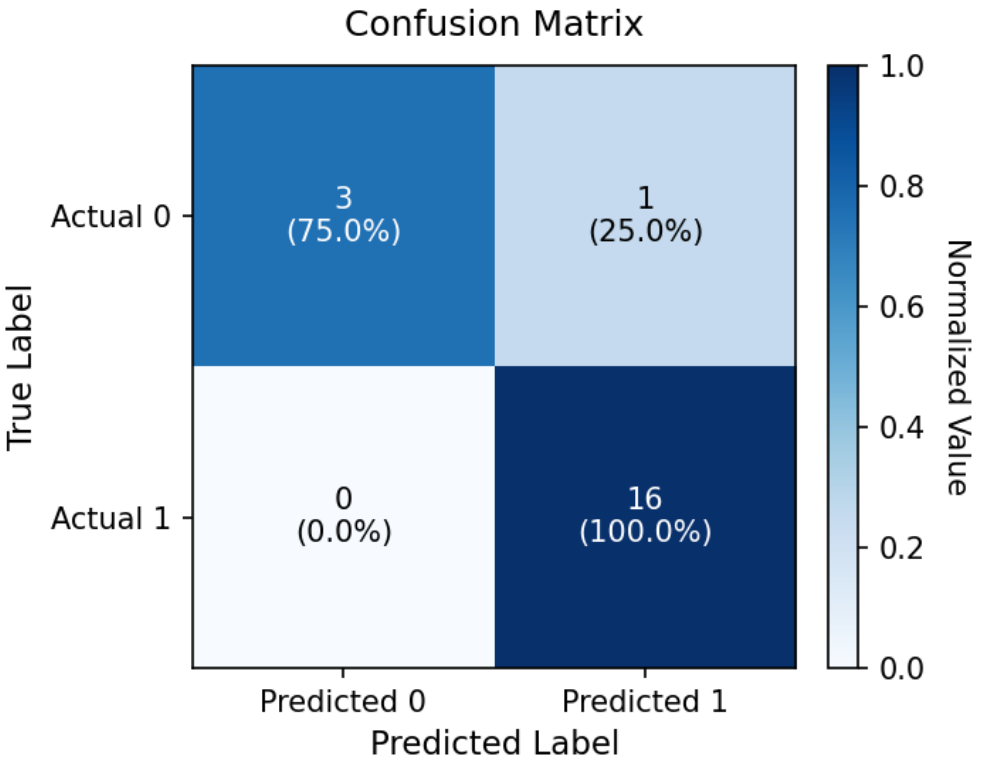

FIGURE 4.9: Confusion Matrix of Random Forest

TABLE 4.8: XGBoost Classifier Performance

Metric	**Value**
Accuracy	0.90

The confusion matrix for XGBoost, shown in Figure 4.8, indicates that the model exhibited a high detection rate for OA-positive subjects but a reduced ability to correctly classify all healthy subjects.

Similar to the Random Forest model, oversampling improved sensitivity for diseased cases but introduced occasional misclassifications in the healthy class.

TABLE 4.9: Comparison of OA Classification Performance

Study and Method	**Input**	**Seg.**	**Dataset**	**Task / Accuracy**
3D CNN [62]	Whole MRI	No	OAI (subset)	OA / non-OA: 83.0%
3D CNN [62]	Whole MRI	No	OAI (subset)	KL (5-class): 54.0%
This Work	Segmented Cartilage Thickness	Yes	OAI-ZIB	OA / non-OA: 80%; 90%

Although several studies have investigated knee osteoarthritis (OA) classification using deep learning, direct quantitative comparison with the proposed approach is not feasible, as none of the existing works performed classification following automated cartilage segmentation, nor were they evaluated on the same dataset used in this study. Most existing methods apply classification directly on raw imaging data, without explicitly isolating anatomical structures through segmentation. For example, Guida et al. employed a 3D convolutional neural network to classify OA severity directly from knee MRI volumes, achieving an accuracy of 83.0% for binary OA/non-OA classification and 54.0% for five-grade Kellgren–Lawrence (KL) classification.

Their approach relied on volumetric MRI inputs and did not incorporate a preceding segmentation step. In contrast, the proposed framework adopts a two-stage pipeline, where cartilage regions are first segmented and subsequently used for classification. This design enables the classifier to focus on anatomically relevant regions associated with OA progression, rather than learning from entire image volumes [62].

While a direct numerical comparison is not possible due to differences in datasets, evaluation protocols, and model objectives, the reported results from existing studies provide a contextual performance reference for OA classification tasks.

Therefore, the contribution of this work lies not in outperforming prior classification accuracies, but in demonstrating the feasibility and effectiveness of classification driven by segmentation-derived features, which has not been explored in prior studies.

4.3 Conclusion

This Chapter presents the implementation and evaluation of the proposed methodology for automated knee MRI analysis. The 3D U-Net model demonstrated highly stable and reproducible segmentation, achieving Dice scores above 97% for bone

and 78–85% for cartilage regions. Cartilage thickness measurements reflected expected anatomical trends, with femoral cartilage being thickest, lateral tibial intermediate, and medial tibial thinnest, consistent with clinical observations of early OA progression.Using these segmentation-derived features, Random Forest and XGBoost classifiers effectively distinguished healthy and OA-affected subjects, achieving accuracy of 80% and 90%, respectively. The results validate that the two stage pipeline segmentation followed by feature-based classification—is robust, clinically relevant, and capable of capturing subtle morphometric variations associated with osteoarthritis.

Chapter 5

Conclusion and Future Work

This chapter presented the experimental outcomes obtained from the proposed deep learning and machine learning framework for automatic analysis of knee joint MRI data. The results demonstrated that the 3D U-Net segmentation model effectively delineated the major anatomical structures of the knee joint, achieving Dice coefficients above 0.97 for bony regions and between 0.78 and 0.85 for cartilage regions. These results confirm the reliability of the segmentation pipeline in accurately identifying both bone and cartilage boundaries, which are crucial for subsequent morphometric analysis. The cartilage thickness evaluation produced physiologically consistent measurements, following the expected anatomical hierarchy: Femoral Cartilage > Lateral Tibial Cartilage > Medial Tibial Cartilage. This pattern is consistent with previous clinical studies and reflects natural variations in cartilage loading and wear associated with osteoarthritis progression. The low standard deviation values indicate stable and reproducible thickness estimations across the test set. The classification experiments further demonstrated that morphometric features, particularly cartilage thickness, can serve as reliable biomarkers for OA severity assessment. Both the Random Forest and XGBoost models achieved an overall accuracy of 90% with Random Forest showing slightly weak 0.80. The application of SMOTE effectively addressed dataset imbalance, improving the models' ability to generalize and accurately classify both healthy and OA affected subjects. Collectively, these findings validate the integration of

deep learning–based segmentation with statistical morphometric analysis and machine learning classification as an effective and automated framework for early osteoarthritis detection. The proposed system provides a reproducible, computationally efficient solution that supports clinical decision-making through accurate structural segmentation, reliable cartilage quantification, and precise OA classification.

5.1 Future Work

Future research should address several areas to enhance the performance and clinical utility of the proposed framework:

i. Expand to Multiclass OA Grading: Extending classification beyond KL 0–1 to include KL 2–4 would improve clinical relevance and align with full OA severity assessment.

ii. Integrate More Biomarkers: Incorporating shape features, texture descriptors, bone-marrow lesions, and meniscal pathology could strengthen classification performance and provide deeper insight into OA progression.

iii. Improve Tibial Cartilage Segmentation: Future models may benefit from attention mechanisms, transformer-based architectures, or multi-stage refinement networks to improve segmentation of thin cartilage layers.

iv. Use Larger and More Diverse Datasets: External validation on multi-center MRI datasets is essential for improving generalization and reducing dataset bias, as emphasized by Rodríguez-Vila et al. (2022).

v. Develop End-to-End Clinical Tools: Integrating the entire pipeline (segmentation → biomarker extraction → classification) into a user-friendly interface could support clinical decision-making and routine OA monitoring.

Through these enhancements, future research can build on the strengths of this

study to develop more accurate, interpretable, and clinically deployable AI solutions for osteoarthritis assessment.

Bibliography

[1] D. J. Hunter and S. Bierma-Zeinstra, "Osteoarthritis," *The Lancet*, vol. 393, no. 10182, pp. 1745–1759, 2019.

[2] A. K. Reddy, J. M. Anderson, H. M. Gray, K. Fishbeck, and M. Vassar, "Clinical trial registry use in orthopaedic surgery systematic reviews," *JBJS*, vol. 103, no. 10, p. e41, 2021.

[3] J. Folkesson, E. Dam, O. Olsen, and B. Carstensen, "Segmentation of articular cartilage from mr images using a voxel classification framework," *IEEE Transactions on Medical Imaging*, vol. 26, no. 1, pp. 106–115, 2007.

[4] L. Shamir, S. M. Ling, W. Scott, M. Hochberg, L. Ferrucci, and I. G. Goldberg, "Early detection of radiographic knee osteoarthritis using computer-aided analysis," *Osteoarthritis and Cartilage*, vol. 17, no. 10, pp. 1307–1312, 2009.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, vol. 9351, 2015, pp. 234–241.

[6] 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.

[7] H. S. Gan, M. H. Ramlee, A. A. Wahab, Y. S. Lee, and A. Shimizu, "From classical to deep learning: review on cartilage and bone segmentation techniques in knee osteoarthritis research," *Artificial Intelligence Review*, vol. 54, no. 4, pp. 2445–2494, 2021.

[8] B. Rodríguez-Vila, P. Valderrama, N. Landinez, M. M. Quinn, J. Yao, and J. E. Goetz, "Democratization of deep learning for segmenting cartilage from mris: Transparent and reproducible pipelines," *Journal of Orthopaedic Research*, vol. 40, no. 7, pp. 1422–1434, 2022.

[9] Ö. Z. L. Ü. Aysun, F. Yaman, and M. A. Leblebicier, *Kas İskelet Sistemi Ağrılarına Multidisipliner Yaklaşım.* Akademisyen Kitabevi, 2020.

[10] J. Fripp, S. Crozier, S. K. Warfield, and S. Ourselin, "Automatic segmentation of the bone and extraction of the bone–cartilage interface from mri," in *Medical Image Computing and Computer-Assisted Intervention.* Springer, 2010, pp. 101–108.

[11] E. B. Dam, M. Loog, C. Christiansen, and M. Nielsen, "Automatic segmentation of high- and low-field knee mris using elastic models," *IEEE Transactions on Medical Imaging*, vol. 34, no. 8, pp. 1587–1596, 2015.

[12] S. Khan, M. A. Khawer, J. Zhong, R. Qureshi, M. Asim, and W. Chen, "Advancing deep learning based knee cartilage segmentation in mri: Innovations, challenges and applications," *Osteoarthritis and Cartilage Open*, p. 100702, 2025.

[13] M. Favero, R. Ramonda, M. B. Goldring, S. R. Goldring, and L. Punzi, "Early knee osteoarthritis," *RMD Open*, vol. 1, no. Suppl 1, 2015.

[14] G. E. Gold *et al.*, "Advanced mri of articular cartilage," *Radiology*, vol. 240, pp. 622–637, 2006.

[15] G. Litjens *et al.*, "Survey of deep learning in medical imaging," *Medical Image Analysis*, 2017.

[16] B. Norman, V. Pedoia, and S. Majumdar, "Automatic segmentation of cartilage and meniscus using u-net cnns," *Radiology*, 2018.

[17] J. Moon, P. Jadhav, and S. Choi, "Deep learning analysis for rheumatologic imaging: current trends, future directions, and the role of human," *Journal of Rheumatic Diseases*, vol. 32, no. 2, pp. 73–88, 2025.

[18] R. Li, X. Yuan, M. Radfar, P. Marendy, W. Ni, T. J. O'Brien, and P. M. Casillas-Espinosa, "Graph signal processing, graph neural network and graph learning on biological data: a systematic review," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 109–135, 2021.

[19] J.-S. Lee and Y.-N. Chung, "Integrating edge detection and thresholding approaches to segmenting femora and patellae from magnetic resonance images," *Biomedical Engineering: Applications, Basis and Communications*, vol. 17, no. 01, pp. 1–11, 2005.

[20] P. M. Anshad, S. Kumar, and S. Shahudheen, "Segmentation of chondroblastoma from medical images using modified region growing algorithm," *Cluster Computing*, vol. 22, no. Suppl 6, pp. 13 437–13 444, 2019.

[21] T. G. Williams, A. P. Holmes, J. C. Waterton, R. A. Maciewicz, C. E. Hutchinson, R. J. Moots, A. F. Nash, and C. J. Taylor, "Anatomically corresponded regional analysis of cartilage in asymptomatic and osteoarthritic knees by statistical shape modelling of the bone," *IEEE Transactions on Medical Imaging*, vol. 29, no. 8, pp. 1541–1559, 2010.

[22] J.-S. Lee and Y.-N. Chung, "Integrating edge detection and thresholding approaches to segmenting femora and patellae from magnetic resonance images," *Biomedical Engineering: Applications, Basis and Communications*, vol. 17, no. 01, pp. 1–11, 2005.

[23] A. Norouzi, N. Habibi, Z. Nourbakhsh, and M. S. M. Rahim, "An iterative two-step method using thresholding and svm to segment bones from knee magnetic resonance images," *Majlesi Journal of Electrical Engineering*, vol. 17, no. 3, 2023.

[24] F. Yi and I. Moon, "Image segmentation: A survey of graph-cut methods," in *2012 International Conference on Systems and Informatics (ICSAI2012)*. IEEE, 2012, pp. 1936–1941.

[25] K. Camilus and G. V K, "A review on graph based segmentation," *International Journal of Image, Graphics and Signal Processing*, vol. 4, June 2012.

[26] S. Y. Ababneh and M. N. Gurcan, "An efficient graph-cut segmentation for knee bone osteoarthritis medical images," in *2010 IEEE International Conference on Electro/Information Technology*. IEEE, 2010, pp. 1–4.

[27] Y. Yin, X. Zhang, R. Williams, X. Wu, D. D. Anderson, and M. Sonka, "Logismos–layered optimal graph image segmentation of multiple objects and surfaces: Cartilage segmentation in the knee joint," *IEEE Transactions on Medical Imaging*, vol. 29, no. 12, pp. 2023–2037, 2010.

[28] S. Kashyap, H. Zhang, K. Rao, and M. Sonka, "Learning-based cost functions for 3-d and 4-d multi-surface multi-object segmentation of knee mri: Data from the osteoarthritis initiative," *IEEE Transactions on Medical Imaging*, vol. 37, no. 5, pp. 1103–1113, 2018.

[29] L. Shan, C. Zach, C. Charles, and M. Niethammer, "Automatic atlas-based three-label cartilage segmentation from mr knee images," *Medical Image Analysis*, vol. 18, no. 7, pp. 1233–1246, 2010.

[30] P. Zarychta, "Atlas-based segmentation in extraction of knee joint bone structures from ct and mr," *Sensors*, vol. 22, no. 22, 2022.

[31] M. Brehler, G. Thawait, J. Kaplan, J. Ramsay, M. J. Tanaka, S. Demehri, J. H. Siewerdsen, and W. Zbijewski, "Atlas-based algorithm for automatic anatomical measurements in the knee," *Journal of Medical Imaging*, vol. 6, no. 2, p. 026002, 2019.

[32] V. Pedoia, S. Majumdar, and T. M. Link, "Segmentation of joint and musculoskeletal tissue in the study of arthritis," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 29, pp. 207–221, 2016.

[33] L. Shan, C. Zach, C. Charles, and M. Niethammer, "Automatic atlas-based three-label cartilage segmentation from mr knee images," *Medical Image Analysis*, vol. 18, no. 7, pp. 1233–1246, 2014.

[34] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain mri segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.

[35] T. Rohlfing, R. Brandt, R. Menzel, and C. R. J. Maurer, "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, pp. 1428–1442, 2004.

[36] F. P. Nikolopoulos, E. I. Zacharaki, D. Stanev, and K. Moustakas, "Personalized knee geometry modeling based on multi-atlas segmentation and mesh refinement," *IEEE Access*, vol. 8, pp. 56 766–56 781, 2020.

[37] S. Mukherjee, O. Bandyopadhyay, A. Biswas, and B. Bhattacharya, "Tracking patellar osteophytes to detect osteoarthritis," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 11, pp. 1–11, March 2023.

[38] A. Aldieri, R. Biondi, A. A. La Mattina, J. A. Szyszko, S. Polizzi, D. Dall'Olio, N. Curti, G. Castellani, and M. Viceconti, "Development and validation of a semi-automated and unsupervised method for femur segmentation from ct," *Scientific Reports*, vol. 14, no. 1, p. 7403, 2024.

[39] J. G. Tamez-Peña, J. Farber, P. C. González, E. Schreyer, E. Schneider, and S. Totterman, "Unsupervised segmentation and quantification of anatomical knee features: data from the osteoarthritis initiative," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 1177–1186, 2012.

[40] H. Seo, M. Badiei Khuzani, V. Vasudevan, C. Huang, H. Ren, R. Xiao, X. Jia, and L. Xing, "Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications," *Medical Physics*, vol. 47, 2020.

[41] B. Felfeliyan, A. Hareendranathan, G. Kuntze, J. Jaremko, and J. Ronsky, "Mri knee domain translation for unsupervised segmentation by cyclegan (data from osteoarthritis initiative (oai))," 2023.

[42] S. Yu, M. Yang, R. Lartey, W. Holden, A. H. Ok, S. Khan, J. Kim, C. Winalski, N. Subhas, V. Chaudhary, and X. Li, "Unsupervised segmentation of knee bone marrow edema-like lesions using conditional generative models," *Bioengineering*, vol. 11, no. 6, p. 526, 2024.

[43] H. Kamper, K. Livescu, and S. Goldwater, "An embedded segmental k-means model for unsupervised segmentation and clustering of speech," 2017.

[44] S. Li, S. Zhao, Y. Zhang, J. Hong, and W. Chen, "Source-free unsupervised adaptive segmentation for knee joint mri," *Biomedical Signal Processing and Control*, vol. 92, p. 106028, Feb 2024.

[45] G. H. Seng, "Knee cartilage segmentation using multi-purpose interactive approach," 2016.

[46] F. Milletari *et al.*, "V-net: Fcn for volumetric segmentation," in *3D Vision Conference*, 2016.

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

[48] A. D. Desai, F. Caliva, C. Iriondo, A. Mortazi, S. Jambawalikar, U. Bagci, and I. S. C. W. Group, "The international workshop on osteoarthritis imaging knee mri segmentation challenge: a multi-institute evaluation and analysis framework on a standardized dataset," *Radiology: Artificial Intelligence*, vol. 3, no. 3, p. e200078, 2021.

[49] V. Juras, G. Chang, and R. R. Regatte, "Current status of functional mri of osteoarthritis for diagnosis and prognosis," *Current Opinion in Rheumatology*, vol. 32, no. 1, pp. 102–109, 2020.

[50] D. A. Kessler, J. W. MacKay, V. A. Crowe, and et al., "The optimisation of deep neural networks for segmenting multiple knee joint tissues from mris," *Computers in Medical Imaging and Graphics*, vol. 86, p. 101793, 2020.

[51] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2013, pp. 246–253.

[52] Y. Yao, J. Zhong, L. Zhang, S. Khan, and W. Chen, "Cartimorph: A framework for automated knee articular cartilage morphometrics," *Medical Image Analysis*, vol. 91, p. 103035, 2024.

[53] E. Panfilov, A. Tiulpin, M. T. Nieminen, S. Saarakkala, and V. Casula, "Deep learning-based segmentation of knee mri for fully automatic subregional morphological assessment of cartilage tissues," *Journal of Orthopaedic Research*, vol. 40, no. 5, pp. 1113–1124, 2022.

[54] F. Renard, S. Guedria, N. De Palma, and N. Vuillerme, "Variability and reproducibility in deep learning segmentation," *Scientific Reports*, 2020.

[55] J. Pineau *et al.*, "Reproducibility in ai research," *Journal of Machine Learning Research*, 2021.

[56] S. Torp-Pedersen, E. M. Bartels, J. Wilhjelm, and H. Bliddal, "Articular cartilage thickness measured with us is not as easy as it appears: A systematic review," *Ultraschall in der Medizin*, vol. 32, no. 1, pp. 54–61, 2011.

[57] R. Kakavand, "Enhancing the efficiency of subject-specific knee joint biomechanical simulations with applications to osteoarthritis," 2024.

[58] V. B. Kraus, M. Nevitt, and L. J. Sandell, "Summary of the oa biomarkers workshop 2009—biochemical biomarkers: Biology, validation, and clinical studies," *Osteoarthritis and Cartilage*, vol. 18, no. 6, pp. 742–745, 2010.

[59] Y. Zhang and J. M. Jordan, "Epidemiology of osteoarthritis," *Clinics in Geriatric Medicine*, vol. 26, no. 3, pp. 355–369, 2010.

[60] Q. Zhang, J. Geng, M. Zhang, T. Kan, L. Wang, S. Ai, H. Wei, L. Zhang, and C. Liu, "Cartilage morphometry and magnetic susceptibility measurement for knee osteoarthritis with automatic cartilage segmentation," *Quantitative Imaging in Medicine and Surgery*, 2023.

[61] J. Antony, K. McGuinness, K. Moran, and N. E. O'Connor, "Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks," in *International Conference on Machine Learning and Data Mining in Pattern Recognition*. Springer, 2017, pp. 376–390.

[62] C. Guida, M. Zhang, and J. Shan, "Knee osteoarthritis classification using 3d cnn and mri," *Applied Sciences*, vol. 11, no. 11, p. 5196, 2021.

[63] T. Wang, K. Leung, K. Cho, G. Chang, and C. M. Deniz, "Total knee replacement prediction using structural mris and 3d convolutional neural networks," in *Proceedings of the International Conference on Medical Imaging with Deep Learning (MIDL), Extended Abstract Track*, London, UK, Jul. 2019.

[64] F. Eckstein, A. Guermazi, G. Gold, J. Duryea, M. P. Hellio Le Graverand, W. Wirth, and C. G. Miller, "Imaging of cartilage and bone: Promises and pitfalls in clinical trials of osteoarthritis," *Osteoarthritis and Cartilage*, vol. 22, no. 10, pp. 1516–1532, 2014.

[65] S. Rani, M. Memoria, A. Almogren, S. Bharany, K. Joshi, A. Altameem, A. U. Rehman, and H. Hamam, "Deep learning to combat knee osteoarthritis and severity assessment by using cnn-based classification," *BMC Musculoskeletal Disorders*, vol. 25, p. 817, 2024.

[66] R. Abdelbasset Brahim, R. Jennane, T. Riad, L. K. Janvier, H. Toumi, and E. Lespessailles, "A decision support tool for early detection of knee osteoarthritis using x-ray imaging and machine learning: Data from the osteoarthritis initiative," *Computerized Medical Imaging and Graphics*, vol. 73, pp. 11–18, Apr 2019.

[67] W. Chen, H. Zheng, B. Ye, T. Guo, Y. Xu, Z. Fu, X. Ji, X. Chai, S. Li, and Q. Deng, "Identification of biomarkers for knee osteoarthritis through clinical data and machine learning models," *Scientific Reports*, vol. 15, no. 1, p. 1703, Jan 2025.