



THE UNIVERSITY OF QUEENSLAND
A U S T R A L I A

Adapting Context Aggregated 3D Network (CAN3D) for Automated Knee Segmentation in Total Knee Replacement (TKR) Patients

by Shaivika Anand

Bachelor of Software Engineering (Honours)



ORCID

*A thesis submitted for the degree of School of Electrical Engineering
and Computer Science at
The University of Queensland in 2025*
School of Electrical Engineering and Computer Science

Shaivika Anand
shaivika.anand@student.uq.edu.au
June 13, 2025

Prof. Michael Bruenig
Head of School
School of Electrical Engineering and Computer Science
The University of Queensland
St Lucia QLD 4072

Dear Professor Bruenig,

In accordance with the requirements of the Degree of Bachelor of Software Engineering (Honours) in the School of Electrical Engineering and Computer Science, I submit the following thesis entitled:

“Adapting Context Aggregated 3D Network (CAN3D) for Automated Knee Segmentation in Total Knee Replacement (TKR) Patients”

This work was performed under the supervision of Dr. Shekhar S. Chandra. I declare that the work submitted in this thesis is my own, except as acknowledged in the text and footnotes, and has not been previously submitted for a degree at The University of Queensland or any other institution.

Yours sincerely,
Shaivika Anand

Declaration by Author

This thesis is composed of my original work, and contains no material previously published or written by another person, or generated by artificial intelligence (AI), except where clearly referenced and acknowledged either in the main text or the preliminary pages of this thesis. I have clearly stated the contribution of others to jointly authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, use of AI tools, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Abstract

Osteoarthritis (OA) is a degenerative joint disease that commonly affects the knee, leading to a progressive cartilage loss and structural deterioration. Early diagnosis is important, but remains a challenge due to the lack of reliable tools and standardised clinical definitions [1]. Magnetic Resonance Imaging (MRI) provides detailed views of soft tissues, making it a promising way for detecting early OA-related changes [2]. However, manual segmentation of MRI data is time-consuming and prone to variability. Recent advancements in deep learning have showed potential for automating this task, but current models often lack validation on complex, real-world datasets [3].

This thesis focuses on the adaptability of CAN3D which is a compact 3D convolutional neural network, initially developed for prostate and hip and eventually trained on brain MRI segmentation, for use in knee OA analysis. The model is first evaluated on the OASIS brain MRI dataset, then retrained on the OAI-ZIB dataset of healthy knees, and finally tested on the OAI Total Knee Replacement (TKR) dataset, which includes patients with severe OA . The study uses both qualitative and quantitative metrics to assess segmentation performance across datasets.

By being the first to apply CAN3D to the OAI-TKR dataset, this research evaluates the model's generalisability across different level of OA. The findings aim to inform the development of efficient, clinically applicable deep learning models for OA monitoring and post-surgical assessment. The outcomes of this work support the integration of CAN3D into clinical workflows, particularly in resource-constrained settings, where compact and efficient models can help with earlier and more accurate assessment. Even though further validation and refinement are necessary, this study lays important groundwork for adapting deep learning tools that can help with osteoarthritis diagnosis and improve patient outcomes.

Publications Included in this Thesis

No publications included.

Other Publications During Candidature

No publications included.

Contributions by Others to this Thesis

This thesis extends the work of Wei Dai, Boyeong Woo, Siyu Liu, Matthew Marques, Craig Engstrom, Peter B. Greer, Stuart Crozier, Jason A. Dowling, and Shekhar S. Chandra, as presented in the paper **“CAN3D: Fast 3D medical image segmentation via compact context aggregation”**, which proposes the CAN3D model for 3D medical image segmentation. Their model served as the foundation for this research. While their work focused on segmentation of large 3D MR volumes across multiple anatomical regions, this thesis adapts and evaluates CAN3D specifically for knee cartilage segmentation. It further explores performance optimisation and generalisability across knee datasets such as Osteoarthritis Initiative (OAI-ZIB) and Osteoarthritis Initiative - Total Knee Replacement (OAI-TKR).

No other contributions were made by these authors to the research design, implementation, or writing of this thesis.

Use of Artificial Intelligence

To ensure transparency and uphold academic integrity, the following table outlines the use of AI tools in the preparation of this thesis:

AI Tool	Purpose of Use	Prompts Used	Thesis Section
ChatGPT 4.0	Helped in refining and clarifying thesis content including explanations, summaries, and rephrasing.	Examples: “Help me to understand the basic difference between model architecture of CAN3D, KAN3D and UNET”, “Rewrite this paragraph for clarity, do not change the original wordings, just try to correct grammar or spell errors”	Introduction, Literature Review, and Methodology
Grammarly	Help with grammar, style, and academic tone.	N/A (automated grammar and style suggestions)	Entire thesis
Zotero + AI plugins	Assisted in managing references and generating citation formats accurately.	N/A	Bibliography and citations throughout thesis

Statement of Responsibility

I declare that I am responsible for the content of this thesis, including all material that has been created, edited, or improved with the assistance of artificial intelligence tools. While tools such as ChatGPT 4.0 and Grammarly were used for clarification, summarisation, and language refinement, all content was critically reviewed and verified by me. No content was blindly copied or directly generated without evaluation, and all AI-assisted contributions were incorporated with academic judgment and integrity. The use of these tools aligns with academic integrity principles and complies with The University of Queensland's guidelines on responsible use of generative AI in research and writing.

Editorial Assistance

No editorial assistance.

Financial Support

No financial support was received to fund this research.

Inclusion of Work Submitted for Another Degree

No works submitted towards another degree or diploma have been included in this thesis.

Research Involving Human or Animal Subjects

No animal or human subjects were included in this research.

Acknowledgments

I would like to express my gratitude to my supervisor, Dr. Shekhar S. Chandra, for his constant support, encouragement, and guidance throughout the course of this thesis. His mentorship has been instrumental in shaping my research skills, helping with my understanding of concepts, and building my confidence as an independent researcher. I would also like to thank the members of Dr. Chandra's research group for their help, support and guidance which helped to the development of this work. Lastly, I am incredibly grateful to my family and friends for their constant encouragement and emotional support throughout this journey. Their belief in me kept me motivated through the most challenging times.

Contents

Declaration by Author	iii
Abstract	iv
Contents	viii
List of Figures	x
List of Tables	xii
List of Abbreviations	xiii
1 Introduction	1
1.1 Background and Context	1
1.1.1 Understanding Osteoarthritis	1
1.1.2 Challenges in Early Diagnosis and Role of Medical Imaging	2
1.2 Problem Statement	3
1.3 Objectives and Research Questions	4
1.4 Significance of the Study	5
2 Literature Review and Background	6
2.1 The Global Burden of Knee Osteoarthritis: Challenges in Diagnosis and the Need for Novel Approaches	6
2.2 Conventional Techniques: Segmentation and Ground Truth Methods	7
2.2.1 Radiography	7
2.2.2 Kellgren-Lawrence Grading System	8
2.2.3 Methodologies for Knee Cartilage Segmentation	9
2.3 MRI Introduction	9
2.3.1 Cartilage Volume as a Biomarker in OA	11
2.4 The Advent of Deep Learning	11
2.4.1 Deep Learning in MRI Acquisition	11
2.4.2 Deep Learning for Automated Cartilage Segmentation	12
2.5 Recent Technological Advancements in Segmentation Models	13

2.5.1	UNET	13
2.5.2	CAN3D	14
2.5.3	KAN3D	16
3	Methodology	18
3.1	Existing Work	18
3.2	Model Architecture	18
3.3	Experiments	21
3.3.1	Experimental Design and Workflow	21
3.3.2	Datasets Used	21
3.3.3	Train/Test Splits	21
3.3.4	Loss Functions	22
3.3.5	Hyperparameters	24
3.3.6	Implementation Details	25
3.3.7	Evaluation Metrics	31
4	Results	32
4.0.1	Validation Set Segmentation Accuracy	32
4.0.2	Performance of the trained CAN3D Model	33
4.0.3	Comparative Volume Analysis Across Cohorts	34
4.0.4	Summary and Implications	36
5	Discussion	38
5.0.1	Interpretation of Tibial and Femoral Cartilage Volume Findings	38
5.0.2	Dataset Domain Shift and Generalisability	39
6	Future Work	41
7	Conclusion	43
	Bibliography	45

List of Figures

1.1	A normal joint vs OA affected knee joint	2
2.1	(a) Grade 0 – Healthy knee (b) Grade I – Minor signs of OA (c) Grade II – Early wear and tear with clearly reduced space between bones. (d) Grade III – Moderate damage with bones getting closer and the surface underneath harden (e) Grade IV – Severe damage with bones touching and clear joint deformities.	8
2.2	MRI scan of the knee joint showing cartilage and bone structures relevant to osteoarthritis	10
2.3	Simple architecture of 2D-CNN	12
2.4	Model architecture for UNET	14
2.5	Model architecture of Context Aggregation Network 3D (CAN3D)	15
2.6	(a) Kaleidoscope transform on a 2D image. (b) Kaleidoscope transform on a 3D image.	16
2.7	The KAN3D model architecture where KT denotes kaleidoscope transform and KT-1 denotes inverse kaleidoscope transform	17
3.1	Overview of the CAN3D-based segmentation pipeline from input MRI volume to segmented output	19
3.2	(a) A convolution block. (b) Convolution filter with different dilation rates. (c) Detailed architecture of CAN-3D model.	20
3.3	Initial noisy outputs	26
3.4	Final output for the OASIS	27
3.5	Different views of the same segmented MRI scan showing cartilage and bone structures.	29
3.6	Qualitative comparison of original MRI scans and predicted segmentation masks across three anatomical views: axial, coronal, and sagittal.	30
4.1	Qualitative cartilage segmentation result from the OAI-TKR dataset for patient ID 9250756.RIGHT_6. The segmentation boundaries show high visual correspondence with cartilage structure.	33
4.2	Qualitative cartilage segmentation result from the OAI-TKR dataset for patient ID 9792327.RIGHT_3. The segmentation is relatively clear, even in the presence of structural degeneration.	33
4.3	Qualitative cartilage segmentation result from the OAI-TKR dataset for patient ID 9485359.LEFT_3. The segmentation appears visually plausible, though the axial view shows potential boundary confusion.	34

4.4 Qualitative cartilage segmentation result from the OAI-TKR dataset for patient ID 9035541_LEFT_3. The axial segmentation is less distinct, possibly due to anatomical complexity or model limitations.	34
4.5 Boxplot comparison of femoral cartilage volume across cohorts. OAI-ZIB shows the highest median, while TKR Unhealthy exhibits the most degeneration and variability . . .	35
4.6 Boxplot comparison of tibial cartilage volume across cohorts. No significant differences are observed	36

List of Tables

3.1	Key Hyperparameters Used in the Model	25
3.2	Sample of merged dataset combining segmentation output with clinical metadata.	31
4.1	Dice Similarity Coefficient (DSC) for femoral and tibial cartilage on the validation set . .	32
4.2	Summary statistics for femoral cartilage volume (in mm ³).	35
4.3	Summary statistics for tibial cartilage volume (in mm ³).	36

List of Abbreviations

Abbreviations	Full Form
CAN3D	Context-Aggregated Network 3D
KAN3D	Knowledge-Aware Network 3D
KT	Kaleidoscope Transform
UNET	U-shaped Convolutional Neural Network
OAI	Osteoarthritis Initiative
TKR	Total Knee Replacement
MRI	Magnetic Resonance Imaging
OA	Osteoarthritis
DL	Deep Learning
AI	Artificial Intelligence
MSE	Mean Squared Error
MAE	Mean Absolute Error
ReLU	Rectified Linear Unit
DSC	Dice Similarity Coefficient
CNN	Convolution Neural Network
DSL	Dice Squared Focal Loss

Chapter 1

Introduction

1.1 Background and Context

Osteoarthritis (OA) is a progressive joint disease that imposes a significant burden on individuals and healthcare systems worldwide, specially with increasing age and lifestyle changes [1]. To understand its clinical presentation, diagnostic challenges, and the coming up of medical imaging. The following subsections talks about the clinical background of OA and highlight the importance of medical imaging in helping with early diagnosis and monitoring of OA.

1.1.1 Understanding Osteoarthritis

Arthritis, derived from the Greek term meaning “disease of the joints,” is defined as acute or chronic inflammation of the joints, often accompanied by pain and structural damage [4]. The etiology of arthritis varies by type. Among the different types of arthritis, osteoarthritis (OA) is the most prevalent, particularly affecting weight-bearing joints such as the knee. Its major risk factors include increasing age, female sex, joint trauma, obesity, and certain genetic mutations — notably in collagen types II, IV, V, and VI [5,6]. Knee OA affects an estimated 19% to 30% of adults over the age of 45, making it a significant public health concern [7]. It is characterized by a degenerative cascade involving progressive cartilage loss, which ultimately leads to structural damage of the bone.

Figure 1.1 shows the structural differences between a healthy knee joint and one affected by osteoarthritis. In a normal joint, key anatomical components such as the cartilage, synovial membrane, synovial fluid, and joint capsule work together to allow smooth and pain-free motion. However, in OA-affected joints, cartilage becomes thinned, bone ends begin to rub together, and bone spurs (osteophytes) may form, leading to pain and joint stiffness.

Typical radiological findings in OA include subchondral cysts, osteophyte formation, and thickening of the subchondral plate [4]. At a molecular level, inflammatory mediators such as interleukins and chemokines (these are signaling proteins which are used by our immune system to coordinate inflammation) play a key role in the progression of osteoarthritis. Specifically, proteins like interleukin-6, monokines, interferon-induced protein-10, and macrophage chemotactic protein stimulate the

release of enzymes that break down joint collagen, resulting in a cartilage degradation [8]. Apart from that, calcification (a process where calcium salts starts to build up in tissues) can stiffen the cartilage, reducing its ability to absorb shock and leading to further deterioration. In osteoarthritis, calcification of the surrounding articular cartilage contributes to the thinning and eventual destruction of the cartilaginous matrix. Aging also impairs chondrocyte function, further increasing vulnerability to osteoarthritic degeneration [4, 8].

Common symptoms of arthritis include joint pain, swelling, stiffness, loss of function, deformity, weakness, and instability. These physical symptoms are often accompanied by fatigue, sleep disturbances, emotional distress, and signs of any underlying systemic illness. Pain tends to worsen with activity and toward the end of the day [4].

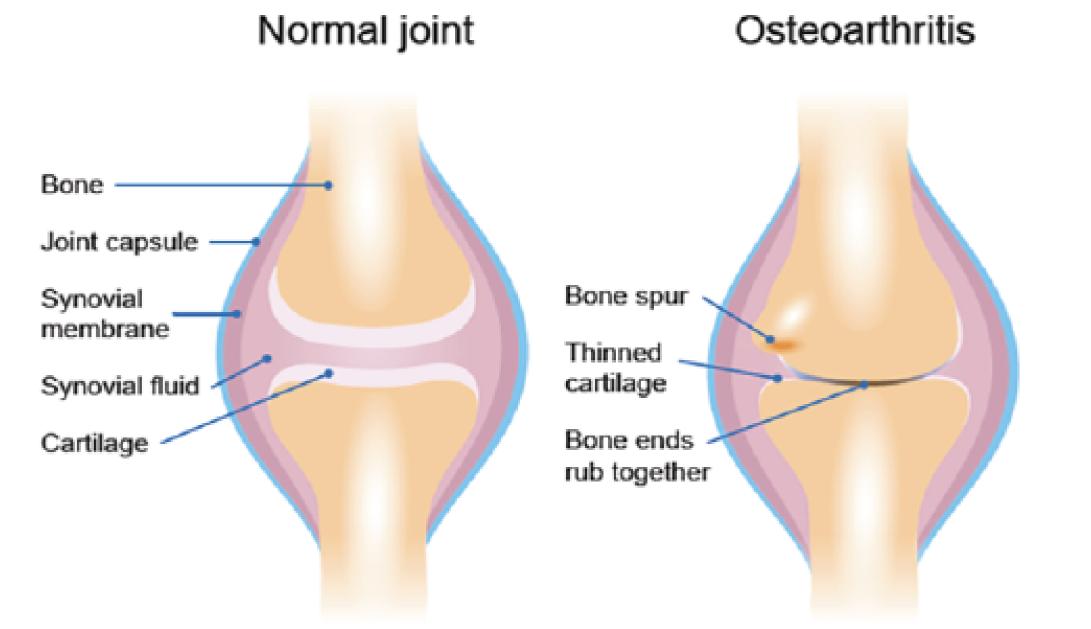


Figure 1.1: A normal joint vs OA affected knee joint

Source: Adapted from [9].

1.1.2 Challenges in Early Diagnosis and Role of Medical Imaging

Early diagnosis of knee osteoarthritis (OA) is essential, as it opens a valuable “treatment window” where early interventions can potentially slow down or alter the progression of the disease [10]. Despite its importance, reliable early diagnostic tools are still missing in clinical practice [5]. People who go on to develop full-blown OA may already be in its early stages years before receiving a formal diagnosis. To identify such individuals, researchers have explored multivariable prediction models that connect current risk factors with the future onset of OA [11, 12]. A few early prediction models have been proposed, combining clinical signs with imaging features [13, 14] or biological markers from lab tests [15].

One of the challenges is that there is no clear standard way for diagnosing knee OA in clinical settings (unlike the classification criteria used in research studies). Because of this, models have been developed using different definitions of OA outcomes, such as diagnosis based on American College

of Rheumatology (ACR) criteria, persistent knee pain, or radiographic evidence of OA [13, 16]. A promising approach to bridge the gap between research and real-world diagnosis is to use a clinical expert consensus, as done in previous studies [17]. A major difficulty in early-stage OA research is figuring out who actually qualifies as having early-stage OA. Definitions vary widely in the literature. For example, using Kellgren–Lawrence (KL) grades less than 2 [15, 18], $KL < 2$ combined with symptoms, or identifying first-time patients in primary care [19].

Currently, no single definition is widely agreed upon [16]. It's also important to differentiate between diagnosis, which happens in clinical practice, and classification, which is used in research to sort patients into study groups [17]. The symptoms in early-stage OA tend to come and go, making both diagnosis and classification based on symptoms alone difficult [14, 16]. This presents an opportunity for imaging to support clinical diagnosis, especially when imaging tools are accessible, fast, and affordable.

Due to the slow progression, it can take years for early-stage OA to develop into its more established form. This delay makes it challenging to evaluate treatment effectiveness or disease trajectory in a timely way. To address this, researchers use short-term measurable surrogate outcomes indicators that act as stand-ins for long-term outcomes. For example, subtle changes in joint structure or early symptom progression, observable through imaging, can suggest how the disease may evolve [5]. These imaging changes are often referred to as imaging biomarkers, which are objective features visible through MRI or X-ray that reflect biological processes or responses to treatment. Like other biomarkers, they need to be thoroughly validated before being used in clinical or research settings. The BIPED classification system which is burden of disease, investigative, prognostic, efficacy, and diagnostic, provides a structured way to assess and categorize them [5]. This thesis explores how deep learning can help with the detection and validation of such imaging biomarkers, potentially improving early OA diagnosis.

Magnetic Resonance Imaging (MRI) plays a key role in assessing cartilage and other soft tissues affected by OA. However, manually segmenting MRI images is time-consuming and prone to variability. In recent years, deep learning techniques have shown great potential in automating this process, with models like U-Net and its 3D variants delivering promising results in brain and musculoskeletal imaging [5, 15]. Despite advancements, there remains a significant gap in applying high-performance 3D segmentation models, such as CAN3D, to knee OA datasets, particularly in post-operative Total Knee Replacement (TKR) patients from the OAI dataset. Previous studies have focused largely on healthy patient datasets or brain imaging tasks [5].

1.2 Problem Statement

AI-driven cartilage segmentation for osteoarthritis still faces several challenges. Automated systems often lack the reliability and precision demonstrated by expert radiologists. Successfully integrating AI into clinical workflows requires rigorous validation across diverse patient populations and the ability to adapt to varying imaging conditions. To bridge these gaps, various models like CAN3D, U-Net, and

KAN3D have been developed. While U-Net excels at pixel-level classification in biological images, models like CAN3D have proven effective for brain MRI segmentation using datasets such as OASIS.

However, their application and generalisation to other medical domains particularly knee MRI segmentation in osteoarthritis using datasets like OAI-TKR—have been largely unexplored. Notably, no prior studies have transferred and adapted CAN3D from brain MRI to segment MRIs from the OAI Total Knee Replacement (TKR) dataset, which includes patients with advanced osteoarthritis. Most existing research has focused on healthy patient data, with minimal work addressing segmentation in post-operative or severely degenerated knees.

This study builds upon the work by Wei [20], who introduced CAN3D which is a compact convolutional neural network designed for 3D medical image segmentation on standard clinical workstations. While their model demonstrated strong performance on pelvic MRIs for prostate cancer, it has not yet been applied to the knee data. This study addresses that gap by being the first to adapt CAN3D, originally trained on the OASIS brain dataset, to the OAI Total Knee Replacement (TKR) dataset. This cohort includes patients with advanced osteoarthritis, presenting additional segmentation challenges such as joint deformities, prosthetic implants, and class imbalance. The adaptation process involved generating new test labels and enabled cross-condition comparisons with previously analyzed healthy knee data. Through this work, we aim to evaluate CAN3D’s generalisability and adaptability across diverse clinical datasets and anatomically challenging cases.

1.3 Objectives and Research Questions

The primary objective of this study is to adapt and evaluate the CAN3D model, originally developed for segmenting pelvic and prostate MRIs, for use in analyzing knee MRIs from patients with advanced osteoarthritis. The model was previously adapted for brain MRI segmentation using the OASIS dataset, which formed the foundation for this work. In this study, we further adapt CAN3D to the OAI Total Knee Replacement (TKR) dataset, which presents unique challenges such as joint deformities, prosthetic implants, and class imbalances. Our goal is to assess the model’s ability to generalize across different anatomical regions and clinical conditions, and to determine its effectiveness in accurately segmenting severely degenerated knee structures for improved OA monitoring and post-surgical analysis

The specific objectives of this research are:

- To get an understanding between different deep learning models, like UNET, CAN3D and KAN3D
- To start by training and evaluating the CAN3D model on the OASIS brain MRI dataset, establishing a reliable baseline for segmentation performance.
- To retrain the model on the OAI-ZIB knee MRI dataset (healthy cases), and evaluate its performance using metrics such as loss, Dice Similarity Coefficient (DSC), and surface distance.

- To further test the model on the OAI-TKR dataset, which includes post-operative cases with advanced OA.
- To perform both qualitative (using ITK-SNAP) and quantitative evaluations by comparing segmentation outputs for TKR patients with those from healthy OAI patients.
- To assess the adaptability and performance of CAN3D when applied across different anatomical sites and disease stages.

With these considerations in mind, this study aims to address the following research questions :

- **Research Question 1 (RQ1):** How well does the CAN3D model, originally trained on brain MRIs, generalise to knee MRI segmentation tasks?
- **Research Question 2 (RQ2):** Can the CAN3D architecture be effectively retrained on the OAI-ZIB dataset to produce accurate segmentation of healthy knee structures?
- **Research Question 3 (RQ3):** How does the model perform on complex cases from the OAI-TKR dataset, which includes post-operative knees with severe osteoarthritis and implants?
- **Research Question 4 (RQ4):** What are the key differences in segmentation performance (qualitative and quantitative) between healthy and TKR datasets?
- **Research Question 5 (RQ5):** Does the compact and efficient design of CAN3D maintain high accuracy and reliability across different anatomical domains and disease severities?

1.4 Significance of the Study

This study addresses a crucial gap in the use of artificial intelligence for knee osteoarthritis (OA) analysis specifically in the context of post-operative Total Knee Replacement (TKR) patients. While models like CAN3D have already proven effective in brain and pelvic MRI segmentation, their application to knee MRIs, especially in complex clinical cases, has not been thoroughly explored. By adapting CAN3D originally trained on the OASIS brain dataset to the OAI-TKR dataset, this research examines how well such models can generalise across very different anatomical regions and disease conditions.

Importantly, this is the first known attempt to apply CAN3D to knee OA segmentation. In doing so, it not only evaluates the model's strengths and limitations but also contributes new test labels for the TKR dataset, adding valuable resources for future research. By comparing the model's performance across both healthy and post-surgical knee MRIs, the study aims to offer insights and direction that may help with better clinical monitoring of OA progression and surgical outcomes. In the long term, the findings can help advance the development of compact, efficient AI tools that work reliably even on standard clinical hardware, which would help to bridge the gap between research and clinical practice.

Chapter 2

Literature Review and Background

This section aims to discuss about how over the time OA worldwide, the drawbacks of conventional diagnosis and treatment approaches, and their limits. Following that, we will discuss how more recent AI-driven methods have transformed diagnosis, especially in the area of knee cartilage segmentation. we further talk about how this thesis aims on ..

2.1 The Global Burden of Knee Osteoarthritis: Challenges in Diagnosis and the Need for Novel Approaches

Osteoarthritis (OA) is one of the most prevalent causes of chronic disability globally, with knee OA being particularly widespread. In the United States alone, OA affects approximately 10% of men and 13% of women over the age of 60 [6]. As of 2020, it is estimated that 595 million people worldwide were living with osteoarthritis, which constitutes about 7.6% of the global population. This marks a 132% increase from 1990, when the number was approximately 256 million [21].

The incidence of OA is rising, as highlighted by the Global Burden of Disease 2019 study, particularly in high socio-demographic index (SDI) countries like Australia, and also in China. Future projections for 2030 suggest that the age-standardized incidence might decrease among women but increase among men [22]. Projections for 2050 indicate that the number of individuals affected by osteoarthritis could approach 1 billion, driven by factors such as aging populations, increasing obesity rates, and longer life expectancies [21].

A significant proportion of new OA cases are early-onset (before age 55), which account for over half of all incident cases and contribute substantially to years lived with disability (YLDs). Specifically, early-onset OA contributes to approximately 26.1% of all YLDs due to OA. Between 1990 and 2019, the incidence, prevalence, and YLDs attributable to early-onset OA doubled, particularly in low- and middle-SDI countries [22].

In 2019, high body mass index (BMI) contributed to 15.9% of YLDs caused by early-onset OA, with the highest burden observed in middle-SDI countries. Over a 30-year period, the contribution

of high BMI to YLDs of early-onset OA has risen by approximately 329%. The economic cost of early-onset OA is estimated at over USD 106.87 billion, with nearly 60% of this cost attributed to indirect productivity losses [22].

Despite its growing burden, OA diagnosis is often delayed and inconsistent. In most regions, especially low-resource regions, over reliance on subjective clinical judgment and limited access to advanced imaging equipment hinder timely and accurate diagnosis. This creates missed opportunities for early intervention [10].

Most new cases occur within the working years of individuals, the effect of OA on labor activity is a critical concern. A systematic literature review by Ching et al. found that OA was a reason for absenteeism (1.4% to 14%) as well as decreased work productivity (presenteeism) most commonly due to joint pain. Physical work demands were linked with higher rates of absenteeism, presenteeism, and early job loss. Moderate-severe joint pain and pain interference were similarly attributed to work changes and early leavers from work, while comorbidities and inadequate support by coworkers worsened these findings [23].

These diagnostic challenges, as well as the increase in the burden and expense of the disease, do showcase the urgent need for new, scalable, and non-bias diagnostic approaches to improve early diagnosis and reduce the increased social and economic burden of OA.

2.2 Conventional Techniques: Segmentation and Ground Truth Methods

2.2.1 Radiography

Radiography is the first-line imaging modality recommended by the European Alliance of Associations for Rheumatology (EULAR) and American College of Radiology (ACR) for the diagnosis of knee osteoarthritis (OA) [2]. It is the most common tool due to its high specificity, cost-effectiveness, and universal availability. Radiographs are typically acquired in two standard planes, lateral and anterior-posterior, which may include weight-bearing or additional views like the patella or Rosenberg view, depending on clinical needs [2]. In clinical settings, radiography plays a central role in assessing joint structure. It allows clinicians to detect characteristic OA features such as joint space narrowing (JSN), osteophyte formation, subchondral sclerosis, and bone cysts [24]. These structural changes are used to determine disease severity and monitor its progression over time. For instance, JSN is a key indicator of cartilage loss, and the presence of bony changes often suggests more advanced stages of the disease [24]. Radiographs also form the basis for widely adopted classification systems such as the Kellgren–Lawrence (KL) grading scale, which assigns OA severity based on visible structural changes [25].

Although useful for diagnosis, radiography has some serious drawbacks. It is not capable of directly imaging soft tissues like cartilage, menisci, or synovium and hence is less sensitive to early degenerative alterations. Further, secondary signs such as JSN or osteophytes appear late in the

disease process, thus compromising its ability to detect OA in its early stages. It can also occur with inconsistency between symptom and radiographic findings in patients, partially due to variation of OA phenotypes and disease pathophysiology [2]. As a result, radiography is often complemented by advanced imaging modalities like MRI when detailed tissue-level assessment is needed [2].

2.2.2 Kellgren-Lawrence Grading System

Radiographs form the basis for the interpretation of the severity of the disease in knee OA, and clinicians use standard classification systems to classify the severity of joint degeneration. For that, the Kellgren–Lawrence (KL) grading scale is the most commonly used tool in clinical and research practice. Introduced by Kellgren and Lawrence in 1957, it divides OA severity into five categories based on radiographs Figure 2.1 [19]. Grade 0 is no OA-specific changes in the joint, Grade I is described as doubtful OA-changes, and Grades II and III are slight and moderate changes, respectively, distinguished by the absence or presence of subchondral sclerosis. Grade IV considers advanced OA stages with joint deformity and very small joint space width (JSW) [19].



Figure 2.1: (a) Grade 0 – Healthy knee (b) Grade I – Minor signs of OA (c) Grade II – Early wear and tear with clearly reduced space between bones. (d) Grade III – Moderate damage with bones getting closer and the surface underneath harden (e) Grade IV – Severe damage with bones touching and clear joint deformities.

Source: Adapted from [19].

While widely used, this approach is marred by the fact that it is completely dependent on the physician's experience and knowledge during the grading process, which could result in misclassification [19]. Besides, because of the slight variation in radiographs, it might be particularly challenging to distinguish between early OA (grades 0–1) [19].

2.2.3 Methodologies for Knee Cartilage Segmentation

Over time, there have been three main methodologies that can be used to diagnose knee cartilage segmentation :

Interactive Methods

Interactive methods involve manually drawing regions of interest (ROI) and subsequently enhancing them using edge detection techniques. Although these methods are time-consuming, they have historically served as benchmarks for more sophisticated approaches in image segmentation [26].

Semiautomatic Techniques

Semiautomatic techniques include methods such as the active contour model, which starts with a manually drawn initial shape and improves it through automated iterations. Additionally, Bayesian probability estimation-based pixel categorisation methods have laid the groundwork for current deep learning models [27].

Automatic Methods

The introduction of Convolutional Neural Networks (CNNs) led a significant breakthrough in image analysis. Early CNNs, however, faced limitations in segmentation accuracy due to their tendency to lose spatial data during convolution and pooling operations. This challenge led to the development of Fully Convolutional Networks (FCNs), which can handle images of any size and eliminate fully connected layers. Despite these advancements, issues such as spatial resolution loss persisted, necessitating further improvements [27].

2.3 MRI Introduction

The introduction of Magnetic Resonance Imaging (MRI) brought a major shift in how knee osteoarthritis (OA) is assessed and understood. While traditional X-rays focus mainly on bone and are useful for identifying joint space narrowing in later stages, MRI provides detailed images of soft tissues. This includes early changes in cartilage, menisci, the C-shaped cartilage pads that act as shock absorbers between the thigh and shin bones synovium, the thin membrane lining the joint capsule that produces lubricating fluid and subchondral bone, the layer of bone just beneath the cartilage that supports and distributes joint load which can appear well before any clear signs are visible on radiographs [28, 29].

The first use of MRI for morphometric analysis of articular cartilage, published in 1994, set the foundation for a new direction in OA research [30]. MRI allowed, for the first time, accurate and direct visualisation of cartilage and other joint structures *in vivo*. This meant that researchers could now quantitatively measure structural changes over time, challenging earlier views that primarily associated OA with cartilage thinning and bone-on-bone friction. As a result, MRI has contributed to a paradigm shift in the understanding of OA as a disease affecting the entire joint [30].

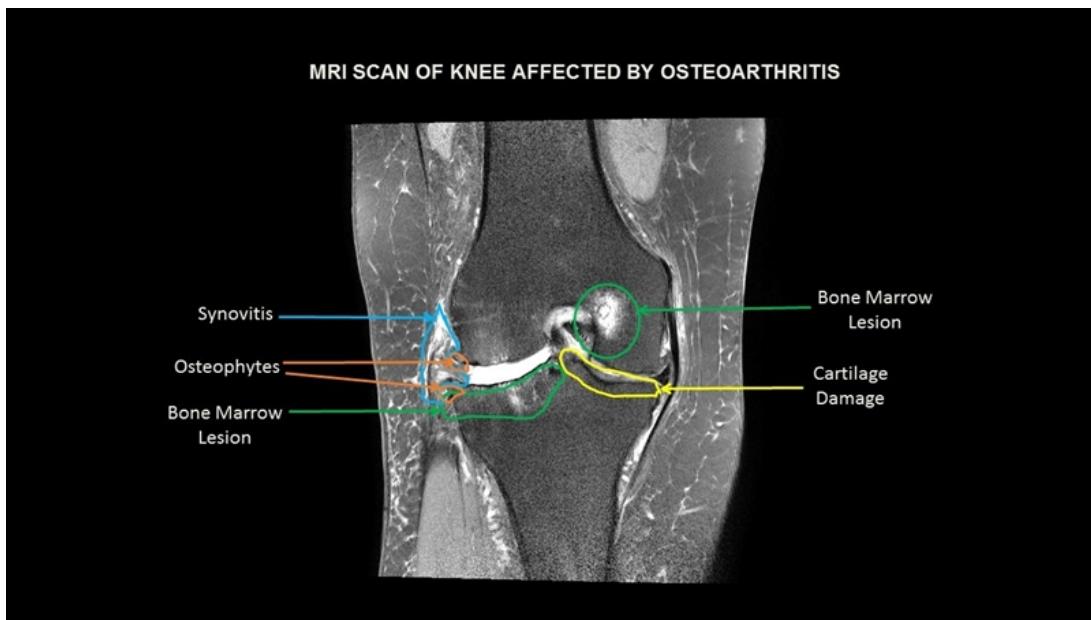


Figure 2.2: MRI scan of the knee joint showing cartilage and bone structures relevant to osteoarthritis

Source: Adapted from [31].

Modern MRI techniques, such as T2 and T_{1ρ} (T1-rho) mapping, go beyond structural imaging to assess the biochemical composition of cartilage. T2 mapping reflects the water content and collagen fiber orientation within cartilage, making it useful for identifying early changes in tissue structure before visible damage occurs. T_{1ρ} mapping, on the other hand, is sensitive to proteoglycan content—an important component of cartilage that depletes during the early stages of OA. Together, these techniques help detect subtle cartilage degeneration and are valuable tools for early diagnosis and monitoring of disease progression [28, 29].

Typical MRI findings in OA include several structural and biochemical abnormalities. Cartilage defects involve thinning, softening, or focal loss of the cartilage layer that normally cushions the joint. Bone marrow lesions (BMLs) are regions of altered signal intensity within the subchondral bone, reflecting inflammation and bone remodeling; when visible on both PD-weighted and T1-weighted images, they are often linked to more advanced stages of OA [29]. Subchondral cysts are fluid-filled sacs that form beneath the cartilage surface, often due to mechanical stress or cartilage breakdown. Osteophytes, or bone spurs, are bony outgrowths that develop at joint margins in response to instability or cartilage loss. Damage to ligaments and menisci, which are critical for joint stability and shock absorption, may also be detected on MRI and often contributes to pain and joint dysfunction. Additional findings such as joint effusion (fluid buildup), synovitis (inflammation of the joint lining), periarticular cysts, and loose intra-articular bodies are also common and may help explain a patient's symptoms even when radiographic findings are minimal [28, 29].

However, despite its strengths, MRI has its downsides. It is generally more expensive and less accessible than X-rays, and sometimes the MRI findings do not align well with a patient's symptoms. Moreover, there is still ongoing research to determine how well MRI results can predict the future course of OA [29].

Even though MRI provides detailed images critical for understanding OA, recent advances in

artificial intelligence, particularly deep learning, have significantly improved the acquisition and analysis of MRI data, leading to more accurate and efficient OA assessment [29].

2.3.1 Cartilage Volume as a Biomarker in OA

Cartilage volume is used as an imaging biomarker in OA research, mainly because of its direct association with structural degeneration [32]. A lot of studies have quantified changes in cartilage volume over time to monitor disease progression. But, the reliability of cartilage volume as a diagnostic indicator remains a subject of debate. Some research has shown that tibial cartilage volume, while easy to segment and measure, often exhibits significant overlap between healthy and OA-affected subjects, particularly in early-stage cases [30]. In contrast, volume loss in the femoral cartilage, did demonstrate a better correlations with OA severity [33]. This distinction suggests that not all volume-based measures are equally informative and points to the need for an evaluation [32]. Later chapters of this thesis revisit this issue in detail, using experimental findings to compare tibial and femoral volume metrics in real-world datasets.

2.4 The Advent of Deep Learning

Even after the coming of MRI for OA diagnosis, manual interpretation is required to interpret those results which is time-consuming and subjective. Deep learning (DL) methods have shown promise in automating this process, improving accuracy and workflow efficiency. These advances span the entire imaging pipeline, from acquisition to segmentation [3].

2.4.1 Deep Learning in MRI Acquisition

Recent DL techniques have revolutionized MRI acquisition by helping with image reconstruction from under-sampled data, producing high-quality images with shorter scan times. This improvement is vital, as segmentation accuracy depends on input image quality. High-quality MRI images are foundational for reliable AI-based cartilage segmentation. Subtle features such as cartilage thinning, lesions, or bone marrow lesions are difficult to detect in low-resolution or noisy images. By enhancing image clarity and spatial resolution through DL-based reconstruction, the diagnostic accuracy of downstream segmentation algorithms is significantly improved [3, 34]. Apart from that, faster scan protocols by DL reduce patient discomfort and motion artifacts, both of which commonly degrade image quality. This leads to more consistent datasets, which is crucial when training for the deep learning models for segmentation or risk prediction. Therefore, over time, the innovations and up-gradations in DL for image reconstruction are not a separate thread, rather they directly help with the effectiveness, reliability, and scalability of AI-driven OA assessment pipelines. As segmentation methods are highly sensitive to the quality of the imaging input, this improvement plays a pivotal role in supporting precise and clinically useful outcomes. With high-quality input data secured, deep learning can then be applied

downstream to automate cartilage segmentation, which is the next crucial step to detect for OA in a patient [3, 34].

2.4.2 Deep Learning for Automated Cartilage Segmentation

Building on the MRI data quality, DL segmentation algorithms automate cartilage lesion detection and quantification, which helps in improving diagnostic objectivity and disease tracking [35].

Artificial intelligence (AI), machine learning (ML), and deep learning have revolutionised [35] OA diagnosis by providing automated tools that reduce subjectivity and increase accuracy. By identifying cartilage lesions and ligament rips with MRI, experimental algorithms currently make it easier to automatically assess the severity of osteoarthritis (OA) from radiographs [36]. These AI-powered segmentation methods improve diagnostic accuracy while enabling ongoing tracking of the course of the illness and the effectiveness of treatment. The way OA is diagnosed and treated may change if these techniques are incorporated into clinical practice. Better patient outcomes are possible with automated procedures because they give more dependable assessments and lessen the effort for physicians [2].

To discuss further we have 2 case studies to talk about:

Liu et al.'s Approach

A technique for identifying cartilage lesions in knee MRIs was created by Liu et al. [35]. Their method involved two different two-dimensional (2D) convolutional neural networks (CNNs) in two steps. 2D CNNs are commonly used in medical imaging for analyzing slice-based data; they apply filters across two spatial dimensions to automatically detect features such as edges, textures, or anatomical structures. The first CNN was devoted to cartilage segmentation, whereas the second CNN concentrated on lesion identification [35].

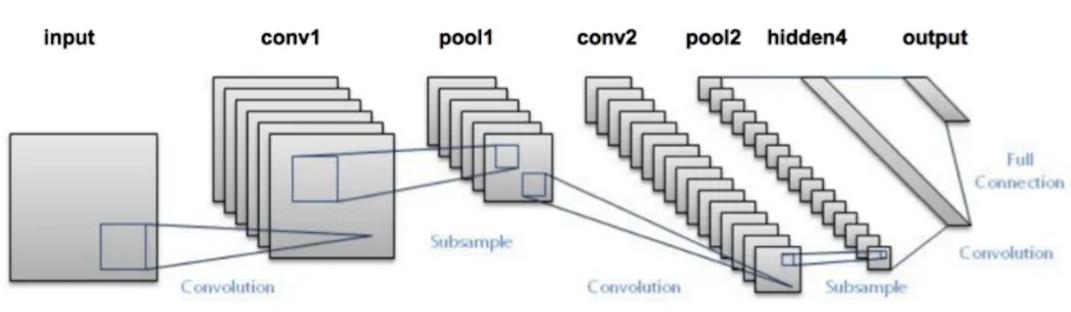


Figure 2.3: Simple architecture of 2D-CNN

Source: Adapted from [37].

This technique outperformed radiologists in cartilage lesion detection, with sensitivity rates of 84.1% and 80.5% when evaluated on 175 knee MRIs [38]. In comparison, radiologists' sensitivity ranged between 60.8% and 80.2%. However, the automated system's specificity was lower (85.2% and 87.9%) compared to the radiologists' specificity of 92.2% to 96.5% [39].

The study also reported that the automated system had superior intra-observer agreement ($\kappa = 0.76$) compared to inter-observer agreement among radiologists ($\kappa = 0.57\text{--}0.73$). The lower specificity of

the automated system was attributed to its reliance on a single 2D sagittal fast spin-echo sequence (fat-saturated T2), while radiologists used three sagittal sequences [38]. This study underscores the capability of DL systems to match and even exceed radiologists in cartilage lesion detection, though trade-offs in specificity must be addressed for clinical deployment [35].

Pedoia et al.'s Approach

Pedoia et al. [36] developed a deep learning system to identify patellar cartilage abnormalities as well as meniscal injuries. Their algorithm utilised a dataset of 1,478 MRIs from 302 patients annotated by radiologists. The dataset included patients with and without OA, as well as individuals with anterior cruciate ligament (ACL) injuries and those post-ACL reconstruction [36].

The modified Whole-Organ Magnetic Resonance Imaging Score (WORMS) [38] was used to grade the meniscus and patello femoral cartilage. WORMS is a comprehensive MRI-based scoring system that evaluates multiple joint structures to assess the severity of osteoarthritis, providing a detailed picture of joint health. Their system used a two-step approach: a three-dimensional (3D) CNN for lesion identification, followed by a 2D U-Net architecture for automatic segmentation. When compared to radiologist annotations as the gold standard, the system achieved a binary cartilage lesion detection sensitivity of 80.0% and specificity of 80.27%. Inter-rater variability was measured by having three radiologists evaluate a subset of 17 MRIs. The average agreement was 89.56% in the absence of cartilage lesions and 79.74% when lesions were present [38]. These findings mark a significant step forward in the accurate and efficient evaluation of cartilage health.

This study demonstrates the potential of deep learning to detect and segment cartilage and meniscal abnormalities with accuracy comparable to expert radiologists, highlighting its promise for a better, large-scale OA assessment in diverse patient populations. Although MRI is widely used for OA diagnosis, interpreting the scans remains a time-consuming and subjective task. Deep learning (DL) offers a solution by automating the analysis pipeline—from acquisition to segmentation, helping with both diagnostic accuracy and workflow efficiency [36].

2.5 Recent Technological Advancements in Segmentation Models

There have been a lot of advancements in segmentation models for medical imaging, particularly for knee cartilage evaluation, over the period of time. This section is a review of major deep learning models, their applications, performance and their contribution towards improving segmentation accuracy.

2.5.1 UNET

U-Net is a convolutional neural network architecture mainly designed for biomedical image segmentation. It was introduced by Ronneberger et al. in 2015, and it follows an encoder–decoder structure where the encoder captures contextual information by successive convolution and pooling layers,

and the decoder enables precise localisation using upsampling layers which are combined with skip connections. These skip connections between the encoder and decoder layers helps to preserve spatial information, which is important for accurate segmentation. U-Net performs well even with limited annotated data and has become the backbone of many medical image segmentation tasks, including cartilage segmentation in knee MRIs, due to its ability to effectively handle fine structures and boundaries [40].

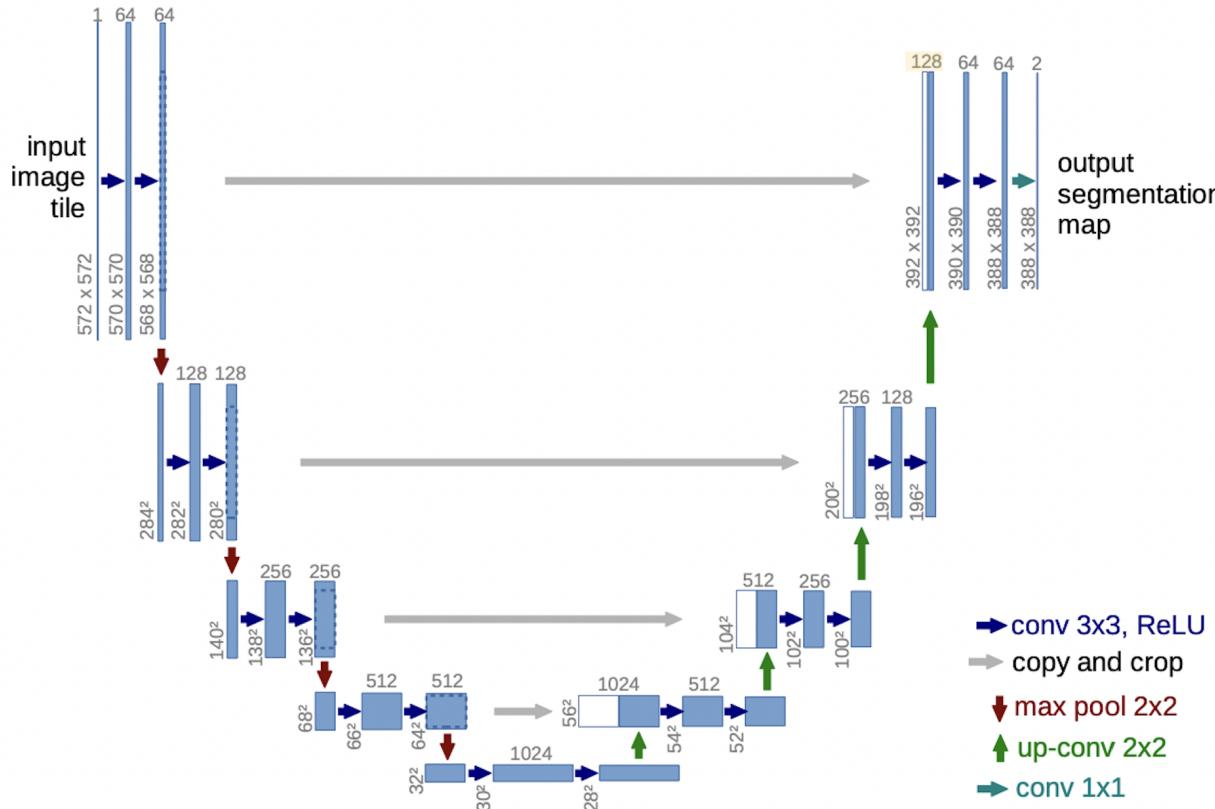


Figure 2.4: Model architecture for UNET

Source: Adapted from [40].

A lot of recent works have used 3D variants of U-Net to address anomaly detection and segmentation in knee MRI scans. For example, anomaly-aware segmentation studies use 3D U-Nets with deep supervision to segment bones and cartilage accurately, guided by error maps highlighting pathological regions [41]. All these approaches demonstrate the model's capability to learn anatomical structures and detect abnormalities directly from raw MRI data. CAN3D, which is discussed in the very next section, builds upon this foundation by improving contextual attention and 3D convolutional feature encoding, resulting in better segmentation performance and faster inference [40].

2.5.2 CAN3D

CAN3D is also a 3D convolutional neural network designed specifically for the semantic segmentation of high-resolution volumetric MRI images. It addresses common challenges in medical imaging such as the class imbalance, limited computational resources, and the need for both fine-grained

detail and global context model through an architecture that could focus on efficiency and as well as scalability [41].

Even though U-Net have proven very useful in medical image segmentation, particularly due to it's encoder-decoder structure and skip connections, CAN3D builds upon this foundation with a more context-focused architecture. Unlike U-Net, which relies on symmetric downsampling and upsampling paths, CAN3D does not adopt the traditional encoder-decoder structure in favor of a more direct and efficient pipeline based on dilated convolutions. At its core, it integrates a 3D Context Aggregation Module (CAM), which allows the model to capture both local features and broader contextual information without requiring large model sizes or excessive GPU memory. By using fewer layers and smaller input patches, it speeds up training and reduces resource usage. Apart from that the incorporation of dilated convolutions which are basically convolutional filters applied with gaps (or dilations) between the sampled points, helping the network to capture information over a larger area without increasing the number of parameters or computations help expand the receptive field (refers to the region of the input image that influences a specific output neuron) without adding extra parameters, allowing the model to capture broader context an important aspect for accurate segmentation in medical images where spatial structure matters. This maximized receptive field helps the model to understand global structure and spatial relationships, which is crucial for precise medical image segmentation [20, 41].

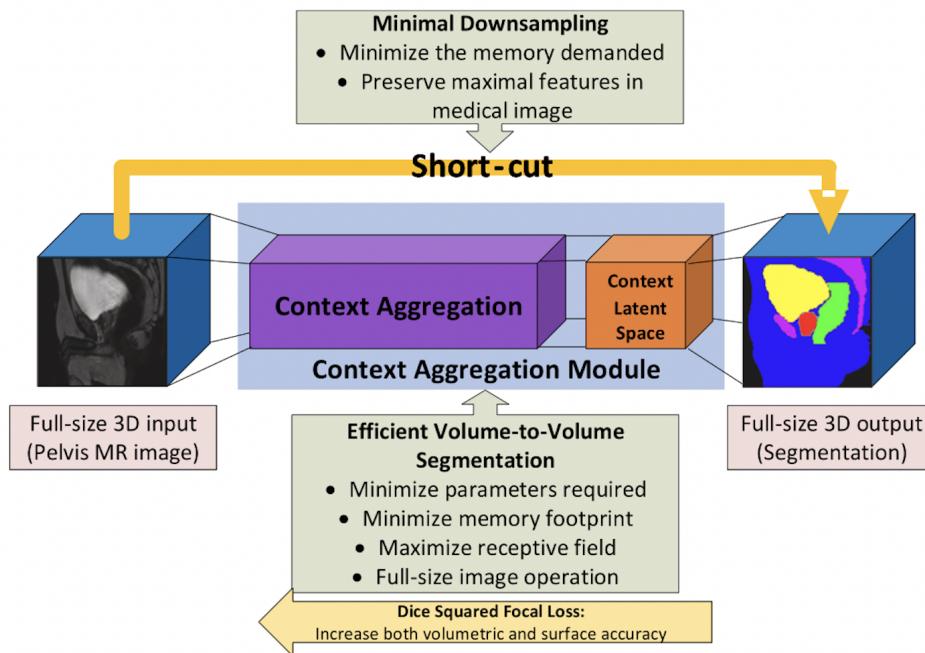


Figure 2.5: Model architecture of Context Aggregation Network 3D (CAN3D)

Source: Adapted from [20].

The model processes large 3D volumes by dividing the input into patches rather than requiring entire full-size volumes at once. This patch-based approach allows the model to operate within limited GPU memory, learn and preserve important spatial context within each patch while keeping computations feasible [41].

The learning process helps to optimise CAN3D and introduces a hybrid loss function known as Dice Squared Focal Loss (DSF), which combines Dice Squared Loss (a modified form of the standard Dice loss) with Focal Loss [20]. This helps in the model’s ability to handle class imbalance and refine predictions around complex structures like cartilage surfaces or soft tissue borders [41].

The model’s compact architecture results in rapid inference, in some cases even producing segmentation outputs within seconds making it suitable for real-world clinical settings where time and resources are limited [20, 41].

2.5.3 KAN3D

KAN3D is a three-dimensional medical image segmentation model that builds upon the Context Aggregation Network (CAN) architecture. The main difference in KAN3D lies in its use of the Kaleidoscope Transform (KT), which acts as a preprocessing and post-processing step to help with the model’s efficiency and accuracy [42]. The Kaleidoscope Transform, originally introduced in the image processing domain, is a method for rearranging image data by downsampling and duplicating the input in a structured manner. Specifically, given a positive integer downsampling factor v and a smear factor σ , the KT generates multiple spatially offset copies of the original image. For the use case in KAN3D, a simplified form of KT with $\sigma = 1$ is used, helping the transformation of a 3D image into a batch of smaller, spatially shifted sub-volumes. For example, a $(v = 4, \sigma = 1)$ transform decomposes the input into $4^3 = 64$ low-resolution variants of the original image, each shifted slightly in space [42].

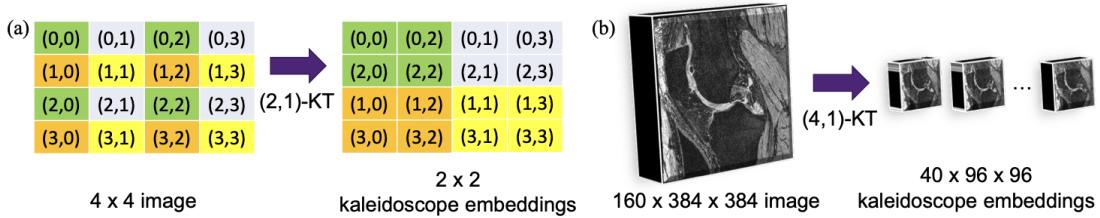


Figure 2.6: (a) Kaleidoscope transform on a 2D image. (b) Kaleidoscope transform on a 3D image.

Source: Adapted from [42].

Instead of concatenating these downsampled versions along spatial axes, KAN3D reshapes them into a batch of 3D images using the `Rearrange` operation from the Einops package [CITE: Rogozhnikov 2022]. This structure allows for efficient parallel processing in 3D convolutional neural networks. After the network processes these 64 sub-volumes, an inverse KT operation is applied to reassemble the outputs back into the original full-resolution image space, thus producing the final segmentation prediction. KAN3D itself is based on a modified 3D implementation of CAN [CITE: Dai et al. 2022], which is known for capturing both local and global context. However, the original CAN does not employ downsampling, which makes it more memory-intensive. To address this, KAN3D introduces two levels of downsampling and upsampling blocks, helping it to operate within GPU memory constraints while retaining the CAN module for context aggregation. Skip connections are

added between the downsampling and upsampling paths to preserve spatial information and stabilize learning [42].

Apart from that, the model has a deep supervision to help with the training stability and improve convergence. This technique involves producing auxiliary segmentation outputs at intermediate layers of the network, which are then up-sampled and merged with the final output. This helps to guide the learning process at multiple levels of abstraction and can also increase convergence [42].

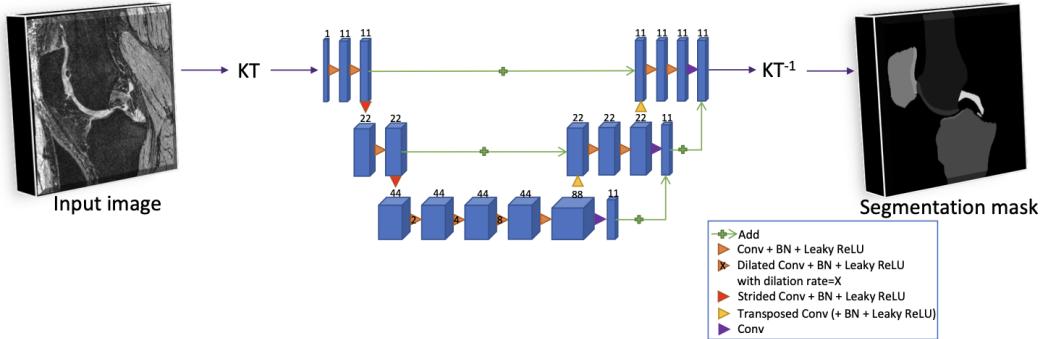


Figure 2.7: The KAN3D model architecture where KT denotes kaleidoscope transform and KT-1 denotes inverse kaleidoscope transform

Source: Adapted from [42].

For optimisation, KAN3D uses the multi-class Dice loss function, which is particularly well-suited for medical image segmentation due to its ability to handle class imbalance effectively [42]. In summary, KAN3D introduces a new combination of spatial rearrangement (via KT), memory-efficient network design, and strong supervision strategies to deliver accurate and efficient segmentation of 3D medical images, particularly suited for knee MRI datasets.

Even though KAN3D was not used in the experimental phase of this thesis, it was explored and studies about during the literature review to understand current trends in 3D medical image segmentation. Reviewing its architecture and performance helped shape the understanding of design choices in modern models, particularly those tackling the challenges of memory efficiency and spatial precision.

Chapter 3

Methodology

This chapter describes the methodology adopted for adapting the CAN3D model to segment knee magnetic resonance imaging (MRI) scans. It covers the model architecture, the experimental setup , the datasets used and train/test split, and the loss functions. Hyperparameter choices, implementation specifics, and evaluation metrics are also outlined.

3.1 Existing Work

To our knowledge, the CAN3D model has primarily been applied to the OASIS dataset for 3D brain MRI segmentation tasks. However, its application particularly to the OAI TKR dataset, has not been explored in prior literature. There is no existing study that evaluates CAN3D’s performance on the OAI TKR subset, which includes MRI scans from patients who have undergone total knee replacement (TKR). Although, the OAI-ZIB dataset does contain segmentation labels for healthy knees, the TKR subset lacks corresponding ground truth annotations. This absence of labels limits the ability to directly train or validate segmentation models on this subset, presenting a practical challenge for model evaluation and comparison.

The original CAN3D implementation did show promising results for segmenting brain structures, but its adaptability to anatomically and visually different domains such as the TKR subset remains unverified. The architectural design, originally tuned for neuro imaging, may or may not generalise well to the TKR cohort. This research aims to bridge that gap by adapting CAN3D for use on the OAI TKR dataset, generating test segmentations, and evaluating its generalisability against labeled non-TKR cases from the same dataset using appropriate metrics such as the Dice Similarity Coefficient.

3.2 Model Architecture

The model used in this study is a compact version of the Context Aggregate Network for 3D data, referred to as CAN3D, originally proposed by Wei Dai et al. [20]. Unlike traditional encoder-decoder models that rely on multiple downsampling and upsampling steps, CAN3D follows a simplified

architecture with only a single downsampling and upsampling stage. It is designed to efficiently process volumetric MRI data by extracting both fine-grained and global contextual features. Following is the model architecture [20] :

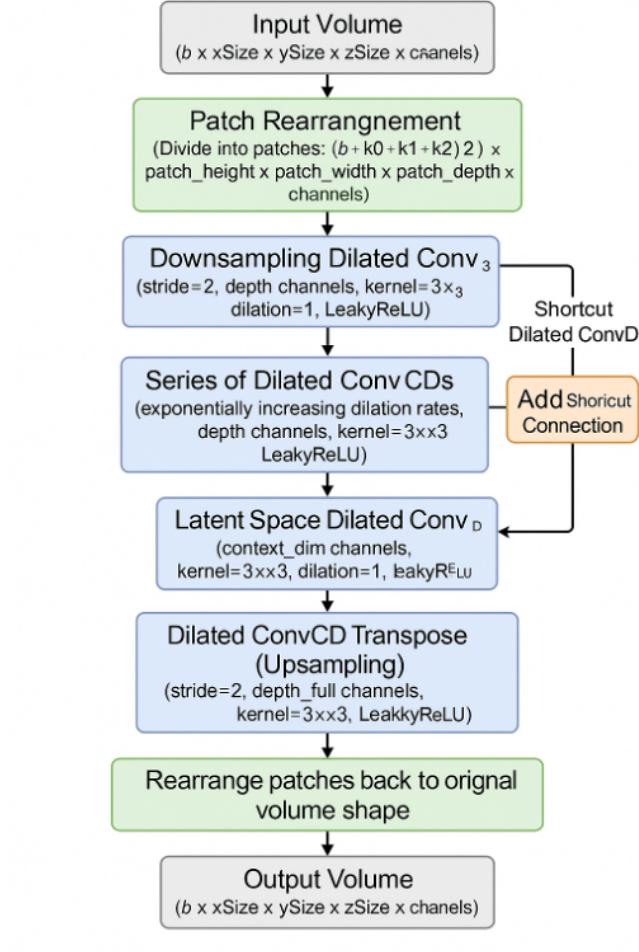


Figure 3.1: Overview of the CAN3D-based segmentation pipeline from input MRI volume to segmented output

- **Patch-based Input Representation:** To process high-resolution 3D MRI volumes without exceeding memory constraints, the input image is first divided into smaller 3D patches. This reshaping ensures localised processing while preserving anatomical detail, allowing the network to scale to large volumetric data.
- **Initial Feature Extraction:** Each patch is passed through a 3D convolutional layer with a dilation rate of 1. This layer captures low-level features such as edges and textures. A LeakyReLU activation function is used to prevent neuron inactivation and improve gradient flow in early stages.
- **Downsampling via Strided Convolution:** Instead of conventional pooling, spatial downsampling is achieved through a strided 3D convolution with dilation. This allows the model to reduce spatial dimensions while still learning robust features, avoiding the information loss commonly associated with pooling.

- **Context Aggregation Module (CAM):** The core of the CAN3D model is the CAM, a sequence of dilated 3D convolutional layers with increasing dilation rates (e.g., 2, 4, 8, 1). These layers exponentially expand the receptive field, allowing the model to integrate both local and global contextual information across the entire volume without sacrificing resolution.
- **Adaptive Normalization and Activation:** Each CAM layer uses Adaptive Instance Normalization (AdaIN), which combines identity mapping with instance normalization, allowing the network to balance normalization strength dynamically during training. This is followed by LeakyReLU to enhance non-linearity.
- **Latent Context Space:** After context aggregation, the feature maps are passed to a latent space layer with expanded channels. This acts as a context-rich bottleneck, summarising both fine structural details and long-range spatial dependencies for accurate segmentation.

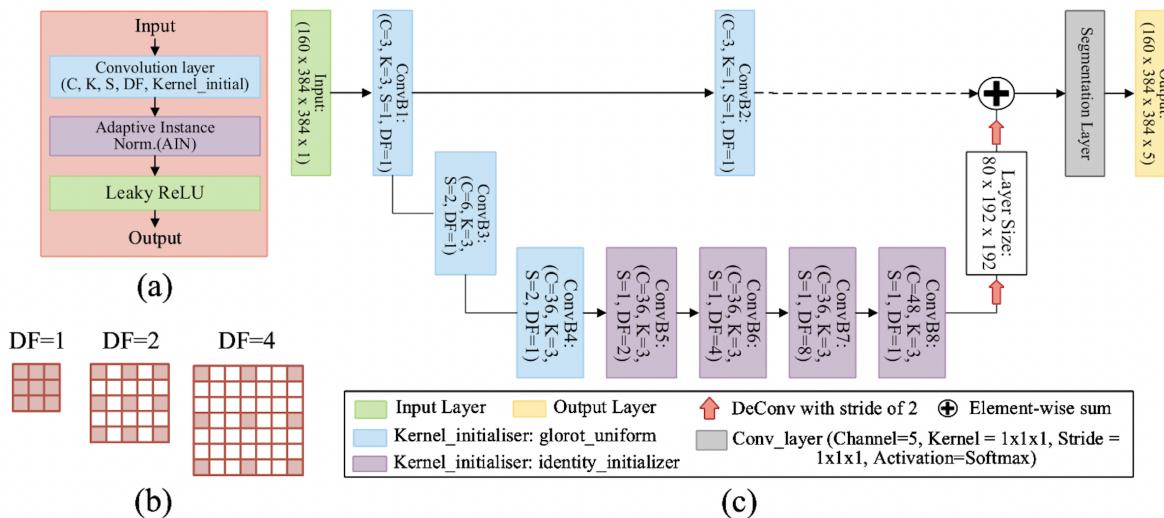


Figure 3.2: (a) A convolution block. (b) Convolution filter with different dilation rates. (c) Detailed architecture of CAN-3D model.

Source: Adapted from [20].

- **Upsampling with Skip Connections:** The latent representation is upsampled using a 3D transposed convolution (deconvolution). A shortcut connection from earlier layers (using a $1 \times 1 \times 1$ convolution) is added to restore high-frequency spatial details, improving segmentation accuracy and network stability.
- **Output Layer and Patch Recombination:** A final $1 \times 1 \times 1$ convolution with Softmax activation is used to generate voxel-wise segmentation probabilities. These are then rearranged to reconstruct the full 3D volume from the segmented patches.

This architecture allows efficient multi-scale feature learning with minimal parameter overhead and is well-suited for clinical use due to its memory efficiency, accurate boundary preservation, and real-time inference capability.

3.3 Experiments

The CAN3D model was evaluated on three datasets: the OASIS brain MRI dataset, the OAI-ZIB knee dataset, and OAI-TKR which is a specific subset of the OAI dataset. The experimental setup and evaluation strategy are described below.

3.3.1 Experimental Design and Workflow

3.3.2 Datasets Used

This study employs three distinct MRI datasets for training, validation, and evaluation:

- **OASIS (Open Access Series of Imaging Studies)** [43]: A brain MRI dataset containing T1-weighted structural images of individuals ranging from healthy adults to those with dementia. The images, acquired with Siemens 1.5T scanners at 1 mm^3 resolution, were preprocessed (e.g., skull stripping, bias correction, registration). Ground truth segmentations include gray matter, white matter, cerebrospinal fluid, and background. This dataset served as the initial testbed for validating the pipeline and CAN3D architecture.
- **OAI-ZIB (Osteoarthritis Initiative)** [1]: A set of 507 3D knee MRIs with expert manual segmentations. Acquired on Siemens 3T Trio systems at $0.36 \times 0.36 \times 0.7 \text{ mm}$ resolution, each scan includes five segmentation classes: background, femoral bone, tibial bone, femoral cartilage, and tibial cartilage. This dataset was used to train and validate the adapted CAN3D model for knee segmentation.
- **TKR (Total Knee Replacement) Subset of OAI** [44]: This underexplored dataset contains 1,717 3D knee MRIs from patients either scheduled for or having undergone total knee replacement surgery. For evaluation, scans were split into:
 - **TKR-Healthy**: Patients with osteoarthritis but no surgical requirement.
 - **TKR-Unhealthy**: Patients who had undergone or were scheduled for total knee replacement.

This allowed comparative testing across OA severity levels.

3.3.3 Train/Test Splits

Each dataset was used for a specific experimental purpose:

- **OASIS**: Used for preliminary validation, the CAN3D model was trained on 337 volumes and tested on 17 volumes to verify functionality and performance on brain MRI segmentation before it's adaptation to knee.

- **OAI-ZIB:** Used for model training and internal validation. Out of the 507 available scans, 492 were used for training and 15 for validation, following a fixed (non-random) split. This decision was not based on standard protocols like 80-20 or 70-30 splits, but rather on the practical limitation of having a relatively small number of annotated knee MRIs. Since this was the model’s first exposure to knee anatomy, maximising training data was prioritised to improve learning effectiveness.
- **OAI-TKR:** Used for final model evaluation. As the TKR dataset was never seen during training or validation, it served as an external test set to assess the model’s generalisation capability. Each MRI scan in this dataset also had corresponding file which indicated about the patient’s KL grade, indirectly reflecting the severity of OA. Based on this, patients were grouped into TKR_healthy and TKR_unhealthy categories. It is important to note that TKR_healthy does not imply the absence of osteoarthritis; rather, it indicates that the condition was not yet severe enough to get a total knee replacement.

3.3.4 Loss Functions

The following loss functions and related formulas are adapted from [20]. The explanations and mathematical expressions closely follow their original presentation.

Dice Similarity Coefficient Loss

The Dice loss function is derived from the Dice Similarity Coefficient (DSC), which is a statistical measure used to quantify the spatial overlap between two binary masks. This property makes it highly suitable for medical image segmentation tasks such as knee cartilage segmentation, where the region of interest occupies only a small portion of the image. Unlike pixel-wise losses such as Binary Cross-Entropy (BCE), Dice loss directly evaluates the overlap between predicted and ground truth masks, which helps to mitigate class imbalance by emphasising the correct prediction of the cartilage region [20, 45].

The Dice Similarity Coefficient (DSC) quantifies how much overlap exists between predicted segmentation results and the manually labeled ground truth. It is widely used in medical image segmentation to assess how closely the predicted and actual masks align [20]:

$$L_{DSC} = 1 - S_{DSC} = 1 - \frac{2 \sum_{i \in I} P_i Q_i}{\sum_{i \in I} P_i^2 + \sum_{i \in I} Q_i^2} \quad (3.1)$$

where P is the ground truth set, and Q is the estimated segmentation set. P and Q have the same shape with a total number of voxels, I .

Although Dice loss does promote overlap between predicted and ground truth regions, it does not account for the magnitude of voxel-wise differences. As a result, it may ignore intensity based errors during training. Hence, there also is a modified version of Dice loss that does incorporates terms

that preserve both regional agreement and absolute voxel values, improving training stability which is explained in the later part of this section [20].

In this study, DSC loss was important for training segmentation models to accurately delineate knee cartilage by maximising the overlap between predicted and ground truth regions. Its sensitivity to spatial alignment made it especially valuable for evaluating the quality of boundary preservation [20]. As cartilage boundaries are often thin and irregular, DSC helped prioritize regional accuracy, laying the foundation for integrating more loss terms like Dice Squared Focal Loss.

Mean Squared Error

The mean squared error (MSE), also known as l_2 , is an error measure that has been used in many fields, from regression problems to image processing and pattern recognition [46]. MSE provides a lot of advantages, such as being convex, symmetric, and differentiable, which are properties that make it very suitable for optimisation problems, as well as being parameter-free and not computationally expensive [46].

In this study, Mean Squared Error (MSE) loss is utilised alongside overlap-based losses to help with the pixel-wise difference between the predicted segmentation mask and the ground truth mask. Unlike losses that only focus on spatial overlap, MSE helps to penalize absolute differences in predicted mask intensities, promoting more precise pixel-level predictions and stabilising the training process.

For example for a given 2 distinct signals $x = \{x_i \mid i = 1, 2, \dots, N\}$ and $y = \{y_i \mid i = 1, 2, \dots, N\}$, where N is the total number of samples in each signal. In the context of this study, these signals correspond to images specifically, the predicted segmentation mask and the ground truth mask. Each x_i and y_i represents the pixel value at position i .

The Mean Squared Error (MSE) between these two signals (images) is computed as follows [20]:

$$\text{MSE}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (3.2)$$

In this study, MSE was used in the reconstruction tasks to assess how closely the generated images resemble the original knee MRI images. Its ability to penalise large deviations makes it effective for preserving structural fidelity, which is important when dealing with fine anatomical details such as cartilage. By minimising MSE, the model learns to generate high-quality, low-error reconstructions, providing a strong foundation for downstream segmentation performance [20, 46].

Dice Squared Focal Loss

DSL, a modified version of the standard DSC loss, has Mean Squared Error (MSE) directly into its formulation. This helps the loss function by not only maximizing spatial overlap but also minimizing the voxel-wise differences between predictions and ground truth [20, 47].

This loss function is mainly relevant to our work as it improves segmentation quality by addressing both regional and voxel-wise differences in cartilage boundaries. Given the complexity of knee cartilage shapes and their subtle variations, the embedded MSE component in DSL helps stabilize

training and reduce noise. It also makes sure that the model learns not just to match regions but also the finer intensity-level variations, which is important for accurate cartilage delineation. As outlined in prior work [20], the Dice loss can be alternatively formulated using similarity-based definitions [20] :

$$L_{DSC} = 1 - \frac{2 \sum_{i \in I} P_i Q_i}{\sum_{i \in I} P_i^2 + \sum_{i \in I} Q_i^2} + \frac{\sum_{i \in I} (P_i - Q_i)^2}{\sum_{i \in I} P_i^2 + \sum_{i \in I} Q_i^2} \quad (3.3)$$

According to [20], the formulation of L_{DSC} inherently includes the squared difference between the two sets. A weighted multi-channel Dice Similarity Loss (DSL) is presented in their work as [20] :

$$L_{DSL} = \frac{\sum_{k \in K} \left(w_k \cdot \left(\Psi + \frac{\Psi}{\sum_{i \in I} P_{i,k}^2 + \sum_{i \in I} Q_{i,k}^2} \right) \right)}{\sum_{k \in K} w_k} \quad (3.4)$$

where $\Psi = \sum_{i \in I} (P_{i,k} - Q_{i,k})^2$ represents the sum of squared differences across voxels between the predicted output for each class and its corresponding one-hot-encoded ground truth label. This component helps capture pixel-wise intensity variations that are important for accurate segmentation.

The FL is shown as [20] :

$$L_{FL} = \sum_{k \in K} -\alpha_k (1 - p_k)^\gamma \log(p_k) \quad (3.5)$$

where the Focal Loss (FL) adjusts the standard cross-entropy loss to emphasise harder-to-classify examples by reducing the loss contribution from correctly predicted voxels. It introduces a modulating factor $(1 - p_k)^\gamma$, where p_k is the predicted probability for class k , and γ controls the strength of down-weighting. The loss also includes a weighting factor α_k to address class imbalance [20].

3.3.5 Hyperparameters

The model was trained with the following key hyperparameters: a learning rate of 1e-3 using the Adam optimizer, softmax activation for the output layer, and categorical cross-entropy loss with one-hot encoded labels. A context dimension of 96 was chosen to capture relevant spatial dependencies in the latent space. These values were selected based on initial testing and alignment with prior studies in knee segmentation.

Table 3.1: Key Hyperparameters Used in the Model

Hyperparameter	Value	Description and Impact
context_dim	96	Latent/contextual dimension in the model’s bottleneck layer. Higher values allow the model to encode more abstract features but may increase overfitting or computational cost.
depth	64	Subsampled depth of convolution layers. Controls feature map size at lower resolution. Increasing depth helps capture richer features but also increases memory usage.
depth_full	32	Depth at full resolution. It impacts the resolution of finer spatial details. Lower values may lead to loss of small-scale segmentation accuracy.
num_layers	4	Number of convolutional layers in the encoder/decoder. More layers improve capacity but may also increase risk of overfitting.
learning_rate	1×10^{-3}	Learning rate for the Adam optimizer. Controls how fast the model updates. Too high may cause instability; too low may slow convergence.
activation	Softmax	Final layer activation function. Used for multi-class segmentation to output class probabilities per voxel.
categorical	True	Indicates that labels are treated as categorical (multi-class). Required for using categorical loss functions like soft Dice or categorical cross-entropy.
optimizer	Adam	Adaptive optimizer used for training. Adjusts learning rate dynamically for each parameter to improve convergence.

3.3.6 Implementation Details

To evaluate the adaptability and performance of the CAN3D model on TKR data, a multi-stage experimental pipeline was followed:

- **Phase 1 – Initial Functionality Check (OASIS):** The original CAN3D model was first trained and tested on the OASIS dataset to validate its basic functionality before adapting it to a completely different anatomical structure. The OASIS dataset consists of 337 MRI volumes, but only 17 of them come with annotated segmentation labels. This limited the scope of quantitative evaluation, but the primary goal of this phase was to make sure that the model could run and generate reasonably accurate outputs on a known modality, brain MRIs. This step was important to understand and build confidence in the model’s implementation and pipeline before moving forward with its adaptation to knee MRI data. This phase also helped better to know and get a qualitative analysis for the CAN3D model.

Initially, a lot of issues were faced while training the model. Early runs produced noisy and unstable segmentations that failed to match the expected anatomical structures. The model output lacked consistency across slices and subjects, often overfitting to background regions or generating completely blank segmentations. These problems highlighted both the model's sensitivity to hyperparameters and the need for careful preprocessing of the input data. To address this, the approach which was adapted was to experiment with data normalization strategies, adjusted learning rates, batch sizes, and loss function configurations. These refinements helped stabilize the model and gradually improved its performance on both training and validation samples.

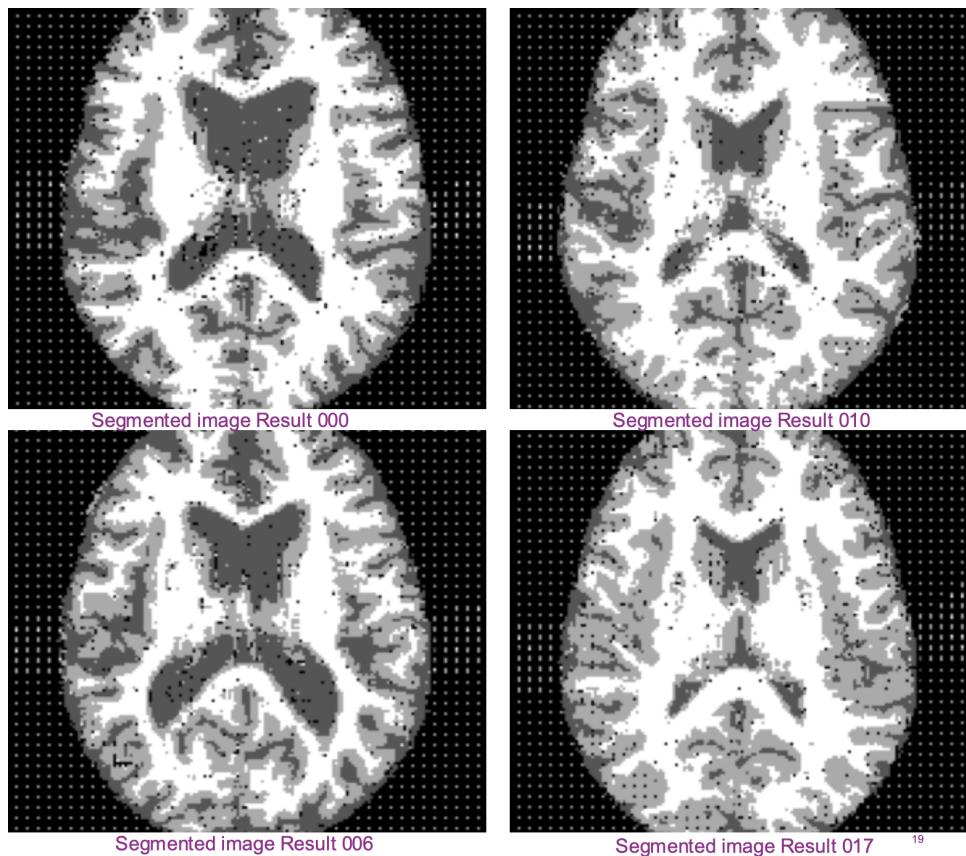


Figure 3.3: Initial noisy outputs

After training and testing the model, it was observed that the model produced less to no noisy outputs that aligned closely with ground truth labels, helping both qualitative visual inspection and limited quantitative analysis.

In addition to model-related issues, setting up the required computational environment also proved challenging. The CAN3D model relies on a GPU-enabled TensorFlow environment to train efficiently, but configuring this environment involved overcoming several compatibility issues between CUDA, python, and TensorFlow versions. At one point, I had to rebuild the entire environment multiple times due to conflicting dependencies and driver mismatches. Eventually, I was able to set up a stable environment that supported GPU acceleration, significantly speeding up training times. This setup process, although time-consuming, was important for handling the

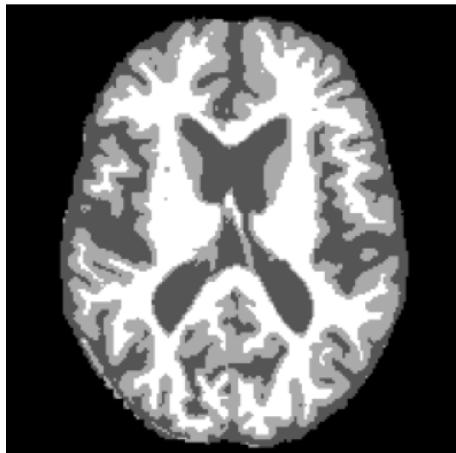


Figure 4-Img_Seg_1

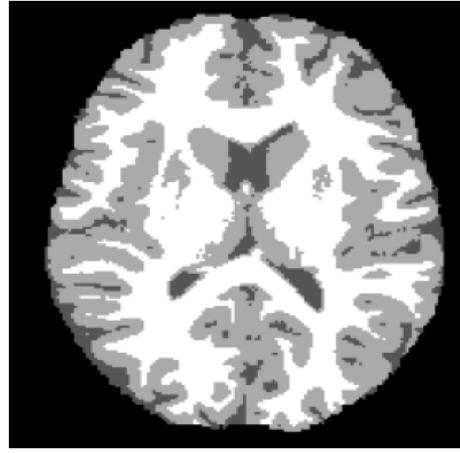


Figure 5-Img_Seg_2



Figure 6-Img_Seg_3

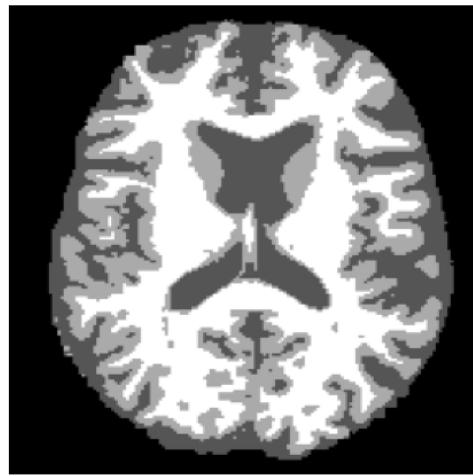


Figure 7-Img_Seg_4

Figure 3.4: Final output for the OASIS

computational demands of 3D MRI segmentation. During model training, it was also observed that both the `channel_weights` and `dsc_weights` arrays were uniformly set to:

```
OAI['channel_weights'] = np.array([1, 1, 1, 1, 1], dtype=np.float32)
OAI['dsc_weights'] = np.array([1, 1, 1, 1, 1], dtype=np.float32)
```

In the context of the OASIS dataset, this uniform weighting implies that all segmentation classes (channels) were treated with equal importance during training and evaluation. Since the dataset primarily includes brain structures with relatively balanced class distributions, this default configuration served as a meaningful baseline.

This observation proved important in later stages of the study. For the OAI and TKR datasets, where anatomical structures and class distributions differ significantly, these weights were adjusted to better reflect the importance and prevalence of different tissues. Thus, the uniform configuration in Phase 1 helped as a controlled reference point for subsequent model adaptations.

- **Phase 2 – Model Training and Validation (OAI-ZIB):** The second phase involved adapting the CAN3D model to the OAI-ZIB dataset, which comprises 507 annotated knee MRI volumes. Out

of these, 492 were used for training and 15 for preliminary validation. This phase is a significant shift from brain MRI segmentation to knee joint segmentation and served as the model’s first exposure to multi-label cartilage structures.

Initially, several implementation challenges were faced while setting up the model to run on the OAI dataset. Since the OASIS dataset used in Phase 1 had only four segmentation classes, the original configuration expected fewer output channels. The OAI-ZIB dataset, however, required five (one being the background) segmentation classes due to its more detailed anatomical labelling—Femoral Bone (FB), Femoral Cartilage (FC), Tibial Bone (TB), Tibial Cartilage (TC), and additional soft tissue structures. This change required an update in both the architecture and loss function settings. The channel and Dice loss weights were modified accordingly:

```
OAI['channel_weights'] = np.array([1, 1, 1, 1, 1], dtype=np.float32)
OAI['dsc_weights'] = np.array([1, 1, 1, 1, 1], dtype=np.float32)
```

This ensured that each anatomical structure such as Femoral Bone, Tibial Cartilage, and others was given equal importance during training. In multi-class medical image segmentation tasks like this, the ground truth labels are often represented using one-hot encoding. This means that for each voxel (3D pixel) in the MRI scan, a vector is assigned with a ‘1’ in the position corresponding to the correct class, and ‘0’s elsewhere [48]. For example, in a six-class setup, a voxel belonging to the Tibial Bone (label 3) would be encoded as [0, 0, 1, 0, 0]. The model is trained to predict a similar vector for every voxel, and the class with the highest probability is selected during inference.

Since one-hot encoding helps the segmentation mask into multiple binary maps which is one per class, the output channel of the model also becomes multi-dimensional. Equal weighting of `channel_weights` and `dsc_weights` (set to [1, 1, 1, 1, 1]) helps to make sure that each class contributes equally to the total loss during optimization. Without this, classes that occupy larger volume in the scan (like bone) could dominate the learning process, making the model neglect smaller but clinically important regions such as cartilage. This balance between the one-hot encoded format and uniform class weighting was important to achieve balanced learning across all tissue types.

Training the model on Rangpur’s GPU-enabled cluster introduced additional complications. While the cluster supported high memory workloads (up to 40GB GPU RAM), the CAN3D model’s complexity and patch-wise 3D inputs led to frequent out-of-memory (OOM) errors during training. These were particularly problematic when using batch sizes larger than one. Despite multiple attempts to optimize memory usage, the training had to be restricted to a batch size of 1 to maintain stability.

Another key modification was reducing the `context_dim` parameter from its original value of 512 (used in Phase 1 with OASIS) to 96. The `context_dim` controls the number of filters or feature channels used in the context aggregation layers of the network. A higher value allows

the model to learn richer, more abstract features, but at the cost of significantly higher memory usage. With six segmentation channels and large 3D patches, keeping `context_dim` at 512 consistently triggered memory errors. Reducing it to 96 helped keep the GPU memory within limits while still allowing the model to retain sufficient contextual learning. This trade-off between representational power and memory efficiency was important for the training to take place.

While the architectural design, loss functions, and optimizers were retained from earlier experiments (as discussed in the previous section), dataset-specific preprocessing steps were employed. These included normalization, spatial cropping, and basic data augmentation such as random flips to improve generalization. Although training proceeded more smoothly after adjustments to batch size, context dimension, and class weights, early validation runs still revealed class imbalance issues, particularly affecting smaller cartilage structures such as femoral and tibial cartilage. These smaller regions were sometimes underrepresented in the predicted segmentations, leading to lower Dice Similarity Coefficient (DSC) scores. To mitigate this, uniform class weighting (via `channel_weights` and `dsc_weights`) was maintained to prevent the model from biasing toward larger anatomical classes like bones. Furthermore, per-class Dice scores were closely monitored during training to assess how well each individual label was being learned. This iterative monitoring helped identify persistent imbalances and guided further fine-tuning of pre-processing and data augmentation strategies in later phases.

After overcoming these constraints with the respective changes, the model did produce consistent and anatomically accurate segmentations across key structures of the knee joint. The preliminary validation results were also promising, with a clear differentiation of the cartilage and bone regions, as shown in Figure 3.5a. This phase validated the model’s adaptability towards the knee structure and prepared it for final evaluation on the TKR dataset.

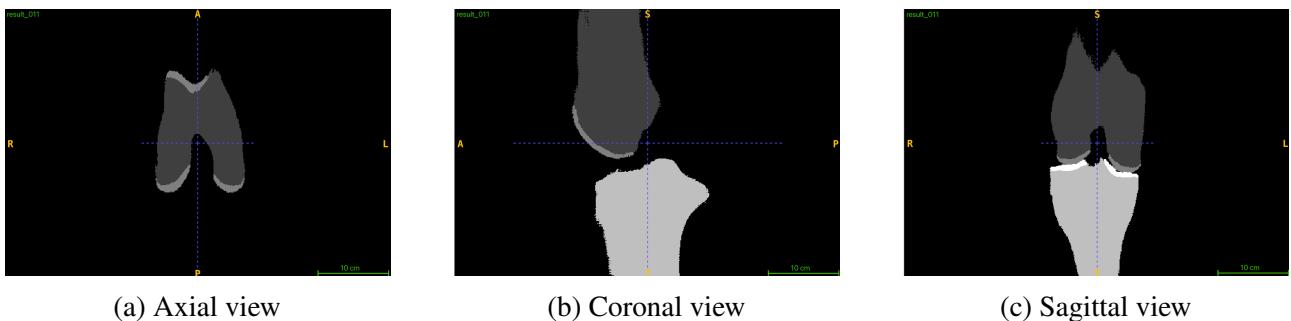


Figure 3.5: Different views of the same segmented MRI scan showing cartilage and bone structures.

- **Phase 3 – Final Testing and Comparative Evaluation (TKR):** In the final phase, the trained model was evaluated on the Total Knee Replacement (TKR) dataset, which includes 1,717 MRI volumes in `.nii.gz` format. This dataset presents a clinically realistic scenario, as it lacks manual ground truth segmentations, which have never, to our knowledge, been generated by anyone. As such, it serves as a valuable benchmark to test the model’s generalizability to new, unseen patient populations without relying on supervised feedback.

To better understand the model's performance across the OA spectrum, the TKR set was divided into two subgroups: est, as these often pose challenges for automated segmentation algorithms.

Due to the absence of ground truth labels, the evaluation during this phase was based on ** qualitative visual inspection**. The selected cases were overlaid with the predicted segmentation masks and manually assessed for anatomical plausibility, label consistency, and boundary sharpness. Representative examples are shown in Figure ?? and ??, which compare the raw MRI scans against the predicted segmentations.

To make sure that the final evaluation was free from any data leakage, a separate validation script was created to verify that no patient identifiers overlapped with those from the OAI-ZIB dataset used during training. **Ten overlapping subjects were identified and excluded from the OAI-ZIB train set:** 9001104, 9002430, 9002817, 9003430, 9004175, 9005075, 9005132, 9006723, 9007827, 9008561. Additionally, a separate test script was also implemented to support this phase. It redefined a lot of custom components originally used during model training, such as:

- `identity_initializer_2D` and `identity_initializer_3D` – used for weight initialization.
- `adaptive_bn_layer` – a custom batch normalization layer tuned for varying input distributions.

These components were important to replicate the architecture during inference, to make sure compatibility with the model checkpoints and preserving the behavior of specialised layers.

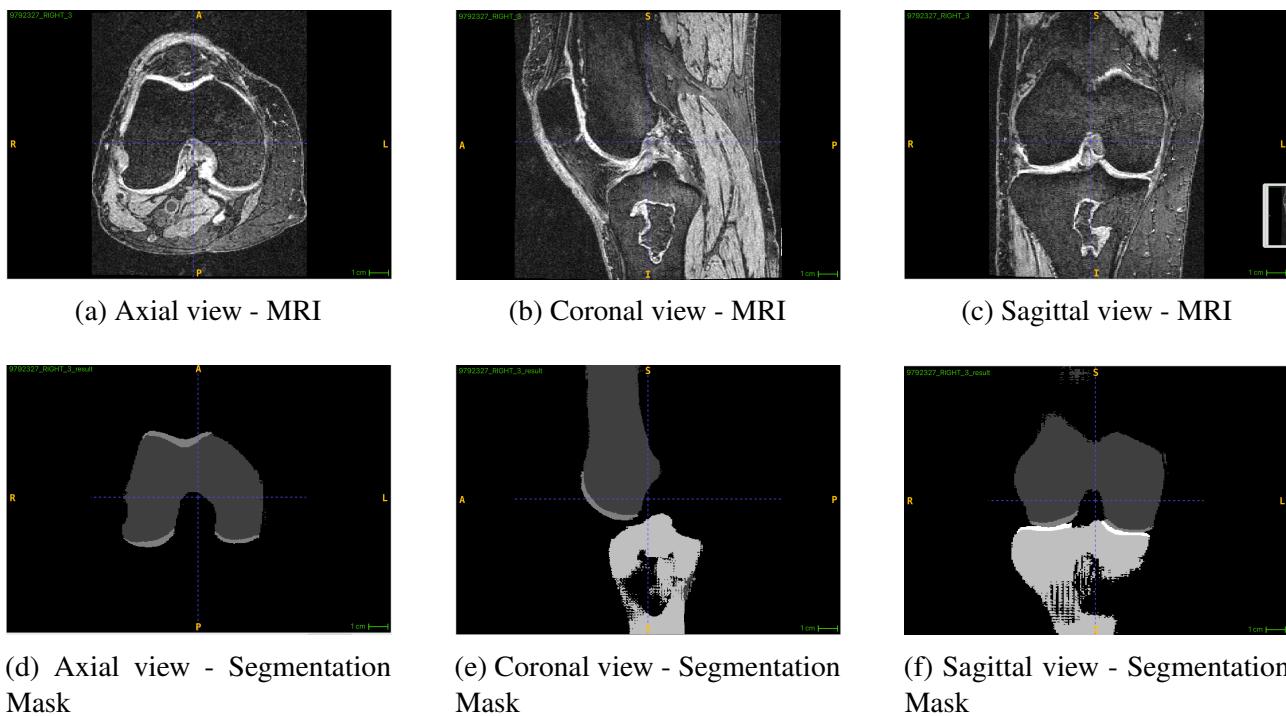


Figure 3.6: Qualitative comparison of original MRI scans and predicted segmentation masks across three anatomical views: axial, coronal, and sagittal.

We can conclude and say that this phase was able to demonstrate the model’s potential for practical deployment in clinical settings by evaluating its ability to produce anatomically coherent segmentation on a large-scale, real-world dataset without manual supervision.

3.3.7 Evaluation Metrics

For the evaluation of the tested segmentation masks, voxel counts and volumes were calculated for each anatomical region segmented in the knee joint. This evaluation was performed on both the OAI and TKR datasets.

Segmentation masks for each patient were visualised using ITK-SNAP, where specific intensity values corresponded to distinct anatomical structures:

- **1 (Red):** Femoral Bone (FB)
- **2 (Green):** Femoral Cartilage (FC)
- **3 (Blue):** Tibial Bone (TB)
- **4 (Yellow):** Tibial Cartilage (TC)

Using custom Python scripts with NiBabel and NumPy, voxel counts were extracted from each patient’s segmentation mask in .nii.gz format. The voxel spacing information was used to convert these counts into physical volumes. For each patient, a CSV file was created containing the patient ID, voxel counts, volumes, and voxel spacing for all segmented regions.

These individual CSV files were then merged with clinical metadata provided by the supervising professor. The clinical data, originally stored in a .pk1 file, was converted into a structured CSV format to facilitate this merging. The combined dataset included a variety of clinical and demographic information such as age, gender, BMI, smoking and drinking habits, comorbidities, medication use, pain and disability scores, Kellgren-Lawrence Grade (KLG), and indicators of osteoarthritis severity and treatment status, including whether the patient underwent total knee replacement (TKR).

This combination of anatomical segmentation data and clinical metadata allowed a detailed evaluation of how imaging features relate to patient health and disease progression.

Example Snapshot from Merged Dataset

Table 3.2: Sample of merged dataset combining segmentation output with clinical metadata.

Patient ID	Vox FB	Vol FB	Vox Cart	Vol Cart	Vox TB	Vol TB	Vox EC	Vol EC	x	y	z	KLG	OA Status
9006407	2025793	188489.59	137559	12799.16	1401525	130404.67	45106	4196.88	0.700	0.365	0.365	3.0	Healthy
9044005	1662511	154687.86	107117	9966.67	1102884	102617.52	28586	2659.78	0.700	0.365	0.365	4.0	Unhealthy

Note: For each patient, individual CSV files containing raw voxel and volume statistics were initially created. These were subsequently merged with clinical metadata to produce the final summary dataset used in the analysis.

Chapter 4

Results

This chapter presents the findings of the application of the CAN3D trained model on the Total Knee Replacement (TKR) dataset. The primary focus is to discuss about results of the evaluation conducted on the model's performance in segmenting knee cartilage and assessing its potential for clinical application. In addition, comparative results using box plots show the distribution of cartilage volumes across cohorts, highlighting differences in median, spread, and potential outliers that indicate disease severity.

4.0.1 Validation Set Segmentation Accuracy

The accuracy of the CAN3D segmentation model was evaluated on a held-out validation set consisting of 15 MRI volumes from the OAI-ZIB dataset. Table 4.1 shows the Dice Similarity Coefficient (DSC) scores computed on a held-out validation set from the OAI-ZIB dataset. The DSC values indicate strong overlap between predicted and ground truth labels for femoral and tibial cartilage, confirming that the model's segmentation outputs are sufficiently accurate for downstream volumetric analysis.

Table 4.1: Dice Similarity Coefficient (DSC) for femoral and tibial cartilage on the validation set

Cartilage Label	Mean DSC	Median DSC	Std. Dev
Femoral	0.89	0.90	0.03
Tibial	0.86	0.86	0.04

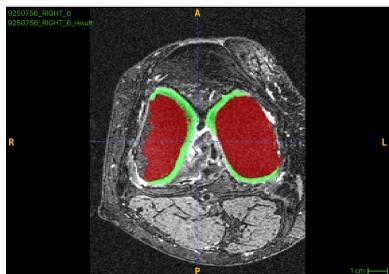
DSC values were estimated from a held-out validation set of 15 MRI volumes from the OAI-ZIB dataset. While test set labels were unavailable, visual inspection and downstream analysis suggest that the model generalised well to unseen TKR data.

4.0.2 Performance of the trained CAN3D Model

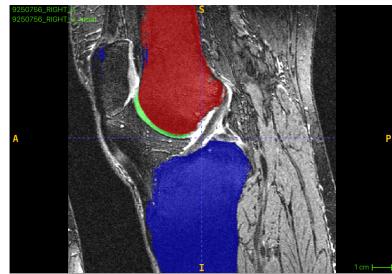
The CAN3D model was trained on the OAI-ZIB dataset consisting 337 image volumes. It achieved a mean Dice Similarity Coefficient (DSC) of **0.9445** and a final training loss of **0.0567**, reflecting upon good segmentation accuracy. The trained model was then applied on the OAI-TKR dataset.

A qualitative assessment using ITK-SNAP was conducted. Visual overlays of predicted masks on corresponding MRI slices demonstrated strong anatomical alignment, supporting the model's reliability for the OAI-TKR dataset.

Below are qualitative results (for some patients) showing cartilage segmentation outputs from the model on MRI scans :



(a) Axial View

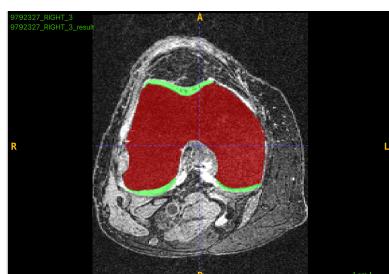


(b) Coronal View



(c) Sagittal View

Figure 4.1: Qualitative cartilage segmentation result from the OAI-TKR dataset for patient ID 9250756_RIGHT_6. The segmentation boundaries show high visual correspondence with cartilage structure.



(a) Axial View



(b) Coronal View



(c) Sagittal View

Figure 4.2: Qualitative cartilage segmentation result from the OAI-TKR dataset for patient ID 9792327_RIGHT_3. The segmentation is relatively clear, even in the presence of structural degeneration.

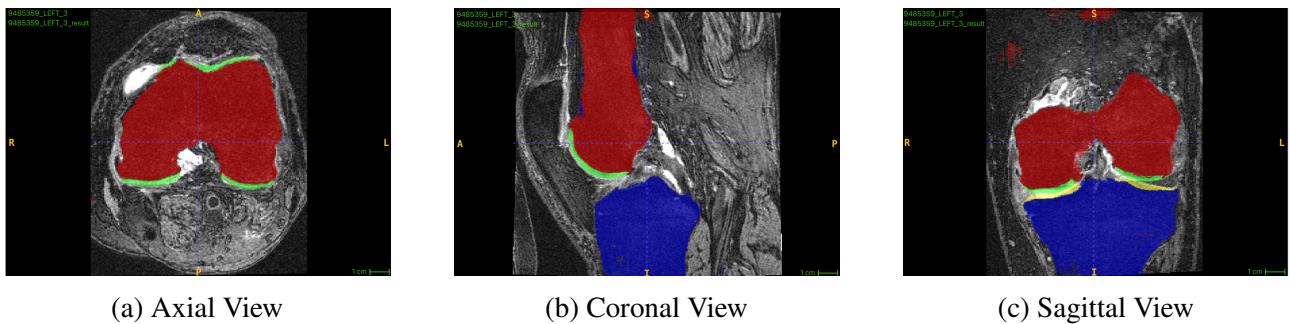


Figure 4.3: Qualitative cartilage segmentation result from the OAI-TKR dataset for patient ID 9485359_LEFT_3. The segmentation appears visually plausible, though the axial view shows potential boundary confusion.

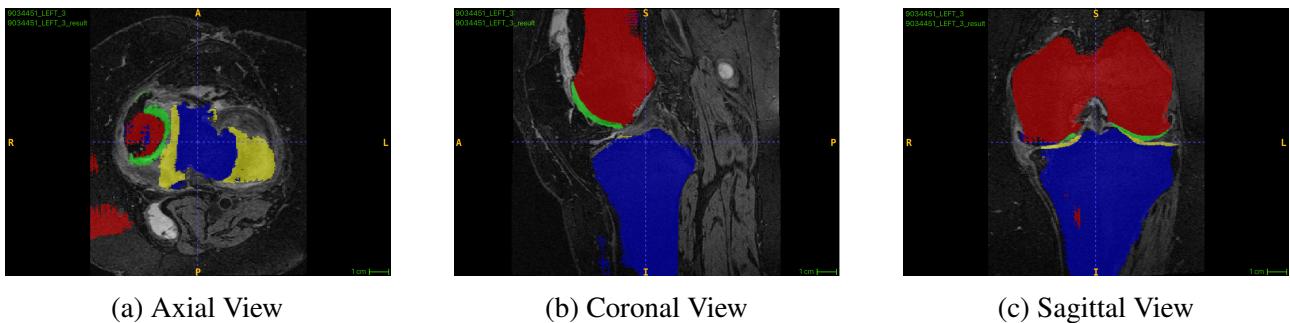


Figure 4.4: Qualitative cartilage segmentation result from the OAI-TKR dataset for patient ID 9035541_LEFT_3. The axial segmentation is less distinct, possibly due to anatomical complexity or model limitations.

The qualitative results demonstrate that the model is generally effective at segmenting cartilage across different patients and degree of OA. Most segmentation results across sagittal and coronal views appear consistent, the axial slices sometimes did show less distinct boundaries. This could be due to anatomical complexity or model limitations. But overall, visual alignment supports the generalisability of the trained CAN3D model.

4.0.3 Comparative Volume Analysis Across Cohorts

To evaluate clinical relevance, cartilage volume distributions were compared between cohorts using statistical summaries and box plots. These results helped us to know about variations like median, spread, and outliers, for both femoral and tibial cartilage helping us to provide the severity of the disease and compare the results and structural differences among groups. To revisit it is a known fact for us that :

- **OAI-ZIB Healthy:** Healthy patients with no OA
 - **TKR Healthy:** Subjects having OA but not severe enough to get a knee replacement
 - **TKR Unhealthy:** Subjects who have OA and require to undergo a total knee replacement

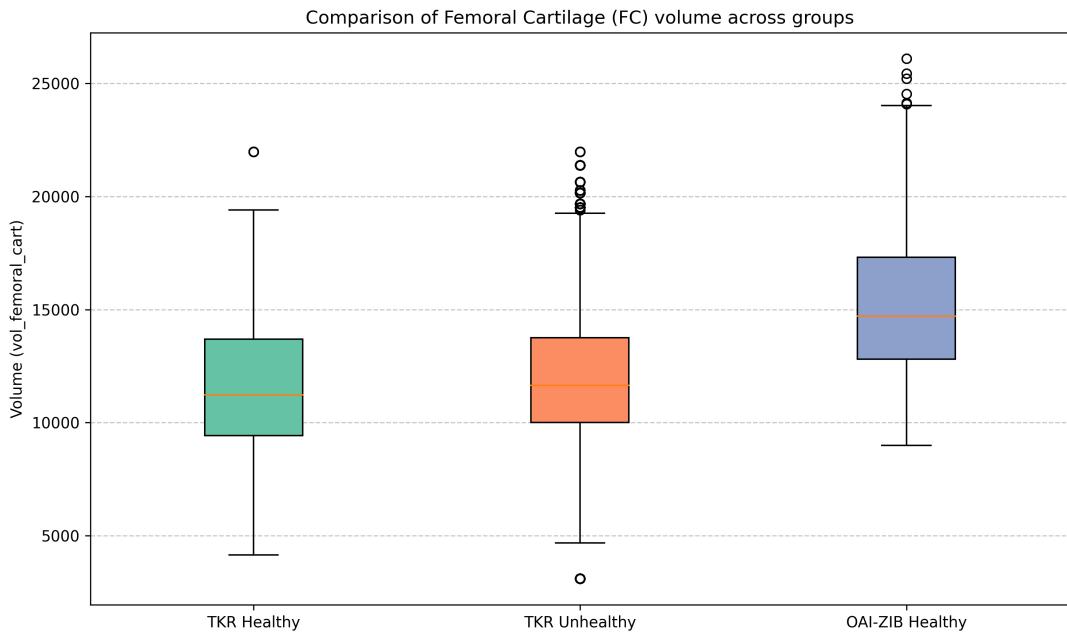


Figure 4.5: Boxplot comparison of femoral cartilage volume across cohorts. OAI-ZIB shows the highest median, while TKR Unhealthy exhibits the most degeneration and variability

Femoral Cartilage Volume

Box plots for femoral cartilage volumes (Label 2) across cohorts are shown in Figure 4.5. The median and range values for each group are summarized in Table 4.2.

Table 4.2: Summary statistics for femoral cartilage volume (in mm³).

Group	Median	IQR	Range	Outliers
OAI-ZIB (Healthy)	~14,500	13,000–17,000	9,000–23,000	>25,000
TKR Healthy	~11,000	10,000–14,000	3,000–19,000	>2,000
TKR Unhealthy	~11,500	9,000–13,500	4,500–19,500	>21,000

As observed from the box plot, the OAI-ZIB cohort has the highest femoral cartilage volume, with a median around 14,500 mm³, compared to 11,000 mm³ for the TKR Healthy group and 11,500 mm³ for the TKR Unhealthy group. The median reflects the central tendency of each cohort, and a higher femoral cartilage volume is generally associated with a healthier knee [49]. This supports the validity of the segmentation output.

The interquartile range (IQR) for the OAI-ZIB group spans from approximately 13,000 to 17,000 mm³, whereas the TKR Unhealthy group ranges from 9,000 to 13,000 mm³, and the TKR Healthy group ranges from 10,000 to 13,500 mm³. This suggests a significant degradation in cartilage volume among TKR Unhealthy patients. The maximum volume in the OAI-ZIB group reaches nearly 23,000 mm³, which aligns with expectations for healthy individuals.

The value of outliers are far from the main distribution, reflecting the individual variability and may indicate either segmentation noise or atypical patient conditions.

Overall, femoral cartilage volume was highest in the OAI-ZIB cohort, while the TKR unhealthy group showed both the lowest median and the widest range, consistent with varied osteoarthritis severity.

Tibial Cartilage Volume

Tibial cartilage volumes (Label 3) was also compared across cohorts. Results are shown in Figure 4.6, with statistics in Table 4.3.

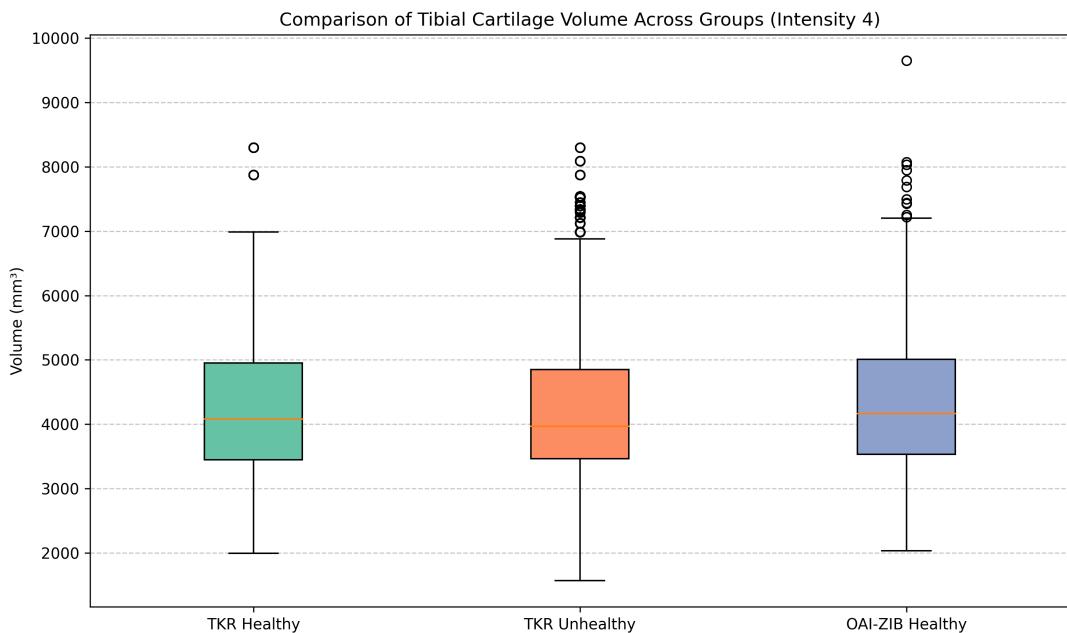


Figure 4.6: Boxplot comparison of tibial cartilage volume across cohorts. No significant differences are observed

Table 4.3: Summary statistics for tibial cartilage volume (in mm³).

Group	Median	IQR	Range	Outliers
OAI-ZIB (Healthy)	~4200	3500–5000	2000–7200	7400, 8000, 9700
TKR Healthy	~4100	3400–4800	2000–7000	7800, 8300
TKR Unhealthy	~3900	3300–4700	1600–6900	7100–8400

Tibial volumes were relatively consistent across cohorts, with considerable overlap in IQR and outliers. This suggests that tibial cartilage degeneration is less distinct or more gradual in OA progression, and thus may not serve as a strong standalone marker for disease classification.

4.0.4 Summary and Implications

The evaluation highlights key findings:

- CAN3D achieves high segmentation accuracy ($DSC \approx 0.9445$) on the OAI-ZIB dataset.

- Femoral cartilage volume serves as a reliable indicator of OA severity, with clear differentiation between healthy and unhealthy cohorts.
- Tibial cartilage volume shows minimal variation between groups and thus provides limited diagnostic value in isolation.

Overall, CAN3D demonstrates potential as a segmentation tool to support OA diagnosis and monitoring, particularly when integrated into broader clinical workflows.

Chapter 5

Discussion

This chapter discusses the main findings of this work in relation to the literature and from a clinical perspective. Although the CAN3D model has shown its favorable performance for knee cartilage segmentation on MRI scans in terms of the cross-data evaluation, there are several salient observations that require further discussion. In particular, the difference in femoral and tibial cartilage segmentation performance, the effects of representation shift between cohorts, and the general medical AI model generalisability are discussed. This chapter also discusses the potential improvements of the present approach and the possible directions for future research that can improve the robustness and clinician-friendly aspects of the model. By relating experimental findings to practical implications, in this review, we hope to close the distance between technical outcomes and medical significance.

5.0.1 Interpretation of Tibial and Femoral Cartilage Volume Findings

The main goal of this study was to evaluate the effectiveness of the CAN3D segmentation model on TKR dataset, and mainly to assess how well the model could generalise to previously unseen clinical data. Although early qualitative analysis and quantitative measurements using Dice Similarity Coefficient (DSC) indicated good accuracy in segmentation, an additional analysis of cartilage volume showed an interesting observation that tibial cartilage volume, a frequently measured feature in OA research did not show a substantial ability to distinguish between those subjects with OA and those without.

There was a large overlap between TKR-healthy, OAI-ZIB and TKR-unhealthy subjects in the box plot of tibial cartilage volume, suggesting that volume differences in the tibia by itself may not serve as a reliable biomarker of disease presence. This observation, while initially unexpected, aligns with prior findings in the literature.

Studies have pointed out that changes in tibial cartilage volume are often subtle and less strongly associated with symptomatic manifestations of OA. This may be because tibial cartilage

degeneration occurs in a more diffuse and variable pattern, and is less susceptible to early and localised cartilage thinning than femoral compartments [33].

In contrast, femoral cartilage in particular the medial femoral compartment, has shown more promise in OA-related structural assessment. As also reflected in some studies, femoral cartilage volume differences were more consistent and showed a clearer stratification between the three subject groups. Research has supported that the medial femoral region bears a higher mechanical load during activities like walking and stair climbing, making it more prone to degeneration in the early stages of OA [32, 33]. These biomechanical stressors result in localized cartilage wear that is easier to detect and quantify via MRI-based segmentation [32, 49].

Apart from that, there are a lot of longitudinal studies like from the Eckstein and Peterfy et al., 2004 have reported that femoral cartilage thinning correlates more directly with pain scores, functional impairment, and structural disease progression over time [50]. These findings carry several implications. At first, they help us to know that volumetric analysis of tibial cartilage alone is insufficient as a diagnostic or monitoring tool for OA. Relying on tibial volume could lead to under-detection or mischaracterise the disease severity. Second, the results underscore the need for regional and compartment-specific analysis in OA imaging research, where medial femoral cartilage may serve as a more robust structural biomarker. Third, from a model development perspective, these outcomes highlight the importance of aligning the evaluation targets with clinically meaningful endpoints, rather than using blanket volume measures across all compartments.

Moreover, the poor performance of tibial cartilage volume as a disease indicator also raises questions about its utility in clinical decision making and predictive modeling. Future work might include to explore whether combining tibial cartilage measures with shape-based, texture-based, or biochemical imaging markers (such as T2 mapping) which could help to enhance discriminatory power. Additionally, evaluating rate of change in volume over time, instead of single timepoint analysis may better capture disease progression, especially in early-stage OA.

At last, these findings emphasise the value of interpretable, clinical-aligned metrics in segmentation research. As models like CAN3D become increasingly effective in segmenting complex anatomy, the challenge shifts from “how well did the model segment?” to “how meaningful is what we segmented?” This shift points toward a broader research agenda that helps us to prioritize biological interoperability and clinical relevance over purely technical performance.

5.0.2 Dataset Domain Shift and Generalisability

An important aspect of this study was evaluating how well the CAN3D model could generalise from one dataset to another. Specifically, the model was worked well on OASIS and then trained on the OAI-ZIB dataset, which consisted of healthy subjects, and then tested on the TKR dataset, which included patients with varying degrees of osteoarthritis. Even after this domain

shift, the model retained a reasonable to good level of segmentation performance, though some performance degradation was observed.

This performance drop can be attributed to several factors. First, batch size variations due to differing hardware memory constraints (batch size of 2 for OASIS and 1 for TKR) which may have affected gradient stability and model generalisation. Second, there were also potential differences in MRI acquisition protocols, scanner models, and overall disease distribution between the two datasets likely contributed to domain shift. This is often seen as a common challenge in medical imaging where even slight protocol variations can make a significant impact in the model’s performance.

But irrespective of that the CAN3D model’s ability to adapt to unseen clinical data without retraining helps us to say that it did yield a promising result with good generalisability. However, these findings also highlight the importance of training on diverse and heterogeneous datasets to build models that are truly deployable in real-world clinical environments.

Chapter 6

Future Work

This study did demonstrate promising results with the CAN3D model, but a lot of things still remain unexplored which could potentially lead to improve the model’s performance and clinical applicability. The following directions outline key areas for further research and development.

Future work will focus on applying the CAN3D model to additional and more diverse datasets to better evaluate its generalisability and performance. The current study tested CAN3D on the OAI-ZIB and TKR datasets, but extending to other large datasets or clinical cohorts, will help validate the model’s performance in more real world scenarios.

An important direction also involves a hyperparameter optimization. The current training used a batch size of 1 due to memory constraints, but given CAN3D’s compact architecture, experimenting with larger batch size could help to improve training stability and model convergence rates. Larger batches typically help with smoother gradient estimates and might enhance the model’s generalisation ability. Apart from that, the context dimensionality (currently fixed at 96) is a key architectural parameter controlling the latent space size of the model. Adjusting this value can help with the influence of the spatial features captured as higher dimensions might allow the network to encode more complex structural relationships, potentially improving segmentation accuracy, while lower dimensions could reduce overfitting and improve efficiency. Figuring out and playing around with this balance through controlled experiments will help to provide even more valuable insights into the model’s capacity limits and ideal configuration for different datasets.

Beyond segmentation accuracy, future studies should also incorporate more detailed quantitative analyses that consider additional OA-related factors. While cartilage volume and thickness are important, OA progression is multifactorial and involves joint space narrowing, osteophyte formation, meniscal damage, bone marrow lesions, and patient symptoms such as pain or even functional impairment. Taking these variables into the evaluation framework would help to provide a more holistic understanding of the model’s clinical relevance and diagnostic value.

Moreover, adapting CAN3D to segment other anatomical regions affected by OA, such as

the ankle joints, represents a promising expansion. Each joint has unique anatomical and pathological features [6], so retraining or fine-tuning CAN3D with region-specific datasets will test its versatility and generalisability.

Another direction for continuation could also involve to evaluate the performance of the dataset on the KAN3D model which is an architectural extension of CAN3D that integrates Kaleidoscope transformation which helps to better capture anatomical symmetries and contextual dependencies.

In summary, while the current study does show us the strengths of the CAN3D model in segmenting knee cartilage and highlights its potential for clinical integration, it also opens up a lot of avenues for further research and exploration. By playing around with architectural tuning, or broader clinical validation, and multi-dimensional evaluation strategies, future research can build upon this foundation to create a more adaptable, accurate, and clinically meaningful segmentation model. Bridging the gap between high-performance models and real-world applicability will be key to making such tools truly impactful in aiding early diagnosis and disease monitoring in osteoarthritis.

Chapter 7

Conclusion

This thesis presented the adaptation and evaluation of the CAN3D model for volumetric segmentation of knee cartilage in TKR dataset. Through experiments on multiple datasets, including OASIS, training on OAI-ZIB and final testing on the TKR, CAN3D demonstrated strong performance in capturing detailed anatomical structures and showed promising generalisability to new clinical data. The comparison between tibial and femoral cartilage segmentation highlighted important insights into osteoarthritis biomarkers, helping us to reinforce the importance of femoral cartilage as a more reliable indicator for disease progression.

The model architecture of CAN3D has a compact yet powerful architecture based on dilated convolutions, which helps it to capture both local and global contextual information while maintaining computational efficiency. Through multiple experiments on different datasets like OASIS, OAI-ZIB and TKR, CAN3D demonstrated strong segmentation performance and promising generalisability between different cohorts.

One of the key finding of this study is the differential diagnostic value of cartilage compartments. While tibial cartilage volume showed limited discriminatory power between OA and non-OA groups, femoral cartilage particularly in the medial compartment came out to be as a more reliable structural biomarker. This exploration when further research upon did aligned with existing clinical literature and underscores the importance of targeted segmentation for meaningful OA assessment. The model's relatively better accuracy in femoral cartilage segmentation suggests that AI-based approaches can complement clinical evaluations by providing detailed, quantitative biomarkers linked to symptom severity and disease progression.

Despite these achievements, there are many things that remain unexplored or have a potential to improve. The observed performance drop when transferring from the OAI-ZIB to the TKR dataset shows us the impact of domain shifts caused by differences in MRI acquisition, patient demographics, and class label definitions. The small batch size used during training, due to memory issues may have also limited the model's ability to fully generalise. To address these issues, larger and more diverse training datasets, improved domain adaptation techniques, and

optimised hyperparameter tuning will be important for clinical translation.

Furthermore, this work focused primarily on segmentation accuracy without any major exploration of integrating other relevant OA factors, such as biochemical markers, pain scores, or patient demographics. Future studies could benefit from multimodal data fusion to improve predictive power and clinical relevance. Additionally, extending the model to segment other joint tissues affected by OA, such as menisci and bone marrow lesions, may provide a more in-depth understanding of disease mechanisms. In conclusion, the CAN3D model represents a significant step forward in automated knee cartilage segmentation, offering a good and efficient tool with potential clinical utility in OA diagnosis and monitoring. This research lays the groundwork for further exploration and innovation in AI-driven medical imaging, ultimately aiming to improve patient outcomes through earlier detection, personalised treatment planning, and better disease management.

Bibliography

- [1] Osteoarthritis initiative (oai) dataset, <https://nda.nih.gov/oai/> (2024).
- [2] J. Ehmig, G. Engel, J. Lotz, W. Lehmann, S. Taheri, A. F. Schilling, A. Seif Amir Hosseini, B. Panahi, MR-imaging in osteoarthritis: Current standard of practice and future outlook 13 (15) 2586, number: 15 Publisher: Multidisciplinary Digital Publishing Institute. doi: 10.3390/diagnostics13152586.
URL <https://www.mdpi.com/2075-4418/13/15/2586>
- [3] H. Woo, et al., Deep learning acceleration of mri acquisition and reconstruction, in: Proceedings of the 2024 Conference on Machine Learning Research, 2024.
URL <https://proceedings.mlr.press/v227/woo24a/woo24a.pdf>
- [4] S. Senthelal, J. Li, S. Ardestirzadeh, M. A. Thomas, Arthritis, in: StatPearls, StatPearls Publishing.
URL <http://www.ncbi.nlm.nih.gov/books/NBK518992/>
- [5] E. H. G. Oei, J. Runhaar, Imaging of early-stage osteoarthritis: the needs and challenges for diagnosis and classification 52 (11) 2031–2036. doi:10.1007/s00256-023-04355-y.
URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10509094/>
- [6] C. M, S. E, H. D, N. S, A. I, F. M, B. L, W. S, G. F, H. Cl, L. Li, J. G, C. F, O. R, V. T, B. R, W. A, M. L, The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study 73 (7), publisher: Ann Rheum Dis. doi: 10.1136/annrheumdis-2013-204763.
URL <https://pubmed.ncbi.nlm.nih.gov/24553908/>
- [7] M. D. Kohn, A. A. Sassoon, N. D. Fernando, Classifications in brief: Kellgren-lawrence classification of osteoarthritis 474 (8) 1886–1893. doi:10.1007/s11999-016-4732-4.
URL <https://doi.org/10.1007/s11999-016-4732-4>
- [8] Y. Yin, X. Zhang, R. Williams, X. Wu, D. D. Anderson, M. Sonka, LOGISMOS–layered optimal graph image segmentation of multiple objects and surfaces: cartilage segmentation in the knee joint 29 (12) 2023–2037. doi:10.1109/TMI.2010.2058861.

- [9] Knee osteoarthritis - AOA | australian orthopaedic association.
URL <https://aoa.org.au/for-patients/patient-information/knee-osteoarthritis>
- [10] Cureus | the pressing need to raise awareness about osteoarthritis care among elderly females in pakistan: A cross-sectional study | article.
URL <https://www.cureus.com/articles/20487-the-pressing-need-to-raise-awareness-!/>
- [11] J. Runhaar, M. Kloppenburg, M. Boers, J. W. J. Bijlsma, S. M. A. Bierma-Zeinstra, and the CREDO expert group, Towards developing diagnostic criteria for early knee osteoarthritis: data from the CHECK study 60 (5) 2448–2455. doi:10.1093/rheumatology/keaa643.
URL <https://doi.org/10.1093/rheumatology/keaa643>
- [12] J. Runhaar, Özbüyük, M. Kloppenburg, M. Boers, J. W. J. Bijlsma, S. M. A. Bierma-Zeinstra, the CREDO expert group, Diagnostic criteria for early hip osteoarthritis: first steps, based on the CHECK study 60 (11) 5158–5164. doi:10.1093/rheumatology/keab111.
URL <https://doi.org/10.1093/rheumatology/keab111>
- [13] F. P. Luyten, S. Bierma-Zeinstra, F. Dell'Accio, V. B. Kraus, K. Nakata, I. Sekiya, N. K. Arden, L. S. Lohmander, Toward classification criteria for early osteoarthritis of the knee 47 (4) 457–463. doi:10.1016/j.semarthrit.2017.08.006.
URL <https://www.sciencedirect.com/science/article/pii/S0049017217300987>
- [14] J. van Tiel, E. E. Bron, C. J. Tiderius, P. K. Bos, M. Reijman, S. Klein, J. A. N. Verhaar, G. P. Krestin, H. Weinans, G. Kotek, E. H. G. Oei, Reproducibility of 3d delayed gadolinium enhanced MRI of cartilage (dGEMRIC) of the knee at 3.0 t in patients with early stage osteoarthritis 23 (2) 496–504. doi:10.1007/s00330-012-2616-x.
URL <https://doi.org/10.1007/s00330-012-2616-x>
- [15] J. W. Liew, L. K. King, A. Mahmoudian, Q. Wang, H. F. Atkinson, D. B. Flynn, C. T. Appleton, M. Englund, I. K. Haugen, L. S. Lohmander, J. Runhaar, T. Neogi, G. Hawker, A scoping review of how early-stage knee osteoarthritis has been defined 31 (9) 1234–1241, publisher: Elsevier. doi:10.1016/j.joca.2023.04.015.
URL [https://www.oarsijournal.com/article/S1063-4584\(23\)00797-5/fulltext](https://www.oarsijournal.com/article/S1063-4584(23)00797-5/fulltext)
- [16] F. W. Roemer, A. Guermazi, S. Demehri, W. Wirth, R. Kijowski, Imaging in osteoarthritis 30 (7) 913–934, publisher: Elsevier. doi:10.1016/j.joca.2021.04.018.
URL [https://www.oarsijournal.com/article/S1063-4584\(21\)00881-5/fulltext](https://www.oarsijournal.com/article/S1063-4584(21)00881-5/fulltext)
- [17] D. J. Hunter, A. Guermazi, G. H. Lo, A. J. Grainger, P. G. Conaghan, R. M. Boudreau, F. W. Roemer, Evolution of semi-quantitative whole joint assessment of knee OA:

- MOAKS (MRI osteoarthritis knee score) 19 (8) 990–1002, publisher: Elsevier. doi:10.1016/j.joca.2011.05.004.
- URL [https://www.oarsijournal.com/article/S1063-4584\(11\)00153-1/fulltext](https://www.oarsijournal.com/article/S1063-4584(11)00153-1/fulltext)
- [18] Distinctions between diagnostic and classification criteria?doi:10.1002/acr.22583.
- URL <https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.22583>
- [19] J. H. Kellgren, J. S. Lawrence, Radiological assessment of osteo-arthrosis 16 (4) 494–502, publisher: Elsevier. doi:10.1136/ard.16.4.494.
- URL [https://ard.eular.org/article/S0003-4967\(24\)49310-2/abstract](https://ard.eular.org/article/S0003-4967(24)49310-2/abstract)
- [20] W. Dai, B. Woo, S. Liu, M. Marques, C. Engstrom, P. B. Greer, S. Crozier, J. A. Dowling, S. S. Chandra, CAN3D: Fast 3D medical image segmentation via compact context aggregation, Medical Image Analysis 82 (2022) 102562. doi:10.1016/j.media.2022.102562.
- URL <https://www.sciencedirect.com/science/article/pii/S1361841522002067>
- [21] The lancet: New study reveals the most common form of arthritis, osteoarthritis, affects 15% of the global population over the age of 30 | institute for health metrics and evaluation.
- URL <https://www.healthdata.org/news-events/newsroom/news-releases/lancet-new-study-reveals-most-common-form-arthritis>
- [22] J. Ehmig, G. Engel, J. Lotz, W. Lehmann, S. Taheri, A. F. Schilling, A. Seif Amir Hosseini, B. Panahi, Mr-imaging in osteoarthritis: Current standard of practice and future outlook, Diagnostics 13 (15) (2023). doi:10.3390/diagnostics13152586.
- URL <https://www.mdpi.com/2075-4418/13/15/2586>
- [23] A. Ching, Y. Prior, J. Parker, A. Hammond, Biopsychosocial, work-related, and environmental factors affecting work participation in people with osteoarthritis: a systematic review 24 (1) 485. doi:10.1186/s12891-023-06612-6.
- URL <https://doi.org/10.1186/s12891-023-06612-6>
- [24] M. G. Piccolo, et al., Imaging of knee osteoarthritis: A review of multimodal diagnostic approach, Quantitative Imaging in Medicine and Surgery 14 (4), accessed online on May 10, 2025 (2024).
- URL <https://qims.amegroups.org/article/view/112507/html>
- [25] L. A, E. Mh, D. Em, C. C, Epidemiology and burden of osteoarthritis 105, publisher: Br Med Bull. doi:10.1093/bmb/lds038.
- URL <https://pubmed.ncbi.nlm.nih.gov/23337796/>
- [26] M. S.MallikarjunaSwamy, M. S. Holi, Knee joint articular cartilage segmentation, visualization and quantification using image processing techniques: A review 42 (19) 36–43.

- doi:10.5120/5803-8151.
URL <http://research.ijcaonline.org/volume42/number19/pxc3878151.pdf>
- [27] F. Ambellan, A. Tack, M. Ehlke, S. Zachow, Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative 52 109–118. doi:10.1016/j.media.2018.11.009.
URL <https://linkinghub.elsevier.com/retrieve/pii/S1361841518304882>
- [28] Osteoarthritis of the knee | radiology reference article | radiopaedia.org.
URL <https://radiopaedia.org/articles/osteoarthritis-of-the-knee>
- [29] E. M. Macri, T. Neogi, M. Jarraya, A. Guermazi, F. Roemer, C. E. Lewis, J. C. Torner, J. A. Lynch, I. Tolstykh, S. R. Jafarzadeh, J. J. Stefanik, Magnetic resonance imaging-defined osteoarthritis features and anterior knee pain in individuals with, or at risk for, knee osteoarthritis: A multicenter study on osteoarthritis 74 (9) 1533–1540, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/acr.24604>. doi:10.1002/acr.24604.
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/acr.24604>
- [30] F. Eckstein, A. E. Wluka, W. Wirth, F. Cicuttini, 30 years of MRI-based cartilage & bone morphometry in knee osteoarthritis: From correlation to clinical trials 32 (4) 439–451, publisher: Elsevier. doi:10.1016/j.joca.2024.02.002.
URL [https://www.oarsijournal.com/article/S1063-4584\(24\)00027-X/fulltext](https://www.oarsijournal.com/article/S1063-4584(24)00027-X/fulltext)
- [31] Understanding the painful progression of osteoarthritis.
URL <https://www.sgul.ac.uk/research/our-impact/transformation-stories/understanding-the-painful-progression-of-osteoarthritis>
- [32] D. J. Hunter, L. March, P. N. Sambrook, The association of cartilage volume with knee pain 11 (10) 725–729. doi:10.1016/S1063-4584(03)00160-2.
URL <https://www.sciencedirect.com/science/article/pii/S1063458403001602>
- [33] A. E. Wluka, R. Wolfe, S. Stuckey, F. M. Cicuttini, How does tibial cartilage volume relate to symptoms in subjects with knee osteoarthritis? 63 (3) 264–268. doi:10.1136/ard/2003.007666.
- [34] R. Stahl, G. Blumenkrantz, J. Carballido-Gamio, S. Zhao, T. Munoz, M. Hellio Le Graverand-Gastineau, X. Li, S. Majumdar, T. Link, Mri-derived t2 relaxation times and cartilage morphometry of the tibio-femoral joint in subjects with and without osteoarthritis during a 1-year follow-up, Osteoarthritis and Cartilage 15 (11) (2007) 1225–1234. doi:10.1016/j.joca.2007.04.018.
URL <https://doi.org/10.1016/j.joca.2007.04.018>

- [35] B. Woo, M. B. Lorenzana, C. Engstrom, W. Baresic, J. Fripp, S. Crozier, S. S. Chandra, Semantic segmentation of 3d medical images through a kaleidoscope: Data from the osteoarthritis initiative, in: Medical Imaging with Deep Learning, PMLR, pp. 905–917, ISSN: 2640-3498.
URL <https://proceedings.mlr.press/v227/woo24a.html>
- [36] J. Abedin, J. Antony, K. McGuinness, K. Moran, N. E. O'Connor, D. Rebholz-Schuhmann, J. Newell, Predicting knee osteoarthritis severity: comparative modeling based on patient's data and plain x-ray images 9 (1) 5761, publisher: Nature Publishing Group. doi:10.1038/s41598-019-42215-9.
URL <https://www.nature.com/articles/s41598-019-42215-9>
- [37] S. Mahapatra, A simple 2d CNN for MNIST digit recognition.
URL <https://medium.com/data-science/a-simple-2d-cnn-for-mnist-digit-recognition-13a2a2f3a2>
- [38] (PDF) a multi-atlas approach for the automatic segmentation of multiple structures in head and neck CT images.
URL https://www.researchgate.net/publication/357565091_A_Multi-atlas_Approach_for_the_Automatic_Segmentation_of_Multiple_Structures_in_Head_and_Neck_CT_Images
- [39] C. M. Myles, P. J. Rowe, R. W. Nutton, R. Burnett, The effect of patella resurfacing in total knee arthroplasty on functional range of movement measured by flexible electrogoniometry 21 (7) 733–739. doi:10.1016/j.clinbiomech.2006.02.008.
- [40] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, arXiv preprint arXiv:1505.04597 Accessed: 2025-06-06 (2015).
URL <https://arxiv.org/pdf/1505.04597>
- [41] Y. Woo, K. Sohn, Y. Zhang, A. Yuille, M.-H. Yang, Learning to score proposals for weakly-supervised object detection, in: K. Chaudhuri, S. Jegelka, L. Song, P. Ravikumar, C. Szepesvári (Eds.), Proceedings of the 39th International Conference on Machine Learning, Vol. 172 of Proceedings of Machine Learning Research, PMLR, 2022, pp. 24106–24125, accessed: 2025-06-06.
URL <https://proceedings.mlr.press/v172/woo22a/woo22a.pdf>
- [42] B. Woo, M. B. Lorenzana, C. Engstrom, W. Baresic, J. Fripp, S. Crozier, S. S. Chandra, Semantic segmentation of 3d medical images through a kaleidoscope: Data from the osteoarthritis initiative, in: I. Oguz, J. Noble, X. Li, M. Styner, C. Baumgartner, M. Rusu, T. Heimann, D. Kontos, B. Landman, B. Dawant (Eds.), Proceedings of the 6th Medical Imaging with Deep Learning Conference, Vol. 227 of Proceedings of Machine Learning Research, PMLR, 2024, pp. 905–917, accessed: 2025-06-06.
URL <https://proceedings.mlr.press/v227/woo24a/woo24a.pdf>

- [43] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, R. L. Buckner, Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented and demented older adults, *Journal of cognitive neuroscience* 19 (9) (2007) 1498–1507.
- [44] Total knee replacement (tkr) dataset - osteoarthritis initiative, <https://nda.nih.gov/oai/> (2024).
- [45] A. D. Yao, D. L. Cheng, I. Pan, F. Kitamura, Deep learning in neuroradiology: A systematic review of current algorithms and approaches for the new wave of imaging technology, *Radiology: Artificial Intelligence* Published Online: March 4, 2020 (Mar. 2020). doi: 10.1148/ryai.2020190026.
URL <https://doi.org/10.1148/ryai.2020190026>
- [46] Z. Barańczuk, P. Zolliker, et al., Mean square error, in: *Advances in Imaging and Electron Physics*, Vol. 161 of *Advances in Imaging and Electron Physics*, Elsevier, 2010, p. Chapter 5.1, available at: <https://www.sciencedirect.com/topics/mathematics/mean-square-error>.
- [47] M. Yeung, E. Sala, C.-B. Schönlieb, L. Rundo, Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation, *Computerized Medical Imaging and Graphics* 88 (2021) 102026. doi:10.1016/j.compmedimag.2021.102026.
URL <https://doi.org/10.1016/j.compmedimag.2021.102026>
- [48] One-hot encoding - an overview, <https://www.sciencedirect.com/topics/computer-science/one-hot-encoding>, accessed: 2025-06-12.
- [49] A. Center, Arthritis news : Association between lower extremity muscle mass and knee cartilage volume.
URL <https://www.hopkinsarthritis.org/arthritis-news/association-between-lower-extremity-muscle-mass-and-knee-cartilage-volume/>
- [50] F. Eckstein, S. Cotofana, W. Wirth, M. Nevitt, M. R. John, D. Dreher, R. Frobell, Painful knees have greater rates of cartilage loss than painless knees after adjusting for radiographic disease stage: Data from the OA initiative 63 (8) 2257–2267. doi:10.1002/art.30414.
URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3149734/>