Novel 3D Deep Learning Models for Knee Osteoarthritis Diagnosis and Prediction from MRI Data

Mohamed Berrimi

▶ **To cite this version:**

Mohamed Berrimi. Novel 3D Deep Learning Models for Knee Osteoarthritis Diagnosis and Prediction from MRI Data. Human health and pathology. Université d'Orléans, 2024. English. ⟨NNT : 2024ORLE1085⟩. ⟨tel-05027456⟩

HAL Id: tel-05027456

https://theses.hal.science/tel-05027456v1

Submitted on 9 Apr 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL Authorization

UNIVERSITÉ D'ORLÉANS

Mathématiques, Informatique, Physique Théorique et Ingénierie des Systèmes (MIPTIS)

Institut Denis Poisson (IDP), UMR CNRS 7013

THÈSE présentée par :

Mohamed BERRIMI

soutenue le : **13 novembre 2024**

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline/ Spécialité : Sciences et Technologies industrielles

Novel 3D Deep Learning Models for Knee Osteoarthritis Diagnosis and Prediction From MRI Data

Nouveaux modèles 3D d'apprentissage profond pour le diagnostic et la prédiction de l'arthrose du genou à partir de données IRM

THÈSE dirigée par :

M. JENNANE Rachid	PU, Université d'Orléans, Directeur

RAPPORTEURS :

M. ADEL Mouloud	PU, Université d'Aix-Marseille
Mme OTHMANI Alice	MCF-HDR, Université Paris-Est Créteil

EXAMINATEURS :

M. CHETOUANI Aladine (Président)	PU, Université Sorbonne Paris Nord
Mme FALL Mame Diarra	MCF, Université d'Orléans
M. JARRAYA Mohamed	AP, Harvard Medical School
Mme TREUILLET Sylvie	PU, Université d'Orléans

Mohamed BERRIMI

Nouveaux modèles 3D d'apprentissage profond pour le diagnostic et la prédiction de l'arthrose du genou à partir de données IRM

Résumé :

L'arthrose du genou, une condition sans remède, exige un diagnostic précis pour une gestion optimale. Bien que l'IRM offre des informations essentielles, les méthodes actuelles se limitent souvent à une vue unique, risquant de manquer des détails critiques. Cette thèse propose des modèles d'apprentissage profond intégrant des architectures multi-vues, multi-étiquettes, multi-instances, multi-modales et multi-tâches pour un diagnostic plus robuste. Entraînés sur la base publique OAI, ces modèles 2D et 3D, utilisant des vues axiales, coronales et sagittales avec diverses séquences IRM, atteignent une précision de 93,20% à 96,70%, surpassant les méthodes existantes. En plus de la classification, des modèles de segmentation anatomique ont été développés pour mieux évaluer l'arthrose, démontrant ainsi le potentiel de l'apprentissage profond à améliorer le diagnostic et ainsi optimiser la prise en charge des patients.

Mots clés : Apprentissage profond, arthrose du genou, Imagerie par résonance magnétique

Novel 3D Deep Learning Models for Knee Osteoarthritis Diagnosis and Prediction from MRI Data

Summary :

Knee Osteoarthritis (OA) is a widespread and debilitating condition affecting millions, with no cure currently available. Accurate and timely diagnosis is vital for effective management and better patient outcomes. Although MRI provides valuable insights into soft tissues, current diagnostic methods typically rely on single-view analysis, possibly missing important information. This thesis introduces advanced deep learning models for robust knee OA diagnosis from multi-view MRI data, incorporating multi-view, multi-label, multi-instance, multi-modal, and multi-task architectures. By leveraging multiple MRI views (axial, coronal, sagittal) and sequences, these 2D and 3D CNN models were trained on the public OAI database, achieving 93.20% to 96.70% accuracy, surpassing current benchmarks. Additionally, segmentation models were developed to highlight anatomical structures key to OA evaluation, underscoring the potential of deep learning to enhance diagnostic precision and patient care in knee OA.

Keywords: Deep learning, knee osteoarthritis, Magnetic Resonance Imaging

INSTITUT ORLÉANS · TOURS Mathématiques & Physique Théorique DENIS POISSON

Institut Denis Poisson, Rue de Chartres
45067 Orléans

AGENCE NATIONALE DE LA RECHERCHE ANR

RÉSUMÉ

Ce travail de doctorat s'inscrit dans le contexte de l'application des techniques d'apprentissage profond au domaine de l'imagerie médicale, et plus particulièrement à la détection et à l'évaluation de l'arthrose du genou (Osteoarthritis, OA). L'essor du deep learning a considérablement modifié la manière d'aborder le diagnostic médical assisté par ordinateur, en améliorant les performances de diverses tâches, de la détection de fractures ou de tumeurs jusqu'à l'analyse de pathologies cardiaques ou neurologiques. Les progrès rapides des algorithmes, l'accès accru à des données cliniques de haute qualité, ainsi que la disponibilité de ressources informatiques puissantes, ont largement contribué à faire du deep learning un outil privilégié pour soutenir les cliniciens, radiologues et chercheurs. Dans le cas spécifique de l'arthrose du genou, une maladie dégénérative du cartilage et d'autres structures articulaires, la mise en œuvre de ces approches s'avère particulièrement prometteuse, bien que de nombreux défis subsistent, notamment la complexité anatomique, la variabilité inter-patient, la rareté des annotations expertes, ainsi que la difficulté à disposer de données suffisamment diversifiées.

Contexte et enjeux de la détection de l'arthrose du genou

L'arthrose du genou constitue un problème majeur de santé publique, conduisant à des douleurs chroniques, une perte de mobilité, et une dégradation progressive de la qualité de vie. Malgré l'importance clinique, ce domaine n'a pas bénéficié d'une attention aussi soutenue que d'autres pathologies dans la communauté du deep learning. Les techniques classiques, telle que la radiographie simple, se révèlent souvent inadéquates pour une détection précoce et précise, car elles ne permettent pas de visualiser les structures cartilagineuses et ligamentaires de façon optimale. À l'inverse, l'imagerie par résonance magnétique (IRM) offre une visualisation détaillée du cartilage, des ménisques, des ligaments et de l'os sous-chondral. C'est cette modalité riche en information et plus sensible aux modifications précoces de l'articulation qui motive le recours à l'apprentissage profond. Cependant, l'obstacle majeur réside dans la construction de jeux de données annotés, suffisamment vastes et diversifiés, ainsi que dans l'élaboration de modèles capables d'intégrer des vues et séquences multiples, provenant de volumes IRM complexes.

La base de données OAI (Osteoarthritis Initiative)

Une ressource clé pour la recherche dans ce domaine est la base de données OAI, un projet de grande envergure lancé en 2002, soutenu par les National Institutes of Health (NIH) et différents partenaires privés. L'objectif fondamental de l'OAI est de mieux comprendre les facteurs de risque, les mécanismes de progression et les options thérapeutiques possibles pour l'arthrose du genou. Le protocole d'étude OAI s'étend sur une période de dix ans et implique plus de 4 700

participants, recrutés dans des conditions variées afin d'assurer une diversité démographique (âge, origines ethniques, contextes géographiques) représentative.

Stratégies et approches de deep learning pour l'OA du genou

Le travail de thèse s'est structuré autour de plusieurs axes conceptuels et méthodologiques, visant à dépasser les limites des approches précédentes. L'objectif général est de développer des modèles plus robustes, plus complets, et capables d'exploiter la richesse des données IRM multi-vues et, dans certains cas, de données issues d'autres modalités (radiographies). Plusieurs défis sont relevés:

- Multi-vues et 3D: L'extraction d'informations tridimensionnelles est essentielle, car l'articulation du genou est une structure complexe dont les lésions arthrosiques peuvent se manifester sous des formes variées selon le plan de coupe. L'intégration de vues multiples (coronales, sagittales, axiales) permet d'obtenir une représentation plus complète de la pathologie.
- Semi-supervision et accroissement des données: La constitution d'ensembles de données annotés demeure coûteuse et complexe. Pour pallier le manque de données, des stratégies semi-supervisées ont été envisagées, permettant au modèle d'exploiter des images non annotées via des pseudo-étiquettes. Ces approches enrichissent le corpus d'apprentissage et améliorent les performances finales.
- Évaluation et protocoles de classification: Le score MOAKS (MRI Osteoarthritis Knee Score) ou encore le KL-grading (Kellgren-Lawrence) sont souvent utilisés pour évaluer la sévérité de l'OA. Cependant, l'attribution de ces scores repose sur des jugements experts et demeure une tâche subjective. Les modèles présentés cherchent à automatiser ce processus, en fournissant une évaluation quantitative et reproductible.
- Multimodalité (IRM + Rayons X): L'un des apports majeurs de cette recherche concerne l'intégration de différentes modalités d'imagerie. Les radiographies, plus courantes et plus économiques, apportent des indices relatifs à la structure osseuse, tandis que l'IRM fournit des informations plus subtiles sur les tissus mous. Leur fusion au sein d'un même modèle vise à compenser les faiblesses de chaque modalité prise séparément.
- Multi-tâches (classification + segmentation): Au-delà de la simple classification (distinguer un genou atteint d'OA d'un genou sain), certains modèles proposent simultanément la segmentation des régions d'intérêt (cartilage, ménisques, etc.). Cette approche multitâche permet de partager les représentations internes du réseau, d'enrichir les connaissances acquises, et d'améliorer la performance en classification.

Contributions spécifiques

Plusieurs modèles et approches ont été développés dans le cadre de cette thèse, chacun apportant une perspective complémentaire:

- Modèle semi-supervisé multi-vue 3D CNN: Un des premiers modèles proposés intègre une architecture 3D CNN capable de traiter simultanément plusieurs vues IRM, afin d'en extraire une représentation plus complète. En utilisant des pseudo-labels, la méthode

s'affranchit partiellement de la nécessité de disposer d'un ensemble entièrement annoté. Les résultats montrent une amélioration sensible en termes de performance, avec un AUC (Area Under the Curve) de l'ordre de 93,20 %, dépassant les approches classiques mono-vue ou celles basées uniquement sur la radiographie.

- M3NET (Multi-modal, Multi-view, Multi-label): M3NET représente une étape supplémentaire en introduisant la multimodalité. Il combine des modèles 3D-ResNet pour l'IRM et 2D-ResNet pour les radiographies, offrant ainsi une analyse conjointe des deux types d'images. En couvrant plusieurs vues IRM et les images aux rayons X, ce modèle s'est montré particulièrement performant pour la détection de différentes lésions du genou (notamment celles affectant le cartilage et le ménisque), dépassant significativement les approches mono-modalité. Le défi reste toutefois de gérer la redondance potentielle des données, ainsi que d'optimiser la stratégie de fusion des caractéristiques afin de tirer le meilleur parti de chaque modalité.

- MVMI (Multiview Multi-instance Network): Ce réseau exploite le concept d'apprentissage multi-instance, où chaque tranche IRM est considérée comme une instance. En faisant appel à une architecture de type Res-Squeeze-Net combinée avec des blocs Squeeze-and-Excitation, MVMI parvient à traiter avec grande finesse chaque tranche, puis à agréger l'information pour fournir un diagnostic global. Cette approche profite pleinement de la richesse tridimensionnelle des données, améliorant ainsi la sensibilité à des signes pathologiques subtiles. Testée sur la base OAI, MVMI atteint un AUC de 94,52 %, validant la pertinence du cadre multi-instance pour ce type d'application.

- Modèle multitâche (segmentation + classification): Dans une autre direction, la thèse propose un modèle exploitant deux tâches complémentaires : la classification de l'OA et la segmentation d'éléments clés (par exemple, le cartilage) sur les mêmes données IRM. Cette approche multi-tâches renforce le pouvoir de discrimination du modèle en mettant en évidence les zones critiques, tout en améliorant la précision diagnostique (jusqu'à 96,70 % d'exactitude). L'avantage réside également dans la réduction du besoin en prétraitement, puisqu'il est possible de former ces modèles à partir d'images IRM brutes, sans étapes complexes de préparation. En outre, cette simplification du flux de travail peut accélérer l'adoption de l'IA dans les milieux cliniques, réduisant le temps et l'expertise nécessaire pour manipuler les données.

Analyse des résultats et performances

Les résultats obtenus au cours de ce travail sont prometteurs et confortent l'idée que le deep learning a un potentiel considérable pour améliorer le diagnostic de l'arthrose du genou. Les modèles développés montrent que l'intégration de multiples vues IRM, voire de données multi-modales (IRM + rayons X), rehausse la performance par rapport aux méthodes traditionnelles. Le recours aux approches semi-supervisées permet de contourner la pénurie de données annotées, un problème récurrent dans le domaine médical. L'idée d'exploiter chaque coupe IRM comme une instance, grâce au multi-instance learning, s'est avérée particulièrement pertinente, et a permis d'obtenir des scores très élevés en détection. Enfin, la fusion des tâches de classification et de segmentation dans un cadre multitâche contribue à extraire des caractéristiques plus discriminantes, améliorant ainsi la sensibilité et la spécificité des modèles.

Défis, limites et perspectives

Malgré les avancées, certaines limites méritent d'être soulignées:

- Généralisation et diversité des données: Les études se sont appuyées sur la base de données OAI, riche mais non exempte de biais. La généralisation des modèles à d'autres populations, d'autres centres hospitaliers ou d'autres protocoles d'acquisition IRM reste à approfondir. Idéalement, des travaux futurs devraient valider ces approches sur des données issues de cohortes plus variées.

- Complexité des modèles et redondance des caractéristiques: Les approches multimodales et multi-vues peuvent introduire des redondances dans l'information. S'il est vrai que l'accumulation de points de vue devrait, en principe, améliorer la robustesse, un filtrage intelligent des données redondantes ou une meilleure sélection des séquences IRM s'avérerait nécessaire. Il est crucial d'élaborer des stratégies de fusion efficaces, capables de combiner les informations pertinentes sans gonfler inutilement la complexité du modèle.

- Interprétabilité et confiance clinique: Au-delà des performances quantitatives, la capacité d'interpréter les résultats est essentielle pour gagner la confiance des cliniciens. Les approches proposées pourraient être complétées par des méthodes d'explicabilité (heatmaps, attention maps) afin d'identifier les régions anatomiques les plus contributives. Cela permettrait de guider les spécialistes, de valider les choix du modèle et de garantir une intégration sûre et efficace dans la pratique médicale.

- Exploitation du multitâche pour d'autres critères cliniques: Si la segmentation conjointe a déjà fait ses preuves, d'autres tâches pourraient être intégrées, par exemple la prédiction de l'évolution de l'OA sur le long terme, l'estimation de la douleur rapportée par le patient, ou la suggestion de stratégies thérapeutiques personnalisées. Cette extension du paradigme multitâche permettrait d'englober l'ensemble du processus clinique, depuis le diagnostic précoce jusqu'au suivi thérapeutique.

La thèse présentée jette les bases d'une nouvelle génération de modèles de deep learning appliqués à l'imagerie du genou, ouvrant des perspectives vers un diagnostic de l'arthrose, plus précis et plus personnalisé. Les contributions majeures résident dans l'intégration de vues multiples d'IRM, l'emploi de techniques semi-supervisées, la fusion de données multimodales (IRM et rayons X), ainsi que l'adoption de paradigmes multitâches. Ces avancées démontrent que l'apprentissage profond n'est pas simplement un outil parmi d'autres, mais un levier puissant pour améliorer considérablement la précision, la robustesse et la portée des diagnostics, fournissant des éléments cruciaux pour la planification du traitement et le suivi du patient. En somme, les travaux menés dans cette thèse constituent un jalon important vers la compréhension des mécanismes de l'arthrose du genou et la mise au point de systèmes d'assistance au diagnostic plus intelligents, plus complets et plus fiables. À long terme, l'implémentation clinique de ces approches pourrait contribuer à alléger la charge des radiologues, réduire les erreurs de diagnostic, et favoriser une prise en charge plus efficace des patients souffrant de cette pathologie débilitante.

Dedication and Acknowledgements

I want to express my heartfelt gratitude to everyone who has stood by me over the past years, offering encouragement and believing in me. A special thank you to my beloved brothers, friends, and my future wife.

I wish to express my gratitude to Prof. Rachid Jennane for his support throughout my PhD journey. His advice and guidance have been invaluable for the success of this work.

I also want to acknowledge the financial support provided by the French National Research Agency (ANR) via the ANR-20-CE45-0013-01 project (MIMOSA project). This support has been instrumental in conducting experiments and acquiring the necessary financial and material resources.

I extend my sincere thanks to the jury members for their valuable feedback and recommendations. Your insights have strengthened this dissertation, and I appreciate the time and attention you've given to this work.

Elhamdolileh, I was ignorant and you taught me, I was lost and you guided me, so all praise is to you.

Mom and Dad and people of Gaza, this work is dedicated to you.

AUTHOR'S DECLARATION

I hereby declare that this work is the result of my own independent work, and that credit has been given to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgment.

SIGNED: *Mohamed Berrimi* DATE: AUGUST 19TH, 2024

TABLE OF CONTENTS

List of Tables

List of Figures

Figure **Page**

CHAPTER 1

DISSERTATION OVERVIEW

Deep learning has emerged as an efficient tool in medical imaging analysis across various modalities, achieving breakthroughs in numerous tasks. It is increasingly applied in computer-aided diagnosis systems, assisting physicians and radiologists in image analysis and decision-making. Deep learning is also utilized to detect and classify various injuries and diseases such as fractures, tumors, cardiac abnormalities, and neurological disorders. This technology not only enhances diagnostic accuracy but also significantly improves the efficiency of treatment planning and patient monitoring. Performance has consistently improved over the past decade, driven by advancements in deep learning techniques, enhanced data availability, and increased computational power [7].

1.1 Context and Motivation

Despite recent advances in deep learning techniques and their widespread application across various medical tasks, significant challenges remain in achieving uniform effectiveness. One primary obstacle is the diversity of medical imaging modalities. Magnetic Resonance Imaging (MRI), Computed Tomography (CT), X-ray, and ultrasound each present unique challenges and specificities, affecting the performance of deep learning models. Additionally, factors such as data scarcity, disease complexity, and the lack of detailed annotations hinder consistent development. Other significant barriers include variations in image quality and resolution, computational resource demands, and regulatory and ethical constraints regarding the use of medical data. These issues collectively hinder the universal applicability of deep learning in medical imaging [8].

In this thesis, we study knee osteoarthritis (OA) detection as a challenging case in deep learning research. Characterized by degenerative changes leading to severe pain and reduced

mobility, Knee OA has received less attention in recent deep learning research compared to other diseases [9]. This oversight is due in part to challenges such as a lack of sufficient data, both in terms of images and annotations, as well as the complexity of the knee structure [10].

Knee OA is a degenerative joint disease that impacts millions of individuals across the globe, leading to substantial economic costs [11]. Annually, the global expenditure on the diagnosis, treatment, and management of knee OA runs into billions of dollars, a significant financial burden given the absence of curative medications for this condition [12]. The costs encompass not only medical treatments but also indirect expenses related to decreased productivity and increased disability.

For diagnostic purposes, clinicians rely on radiographic imaging, particularly X-rays, to evaluate the condition of the knee [13]. During diagnosis, a severity score from 0 to 4 is assigned based on the X-ray findings, where a score of 0 indicates a healthy knee and a score of 4 reflects late-stage OA, also known as the Kellgren and Lawrence system [14]. This classification system is crucial for determining the extent of joint damage and guiding subsequent treatment strategies, helping healthcare providers manage the progression of the disease and mitigate its impact on patients' lives.

Magnetic resonance scans provide comprehensive insights into the knee joint's structure, enhancing the visualization of soft tissues and thereby improving diagnostic accuracy [15, 16]. However, the majority of research efforts in analyzing knee OA have focused on X-rays rather than MRIs [9]. This preference is primarily due to the availability of X-ray data and the challenges associated with processing MRIs, including the complexity of its volumetric nature and the computational demands of training three-dimensional (3D) models, compared to two-dimensional (2D) for X-rays data.

Deep learning models could be particularly beneficial in the context of knee OA diagnosis by automating the interpretation of radiographic images and MRI scans [15]. Deep learning models can enhance the accuracy and consistency of severity scoring, reduce diagnostic times, and potentially lead to earlier interventions that might improve patient outcomes.

This study aims to develop and refine effective deep learning models and training strategies, specifically tailored for knee MRI data, to enhance the accuracy of detecting Knee OA. The research will address challenges including knee OA analysis from volumetric data, high dimensionality of MRI data, the significant computational resources required, and the design of effective networks for knee MRI classification and segmentation.

1.2 Manuscript organisation

The content of the corresponding chapters is largely based on our published papers during the PhD thesis period, with minor modifications in some chapters, and removing the overlapping between sections. The order and content of the chapter's sections were occasionally adjusted for

consistency, and the related works for each paper were removed and consolidated in Chapter 3. These changes were made to ensure coherence within the thesis as a whole. It is important to note that we have also included new contributions to some chapters that we judged to be important to add.

Chapter 01: Dissertation overview

In this chapter, we present the research motivation and problem statement, along with a summary of the contributions made to advance the field of knee OA detection using MRI data. Additionally, a list of publications is provided.

Chapter 02: Knee Osteoarthritis

In this chapter, we discuss knee OA and related injuries as they appear in MRI data. We also address the available datasets, the challenges posed by limited data, and provide an overview of the knee MRI sequences in the OAI database. Additionally, we cover the various knee scoring systems and outline the specific OA definition that will be used in our experiments.

Chapter 03: Literature review and related works

In this chapter, we briefly mention some key concepts in deep learning applied for medical imaging analysis, including image classification, object detection, and image segmentation techniques, along with some learning paradigms and frameworks that have been used in the corp of the manuscript. Additionally, we mention some of the related works in the literature review that study knee OA analysis using deep learning in.

Chapter 04: A Semi-Supervised Multiview-MRI Network for the Detection of Knee OA

In this chapter, we present one of our research papers in which we introduce a semi-supervised multi-view framework and a new 3D CNN model for detecting knee OA using 3D MRI scans. We discuss the proposed semi-supervised learning approach in which we combine labeled and unlabeled scans to improve the performance and generalizability of the proposed model. Experimental results show the effectiveness of this approach in detecting knee OA from 3D MRI scans using a large cohort of 4,297 subjects.

Chapter 05: M3NET: A 3D Joint Multiview Multi-task Multimodal Network for Knee Injuries Classification

This chapter explains how we design a robust learning network called M3NET for meniscus and cartilage classification which uses multiview knee scans, and multimodal knee imaging.

In particular, M3NET is a joint architecture for meniscus and cartilage damage analysis that learns from representative features using 3D knee MRI scans and 2D X-rays. We first train a 3D customized YOLOv8 model to automatically detect the knee joint region from MRI scans. Furthermore, we design a feature fusion architecture to effectively handle both X-ray and MRI modalities and use multiple feature aggregation techniques to combine the learned image embeddings for consistent classification. Our proposed M3NET architecture is evaluated on a large number of MR scans from the osteoarthritis initiative (OAI) containing 1519 scans representing axial, sagittal, and coronal views, with additional X-rays of the same subjects.

Chapter 06: MVMI: Multiview Multi-instance Network for the Detection of Knee OA

In this chapter, we describe a key contribution of our thesis which focuses on designing a neural network architecture to effectively train 2D-based models on volumetric knee MRIs.

This approach processes in parallel multiple views of knee MRIs, treating each slice as an independent instance. Central to the network's architecture is the Res-Squeeze-Net, a CNN model that incorporates residual and squeeze-and-excitation blocks for enhanced feature extraction and emphasis on relevant diagnostic signals. This MultiView Multi-Instance (MVMI) strategy enhances the model's ability to identify and leverage distinct pathological features from different knee views, leading to a more robust and nuanced knee OA features analysis. The experiments revealed that MVMI outperforms volumetric and single-view-based models in knee OA detection.

Chapter 07: Multi-task network for knee joint segmentation and knee OA detection

In this chapter, we discuss how we propose a multi-tasking framework for both knee OA detection from 2D MRI slices extracted from 3D volumes, and knee joint segmentation using advanced custom CNN models, and how we leverage multi-tasking to boost the models' precision for accurately detecting knee OA.

1.3 Research Contributions

The overarching goal of this thesis is to advance the detection and characterization of knee OA and associated injuries through MRI data. This work is dedicated to developing and enhancing deep learning models that significantly improve the precision of diagnosing knee OA. In pursuit of these aims, we have made several key contributions including the design and optimization of deep learning models that are proficient at extracting and learning critical imaging features from multi-view MR scans. These advancements facilitate more accurate identification and understanding of knee OA and related injuries, setting the stage for detailed discussions of each contribution in the following subsequent sections.

1.3.1 Efficient models for MRI analysis

Detecting relevant features from large volumetric data is challenging primarily due to the high dimensionality of the data and the intensive computational resources required. CNNs, as mechanisms for feature extraction, have become role models in recent image classification tasks by enabling neural networks to automatically learn and detect decisive features for classification tasks.

Furthermore, most image classification tasks employing widely used ImageNet models [17] are inherently designed for 2D data (known as transfer learning). Consequently, their fine-tuning is limited on 3D volumetric data. This limitation stems from the fact that these models were originally optimized for 2D input and do not accommodate the spatial correlations present along the third dimension, which are critical in volumetric datasets like those used in knee MRI.

In this thesis, to address these issues, we present various deep learning architectures across multiple papers and sections tailored for processing and extracting relevant and decisive features from a range of knee MRI acquisitions. The proposed CNN models demonstrated significant improvements, achieving an increase of +8.1 AUC up to +10% for knee OA detection using MRI data, outperforming classical baselines and related works from literature reviews.

1.3.2 Multi-view Multi-modal multi-label model for Knee injuries analysis

Knee MRI are valuable imaging tools that produce comprehensive images of the structures within the knee joint. Typically, MR scans provide images from multiple views, offering a detailed visualization of the knee structure from various perspectives. When captured from two or more different angles, these scans can yield complementary information, enhancing the analysis of the knee structure and the interpretation of various tissues. Moreover, MRI facilitates the interpretation of multiple soft tissues and their related injuries. However, training individual models for each specific task and injury proves to be quite limited; general models capable of detecting multiple injuries simultaneously are significantly more effective.

In this thesis, we develop deep learning models designed to support knee analysis from a multi-view perspective, leveraging the distinct advantages of various MRI views and acquisitions. This is the first time a multi-task, multi-view model has been trained for multiple knee injuries within a single framework, utilizing over 1500 multi-view MRI scans. Across multiple experiments, we found that integrating multi-view knee MRI into the training models has increased the model's ability to detect knee OA correctly.

Additionally, we propose a multi-label multimodal learning approach for knee injury classification, leveraging both multi-view MRI and X-ray data. This approach is the first of its kind to combine these two imaging modalities within a single model for the simultaneous classification of meniscus tears and cartilage loss. By integrating information from both MRI and X-ray, the model can leverage the strengths of each modality, potentially leading to improved diagnostic ac-

curacy. This multi-task learning strategy improves the model's generalizability and performance compared to training separate models for each injury.

1.3.3 Multi-Instance Networks for 3D Knee MRI Analysis

The volumetric nature of 3D MRI presents unique challenges for training deep learning models. While 2D-based models often demonstrate superior performance due to their compatibility with ImageNet-based fine-tuning and established architectures, effectively leveraging them for 3D data analysis requires innovative approaches.

In this thesis, we introduce the use of multi-instance networks (MIL) to effectively train 2D-based CNN models on 3D MRI scans for detecting knee OA. This approach addresses the inherent challenges of 3D data analysis by treating each slice within the MRI volume as a separate instance. This strategy allows us to leverage the advantages of well-established 2D CNN architectures while effectively capturing information from the entire 3D volume.

MIL offers several key advantages for 3D MRI analysis. First, it effectively handles the variability and complexity of 3D data by considering each slice as an individual instance within a larger set. This allows the model to learn from a more diverse range of representations, improving its generalizability across various imaging conditions. Second, MIL enhances the model's robustness by isolating the most salient features for classification from each slice. By focusing on the most informative slices, the model can achieve higher classification precision. This approach effectively bridges the gap between the power of 2D CNNs and the complexities of 3D MRI data, enabling more accurate and efficient knee OA detection.

1.3.4 Segmentation-Based Multi-Task Models for knee Osteoarthritis analysis

Image segmentation is pivotal in medical imaging, especially in tasks that require precise anatomical delineation. This technique divides an image into multiple segments, isolating specific anatomical regions to simplify image analysis and enhance diagnostic accuracy.

In this thesis, we propose multi-tasking models that combine image segmentation with classification to enhance knee OA detection accuracy from MRI slices. By using detailed anatomical data from segmentation, these models improve the precision of classifying different stages of knee OA. We propose advanced models that utilize the segmentation of knee joints from large databases to classify knee MRI effectively. This approach increases the reliability and accuracy of MRI-based diagnostics in clinical settings.

Specifically, by incorporating segmentation into our multi-task models, we achieved a significant improvement of +3.3% in AUC compared to our previous volume-based (3D CNNs) models without segmentation. This improvement highlights the value of integrating detailed anatomical information derived from segmentation for enhancing knee OA detection accuracy. This enhanced accuracy can potentially lead to more precise diagnoses and personalized treatment planning for

patients, ultimately contributing to improved clinical outcomes and more effective management of knee OA.

1.4 Publications

1. Berrimi, M., Oussalah, M., & Jennane, R. (2023, October). A 3D Deep Learning Approach for Meniscus Tear Severity at the Region-level. In 2023 Twelfth International Conference on Image Processing Theory, Tools and Applications (IPTA) (pp. 1-6). IEEE.

2. Berrimi, M., Hans, D., & Jennane, R. (2024). A semi-supervised multiview-MRI network for the detection of Knee Osteoarthritis. Computerized Medical Imaging and Graphics, 114, 102371.

3. Berrimi, M. Teoh, Y X, Chetouani, A., Houam, L.,& Jennane, R. (2024) A Multi-Instance Learning Approach for Improving Knee Osteoarthritis Diagnosis from MRI Data. In 21st International Conference on Content-based Multimedia Indexing (CBMI)

4. Teoh, Y X, Berrimi, M. Lai, K W., Li Goh, Siew. , & Jennane R. Improving Knee Osteoarthritis Detection through a Multitask Approach on 2D MRI Slices (**Under-review**)

5. Berrimi, M. Jennane, R. (2024) MVMI: Multiview Multi-instance Network for the Detection of Knee Osteoarthritis. (**preprint**)

6. Berrimi, M. Anwar, S.M, & Jennane, R. (2024), M3NET: A 3D Joint Multiview Multilabel Multimodal Network for Knee Injuries Classification. (**preprint**)

CHAPTER 2

KNEE OSTEOARTHRITIS

In this chapter, we explore the essential role of imaging and semi-quantitative scoring systems in evaluating knee OA and related injuries. We discuss the role of MRI and explain various MRI sequences for knee joint analysis. Accurate assessment of knee conditions is critical for diagnosing and managing OA, and scoring systems like WORMS, BLOKS, and MOAKS provide a structured approach to quantifying damage across different knee regions. These systems are necessary to standardize evaluations, guide treatment decisions, and support research. Additionally, we discuss the available datasets and resources that are crucial for training deep learning models, highlighting their importance in advancing OA research and improving diagnostic accuracy.

2.1 The knee

The knee is one of the most critical and complex joints in the human body, essential for both weight-bearing and facilitating mobility. Comprising bones, cartilage, ligaments, and tendons, the knee is responsible for enabling fundamental movements such as walking, running, and jumping. This joint must constantly manage substantial mechanical loads while maintaining a delicate balance between flexibility and stability. However, the knee's intricate structure and its continuous exposure to high levels of stress render it particularly susceptible to injury, especially in contexts involving sports and physical labor [18]. The soft tissues of the knee, including ligaments and cartilage, are vulnerable to damage from repetitive strain or acute trauma, which can lead to tears, sprains, and other injuries. Over time, these injuries may contribute to the development of degenerative conditions such as knee OA, resulting in pain, stiffness, and compromised joint function [19]. Consequently, the knee, despite its robust nature, requires careful management to prevent injury and long-term degeneration [20].

In this thesis, we are interested in the diagnosis of particular knee injuries that could potentially lead to knee OA, and also we study the detection of knee OA from multimodal multiview knee imaging data.

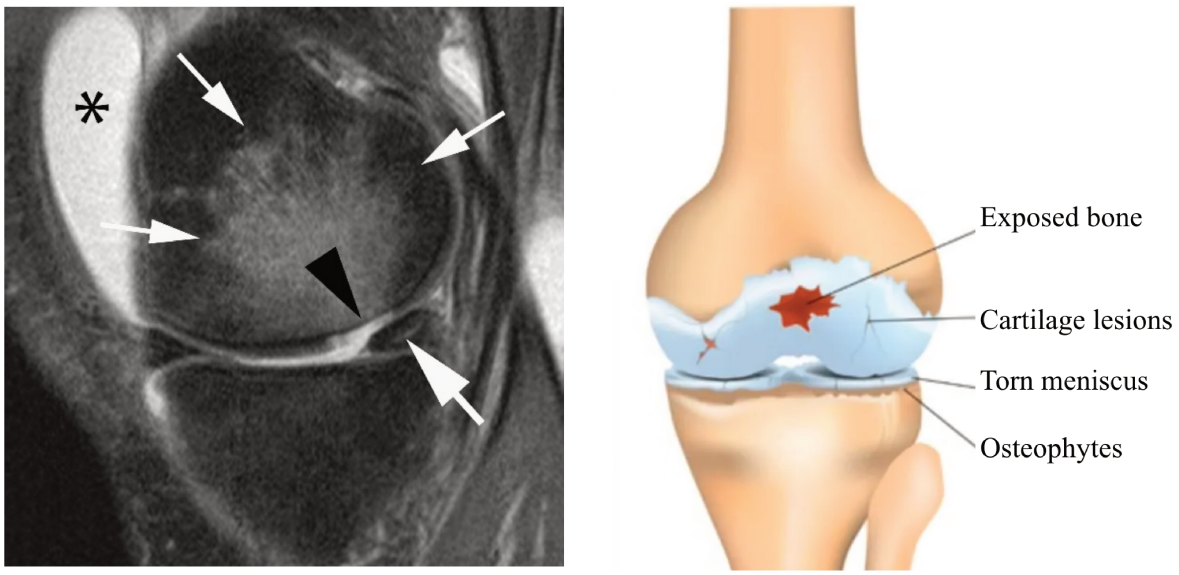


Figure 2.1: Illustration of knee OA and damaged knee compartments. Cartilage damage (black arrowhead), torn meniscus (large arrow), signs of inflammation (asterisk) or bone marrow lesions (small arrows) [1].

2.2 Knee Osteoarthritis

Knee osteoarthritis (OA) is a degenerative joint disease characterized by the breakdown of articular cartilage, leading to pain, stiffness, and impaired movement. Clinically, knee OA presents with symptoms such as joint swelling, reduced range of motion, and crepitus during movement [21]. Patients may also experience morning stiffness that improves with activity [22]. Over time, the disease progresses, causing increased pain, joint deformity, and disability. Imaging techniques like MRI are essential for diagnosing and monitoring the progression of OA, providing detailed visualization of the joint's structural changes [23, 24].

- **Cartilage Degradation:** The hallmark of knee OA is the breakdown of articular cartilage, which is primarily composed of collagen and proteoglycans. Mechanical stress and biochemical changes lead to the loss of these key components, resulting in cartilage thinning and the formation of fissures. As the cartilage wears away, the joint loses its protective cushion, leading to increased friction and pain during movement [25].

- **Inflammation:** Although OA is primarily considered a degenerative disease, inflammation plays a significant role in its progression. Synovitis, or inflammation of the synovial mem-

brane, can occur due to cartilage breakdown products being released into the joint space [4].

- **Changes in Subchondral Bone:** As cartilage degenerates, the subchondral bone (the layer of bone just beneath the cartilage) undergoes changes in response to altered load distribution. These changes include bone sclerosis (hardening) and the formation of osteophytes, which contribute to joint pain and deformity. Additionally, bone marrow lesions can develop, further contributing to pain and inflammation within the joint [20].

2.2.1 Impact of Knee OA: Personal and Economic Perspectives

Knee OA has a profound effect on both individuals and society. On a personal level, the pain and stiffness caused by knee OA can make everyday tasks like walking, climbing stairs, or even standing difficult [22]. This often leads to a less active lifestyle, which can cause other health problems like weight gain and heart issues. The constant pain can also lead to feelings of sadness, anxiety, and a lower quality of life. Social activities may become harder to participate in, causing people with knee OA to feel isolated.

The economic impact of OA is equally significant. Individuals often face high medical costs, including doctor visits, medications, physical therapy, and sometimes surgery. Knee replacements, which are common for severe OA, can be very expensive. Moreover, people with knee OA might have to take time off work or even retire early, leading to loss of income [26].

The economic burden of OA is substantial across several countries, with costs reflecting both direct medical expenses and indirect costs such as lost productivity. In the United States, the annual cost of knee and hip replacements rose significantly from $7.9 billion in 1997 to $22.6 billion by 2004, reflecting a dramatic increase in the number of hospital admissions related to OA [27, 28]. Additionally, job-related OA costs in the U.S. are estimated to range between $3.4 billion and $13.2 billion annually, with absenteeism due to OA contributing an additional $10.3 billion in indirect costs [29, 30]. In Canada, a 2005 survey found that the indirect costs for individuals over 55 with knee or hip arthritis could be as high as $12,990 annually, largely driven by lost employment time and the need for unpaid caregivers [31]. In France, the COART study estimated that the direct costs of OA in 2002 exceeded 1.6 billion Euros, accounting for 1.7% of the French health system's expenses. This included 570 million Euros in medication costs and 820 million Euros for inpatient treatment, highlighting the significant financial impact of OA on national healthcare systems [32].

The economic impact of OA in France was evaluated in a study [33], focusing on patient costs associated with hip and knee OA. In 2010, 90,946 patients were hospitalized for hip OA and 84,848 for knee OA, with a significant percentage undergoing joint replacement surgery. The annual hospitalization costs amounted to €9,797 per patient with hip OA and €11,644 per patient with knee OA, totaling €1.955 billion. Additionally, the study followed 18,976 community patients and found that annual costs per patient were €715 for hip OA and €764 for knee OA,

resulting in an estimated total cost of €3.5 billion for 4.6 million OA patients, with €2 billion attributed to the healthcare system.

These statistics underscore the substantial economic strain that OA places on both individuals and healthcare systems globally.

2.2.2 The Role of Imaging in the Diagnosis of Knee OA

Imaging is a pivotal component in the diagnosis and evaluation of knee OA, providing critical insights into the structural changes within the joint. The primary imaging modalities utilized are X-ray and MRI, each contributing distinct advantages and limitations to the diagnostic process.

X-ray imaging is the most widely employed technique for the initial diagnosis of knee OA. It is valued for its broad availability, rapid acquisition time, and cost-effectiveness [34]. X-rays are particularly effective in detecting osteoarthritic changes in the bone, such as joint space narrowing, subchondral sclerosis, and osteophyte formation. These radiographic features are traditionally used as diagnostic criteria for OA [14]. However, X-rays are inherently limited in their ability to visualize soft tissue structures, including articular cartilage, menisci, and ligaments. This limitation becomes particularly pertinent in early-stage OA or cases where the disease primarily affects the cartilage, as these changes may not be apparent on X-ray images despite the presence of clinical symptoms [35]. Consequently, the reliance on X-ray alone can result in underdiagnosis or delayed diagnosis of OA, especially in its nascent stages [36].

MRI, on the other hand, offers a comprehensive assessment of the knee joint, encompassing both osseous and soft tissue components [36]. This capability is particularly advantageous in detecting early degenerative changes, such as cartilage thinning, bone marrow lesions, and meniscal tears, which may not be evident on X-ray. Moreover, MRI is instrumental in cases where clinical symptoms are discordant with X-ray findings, providing a more nuanced understanding of the joint pathology [24]. Despite its diagnostic superiority in these contexts, MRI is not without its limitations. It is less accessible than X-ray due to higher costs, longer scanning times, and the requirement for specialized equipment and expertise. These factors can limit its routine use, particularly in resource-constrained settings [37].

2.3 Knee MRI

MRI has emerged as an indispensable tool in the assessment of knee OA, offering superior soft tissue contrast and the ability to visualize all joint structures comprehensively [23, 38, 39]. Unlike other imaging modalities, MRI allows for a detailed evaluation of cartilage morphology, meniscal integrity, ligamentous structures, and bone marrow lesions. These capabilities make MRI particularly valuable for the early detection of OA, monitoring disease progression, and assessing the efficacy of therapeutic interventions.

Advanced MRI techniques, such as T2 mapping, provide deeper insights into cartilage composition and the early biochemical changes that precede visible structural damage [4]. MRI enables the quantification of cartilage water content and collagen fiber orientation, which are critical markers of cartilage health [40]. Detecting these subtle changes early on can be pivotal in preventing further joint degeneration and optimizing treatment strategies.

The non-invasive nature of MRI, combined with its superior soft tissue contrast and absence of ionizing radiation, has solidified its role as the preferred imaging modality for a wide range of knee injuries and conditions. MRI is particularly effective in assessing:

- Cartilage Damage: MRI can detect even subtle cartilage lesions, including fraying, thinning, and full-thickness loss, which are hallmarks of OA progression. The high sensitivity of MRI in detecting cartilage abnormalities allows for early intervention, potentially delaying the onset of severe OA symptoms [41].

- Meniscus Tears: Tears in the menisci, the C-shaped shock-absorbing structures within the knee joint, are readily visualized on MRI [24]. The specificity of MRI for detecting meniscal tears is high, making it a reliable tool for diagnosing these common knee injuries, which often coexist with OA.

- Bone Marrow Abnormalities: MRI is sensitive to changes in bone marrow composition, such as edema (swelling) and bone marrow lesions, which can be indicative of OA or other underlying conditions [4]. These lesions are often associated with pain and are important markers of disease severity.

- Osteophytes (Bone Spurs): The presence and extent of osteophytes, bony projections that commonly develop in OA, can be accurately assessed using MRI [42]. MRI's ability to visualize these bony changes alongside soft tissue structures provides a comprehensive view of the joint environment in OA.

The strength of the MRI magnet, measured in Tesla (T), significantly impacts the quality of the images produced. Clinical MRIs commonly operate at either 1.5T or 3T [43]. While both strengths are capable of diagnosing knee OA, there are distinct advantages associated with each:

- 1.5T MRI: This is the standard field strength for most clinical MRI exams. It provides sufficient image quality for diagnosing a wide range of knee pathologies, including OA. 1.5T MRIs are more widely available, cost-effective, and less prone to certain artifacts, such as susceptibility artifacts, which can affect image clarity.

- 3T MRI: The 3T MRI offers higher resolution images and greater signal-to-noise ratio (SNR) compared to 1.5T MRI. This increased clarity allows for more detailed visualization of small structures within the knee, such as fine cartilage defects and subtle meniscal tears

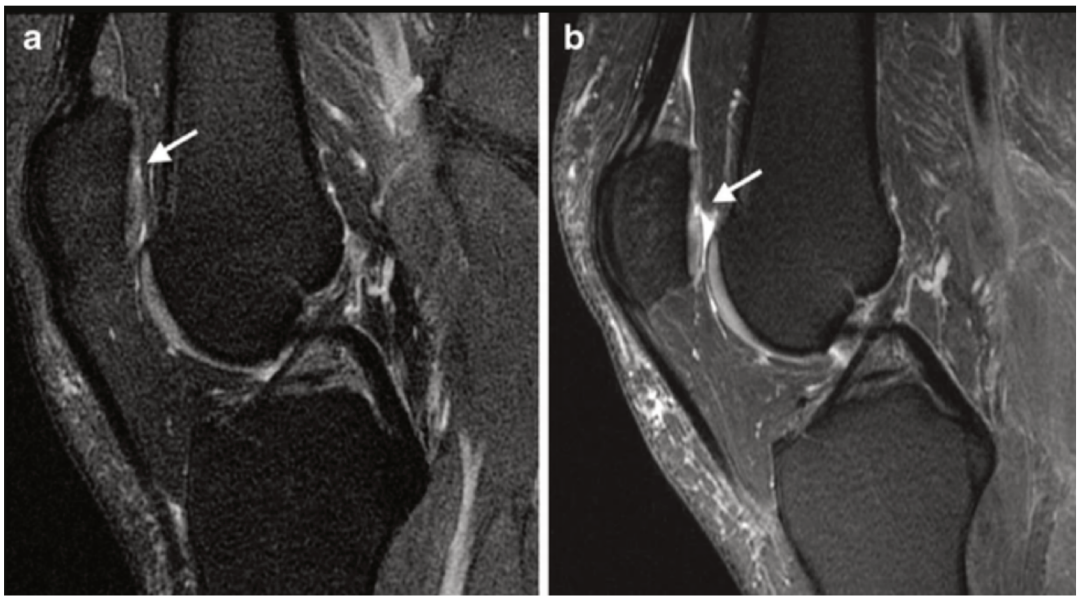


Figure 2.2: 1.5T and 3T sagittal MRI. **(a)** sagittal fs IW FSE sequence (1.5 T) shows a **questionable signal** change in the patellar cartilage (indicated by the arrow). **(b)** The corresponding sequence at 3T **clearly identifies** a patellar cartilage lesion exceeding 50% (indicated by the arrow), as confirmed by arthroscopy as the reference standard [2].

(as the example in Figure. 2.2). However, 3T MRI is more expensive and may not be as readily available in all clinical settings. Despite these limitations, the enhanced sensitivity and specificity of 3T MRI make it particularly valuable in complex cases where detailed anatomical assessment is crucial.

The choice between 1.5T and 3T MRI depends on various factors, including the specific diagnostic needs, patient considerations, and resource availability. While 1.5T MRI is adequate for most routine evaluations, 3T MRI is preferable when the highest image resolution is required, such as in the early detection of OA or the assessment of intricate knee injuries [44].

Overall, MRI's high sensitivity and specificity in detecting both soft tissue and bone abnormalities make it a cornerstone in the diagnosis and management of knee OA [45]. As technology advances, the integration of these imaging capabilities with AI-driven analysis tools holds great promise for further improving diagnostic accuracy and patient outcomes.

In the next section, we discuss multiple MRI sequences used to capture knee structures and pathologies. Additionally, we will review available datasets essential for studying knee OA and related injuries, providing a foundation for the development of AI-driven analysis tools.

2.4 Datasets

In the field of knee OA analysis, only a limited number of datasets have been publicly released for research purposes. This scarcity of available data has significantly impacted the advancement of research in this area, particularly in the development and validation of AI-driven models for knee OA diagnosis and progression monitoring. The lack of diverse and comprehensive datasets hinders the ability of researchers to train robust algorithms, limits the reproducibility of findings, and slows the pace of innovation in clinical applications.

The limited availability of datasets has also led to challenges in the generalization of deep learning models across different populations and clinical settings. Models trained on a single dataset may not perform well when applied to new, unseen data from different sources, highlighting the importance of having access to diverse datasets for training and validation purposes.

2.4.1 Osteoarthritis Initiative (OAI) database

This study utilizes knee MRI data from the Osteoarthritis Initiative (OAI), a landmark research project established in 2002 to deepen our understanding of OA and improve strategies for its prevention and treatment [46]. The OAI is a collaborative effort between the National Institutes of Health (NIH) and various private funding entities, aimed at creating a comprehensive dataset from a large cohort of individuals either at risk for developing OA or already diagnosed with the condition. By conducting longitudinal observations of these participants, the OAI seeks to uncover the factors influencing the onset, progression, and potential treatment of OA.

The OAI database is a valuable resource for researchers in both clinical and AI fields interested in knee OA and related injuries. The OAI is a multi-center, ten-year observational study of men and women, sponsored by the NIH, which is part of the U.S. Department of Health and Human Services. The primary goal of the OAI is to provide data that enhances the understanding of knee OA, a leading cause of disability in adults.

This extensive dataset includes clinical data, patient-reported outcomes, biospecimen analyses, quantitative imaging analyses, radiographs (X-rays), and magnetic resonance images (MRIs) collected over the course of the study. The dataset covers 4,796 participants, with longitudinal data spanning multiple time points—baseline, 12 months, 36 months, 48 months, 72 months, and 96 months. In total, the archive contains data from over 431,000 clinical and imaging visits and nearly 26.6 million images. Over 400 research manuscripts have been published based on this rich dataset [46].

Given the vast scope of the OAI database, this thesis focuses specifically on the baseline MRI data, which includes scans from 4,796 patients. Each patient underwent imaging of both knees, resulting in over 9,500 scans. These scans encompass multiple MRI sequences, including COR T1 3D FLASH, SAG IW TSE, COR MPR, MP LOCATOR, SAG T2 MAP, AX MPR, and COR IW TSE. Each patient's data is organized into two main folders (left and right knee), with subfolders

corresponding to the different MRI sequences, and X-rays are also included.

The study cohort of over 4,700 participants is demographically diverse, representing a wide range of ages, backgrounds, and geographic locations. This diversity allows for an in-depth exploration of the various risk factors and demographic correlates associated with OA. Participants undergo regular clinical evaluations, imaging procedures, and biological specimen collections, providing a rich and varied dataset for analysis.

The OAI's comprehensive repository of clinical, radiographic, and biochemical data has significantly advanced research in the field of OA. This extensive dataset enables researchers to identify patterns, risk factors, and potential biomarkers associated with OA. Additionally, it supports the evaluation of therapeutic interventions and has accelerated the development of new imaging modalities, biomarkers, and clinical outcome measures. These contributions are crucial for enhancing the diagnosis and monitoring of OA, ultimately leading to better management strategies for this debilitating condition.

2.4.2 MRNet Dataset

The MRNet dataset [47] comprises 1,370 knee MRI exams conducted at Stanford University Medical Center, designed to aid research in knee pathology with a particular focus on AI applications. The dataset includes 1,104 abnormal exams (80.6%), with 319 (23.3%) containing ACL tears and 508 (37.1%) showing meniscal tears. These labels were manually extracted from clinical reports.

The MRI exams were performed using GE scanners (GE Discovery, GE Healthcare, Waukesha, WI) equipped with standard knee MRI coils. The protocol included three key MRI sequences: coronal T1-weighted, sagittal proton density (PD)-weighted, and axial PD-weighted with fat saturation. Of the exams, 775 (56.6%) were conducted using a 3.0-Tesla magnetic field, while the remainder were done at 1.5-Tesla, with the 3.0T offering better image resolution and signal-to-noise ratio.

The dataset is split into a training set with 1,130 exams from 1,088 patients, a validation set (referred to as the tuning set) with 120 exams from 111 patients, and a hidden test set (referred to as the validation set) with 120 exams from 113 patients. Stratified random sampling was used to ensure that each set contained at least 50 positive examples of abnormal exams, ACL tears, and meniscal tears, with all exams from the same patient kept within the same split.

2.4.3 KneeMRI Dataset

The KneeMRI dataset [48] is a retrospectively collected dataset from MRI examinations conducted at the Clinical Hospital Centre Rijeka, Croatia, spanning the years 2006 to 2014. The imaging was performed using a Siemens Avanto 1.5T MR scanner, utilizing a proton density-weighted fat suppression technique, which is commonly employed for detailed morphological assessments of the knee. This dataset comprises 944 12-bit grayscale MRI volumes of either the left or right knee, each labeled according to the condition of the anterior cruciate ligament (ACL): healthy,

partially injured, or fully ruptured. The labeling was conducted in a double-blind fashion to ensure objectivity and accuracy.

The MRI scans were acquired with an in-plane resolution of 0.56 mm and a slice thickness of 3 mm, with 3.6 mm spacing between slices. Initially, 969 volumes were collected, but after excluding 25 volumes due to data corruption or the presence of abnormal physiological characteristics—such as knees post-ACL reconstruction or those with severe osteoarthritis—the final dataset consists of 944 valid MRI volumes.

Given that the sagittal plane is most informative for assessing ACL conditions, the dataset focuses exclusively on sagittal plane images, consistent with the clinical practices at the hospital. The volumes vary in size, ranging from $290 \times 300 \times 21$ to $320 \times 320 \times 60$ voxels, with median dimensions of $320 \times 320 \times 32$ voxels. To reduce intra-class variation and ensure consistency across the dataset, all MRI volumes of the right knee were mirrored to match the orientation of the left knee volumes.

This dataset provides a robust resource for research focused on the analysis and classification of ACL injuries using MRI, with the high-resolution, proton density-weighted images serving as a strong foundation for both clinical studies and the development of AI-based diagnostic tools.

2.4.4 MOST (Multicenter Osteoarthritis Study) Dataset

The Multicenter Osteoarthritis (MOST) Study [49] is a comprehensive, multicenter epidemiological study focused on identifying risk factors for knee OA and understanding the structural changes in the knee through MRI assessments. MOST is an observational, longitudinal study designed to follow a large cohort of participants, typically aged 50 to 79 years, over an extended period. The study specifically targets individuals with symptomatic knee OA—defined by frequent knee pain and X-ray evidence of OA at baseline—or those at high risk of developing OA due to the presence of various risk factors.

Participants in MOST undergo thorough evaluations at regular intervals, including baseline and follow-up visits, where data is collected on demographics, medical history, physical activity, genetics, and detailed imaging studies of the knee joint. The study employs a multidisciplinary approach, integrating imaging techniques, biomechanical assessments, clinical evaluations, and patient-reported outcomes to build a comprehensive understanding of OA. This broad and long-term data collection aims to answer critical questions about the progression of OA and to support the development of more effective prevention and management strategies.

The imaging component of MOST is particularly robust, with MR images being captured at multiple time points across the study:

- Baseline MR Images: 5,029 knees from 2,597 participants
- 15-month MR Images: 820 knees from 575 participants
- 30-month MR Images: 4,244 knees from 2,224 participants

- 60-month MR Images: 3,031 knees from 1,648 participants
- 84-month MR Images: 2,753 knees from 1,512 participants

These MRIs provide crucial data on the structural changes in the knee joint over time, allowing for an in-depth analysis of the progression of OA. The study's wide geographic reach includes contributions from Boston University, the University of Alabama at Birmingham, the University of Iowa, and Johns Hopkins University, among others.

2.5 MRI Sequences

In this thesis, we focused on four primary MRI sequences from the OAI dataset, carefully selecting them based on their availability across all patients records. Although additional sequences, such as the Coronal 3D Flash and Coronal TSE, were present in the dataset, we opted not to include them in our analysis. The decision to exclude these sequences was driven by their incomplete availability across the entire patient cohort, which could have introduced inconsistencies and reduced the robustness of our findings. By concentrating on the most consistently available sequences, we aimed to ensure the reliability and comparability of the results presented in this research.

Table 2.1 outlines the details of multiple MRI sequences from the OAI database for 2400 patients.

- **3D-DESS:** The Double Echo Steady State (DESS) sequence in MRI is highly effective for visualizing cartilage, offering exceptional sensitivity to subtle cartilage defects. This makes it particularly valuable in the early detection and assessment of OA changes. Due to its high-resolution capabilities, DESS is extensively utilized for cartilage mapping—a technique that quantitatively evaluates the thickness and volume of cartilage [4].

  In the context of knee injuries, sagittal knee MRI using the DESS sequence provides detailed images that are crucial for diagnosing and understanding the extent of the injury. The high resolution of DESS is advantageous for examining the complex structures of the knee joint, including ligaments and menisci, in addition to cartilage [50]. Within the OAI

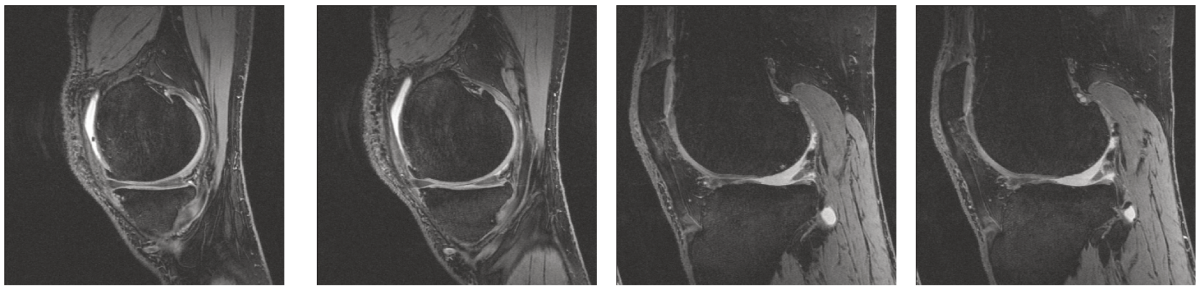


Figure 2.3: Left knee 3D DESS Knee MRI slices

Table 2.1: Comparison of Different MRI Sequences on 2400 samples from the OAI database.

	DESS	**IW TSE**	**COR MPR**
Resolution	0.36 mm × 0.36 mm	0.36 mm × 0.36 mm	0.36 mm × 0.36 mm
Slice Dimension	384 × 384	442 × 448	384 × 384
Slice Thickness	0.7 mm	3 mm	1.5 mm
Number of Slices	160	35-43	76
Sex (Female; Male)	1489; 910	1487; 909	1489; 910
Age (Mean ± SD)	61.88 ± 8.87	61.89 ± 8.86	61.88 ± 8.87

Group 1: DESS, IW TSE, COR MPR

	AX MPR	**MP LOCATOR**	**SAG T2 MAP**
Resolution	0.36 mm × 0.36 mm	0.36 mm × 0.36 mm	0.36 mm × 0.36 mm
Slice Dimension	384 × 384	512 × 512	384 × 384
Slice Thickness	1.5 mm	5 mm	3 mm
Number of Slices	76	23	232
Sex (Female; Male)	1489; 910	1489; 910	1489; 910
Age (Mean ± SD)	61.88 ± 8.87	61.88 ± 8.87	61.88 ± 8.87

Group 2: AX MPR, MP LOCATOR, SAG T2 MAP

	SAG IW TSE	**COR T1 3D FLASH**
Resolution	0.36 mm × 0.36 mm	0.36 mm × 0.36 mm
Slice Dimension	444 × 448	512 × 512
Slice Thickness	3 mm	1.5 mm
Number of Slices	37-38	80
Sex (Female; Male)	1489; 910	1489; 910
Age (Mean ± SD)	61.88 ± 8.87	61.88 ± 8.87

Group 3: SAG IW TSE, COR T1 3D FLASH

database, the sagittal 3D DESS scans consist of 160 slices, representing the aquisition with the highest number of slices. This extensive slicing enhances the ability to capture detailed cross-sectional images of the knee, facilitating a more comprehensive analysis of knee health and aiding in the early detection of degenerative changes and other pathologies.

- **IW-TSE:** The Intermediate-weighted Turbo Spin Echo (IW-TSE) sequences in MRI are designed to provide a balanced contrast between various tissues, enhancing their utility in evaluating diverse structures such as menisci, ligaments, and bone marrow. These intermediate-weighted sequences are crucial in the standard MRI protocols for assessing cartilage morphology and other structures relevant to OA [23]. Also, IW-TSE sequence is considered as the optimal option to evaluate ligaments [4].

In the evaluation of knee injuries, the IW-TSE sequences are particularly valuable. They offer detailed views of the internal architecture of the knee, allowing for a comprehensive assessment of the joint's health and the early detection of degenerative changes. This balanced approach helps in distinguishing between healthy and pathological tissues, which is essential for accurate diagnosis and treatment planning. In terms of data specificity, within the OAI database, the sagittal IW-TSE scans comprise 37 slices.

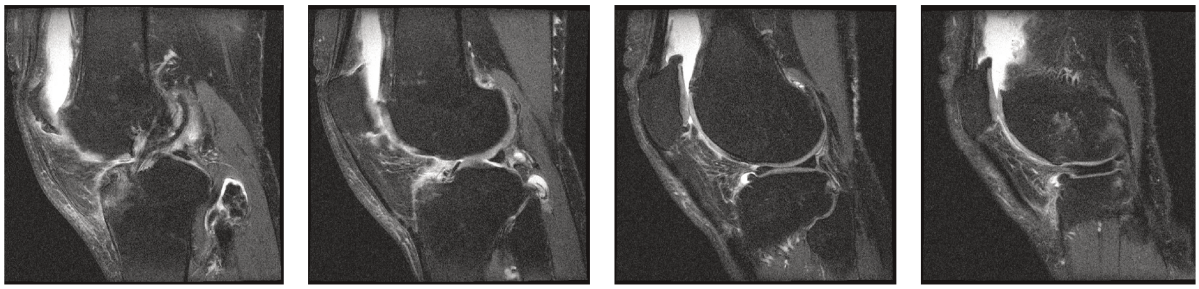


Figure 2.4: Left knee Sagittal IW-TSE Knee MRI slices

- **COR MPR:** Coronal multiplanar reconstruction (MPR) images, often derived from the 3D-DESS sequence, provide a comprehensive view of the cartilage across the entire knee joint. This imaging approach is invaluable for assessing not only the cartilage [51] but also the meniscus, providing crucial insights into the structural integrity and health of these components [52]. MPR is particularly beneficial in cases of internal derangement, where

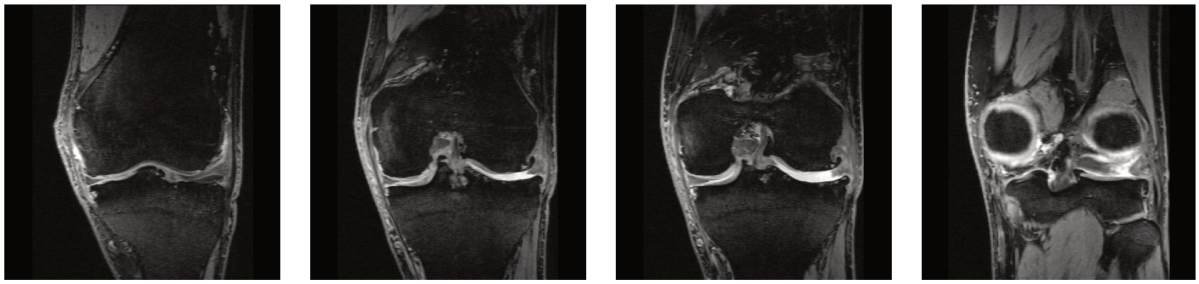


Figure 2.5: Left knee coronal MPR Knee MRI slices

precise imaging is vital for diagnosis and management. This technique helps visualize tears, ligament injuries, and other conditions that may not be as clearly defined in standard MRI formats. In the OAI database, the coronal MPR scans have from 70 to 76 slices.

- **AXR MPR:** The Axial Multiplanar Reconstruction (MPR) technique, similar to its coronal counterpart (COR MPR), provides detailed cross-sectional views of the knee joint. This imaging method is particularly effective in assessing meniscal tears, ligament injuries, and bone marrow abnormalities. It allows for precise visualization of the intricate structures within the knee, facilitating accurate diagnosis and effective management of knee disorders [53, 54].

In addition to its general utility, the axial MPR images are especially valuable for examining the various compartments of the knee joint that can be visualized from the axial plane. These include the medial and lateral compartments, where the femur, tibia, and menisci interact, and the patellofemoral compartment, which involves the interaction between the patella and the femur [54, 55]. This detailed visualization is crucial for diagnosing specific types of injuries such as meniscal tears, which are more common in the medial and lateral compartments, and patellar tracking disorders that affect the patellofemoral compartment. In the OAI database, the Axial MPR scans has from 70 to 76 slices.

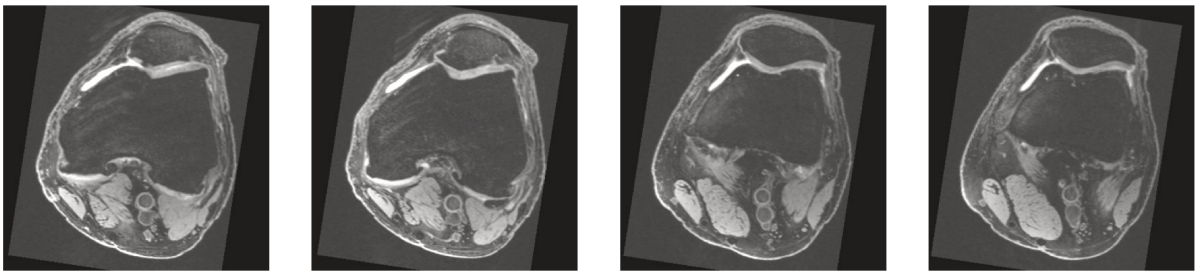


Figure 2.6: Left knee Axial MPR Knee MRI slices

A potential limitation of this study is the lack of explicit consideration for the influence of MRI sequence parameters for model's training. Factors such as slice thickness and specific MRI sequences could potentially affect image characteristics and, consequently, the performance of our deep learning models. However, a detailed investigation of these parameters was beyond the scope of this thesis, which primarily focused on model development and evaluation. It remains unclear whether incorporating these parameters into our analysis would have significantly impacted the results. This could be an interesting avenue for future research to explore further.

2.6 MRI Scoring Systems

Standardized scoring methods are essential in evaluating knee OA using MRI, as they help clinicians and researchers measure the extent of structural changes within a joint. These methods offer a systematic approach to analyzing MRI results, ensuring a uniform assessment of OA progression. There have been efforts towards proposing reliable semi-quantative systems recently to understand and grade different knee injuries within different parts of the knee [3].

A semi-quantitative scoring system is a method used to evaluate the extent of damage in the knee by dividing it into multiple regions and assigning a value that reflects the level of damage in each specific area. This approach allows for a detailed assessment of various knee structures by scoring different parts based on the severity of observed abnormalities.

However, it's important to recognize some limitations of semi-quantitative assessment. These methods require extensive training and calibration before they can be accurately applied. Additionally, they can have limited sensitivity to detect subtle changes, particularly when coding

changes over larger areas. The process can also be costly and time-consuming, especially when considering the need for specialized readers who are trained in these methods [56].

2.6.1 Whole-Organ Magnetic Resonance Imaging score (WORMS)

The Whole-Organ Magnetic Resonance Imaging Score (WORMS) [57] is one of the most widely used semi-quantitative scoring systems in knee OA research. WORMS was first introduced over a decade ago and has since been applied extensively in numerous studies. This system divides the knee into 15 distinct regions and evaluates 14 different features of the knee joint using MRI, typically performed with a 1.5 T system.

The features assessed by WORMS include articular cartilage integrity, subarticular bone marrow abnormalities, subarticular cysts, subarticular bone attrition, marginal osteophytes, meniscal integrity (both medial and lateral), cruciate ligament integrity (anterior and posterior), collateral ligament integrity (medial and lateral), synovitis/effusion, intraarticular loose bodies, and periarticular cysts/bursitis. Each of these features is scored independently, and the scores can be correlated with each other using statistical methods like Spearman Rho to understand the relationships between different types of knee damage.

WORMS employs a subregional approach rather than focusing on individual lesions, particularly for structures like cartilage, bone marrow lesions, and subchondral cysts. This method provides a single score per subregion, which encompasses multiple lesions within that area, potentially simplifying the reading and analysis process. This approach is beneficial in longitudinal studies where lesions might merge, split, or appear adjacent to one another, complicating the assessment if a lesion-based approach were used.

Notably, WORMS is the only semi-quantitative scoring system that specifically evaluates subchondral bone attrition—characterized by flattening or depression of the articular surface that is not related to trauma. This unique aspect of WORMS highlights its comprehensive nature in assessing knee joint health [56].

2.6.2 Boston Leeds Osteoarthritis Knee Score (BLOKS)

The Boston-Leeds Osteoarthritis Knee Score (BLOKS) [58] is a semi-quantitative scoring method designed to evaluate nine intra-articular regions of the knee. BLOKS focuses on eight key features: bone marrow lesions, cartilage, osteophytes, synovitis, effusions, and ligaments. To refine the initial scoring items, a series of iterative reliability exercises were conducted, leading to the final version of BLOKS. The reliability of this scoring system was thoroughly examined, ensuring that it provides consistent assessments across different users.

In addition to establishing reliability, BLOKS was also evaluated for construct validity—its ability to relate to knee pain—and longitudinal validity, which is its capacity to predict cartilage loss over time. These assessments were conducted by comparing BLOKS to another widely used

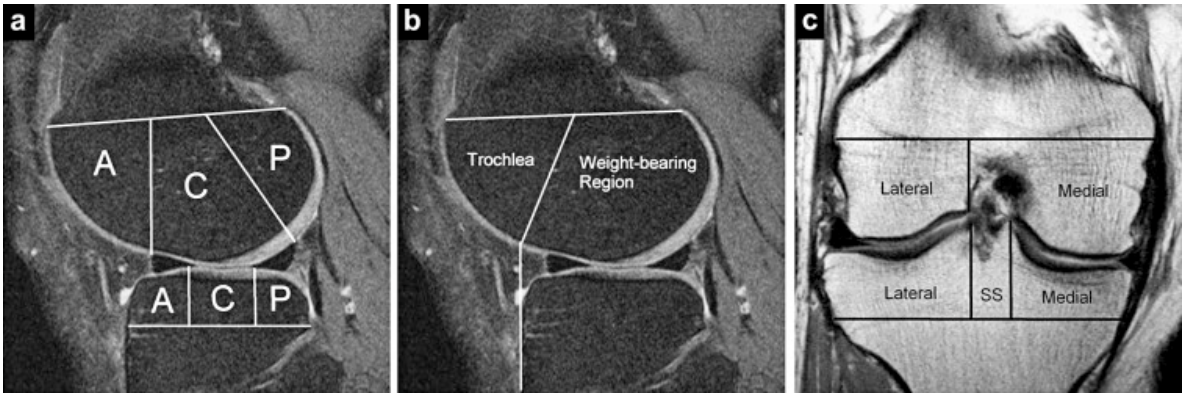


Figure 2.7: Comparison of anatomical subregion locations in WORMS and BLOKS: **(a)** WORMS divides the tibia and femur into anterior, central, and posterior subregions, **(b)** BLOKS divides the femoral condyle into trochlear and weight-bearing subregions, and **(c)** the medio-lateral division of the femur as defined by WORMS, which was also applied in BLOKS scoring for this study [3].

scoring system, the Whole-Organ Magnetic Resonance Imaging Score (WORMS), specifically focusing on bone marrow lesion (BML) scores.

Recent studies have compared the validity and reliability of these two scoring systems [3], offering insights into their respective strengths and weaknesses, particularly in relation to key features like cartilage, meniscus, and BMLs. Each system has its own limitations; for example, the WORMS meniscal scoring method combines several constructs into a single score, which may complicate interpretation. On the other hand, the BLOKS scoring system for BMLs can be cumbersome and complex, with some components appearing redundant (see Figure. 2.7).

Using the OAI database, WORMS and BLOKS were utilized to assess 115 patients, with results published in May 2011.

2.6.3 MRI Osteoarthritis Knee Score (MOAKS)

The MRI Osteoarthritis Knee Score (MOAKS) [4] is an advanced scoring system derived from earlier methods like BLOKS [58]. It refines the evaluation by offering detailed assessments of bone marrow lesions (BMLs), cartilage damage, and meniscal changes, enhancing the granularity of the scoring. This system also streamlines the process by eliminating some overlapping elements from previous scoring methods, focusing on targeted, high-priority areas in the knee.

MOAKS was used to evaluate 600 patients from OAI cohort as part of the FNIH project, with findings released in April 2015 [1].

MRI scans of both knees were performed using a 3 Tesla Siemens Trio system, with a specialized knee coil. To confirm the reliability of the MOAKS system, initial training on MRI evaluation was conducted with two experienced radiologists who later independently analyzed

[1] https://fnih.org/our-programs/osteoarthritis-initiative/

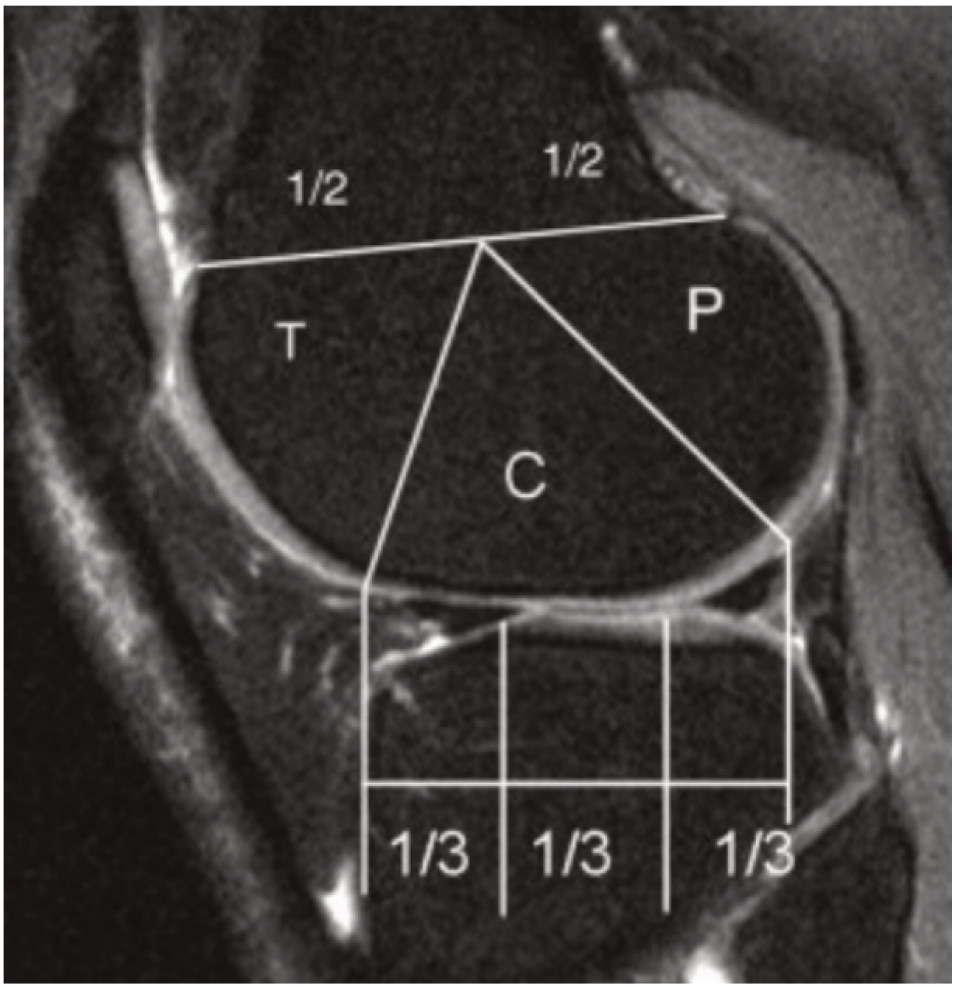


Figure 2.8: Anatomical segmentation of the femur into trochlea (T), central (C), and posterior (P) regions using a sagittal projection. This projection also illustrates the division of the tibia into anterior, central, and posterior sub-regions, segmented into equal thirds. [4]

MRI scans from a study. This included a re-assessment by one radiologist to check for consistency over time.

2.6.3.1 Regions and Scoring in MOAKS

1. **Subregions**: The MOAKS scoring system's division of the knee into 14 distinct subregions is critical for the precise assessment of OA and other injuries severity, as presented in Table 2.2. This subregional approach enhances diagnostic accuracy by allowing for the targeted grading of localized pathologies, such as bone marrow lesions and cartilage defects, which may have varying clinical implications depending on their location. Moreover, this detailed framework facilitates standardized assessments, ensuring consistency across clinical evaluations and research studies. Ultimately, the subregional analysis provided by MOAKS supports the development of more tailored, evidence-based treatment strategies

for knee OA.

Region	Subregions
Patella (Axial View)	Medial Patella (including median patellar ridge), Lateral Patella
Femur	Medial and Lateral Trochlea, Medial and Lateral Central Femur, Medial and Lateral Posterior Femur
Tibia	Anterior, Central, and Posterir Medial Articular Cartilage, Anterior, Central, and Posterior Lateral Articular Cartilage
Subspinous Region	Tibial Spines

Table 2.2: MOAKS Scoring System Subregions

2. **Bone Marrow Lesions (BML)**: Cysts are well-defined fluid-filled regions located adjacent to the subchondral bone plate, while Bone Marrow Lesions (BML) are ill-defined areas within the trabecular bone that exhibit a low signal on T1-weighted MRI sequences and a high signal on T2 fat-saturated (T2FS) MRI sequences, indicating the presence of bone marrow edema or damage.

   The MOAKS scoring system evaluates these bone marrow lesions and cysts by assessing the percentage of the subregional volume that they occupy, as well as the proportion of the lesion that is attributed to bone marrow versus cyst. The severity of these lesions is classified into grades from 0 to 3, with Grade 0 indicating no lesion and Grade 3 indicating that more than 66% of the subregional volume is affected.

   Specifically, the percentage values refers to the proportion of the affected subregion within the joint that is occupied by the lesion. This detailed scoring system, as presented in Table 2.3 and Figure. 2.9, allows for a standardized and quantifiable assessment of joint abnormalities in MRI, which is crucial for determining the extent of injury or degeneration. This structured approach not only facilitates consistent reporting but also aids in the evaluation and monitoring of the progression of joint diseases such as osteoarthritis.

Scoring	Grade 0	Grade 1	Grade 2	Grade 3
Subregional Volume of BML (including cysts)	None	<33%	33-66%	>66%
Percentage of Lesion that is BML vs Cyst	None	<33%	33-66%	>66%

Table 2.3: MOAKS Scoring for Bone Marrow Lesions and Cysts

3. **Articular cartilage**: Articular cartilage is a primary tissue affected by knee OA, a degenerative condition that impacts various joint structures and significantly contributes to the symptoms of the disease. In the MOAKS system, cartilage loss is meticulously assessed across different regions of the knee, excluding the subspinous region, as depicted in Table 2.4. This assessment is crucial as cartilage loss is a central feature of OA, leading to joint pain, stiffness, and reduced function.

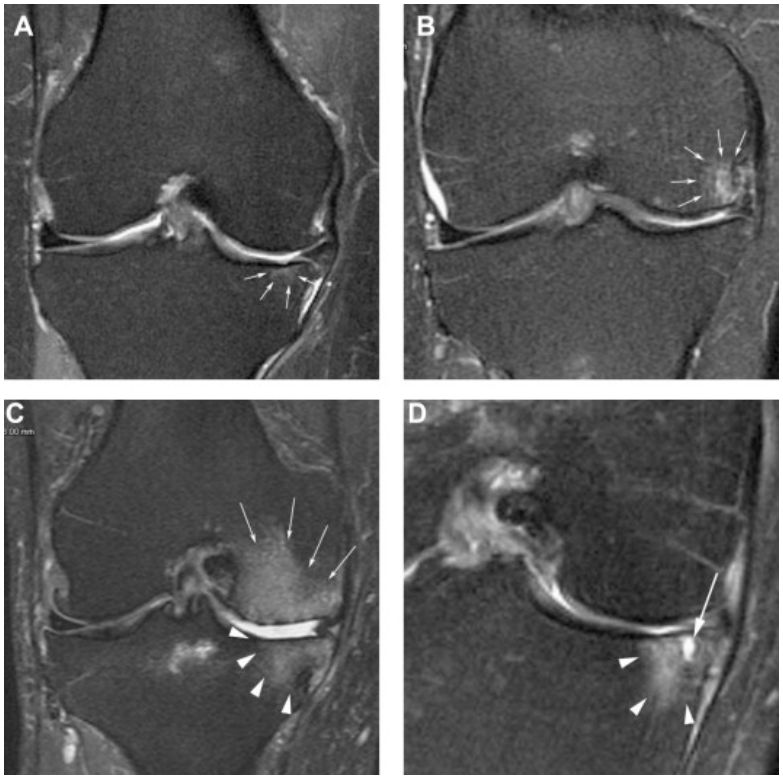


Figure 2.9: Grade 0 represents no bone marrow lesion (BML), Grade 1 indicates less than 33% involvement of the subregional volume, Grade 2 corresponds to 33–66% of the subregional volume, and Grade 3 involves more than 66% of the subregional volume. **(A)** Coronal T2-weighted image demonstrating a small Grade 1 BML in the central subregion of the medial tibia. **(B)** A Grade 2 BML observed in the central subregion of the medial femur. **(C)** Grade 3 BMLs identified in the central subregions of the medial femur (indicated by arrows) and the central medial tibia (indicated by arrowheads). **(D)** Coronal image depicting a BML with both a non-cystic/ill-defined component (arrowheads) and a cystic component (arrow) [4].

The MOAKS system evaluates cartilage loss in two distinct ways. First, it measures the size of the cartilage loss as a percentage of the total surface area within each specific sub-region, which helps in understanding the overall extent of cartilage degradation. Second, it assesses the percentage of the sub-region that exhibits full-thickness cartilage loss, which indicates the most severe form of damage where the cartilage is completely worn away, exposing the underlying bone.

The grading system categorizes cartilage loss into four grades based on the extent of cartilage affected within a specific subregion of the joint. Grade 0 indicates no cartilage loss, meaning the cartilage is fully intact. Grade 1 represents less than 10% loss, indicating that a small portion of the cartilage within the subregion is affected, but the majority remains intact. Grade 2 reflects a more significant loss, with 10-75% of the subregional cartilage being affected, showing a substantial area of cartilage degradation. Grade 3 indicates a

Figure 2.10: Articular cartilage in sagittal knee MRI [5].

severe level of damage, with more than 75% of the cartilage in the subregion lost, often resulting in exposure of the underlying bone.

This detailed evaluation allows clinicians to assess both partial and full-thickness cartilage loss, which is crucial for understanding the severity of OA and how the disease progresses in different regions of the joint. The percentage values provide a standardized way to quantify the extent of cartilage damage, enabling the development of effective treatment strategies aimed at preserving joint function and managing OA symptoms. Understanding the precise degree of cartilage loss is essential for tailoring interventions to slow the progression of the disease and maintain the quality of life for patients [4].

Scoring	**Grade 0**	**Grade 1**	**Grade 2**	**Grade 3**
Cartilage Loss (% of Subregion)	None	<10%	10-75%	>75%
Full-Thickness Loss (% of Subregion)	None	<10%	10-75%	>75%

Table 2.4: MOAKS Scoring for Articular Cartilage Lesions

4. **Osteophytes**: Osteophytes are bony outgrowths that form at the edges of joints affected by OA, resulting from a process known as endochondral ossification. These growths are recognized in radiographic studies as key indicators of the disease [42]. In the MOAKS system, osteophytes are carefully evaluated at each of the 12 specific locations within the joint (see Table 2.5 and Figure. 2.11. Each osteophyte is graded based on its size, with the

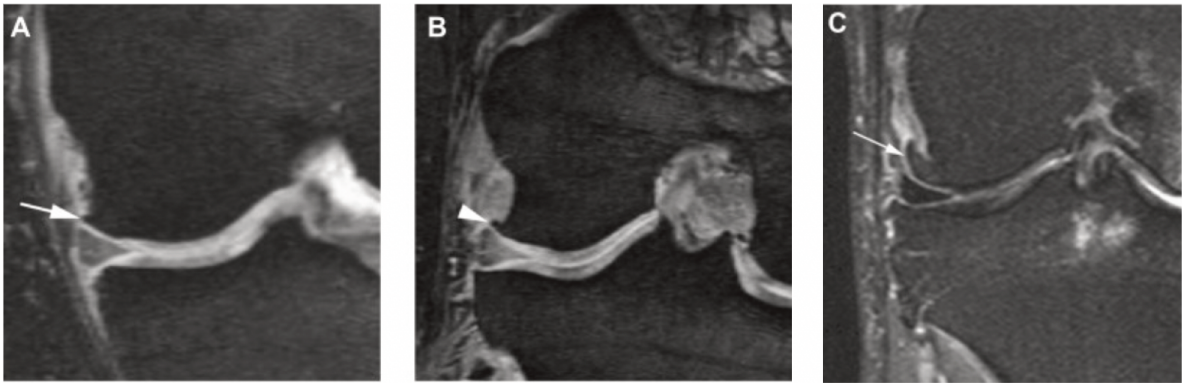


Figure 2.11: Osteophytes are graded by their protuberance from the joint surface: **(A)** Grade 1 osteophyte on the medial femur, **(B)** Grade 2 osteophyte on the lateral femur, and **(C)** Grade 3 osteophyte on the lateral femur. Grading reflects the extent of protrusion rather than the total volume of the osteophyte. [4].

scoring system as follows: Grade 0 signifies no osteophyte presence; Grade 1 denotes a small osteophyte; Grade 2 indicates a medium-sized osteophyte; and Grade 3 represents a large osteophyte [4].

The size grading for osteophytes focuses on the protuberance, or how far the osteophyte extends from the joint surface, rather than the total volume of the osteophyte. Within each location, the largest osteophyte is the one that is scored, ensuring that the most significant growths are assessed for their potential impact on joint function and health.

In our thesis, we assess the presence of osteophytes at the scan level. This means that if osteophytes are detected on either the lateral or medial side of the joint in any scan, we classify the entire scan as indicating the presence of osteophytes. This approach allows us to broadly determine whether osteophytes are affecting the joint, which is crucial for understanding the overall severity and progression of OA in the studied cases.

Region	**Subregion**	**Grade 0**	**Grade 1**	**Grade 2**	**Grade 3**
Anterior Femur	Medial, Lateral (sagittal/axial)	None	Small	Medium	Large
Posterior Femur	Medial, Lateral (sagittal/axial)	None	Small	Medium	Large
Central Femur	Medial, Lateral (coronal)	None	Small	Medium	Large
Patella	Superior, Inferior (sagittal); Medial, Lateral (axial)	None	Small	Medium	Large
Tibia	Medial, Lateral (coronal)	None	Small	Medium	Large

Table 2.5: MOAKS Scoring for Osteophytes in Various Regions

5. **Hoffa's synovitis and synovitis-effusion**: Table 2.6 presents the MOAKS scoring system for mild chronic synovitis and effusion-synovitis, two key indicators of inflammation in OA. Chronic synovitis is graded on T2/PD FS sequences in the sagittal view based on the signal intensity within Hoffa's fat pad, ranging from 0 (normal) to 3 (severe). Effusion-synovitis is

assessed in the axial view, with grades from 0 (physiological volume) to 3 (large, indicating capsular distension). These scores are crucial for quantifying the extent of inflammation, which is frequently observed in OA and has been linked to pain and other clinical outcomes, though the literature on this correlation is somewhat conflicting.

Quantitative MRI markers of synovitis include the measurement of synovial tissue thickness or volume and the rate of synovial enhancement following the intravenous injection of contrast material. To accurately evaluate regions affected by Hoffa-synovitis, T2/IW- or PD-weighted fat-saturated images in the mid-slices of the sagittal plane are suggested. These imaging sequences are essential for detecting and grading synovitis and effusion, thereby providing a comprehensive understanding of the inflammatory processes that contribute to the symptoms and progression of OA [4].

Scoring	**Grade 0**	**Grade 1**	**Grade 2**	**Grade 3**
Chronic Synovitis	Normal	Mild	Moderate	Severe
Effusion-Synovitis	Physiological	Small	Medium	Large

Table 2.6: MOAKS Scoring for Synovitis and Effusion-Synovitis

6. **Meniscus**: The meniscus is a vital structure in the knee, contributing to load-bearing, shock absorption, joint stability, and lubrication. Changes in meniscal position, such as subluxation or extrusion, and morphological alterations like tears or substance loss, are strongly associated with an increased risk of cartilage damage [59]. In the MOAKS system, meniscal position and morphology are systematically graded to assess these critical changes.

   Meniscal extrusion, a key positional change, is graded based on the extent of the meniscus extruding beyond the tibial margin. The grading ranges from less than 2 mm (Grade 0) to more than 5 mm (Grade 3), and it is evaluated in four specific locations: medial meniscus extrusion relative to the medial tibial margin (coronal), maximal anterior extrusion (sagittal), lateral meniscus extrusion relative to the lateral tibial margin (coronal), and maximal anterior extrusion (sagittal). Morphological features of the meniscus are also carefully evaluated and scored as present (yes) or absent (no) in both the medial and lateral menisci, focusing on the anterior and posterior horns (sagittal) and the body (coronal). Specific features scored include various types of meniscal tears—such as vertical, horizontal, radial, complex, and root tears—along with partial and complete maceration, meniscal cysts, and hypertrophy, as illustrated in Figure. 2.12 and Table. 2.8.

   In MOAKS, meniscus tears are specifically assessed across six distinct locations, divided equally between the medial and lateral menisci. The anterior and posterior horn regions are scored using sagittal MRI sequences, while the body region is assessed using coronal sequences. This comprehensive approach ensures that all significant meniscal alterations

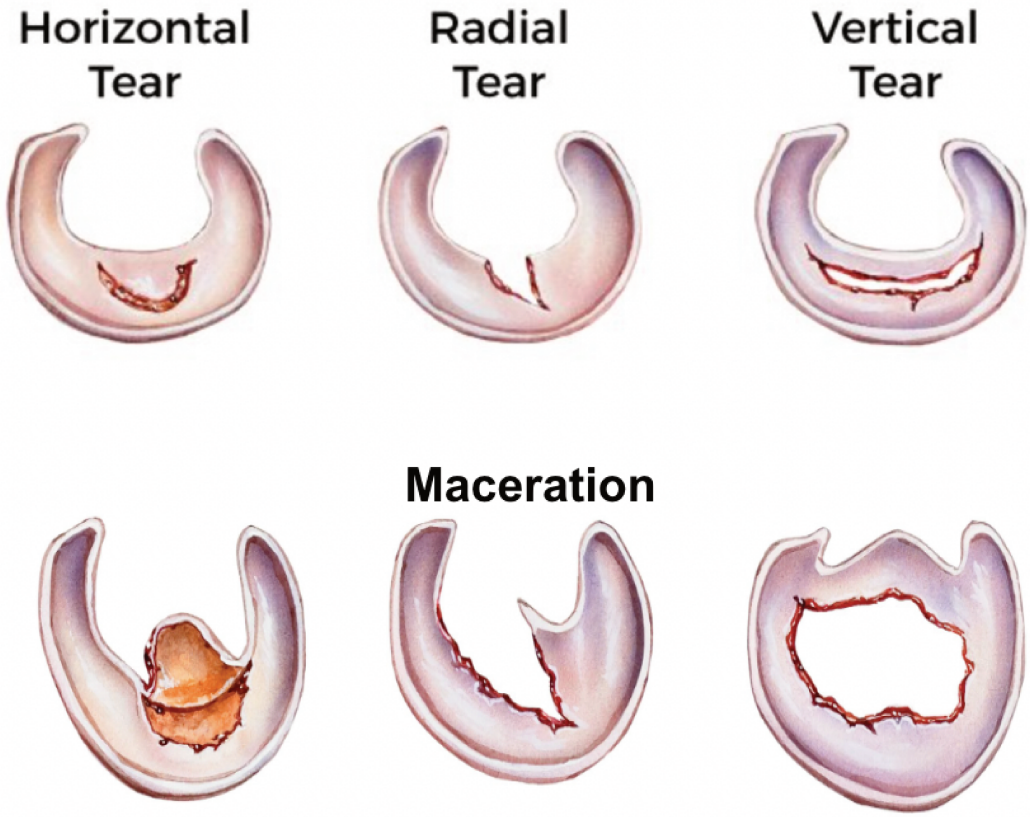


Figure 2.12: Meniscus tears levels

are captured, providing a detailed understanding of the meniscus's role in joint health and its impact on cartilage integrity [4].

Scoring	**Grade 0**	**Grade 1**	**Grade 2**	**Grade 3**
Medial Meniscus Extrusion (Coronal)	<2 mm	2-2.9 mm	3-4.9 mm	>5 mm
Maximal Anterior Extrusion (Sagittal)	<2 mm	2-2.9 mm	3-4.9 mm	>5 mm
Lateral Meniscus Extrusion (Coronal)	<2 mm	2-2.9 mm	3-4.9 mm	>5 mm
Maximal Anterior Extrusion (Sagittal)	<2 mm	2-2.9 mm	3-4.9 mm	>5 mm

Table 2.7: MOAKS Scoring for Meniscus Extrusion in Various Locations

Abnormalities of the meniscus in MOAKS are scored as follows:

7. **Ligaments and Tendons** The MOAKS scoring system also includes a detailed evaluation of the ligaments and tendons in the knee, particularly focusing on the anterior cruciate ligament (ACL), posterior cruciate ligament (PCL), and patellar tendon (see Figure. 2.13). These structures are crucial for knee stability and function, and their integrity is often compromised in OA and other knee pathologies [60].

   For the ACL and PCL, the scoring differentiates between a normal ligament and a complete

Score	Description
0	Normal Meniscus
1	Signal abnormality not severe enough to be considered a meniscal tear
2	Radial Tear
3	Horizontal Tear
4	Vertical Tear
5	Complex Tear
6	Partial Maceration
7	Progressive Partial Maceration (used for follow-up visit scores)
8	Complete Maceration

Table 2.8: MOAKS Scoring for Meniscus Abnormalities

tear, which is a critical distinction as complete tears can lead to significant joint instability and are commonly associated with traumatic injuries. The scoring also accounts for the presence of associated bone marrow lesions or cysts at the sites of insertion or origin, which can indicate chronic stress or injury to these ligaments.

In cases where the ACL has been repaired, this is specifically noted, as it affects both the prognosis and the approach to treatment. The patellar tendon is evaluated for signal abnormalities, which may indicate tendinopathy or other degenerative changes that can contribute to anterior knee pain. This scoring system is important because it allows clin-

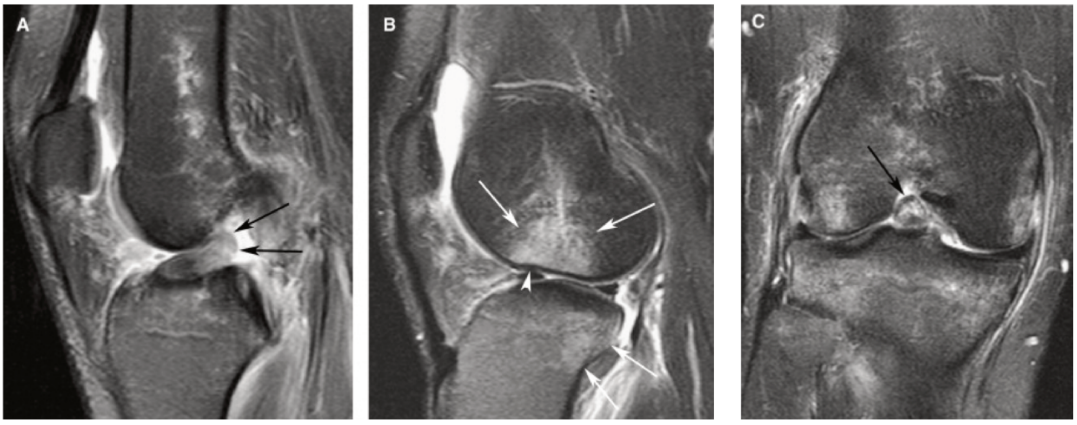


Figure 2.13: Acute Complete Anterior Cruciate Ligament (ACL) Tear. **(A)** Mid-sagittal fat-saturated fast spin echo (FSE) T2-weighted image displaying the distal portion of the ruptured ACL, which appears thickened, hyperintense, and horizontally oriented (arrows). The ligament's axis is abnormally flattened, deviating from Blumensaat's line. **(B)** Lateral parasagittal fat-saturated FSE T2-weighted image showing bone contusions (arrows) on the lateral femoral condyle and posterolateral tibial plateau. A deepened lateral femoral sulcus (arrowhead) and anterior tibial translation are also observed. **(C)** Coronal fat-saturated FSE T2-weighted image highlighting the thickened ACL with increased signal intensity within its fibers (arrow) [6].

icians to systematically assess the condition of these key structures, ensuring that any

abnormalities are identified and factored into the overall diagnosis and treatment plan for knee OA or other conditions.

Structure	**Scoring**
Anterior Cruciate Ligament	Normal vs Complete Tear Associated Bone Marrow Lesion/Cyst at Insertion/Origin ACL Repair
Posterior Cruciate Ligament	Normal vs Complete Tear Associated Bone Marrow Lesion/Cyst at Insertion/Origin
Patellar Tendon	Normal Signal vs Signal Abnormality

Table 2.9: MOAKS Scoring for Ligaments and Tendons

2.6.3.2 Knee OA definition in MOAKS

Tibiofemoral knee OA definition in MRI was presented by researchers using MOAKS to facilitate the analysis of this condition [16].

The definition mentions that the determination of knee OA is based on the presence of specific features categorized into two groups, as outlined in Tables 2.10 and 2.11.

Table 2.10: MOAKS System Features for Diagnosing Knee OA

Criteria	**Group A Features**	**Group B Features**
Defining Features	• Definite osteophyte formation ≥ 2 • Full-thickness cartilage loss ≥ 3	• Subchondral bone marrow lesion or cyst ≥ 1 • Meniscal subluxation, maceration, or degenerative tear ≥ 2 • Partial-thickness cartilage loss (where full-thickness loss is not present) ≥ 1

Table 2.11: Diagnosis Criteria for Tibiofemoral OA

Diagnosis Criteria for Tibiofemoral OA	**Condition**
Criterion 1	Presence of both Group A features.
Criterion 2	Presence of one Group A feature and two or more Group B features.

In this work, we opted for the definition of group A. We could identify 450 patients from the OAI database that meet the definition criteria.

This objective and quantifiable definition of knee OA using the MOAKS system provides a robust framework for training and evaluating deep learning models, enabling us to develop algorithms capable of accurately identifying and classifying OA severity based on MRI findings.

2.7 Conlusion

In this chapter, we examined the pivotal role of imaging and semi-quantitative scoring systems in the evaluation and management of knee OA and related knee injuries. Semi-quantitative scoring systems provide a detailed assessment of the knee by dividing the joint into multiple regions and assigning scores based on the severity of abnormalities. While these systems are powerful tools, they come with certain limitations, such as the need for extensive training, potential issues with sensitivity, and the costs associated with specialized interpretation.

We discussed several key scoring systems, including the Whole-Organ Magnetic Resonance Imaging Score (WORMS), the Boston-Leeds Osteoarthritis Knee Score (BLOKS), and the Magnetic Resonance Imaging Osteoarthritis Knee Score (MOAKS). WORMS offers a subregional approach that is particularly useful for evaluating cartilage, bone marrow lesions, and subchondral cysts. BLOKS, in contrast, focuses on nine intra-articular regions and assesses features such as bone marrow lesions, cartilage integrity, and ligament conditions.

Additionally, we explored MOAKS, another important scoring system that has been widely used in knee OA research. MOAKS provides a comprehensive framework for scoring various knee injuries, including cartilage damage, bone marrow lesions, and meniscal tears. By incorporating MOAKS into our discussion, we highlighted how different knee injuries are assessed and scored, further emphasizing the importance of these systems in understanding the progression of OA and related conditions.

CHAPTER 3

BACKGROUND AND LITERATURE REVIEW

This chapter presents an overview of deep learning techniques used in medical imaging analysis, including image classification, segmentation, and object detection. We briefly introduce some key concepts essential for understanding the contributions presented in the manuscript. Additionally, this chapter reviews relevant research related to the thesis contributions.

3.1 Deep learning in medical imaging

Medical imaging analysis, crucial for diagnosis and treatment planning, often requires expert interpretation of complex image data [61]. Computer vision techniques, particularly deep learning models, offer powerful tools for automating this process [62]. More specifically, Convolutional neural networks (CNNs) have shown remarkable success in extracting relevant features from various types of medical images across multiple tasks like segmentation, classification, and anomaly detection [7].

However, multiple challenges face researchers when applying deep learning to this field [63]. Medical image datasets are often characterized by high variability, stemming from differences in acquisition protocols, scanners used, and patient demographics [64]. This variability, coupled with the inherent imbalance often seen in medical conditions (some being far rarer than others) [65], presents challenges for training robust and generalizable models [8]. Furthermore, deep learning algorithms typically require larger annotated datasets for optimal performance [66]. In the medical domain, where expert annotation is essential, generating such datasets can be prohibitively expensive and time-consuming [67].

Beyond these challenges, extracting clinically relevant features from medical images presents its own unique difficulties [68, 69]. Unlike natural images, subtle variations in texture, shape, or

intensity can hold significant diagnostic weight. This specificity necessitates architectures and training strategies specifically tailored to the complexities inherent in medical image data [70].

In this thesis, we target multiple challenges faced by researchers in training deep learning models effectively across various medical imaging analysis tasks.

3.1.1 Fundamentals of Deep Learning

Deep learning, a subfield of machine learning, has revolutionized various fields, including medical image analysis, by enabling computers to learn intricate patterns and representations directly from data [71]. This learning process is driven by artificial neural networks (ANNs), computational models inspired by the biological nervous system, which consist of interconnected neurons organized in layers.

It is worth noting that the following concepts have been widely applied for deep learning research, though we decided in this manuscript to briefly mention some of the key concepts and refer the reader to cited references for a deeper understanding of these concepts.

3.1.1.1 Neural Networks and Architectures

A fundamental concept in deep learning is the artificial neuron, a mathematical function that mimics the behavior of biological neurons. Each neuron receives input signals, applies weights to these signals, sums them up, and then passes the result through an activation function to produce an output [71]. The activation function introduces non-linearity, allowing the network to learn complex relationships within the data. Common activation functions include:

- Sigmoid: Squashes the output to a range between 0 and 1, often used for binary classification.
- ReLU (Rectified Linear Unit): Outputs the input directly if positive, otherwise outputs 0, known for its computational efficiency and effectiveness in many applications.
- Softmax: Normalizes the output to a probability distribution over multiple classes, commonly used for multi-class classification.

Neural networks are organized in layers, with each layer performing a specific transformation on the input data. Different network architectures are designed for various tasks and data types. Common architectures include:

- Feedforward Neural Networks: These networks process data in one direction, from input to output, without any feedback loops. They are often used for simple tasks, such as linear regression and classification.
- CNNs: CNNs are particularly well-suited for image analysis due to their ability to learn spatial hierarchies of features [72]. They utilize convolutional layers, which apply learned

filters to the input image to extract relevant features. These filters act as feature detectors, capturing local patterns such as edges, textures, and shapes. By stacking multiple convolutional layers, CNNs can learn hierarchical representations, progressing from low-level features to more abstract and complex patterns. CNNs also typically incorporate pooling layers, which downsample feature maps to reduce dimensionality and make the network more robust to variations in the input. Common pooling operations include max pooling and average pooling.

- Recurrent Neural Networks (RNNs): RNNs are designed for sequential data, such as time series and natural language. They have feedback loops, allowing them to retain information from previous inputs and capture temporal dependencies [73].

In this thesis, we primarily focus on CNNs due to their demonstrated effectiveness in medical image analysis, particularly for tasks involving knee MRI data [20, 47, 74–76].

3.1.1.2 Training Deep Learning Models

Training a deep learning model involves optimizing its parameters to minimize the difference between its predictions and the actual values (ground truth) on a given task. This optimization process relies on several key components:

- Dataset: A large, well-annotated dataset is crucial for training a deep learning model. The dataset provides examples for the model to learn from and allows it to generalize to unseen data.

- Loss Function: The loss function quantifies the difference between the model's predictions and the ground truth. Common loss functions for classification tasks include cross-entropy loss and mean squared error. The choice of loss function depends on the specific task and the nature of the data.

- Optimizer: The optimizer is an algorithm that adjusts the model's parameters (weights and biases) to minimize the loss function. Common optimizers include stochastic gradient descent (SGD), Adam, and RMSprop. The optimizer determines how the model learns from the data and how it updates its parameters.

- Backpropagation: Backpropagation is the algorithm used to calculate the gradients of the loss function with respect to the model's parameters. These gradients are then used by the optimizer to update the parameters. Backpropagation allows the network to learn by adjusting its weights based on the errors it makes during training [73].

- Hyperparameters: Hyperparameters are settings that control the training process, such as the learning rate, batch size, and number of epochs. Tuning hyperparameters is crucial for optimal model performance.

3.1.2 Applications in Medical Imaging

3.1.2.1 Image Classification

This dissertation investigates the utilization of deep learning methodologies for the classification of medical images, specifically concentrating on the detection of knee OA and associated injuries using MRI. Although the primary emphasis of this research is the analysis of MRI scans, the underlying principles and techniques are applicable to a wider spectrum of 2D image classification tasks prevalent across different medical imaging modalities, such as X-rays [77] and computed tomography (CT) scans [78].

Neural networks recognize complex patterns within image data, making them powerful tools for 2D and 3D medical image analysis [7]. By leveraging deep learning models, we aim to improve the accuracy and efficiency of diagnosing knee OA and other musculoskeletal conditions, ultimately leading to more effective treatment and management approaches.

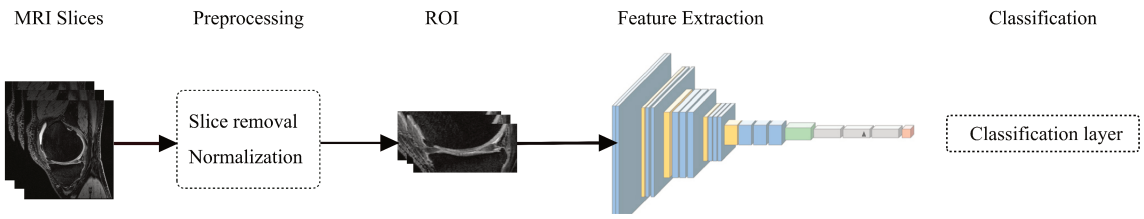


Figure 3.1: Generic overflow of image MRI classification pipleline.

Formally speaking, the pipeline for classifying knee MRI volumes begins with preprocessing, which includes normalizing the sizes and intensities of the MRI scans and enhancing image quality through techniques such as denoising [79], we will discuss in the following sections some preprocessing steps (see Fig. 3.1). This step is crucial for preparing the data for effective learning and classification by the model. Following the preprocessing stage, an appropriate deep learning model is selected and meticulously trained to identify and categorize the intricate patterns present in knee MRI scans. The training regimen entails supplying the model with a substantial dataset of knee MRI images, enabling it to discern the characteristic features of knees afflicted with and without OA. This methodical approach facilitates the model's ability to accurately classify the scans based on the presence or absence of OA indicators.

3.1.2.2 Image Segmentation

Image segmentation in medical imaging involves dividing an image into multiple segments to isolate specific areas or structures, facilitating detailed analysis and treatment planning [80]. The pipeline for this process begins with preprocessing the images to normalize sizes and intensities, using techniques like augmentation to enhance image quality. A specialized deep learning model is then trained to identify and delineate different regions within the images (as illustrated in

Figure. 3.2). This model is meticulously trained to recognize the boundaries and features relevant to specific anatomical or pathological areas. After the model has been trained, it is validated and tested on separate datasets to assess its precision in segmenting new, unseen images. Based on these performance evaluations, adjustments and fine-tuning are made to optimize the model's accuracy and efficiency.

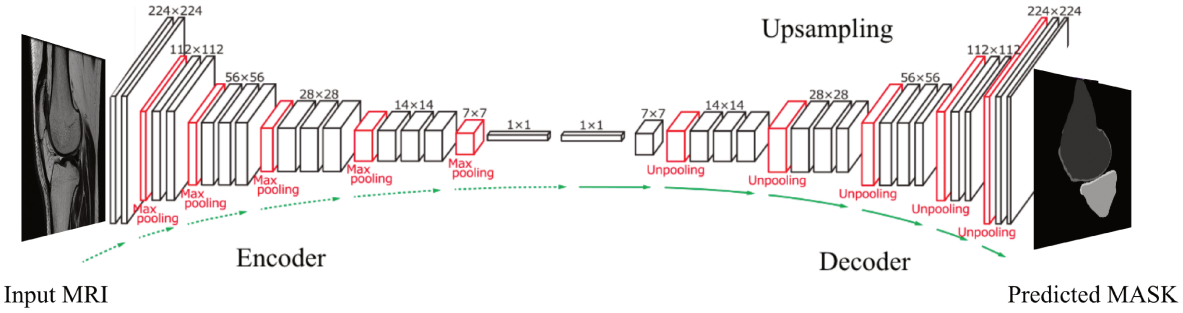


Figure 3.2: Generic overflow of image MRI segmentation pipleline.

Image segmentation is particularly useful in medical imaging because it allows for precise measurements and visualizations of structures, which are vital for surgical planning and disease monitoring [80]. Common use cases include tumor delineation in oncology [81], organ segmentation for transplant preparation [82], and vascular segmentation for cardiovascular assessments [83].

knee MRI segmentation, where the model segments the MRI to accurately outline structures such as bones, cartilage, and ligaments. This is crucial for assessing joint health, diagnosing conditions like OA, and planning orthopedic surgeries [84–86].

3.1.3 Advanced Deep Learning Concepts

Multiple neural networks have been proposed in the last decade to advance computer vision tasks [72, 87, 88]. In this thesis, we are particularly interested in specific neural networks that have reported consistent performance across various experiments.

3.1.3.1 3D CNNs

The inherent complexity of knee MRI data, characterized by its high-dimensional spatial structure, poses significant analytical challenges. MRI scans, being volumetric, encapsulate intricate anatomical details across three dimensions, which necessitates sophisticated computational approaches for effective analysis. 3D CNNs are particularly adept at exploiting this spatial complexity, enabling robust feature extraction crucial for both classification and segmentation tasks in medical imaging [89].

MRI generates volumetric data where each volume consists of a sequence of slices, each representing a detailed cross-section of the examined anatomy. This volumetric representation

captures complex spatial relationships that are not fully exploitable by traditional 2D image processing methods. The continuity and the spatial context provided across the slices are crucial for understanding the structural and functional characteristics of tissues, making 3D analysis indispensable.

3D CNNs extend the capabilities of their 2D counterparts by incorporating an additional dimension into their convolutional filters, allowing for simultaneous processing along the x, y, and z axes. This adaptation is crucial for handling the three-dimensional data from MRI scans. The convolution operation in 3D CNNs can be formulated as follows:

$$V'_{xyz} = \sum_{i,j,k} (K_{ijk} \cdot V_{x+i,y+j,z+k})$$

where the variables are defined as follows:

- V represents the input volume, which in the context of MRI, is the 3D matrix of the scanned images.
- K denotes the 3D convolutional kernel, a smaller cube-like matrix used to filter and process the input volume.
- V' is the output volume or feature map, which results from applying the convolution operation.
- x, y, z are the spatial coordinates in the input volume, indicating the position of a voxel (3D pixel) in the volume.
- i, j, k traverse the dimensions of the kernel K, allowing the kernel to be applied over the entire volume.

It's important to note that the term "*CNN model*" encompasses more than just convolutional layers. It refers to a sophisticated architecture often comprising pooling layers (which, in the case of 3D CNNs, are also implemented in 3D), batch normalization, activation functions, dropout layers, and more. These components work in concert with convolutional layers to facilitate efficient and effective feature extraction. When adapting these models for 3D medical images like MRIs, these additional layers must also be adjusted and implemented in 3D to accommodate the volumetric nature of the input data.

Developing custom models tailored to the specific demands of MRI processing often proves beneficial. This process, however, is not without its complexities. It involves a careful selection of architectural elements, such as the number and type of convolutional layers, pooling strategies, and regularization techniques. Rigorous experimentation and validation are crucial, requiring researchers to carefully tune hyperparameters and evaluate performance metrics on relevant datasets. Despite these challenges, the potential to design models specifically attuned to the

nuances of medical image data makes this a compelling avenue for exploration, often leading to improved accuracy and efficiency in clinical applications.

It should be noted also that most studies in the literature review in OA detection are based on CNNs [9].

3.1.3.2 Handling 3D Data

Analyzing 3D medical images, such as MRI scans, presents unique challenges due to the high dimensionality and complexity of the data. Traditional 2D CNNs, while effective for image analysis, may not fully capture the spatial relationships and contextual information present in 3D volumes.

Several approaches can be employed to handle 3D data effectively:

- 3D CNNs: As discussed earlier, 3D CNNs directly process volumetric data using 3D convolutional filters, enabling the extraction of features that capture spatial correlations across all three dimensions.
- Slice-based Analysis: This approach involves extracting 2D slices from the 3D volume and processing them individually using 2D CNNs. The results from each slice can then be aggregated to obtain a 3D representation. This method can be computationally less demanding than 3D CNNs but may not fully capture the inter-slice dependencies.
- Hybrid Approaches: Combining 2D and 3D CNNs can leverage the strengths of both approaches. For example, a model might use 2D CNNs to extract features from individual slices and then use a 3D CNN to aggregate these features and capture inter-slice relationships.

Choosing the appropriate approach depends on the specific task, the available computational resources, and the characteristics of the data.

3.1.3.3 ImageNet Models

Pretrained models on ImageNet [17], have proven invaluable for numerous computer vision tasks [90, 91]. These models, often comprising deep convolutional architectures, learn rich feature representations that can be beneficial even for domains distinct from natural images, such as medical imaging.

However, directly fine-tuning a 2D ImageNet model on 3D medical data like MRI is often suboptimal [92]. The inherent structural differences between 2D and 3D data, coupled with the domain shift between natural images and medical scans, limit the effectiveness of this approach.

A more promising strategy involves adapting the architectural principles behind successful ImageNet models to the 3D medical imaging domain. Instead of direct fine-tuning, we can leverage

the proven architectural designs of models like ResNet [93], InceptionV3 [94], and others, and train them from scratch on 3D MRI scans. This allows us to capitalize on the strengths of these architectures while ensuring compatibility with the volumetric nature of our data.

Below, we briefly outline the ImageNet-inspired models employed in our experiments, highlighting their key characteristics and how we adapted them for 3D medical image analysis:

- **VGG-16**: Developed by Karen Simonyan and Andrew Zisserman [95] in 2014, VGG16 is a cornerstone in the evolution of deep learning for visual understanding. Its defining characteristic is its deep architecture, comprising 16 convolutional layers. This depth, a significant departure from shallower networks of the time, enabled VGG16 to learn a richer hierarchy of features, leading to improved accuracy in image classification tasks. VGG16 secured second place in the ImageNet competition, achieving a 7.3% error rate. The model's success stems from its consistent use of small 3x3 convolutional filters, promoting parameter efficiency and enabling deeper architectures.

  For processing 3D medical images, we adapt VGG16 by substituting its 2D convolutional layers with their 3D counterparts. This modification allows the network to effectively analyze volumetric data, capturing spatial information across all three dimensions. The resulting 3D VGG16 retains the original model's strong feature extraction capabilities while accommodating the depth dimension essential for medical image analysis. This adaptation has been successfully applied to tasks such as MRI brain tumor segmentation and lung nodule detection.

- **ResNet**: Introduced by Kaiming He et al. [93] in 2015, ResNet revolutionized deep learning by addressing the vanishing gradient problem that hindered the training of very deep networks. This breakthrough was achieved through the introduction of residual connections, also known as skip connections. These connections allow gradients to flow more easily through the network during backpropagation, enabling the training of networks with hundreds or even thousands of layers. ResNet achieved a remarkable 3.57% error rate on the ImageNet competition, setting a new benchmark for image classification accuracy. The key innovation of residual blocks allows the network to learn residual functions, making it easier to optimize and learn complex mappings.

  The adaptation of ResNet for 3D medical imaging involves replacing the 2D convolutions with 3D convolutions. This modification enables ResNet to process volumetric MRI data while preserving the benefits of residual connections. The 3D ResNet effectively captures spatial relationships within the data, critical for accurate medical imaging diagnostics, including tasks such as organ segmentation and disease classification.

- **InceptionV3**: Developed by Szegedy et al. [94], InceptionV3 marked a significant advancement in convolutional neural network architecture. Its distinctive feature is the use of

inception modules. These modules allow the network to concurrently process information at multiple scales, capturing both fine-grained details and larger contextual information. This multi-scale analysis significantly improved upon previous models, achieving a top-5 error rate of 5.6% on ImageNet. Inception modules are computationally efficient due to their parallel structure and the use of 1x1 convolutions for dimensionality reduction.

The 3D adaptation of InceptionV3 incorporates 3D convolutions within its inception modules. This enables the network to effectively process volumetric data, such as MRI scans, extracting nuanced features from the complex spatial relationships inherent in medical images. This adaptation is particularly beneficial for tasks requiring detailed and layered medical analysis, such as tissue classification and anomaly detection.

- **DenseNet**: Introduced by Huang et al. [96] in 2017, DenseNet presents a novel architecture characterized by dense connectivity. In DenseNet, each layer is directly connected to every subsequent layer in a feed-forward fashion. This dense connectivity pattern promotes feature reuse and strengthens feature propagation throughout the network. This design proved highly effective, achieving a 4.51% error rate on ImageNet. The dense connections also mitigate the vanishing gradient problem, allowing for deeper networks.

  Adapting DenseNet for 3D MRI data involves implementing 3D convolutions while maintaining the network's characteristic dense connectivity. This adaptation enhances feature propagation and minimizes information loss across the network, which is crucial for precise segmentation and analysis of medical images. The 3D DenseNet's ability to learn complex relationships within volumetric data makes it well-suited for tasks like brain tumor segmentation and organ boundary delineation.

- **EfficientNet**: Proposed by Mingxing Tan and Quoc V. Le [97] in 2019, EfficientNet represents a paradigm shift in model scaling. It achieves state-of-the-art accuracy on ImageNet, with a top-1 error rate of 15.5% and a top-5 error rate of 2.3%, while utilizing significantly fewer parameters and computational resources compared to its predecessors. EfficientNet's efficiency stems from its compound scaling method, which uniformly scales the network's depth, width, and resolution using a set of carefully determined scaling coefficients. This balanced scaling approach ensures optimal resource utilization and maximizes accuracy.

  Our 3D adaptation of EfficientNet extends these principles to the realm of volumetric data. By employing 3D convolutions and adhering to the compound scaling methodology, the 3D EfficientNet efficiently manages the spatial complexity inherent in MRI data. This adaptation ensures both computational efficiency and diagnostic efficacy, making it a promising architecture for resource-constrained medical imaging applications.

In the subsequent chapters, we will delve into various custom CNN models that we have specifically developed and trained from the ground up. These models have been tailored for a range

of tasks and have undergone extensive experimentation. The discussion will provide detailed insights into the architectural decisions, development processes, and training methodologies. Additionally, we will assess the performance of these models across diverse datasets and scenarios.

3.1.4 Data Preparation and Region of Interest (ROI) Detection

3.1.4.1 Data preparation

Image processing plays a crucial role in the fields of knee MRI classification and segmentation, enhancing the performance and accuracy of diagnostic models. In our experiments, to increase the models' ability for better feature extraction, we performed the following preprocessing steps:

- **Slice Normalization**: A standard practice in image processing and deep learning is to scale the range of pixel intensity values of the input images. This normalization ensures that the backpropagation algorithm, used to train the models, is more stable and converges faster. This process involves changing the range of pixel intensity values. In this study, we considered two intensity normalization techniques and evaluated the impact of each of them. The first technique, Max-min normalization (see (3.1)), consists of scaling the intensities of MRI slices in the range [0-1]. The second technique, referred to as whitening, consists of centering slice intensities and reducing their variance to the unit.

  $$I_{\text{norm}} = \frac{I - \text{min\_intensity}}{\text{max\_intensity} - \text{min\_intensity}} \tag{3.1}$$

  where the value of I represents the original intensity value of a pixel in the slice.

- **Reduction of Dimensionality**: MRI data typically consist of high-resolution 3D volumes that can be computationally intensive to process. Image processing techniques such as cropping and slicing are used to reduce the dimensionality and focus on relevant sections of the scans, thereby decreasing the computational load and speeding up the analysis without sacrificing important anatomical details.

In medical image analysis, focusing computational resources on the most informative regions can significantly enhance both efficiency and accuracy. This is where ROI detection becomes crucial. By localizing the knee joint within the MRI scan, our models focus on feature extraction from the most pertinent areas.

3.1.4.2 ROI Detection

Particularly for knee MRI analysis, the joint region harbors the key anatomical structures implicated in conditions like knee OA, meniscus tears, and cartilage defects. ROI detection, therefore, allows us to narrow our focus, minimizing the influence of extraneous image data and allowing the model to learn more discriminative features related to the knee joint's pathology. This targeted approach leads to a more efficient use of computational resources and often translates to

improved diagnostic accuracy.

Another reason we perform ROI detection is due to the *Out-of-Plane Artifact* [98], which leads to the phenomenon of first slices and ending slices not showing any relevant information related to the knee joint or full black images (see Fig. 3.3). These slices represent +30% of overall slices in some knee MRI acquisitions, adding noisy features to the models' training. Additionally, due to factors such as patient movement, incorrect slice planning, or technical limitations, the middle slices in a knee MRI sometimes do not illustrate the knee structure clearly, as illustrated in `Slice N 130`.

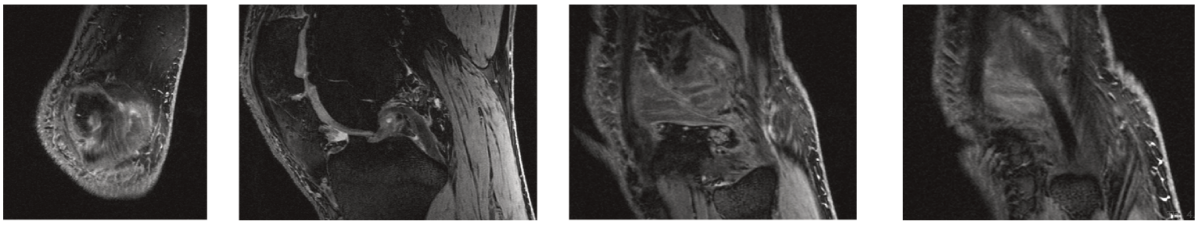


Figure 3.3: Samples of irrelevant MRI slices that do not show the knee joint area from SAG 3D-DESS scans.

In our thesis, for knee ROI detection we propose two approaches that we initially perform before passing the MRI data to model training:

Semi-automatic ROI detection Our initial approach employed a semi-automatic method for ROI localization (see Fig. 3.4). By examining a representative subset of MRI scans, we manually established slice ranges encompassing both irrelevant (primarily empty) slices and those containing the knee joint structure by reviewing multiple MRI scans. This process was tailored to each MRI view and acquisition protocol, as variations in imaging parameters and slice planning necessitate distinct regions of interest.

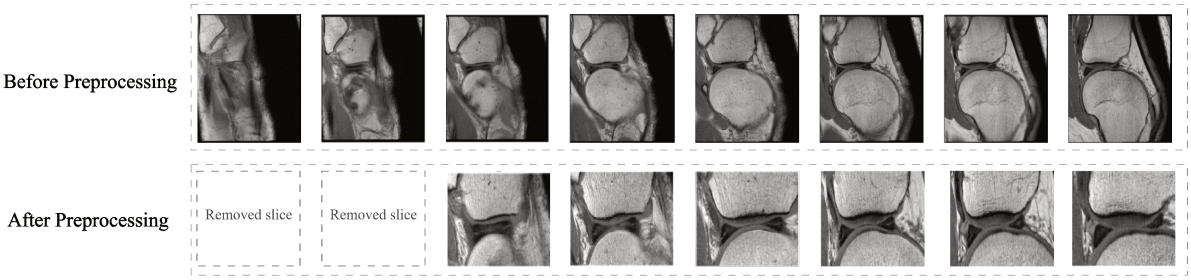


Figure 3.4: Semi-automatic way for knee joint detection and irrelevant slices removal.

Once these slice ranges and knee joint regions were defined on our representative samples, we implemented an automated workflow to process the remaining training data using the obtained

range. This workflow leveraged the predefined regions, ensuring consistency in ROI extraction across all scans for the specific acquisition protocol.

An advantage of this approach is that it's lightweight and doesn't require any model tuning or training to detect and crop the ROI. However, it is important to note that not all MRI scans have the same range of irrelevant slices at the beginning and end of the imaging protocol, thus leading to the removal of relevant slices or missing out noisy slices for certain scans.

Automatic ROI detection Our second approach leveraged the power of object detection for automated ROI extraction. We began by constructing a dataset of 500 MRI slices, encompassing all three standard views (sagittal, axial, and coronal). Each slice was meticulously annotated with ROI using the LabelImg software[1]. These annotations delineated the precise boundaries of the knee joint region within each slice.

To automatically identify and extract these ROIs, we turned to YOLOv8 [2], a SOTA real-time object detection model celebrated for its speed and accuracy (the model was the latest at the time of conducting the experiments). YOLOv8's architecture is designed for efficient and effective object detection. Its backbone network, built on a series of convolutional layers with cross-stage partial connections, extracts rich feature representations from input images. These features are then passed through a neck section, which aggregates information from different scales, further enhancing detection accuracy. Finally, the head of the network makes predictions, outputting bounding boxes that localize objects within the image, along with objectness scores and class probabilities.

While pretrained YOLOv8 models offer a strong starting point, training a custom version tailored to our specific task proved essential [99]. The visual characteristics of MRI slices, with their grayscale nature and subtle anatomical features, differ significantly from the natural images typically used to train generic object detection models. By training YOLOv8 on our annotated dataset of knee MRI slices, we enabled it to learn the specific visual cues and patterns relevant for accurately identifying knee joint ROIs within this unique imaging modality (see Fig. 3.5).

Model performance was rigorously evaluated using the mean Average Precision (mAP) (discussed later in chapter 04), a standard metric for object detection tasks. A higher mAP score indicates superior accuracy in ROI detection.

Once satisfied with the performance of our custom YOLOv8 model, we integrated it into our 3D MRI processing pipeline. For each 3D MRI volume, the model individually analyzes each slice, predicting the location of the knee joint ROI. These individual ROI predictions are then aggregated to generate a comprehensive 3D cropped volume, effectively isolating the knee joint region for further analysis by our models.

[1] https://github.com/HumanSignal/labelImg

[2] No official paper has been announced for YOLOv8: https://blog.roboflow.com/whats-new-in-yolov8/

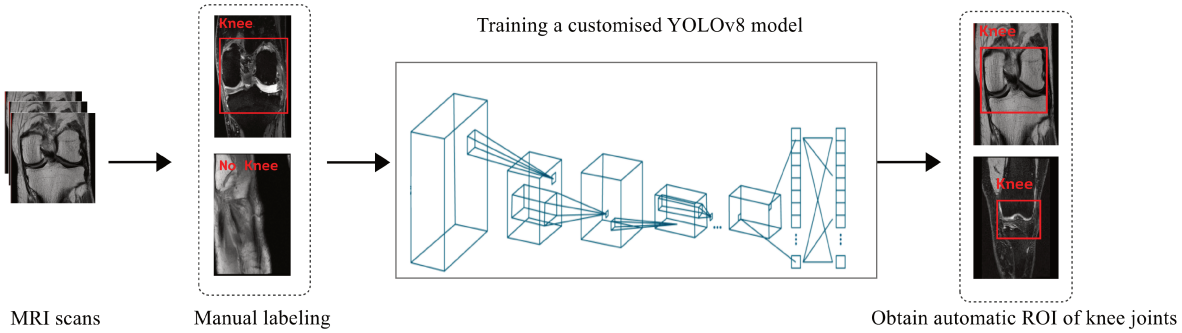


Figure 3.5: Custom YOLOv8 designed for automatic knee joint detection.

3.1.5 Performance Metrics and Evaluation

In medical imaging applications, such as MRI classification, segmentation, and object detection, various metrics are employed to evaluate model performance. This section outlines the key metrics used for different tasks, focusing primarily on binary MRI classification.

3.1.5.1 Classification Metrics

In our experiments along the thesis contributions, we performed multiple tasks, including binary image classification in which the models are trained to distinguish between healthy and diseased states, image segmentation to segment the knee compartment around the knee joint, and image reconstruction. Here we list the evaluation metrics reported along the paper for each task:

1. **Area Under the Receiver Operating Characteristic Curve (AUC-ROC):** AUC-ROC measures the model's ability to distinguish between classes across various threshold settings. It is particularly useful for handling imbalanced datasets.

$$AUC = \int_0^1 TPR(FPR^{-1}(t))\,dt$$

where TPR is the True Positive Rate and FPR is the False Positive Rate. The True Positive Rate (TPR) and False Positive Rate (FPR) are crucial metrics in constructing the Receiver Operating Characteristic (ROC) curve, which is used to compute the AUC-ROC. Here is how these rates are calculated:

- **True Positive Rate (TPR)**: Also known as sensitivity or recall, it measures the proportion of actual positives that are correctly identified by the model. It is calculated as:

$$\text{TPR} = \frac{TP}{TP + FN}$$

where TP are the true positives and FN are the false negatives.

- **False Positive Rate (FPR)**: It measures the proportion of actual negatives that are incorrectly classified as positives by the model. It is calculated as:

$$\text{FPR} = \frac{FP}{FP + TN}$$

where FP are the false positives and TN are the true negatives.

These metrics are plotted against each other to form the ROC curve, from which the AUC is derived.

2. **Accuracy (ACC):** Accuracy calculates the proportion of correct predictions (both true positives and true negatives) out of the total number of cases examined.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

- **True Positives (TP):** Correct predictions of the positive class (e.g., diseased)
- **True Negatives (TN):** Correct predictions of the negative class (e.g., healthy)
- **False Positives (FP):** Incorrect predictions as positive (e.g., healthy cases wrongly identified as diseased)
- **False Negatives (FN):** Incorrect predictions as negative (e.g., diseased cases wrongly identified as healthy)

3. **F1 Score:** The F1 score is the harmonic mean of precision and recall, providing a balanced single score.

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

where:

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

- **Precision:** Measures the accuracy of positive predictions.
- **Recall:** Measures the ability to identify all relevant instances (true positives).
- High **precision** indicates a low rate of false positives, while high **recall** indicates capturing a large proportion of positives.

3.1.5.2 Segmentation Metrics

For segmentation tasks, such as tumor delineation or organ segmentation, the following metrics are used:

1. **Dice Coefficient:** This measures the overlap between the predicted segmentation and the ground truth.
$$Dice = \frac{2|X \cap Y|}{|X| + |Y|}$$
2. **Binary Cross-Entropy (BCE) Loss:** Commonly used as a loss function for training models in binary segmentation tasks.
$$\mathscr{L}_{\text{seg}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]$$

3.1.5.3 Object Detection Metrics

For tasks like identifying and localizing lesions, the following metrics are crucial:

1. **Intersection over Union (IoU):** Measures the overlap between the predicted bounding box and the ground truth bounding box.
$$IoU = \frac{\text{area}(B_p \cap B_{gt})}{\text{area}(B_p \cup B_{gt})}$$
   - B_p: The predicted bounding box, which is the region identified by the model as containing the object.
   - B_{gt}: The ground truth bounding box, which is the actual region where the object is located, according to dataset annotations.
   - Area($B_p \cap B_{gt}$): The intersection area, where both B_p and B_{gt} overlap. This represents the correctly identified and localized part of the object by the model.
   - Area($B_p \cup B_{gt}$): The union area covered by both B_p and B_{gt}, which includes all the space covered by either bounding box.

   The IoU is calculated using the following formula:
$$\text{IoU} = \frac{\text{Area}(B_p \cap B_{gt})}{\text{Area}(B_p \cup B_{gt})}$$
   This metric provides a score ranging from 0 to 1, where higher values indicate better localization accuracy of the model.
2. **Mean Average Precision (mAP):** The mean of the Average Precision (AP) calculated for each class.

3.2 Related Works

This section provides a thorough review of the existing literature on the application of deep learning in the diagnosis and assessment of knee OA, with a particular emphasis on the use of MRI. The review is organized into key thematic areas to facilitate a comprehensive understanding of the field.

3.2.1 Deep Learning for Knee OA Assessment using MRI

MRI has emerged as a valuable tool for assessing knee OA, offering superior visualization of soft tissues such as cartilage, menisci, and ligaments, which are crucial for early OA detection and monitoring. Several studies have focused on utilizing deep learning techniques to enhance the diagnostic capabilities of MRI in knee OA.

Guida et al. [76] conducted a pioneering study using CNNs to detect knee OA from DESS MRI scans obtained from the OAI database. Their results demonstrated that MRI outperforms traditional X-rays in detecting OA, particularly in its early stages. This study underscored the potential of deep learning in improving the accuracy of OA diagnosis by leveraging the rich information available in MRI scans.

Similarly, Pedoiat et al. [75] explored the application of deep learning for T2 relaxometry in knee OA detection. By analyzing 4,384 MRI scans, they compared traditional relaxometry methods with deep learning-based approaches. Their findings indicated that deep learning techniques show considerable promise in automatically quantifying cartilage changes associated with OA, providing a more objective and efficient assessment method than conventional techniques.

Building on these advancements, Namiri et al. [100] developed a CNN-based model designed to classify various knee phenotypes, including bone structure, meniscus/cartilage condition, inflammation, and hypertrophy. Their study, which involved MRI data from 4,791 participants in the OAI database, achieved high Area Under the Curve (AUC) values across all phenotypes. This work highlighted the potential of deep learning to distinguish between different pathological features of knee OA, thus aiding in more precise diagnosis and personalized treatment planning.

Furthering the field of OA assessment, Panfilov et al. [101] proposed an innovative architecture that combines 2D CNNs with Transformers for evaluating OA progression using 3D MRI. Their study, which included a large dataset of 4,866 subjects, demonstrated that this hybrid approach outperforms conventional 2D and 3D CNN models in predicting OA progression. This work is particularly notable for its use of 3D MRI data, which captures more detailed anatomical information than 2D imaging, thereby offering a more comprehensive assessment of OA progression.

Another significant contribution to OA progression assessment is the study by Anonymous et al. [102], which utilized a U-Net model for segmenting bone and cartilage tissues from 593 DESS and IW-TSE MRI sequences. The study integrated three different classifiers—ResNet, DenseNet, and CVAE—with clinical data using a Logistic Regression Classifier to detect early knee OA progression. This multi-classifier approach, combined with advanced segmentation techniques, represents a step forward in the accurate and early identification of OA progression.

In addition to early detection and progression assessment, deep learning has also been applied to predict the severity of OA and the need for Total Knee Replacement (TKR). Tolpadi et al. [103] employed a 3D DenseNet architecture to predict both the severity of knee OA and the likelihood of requiring TKR using MRI images. Their results suggested that 3D MRI data provides more

valuable information for TKR prediction compared to traditional 2D radiographs, emphasizing the importance of 3D imaging in clinical decision-making.

Moreover, Hu et al. [104] investigated the use of deep learning models combined with various MRI sequences, such as 3D-DS-WE and T2 mapping, to evaluate knee cartilage injury. They proposed a multiscale wide residual network model that demonstrated high diagnostic accuracy in visualizing and diagnosing cartilage lesions in patients with knee OA. This study highlights the potential of advanced deep learning architectures in enhancing the diagnostic capabilities of MRI for cartilage injury assessment, a critical aspect of OA management.

3.2.2 Deep Learning for Knee Injury Detection using MRI

While the primary focus of this dissertation is knee OA, it is essential to consider the broader application of deep learning in knee injury detection using MRI. The detection and analysis of knee injuries, such as meniscus tears and ACL injuries, are critical for comprehensive knee health assessment and management.

Bien et al. [47] introduced MRNet, a deep learning framework designed to detect meniscus tears, ACL injuries, and other knee abnormalities using multi-view MRI data. The framework processes slices from different views—sagittal, axial, and coronal—and aggregates the information using max-pooling. This multi-view approach enhances the model's ability to detect abnormalities across different planes, offering a more robust diagnostic tool for knee injuries.

Expanding on the MRNet framework, Azcona [105] proposed a modified architecture incorporating ResNet as a feature encoder. The study explored four different model variations and transfer learning strategies to improve knee injury detection. The modifications led to significant improvements in detection accuracy, particularly in challenging cases where injuries are subtle or distributed across multiple imaging planes.

Additionally, Fritz et al. [106] focused specifically on meniscus tear detection, employing a deep 3D CNN trained on MRI data. Their model achieved promising results for detecting both medial and lateral meniscus tears, demonstrating the effectiveness of deep learning in identifying specific types of knee injuries. This work is particularly relevant for clinicians aiming to accurately diagnose and treat meniscal injuries, which are common in both OA and traumatic knee conditions.

Du et al. [107] explored the predictive potential of hidden biomedical information from knee MR images for OA progression. The study computed a Cartilage Damage Index (CDI) from 36 locations on the tibiofemoral cartilage and used Principal Component Analysis (PCA) for feature reduction. Four machine learning classifiers (ANN, SVM, Random Forest, and Naïve Bayes) were used to predict OA progression measured by KL grade, Joint Space Narrowing Medial (JSM), and Lateral (JSL) grades. The medial feature set outperformed the lateral, with the combined 36-D set yielding the best results. ANN achieved the highest AUC (0.761) for KL grade prediction,

Random Forest excelled in JSM prediction (AUC = 0.785), and ANN performed best for JSL prediction (AUC = 0.695).

A study by Yeoh et al. [108] explored the effectiveness of 3D convolutional neural networks across 13 architectures for knee OA diagnosis, using transfer learning from 2D pre-trained weights. The study found that transfer learning notably enhanced model performance, particularly for ResNet and DenseNet architectures. ResNet34 achieved the highest accuracy (0.875) and F1 score (0.871), while shallower networks like ResNet18, DenseNet121, and VGG11 showed superior AUC values of 0.945, 0.914, and 0.928, respectively. The study utilized 400 3D DESS MRI scans with KL grades as the scoring system.

Another study by Schiratti [109] utilized 9,280 knee MR images from the OAI database to develop a deep learning model that predicts cartilage degradation, measured by joint space narrowing (JSN) at 12 months, using MR images and clinical variables such as BMI. The model achieved a ROC AUC of 0.65, outperforming trained radiologists who scored 0.587. Additionally, the model predicted pain grade with a ROC AUC of 0.72, with attention maps highlighting distinct areas relevant for JSN progression and pain prediction.

Liu et al. [110] developed a fully automated deep learning system for cartilage lesion detection using segmentation and classification CNNs. The study analyzed fat-suppressed T2-weighted fast spin-echo MRI data from 175 patients, with a musculoskeletal radiologist's interpretations serving as the reference standard for training. The system evaluated 17,395 image patches from the femur and tibia, achieving sensitivities of 0.841 and 0.805 and specificities of 0.852 and 0.879 in two evaluations. The areas under the ROC curve were 0.917 and 0.914, respectively, indicating high diagnostic accuracy, with a value of 0.76 demonstrating good intraobserver agreement.

3.2.3 Deep Learning for Knee OA Assessment using X-ray

X-ray imaging remains one of the most commonly used modalities for knee OA assessment due to its accessibility and cost-effectiveness. However, X-rays primarily visualize bone structure and provide limited information about soft tissues, leading to potential limitations in OA diagnosis and monitoring.

Tiulpin et al. [111] leveraged data from the OAI database to train a CNN for predicting knee OA progression using both plain radiographs and clinical data. This multimodal approach demonstrated the potential of integrating clinical and imaging data to improve the accuracy of OA progression predictions.

Schwartz et al. [112] further explored the application of CNNs for analyzing knee plain radiographs, assessing the model's ability to identify critical features and grade the severity of knee OA. Their study highlighted the potential of deep learning to automate the grading process, reducing the reliance on manual assessments and potentially improving consistency in OA diagnosis.

Guan et al. [113] conducted a comparative study on the efficacy of two CNN architectures, VGG16 and DenseNet, trained on knee radiographs, against a support vector machine (SVM) model utilizing non-image data. Their findings suggested that deep learning models outperformed traditional machine learning approaches, particularly in cases where imaging data was the primary source of information.

Building on this work, Tiulpin et al. [114] utilized a CNN pre-trained on ImageNet to perform multiclass classification for predicting OA progression from knee radiographs. Their results were compared with a logistic regression model, demonstrating the advantages of using deep learning for more complex classification tasks.

Further contributions by Guan et al. [115] included the development of a model to predict OA based on the progression of medial joint space loss in knee radiographs. Their approach combined two deep learning architectures inspired by the YOLO model and DenseNet, which enabled more accurate tracking of joint space changes, a critical marker for OA progression.

In addition to progression prediction, the classification and grading of OA severity have been major focuses of research. Zhe et al. [116] proposed a novel approach using Vision Transformer (ViT) models, enhanced with Selective Shuffled Position Embedding (SSPE) and key-patch exchange techniques, for knee OA classification. This study represents a significant advancement in applying transformer-based models to medical imaging, which traditionally relies on CNNs.

Liu et al. [117] applied Faster R-CNN for simultaneous knee joint detection and KL-grading-based classification from plain radiographs. Their approach introduced a novel loss function and larger anchors to address class imbalance and large image sizes, which are common challenges in medical image analysis.

Nasser et al. [118] developed a Discriminative Shape-Texture Convolutional Neural Network (DST-CNN) for the automatic diagnosis of early knee OA. Their model incorporated a discriminative loss and a novel Gram Matrix Descriptor (GMD) block, which enhanced the network's ability to differentiate between subtle texture variations associated with early OA.

Schwartz et al. [119] demonstrated that CNNs could effectively identify features and assess the severity of knee OA from knee radiographs. This work further supports the growing body of evidence that deep learning can automate and enhance traditional OA grading methods.

In a related study, an anonymous team [120] developed a deep learning model for the detection, localization, and classification of knee joint areas in X-ray images using YOLOv3. This model is notable for its ability to perform multiple tasks simultaneously, potentially reducing the need for separate models for each task.

Lastly, Chen et al. [121] introduced two deep convolutional neural network (CNN) methodologies tailored to the Kellgren-Lawrence (KL) grading system. Their work focused on accurately localizing knee joints using YOLOv2 and classifying them using enhanced versions of ResNet, DenseNet, InceptionV3, and VGG, providing a comprehensive approach to both localization and classification tasks.

3.2.4 Deep Learning for Knee Structure Segmentation

Segmentation of knee structures, such as cartilage and menisci, is a critical task in OA assessment and other knee-related conditions. Deep learning has been widely applied to automate and improve the accuracy of segmentation tasks.

Norman et al. [122] employed a 2D U-Net architecture to segment six knee subcompartments, including articular cartilages and the meniscus, from 3D-DESS image datasets. Their work demonstrated the utility of U-Net models in accurately segmenting complex anatomical structures from high-resolution MRI data.

Si et al. [123] extended this approach by utilizing the 2D U-Net model to segment the femur, tibia, patella, and their corresponding cartilages from MR images. Their study focused on measuring cartilage thickness, a key indicator of OA progression, and highlighted the importance of accurate segmentation in assessing disease severity.

Wirth et al. [124] used 2D U-Net for the segmentation of femorotibial cartilages from coronal FLASH and sagittal DESS images. Their work assessed the reproducibility of cartilage morphometry over time, which is essential for monitoring OA progression in longitudinal studies.

In addition to U-Net-based approaches, Kompella et al. [125] integrated Mask R-CNN, utilizing a ResNet-50 backbone, for the automated segmentation of femoral cartilage from ultrasound 2D image scans. This study is significant for its exploration of ultrasound imaging, a less commonly used modality in knee OA research, but one that offers potential advantages in terms of cost and accessibility.

Liu et al. [126] proposed a two-step process for knee structure segmentation and lesion detection. Their approach involved using a 2D U-Net architecture for the automatic segmentation of the meniscus and cartilage, followed by a 3D CNN for detecting lesions. This combined approach underscores the importance of integrating multiple deep learning models to tackle complex medical imaging tasks.

The segmentation of knee structures, such as cartilage and menisci, is a critical task in knee OA assessment and other knee-related conditions. Deep learning models have been widely applied to automate and enhance the accuracy of segmentation tasks, with various approaches demonstrating significant advancements in this area.

Prasoon et al. [127] introduced a method utilizing multi-planar CNNs, where three separately trained 2D CNNs were applied to each orthogonal plane of knee MRIs—sagittal, coronal, and axial. This approach demonstrated the superiority of learned features from CNNs, achieving a Dice Similarity Coefficient (DSC) of 0.8249 in volume overlap. The use of multi-planar CNNs allowed for the incorporation of information from different anatomical planes, leading to improved segmentation performance across the knee joint.

Norman et al. [128] employed a 2D U-Net architecture to segment various sub-compartments of the knee, including articular cartilage and the meniscus. Their study focused on the OAI dataset, where the 2D U-Net achieved a mean validation DSC of 0.867 across 37 subjects. The

dataset was divided into two distinct groups: patients with OA (KL score >1) and patients whose condition remained longitudinally stable (KL score = [0-1]). The 2D U-Net was specifically designed for pixel-wise classification of high-resolution images, which is critical for capturing the fine details necessary for accurate segmentation in OA research.

Expanding on the segmentation framework, Zhou et al. [129] were inspired by the work of Liu et al. [126], combining a CNN model with a Conditional Random Field (CRF) as a post-processing step to refine the segmentation labels. The inclusion of the CRF, which leverages spatial proximity, led to improved segmentation results, with the authors reporting DSC values exceeding 0.80 for all cartilages. This integration of CRF with CNNs highlights the importance of incorporating spatial relationships into segmentation models to enhance their accuracy.

Ambellan et al. [130] advanced the segmentation field by combining 2D and 3D U-Nets and applying them to two different OAI image datasets: OAI-Imorphics and OAI-ZIB, comprising 176 and 507 volumes, respectively. Their method achieved DSC values ranging from 0.858 to 0.90 on the OAI-Imorphics dataset and from 0.856 to 0.899 on the OAI-ZIB dataset [131]. Recognizing the limitations of 2D convolutional filters in capturing spatial consistency in 3D MRIs, Ambellan et al. [130] enhanced the spatial consistency of their segmentation outcomes by incorporating eight adjacent slices to the individual slice input, thus training the CNNs with 17 input channels. This approach significantly improved the model's ability to maintain spatial coherence across the 3D structure of the knee, which is crucial for accurate and reliable segmentation in clinical applications.

A recent study by Nunes et al. [132] developed a multi-task deep learning approach to automate the classification and segmentation of cartilage lesions and bone marrow edema lesions from knee MRI. Utilizing two V-Net architectures, the models first segmented the cartilage compartments and then classified lesion severity. The outputs from both tasks, combined with demographic data, were integrated using a Gradient Boost classifier, achieving an overall accuracy of 0.867 for lesion severity staging. This multi-task approach significantly enhances the efficiency and accuracy of knee OA assessment.

3.2.5 Limitations of Existing Works

Despite the significant progress made in applying deep learning to knee OA assessment, several limitations in the existing research highlight the need for further investigation.

Firstly, the underutilization of MRI in knee OA assessment remains a notable gap. While X-ray imaging is widely used due to its accessibility and cost-effectiveness, it primarily visualizes bone structure and provides limited information about soft tissues like cartilage, menisci, and ligaments. These soft tissues are crucial for early OA diagnosis and monitoring progression, and MRI, with its superior soft tissue contrast, offers a more comprehensive evaluation of these structures. Our research aims to leverage the richness of MRI data to develop more accurate and sensitive models for knee OA assessment.

Secondly, many existing models rely on single-view imaging, which may not capture the full complexity of knee OA. Single-view models are limited in their ability to assess the entire joint, as OA affects various structures within the knee that may not be fully visible in a single imaging plane. Multi-view models, which integrate information from sagittal, coronal, and axial planes, can provide a more holistic representation of the knee joint, potentially improving diagnostic accuracy. Our research explores the use of multi-view data to develop models that better capture the complexity of knee OA.

Another limitation is the prevalence of separate models for specific tasks, such as classification, segmentation, or progression prediction. While these models are effective for their individual purposes, they can be computationally expensive and require a larger amount of annotated data for each task. Multi-task and multi-label models, which can handle multiple tasks simultaneously, offer a more efficient and potentially more robust solution. We investigate the feasibility of multi-task learning for knee OA assessment, aiming to improve efficiency and potentially discover relationships between different OA-related features.

Lastly, the computational cost of 3D models poses a significant challenge. Although 3D models have shown promising results for knee OA assessment, they are computationally demanding and require significant resources for training and deployment. This limitation restricts their applicability in clinical settings, particularly in resource-limited environments. Our research focuses on developing lighter and more efficient models, potentially by incorporating 2D analysis techniques or exploring novel architectures, while maintaining or even improving performance compared to existing 3D models.

In this thesis, we present our efforts to address these aforementioned limitations by developing efficient and accurate deep learning models for knee OA assessment using MRI data. We explore novel architectures that aim to improve performance while minimizing computational demands, paving the way for more practical solutions in knee OA diagnosis and assessment.

3.3 Summary

This chapter laid the groundwork for our exploration of deep learning in knee OA assessment using different screening, mainly focusing on MRI. We started by explaining some fundamentals of deep learning used for medical imaging analysis. A comprehensive review of related work highlighted the growing, yet still nascent, field of deep learning in knee imaging. While several studies have demonstrated promise in the automated detection of knee injuries and OA using various MRI sequences, there remains a significant need for further research, particularly in leveraging 3D MRI data for early OA detection and progression prediction. This chapter establishes the context, motivation, and foundational knowledge upon which we build our novel deep learning framework for knee OA assessment.

In the following chapters, we will discuss our contributions to enhancing the field of detecting knee OA using MRI, and various approaches to advance knee analysis.

CHAPTER 4

SEMI-SUPERVISED MULTIVIEW-MRI NETWORK FOR KNEE OA DETECTION

This chapter addresses the pressing need for accurate and efficient knee OA detection, a widespread and debilitating joint disorder. Building upon the limitations of current deep learning methods, which often overlook the comprehensive definition of knee OA and the benefits of multi-view imaging, we propose a novel approach leveraging semi-supervised learning and a specifically designed 3D CNNs.

Our contribution is threefold. First, we introduce the 3D-ResCNN-GAP model, a novel deep learning architecture incorporating stacked convolutional blocks, Global Average Pooling, and residual connections. This innovative design enables the model to effectively extract discriminative features from 3D MRI data. Second, we implement a semi-supervised learning strategy using pseudo-labeling to expand the training dataset with unlabeled MRI scans. This augmentation addresses the scarcity of labeled data and enhances the model's generalizability. Finally, we leverage multi-view MRI acquisitions, incorporating sagittal and coronal planes, to provide a more comprehensive representation of the knee joint and further improve OA detection accuracy.

4.1 Introduction

In recent years, several studies have focused on using deep learning to tackle the challenges of the detection of knee OA [133, 134], with the aim of developing more accurate, reliable, and cost-effective diagnostic tools. These studies have explored a range of deep learning approaches, such as CNNs [74, 135, 136] and Generative Adversarial Networks (GANs)[137], to analyze and interpret imaging data from a variety of sources, including MRI [60, 76], CT scans [138] and X-ray [77, 139, 140]. These approaches have been shown to be highly effective at detecting early

signs of knee OA, such as cartilage degeneration, bone erosion, and joint swelling, allowing for earlier intervention and more effective treatment of the condition.

Proposing effective solutions for the detection of knee OA has been an active area of research due to the high prevalence and significant impact of the disease on individuals [134]. However, recent methods have several limitations. One limitation is that these methods often do not directly target knee OA, but rather focus on specific knee symptoms (meniscus [84], ACL [47], etc.). This can lead to incomplete or inaccurate diagnoses of knee OA. Additionally, many of the methods rely on limited datasets and do not consider the use of multiple knee views. This can potentially result in the omission of important features and information that might only be visible from specific views, which limits the generalizability of the findings. Another limitation is that some approaches operate on 2D slices or X-rays [141], which can potentially lead to important feature correlations in the 3D space remaining undetected.

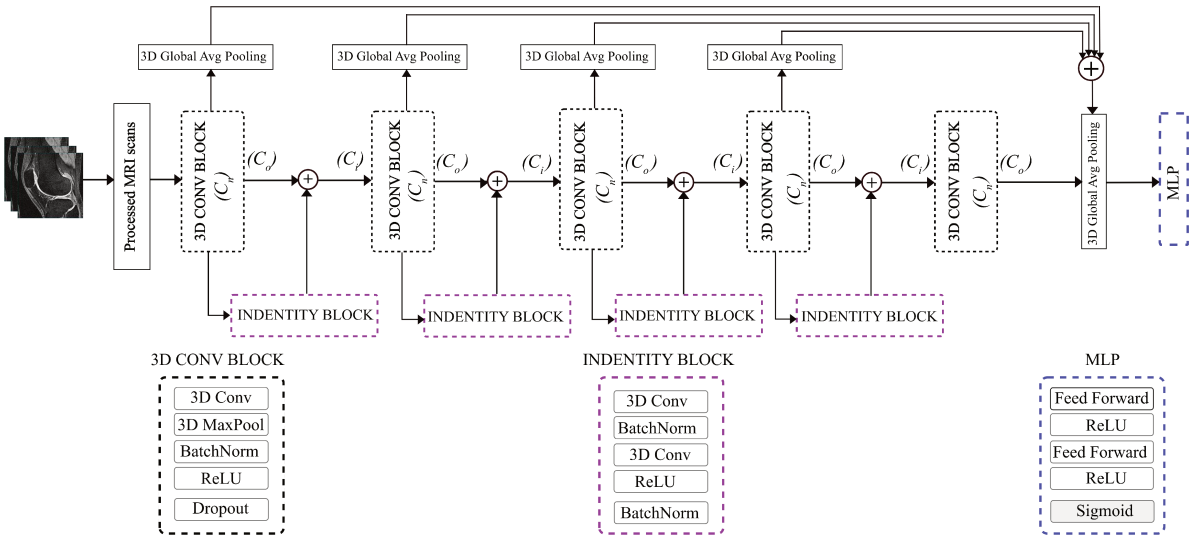


Figure 4.1: Proposed 3D-ResCNN-GAP model architecture. The input to each 3D GAP layer comes from the abstraction produced by the adjacent CNN block, and the output of all 3D GAP layers are summed up with the final GAP layer. The identity block includes CNN layers to enhance the model's ability to capture local features at a specific CNN block. $\oplus$ denotes the concatenation operation.

To address these limitations, it is necessary to develop novel approaches for knee OA detection that consider multiple knee imaging views, a reliable knee OA definition (explained in 2.6), and a larger dataset.

Furthermore, recent advances in deep learning and computational resources have enabled the development of more sophisticated and efficient models for knee OA detection, such as multi-view models that combine information from multiple views [142], and semi-supervised models that can learn from large and diverse datasets without the need for extensive manual labeling [139, 143].

These models are expected to have a major impact on the field of knee OA detection.

This chapter introduces a novel deep learning framework, a 3D deep residual convolutional model with concatenated GAP features (3D-ResCNN-GAP), to enhance the detection of knee OA from 3D MR scans. Our approach is based on the comprehensive definition of knee OA, which includes the identification of osteophytes and cartilage loss as outlined in [16], using MOAKS scores [4] for screening analysis. The 3D-ResCNN-GAP model innovatively combines stacked convolutional layers, Global Average Pooling (GAP), and residual and skip connections. This integration facilitates the efficient analysis of various views obtained from MR images, significantly improving the accuracy and reliability of knee OA detection.

In addition, we incorporate a semi-supervised learning (SSL) approach, which uses a combination of labeled and unlabeled MRI data to improve the performance and generalizability of the proposed model on a large cohort. SSL has several advantages over traditional supervised learning models, including (i) the ability to learn from a larger and more diverse dataset, (ii) improved resistance to overfitting, and (iii) the ability to use unlabeled data that may be abundant but difficult or expensive to annotate. Our proposed model was evaluated on a large dataset of 3D MR images from the publicly available OAI database [46], and its performance was compared to classical CNN models.

The overall contributions of this chapter are as follows:

- We introduce a novel deep learning model, 3D-ResCNN-GAP, composed of consecutive 3D CNN blocks, residual CNN, and GAP layers for the detection of knee OA from 3D MR images, and also compared its performance to classical CNN models.
- We propose a semi-supervised approach to increase the training dataset and improve the detection performance of the 3D-ResCNN-GAP model.
- We evaluate the model on a multiple and large MRI dataset to ensure generalizability on larger populations and to avoid the model's performance being affected by variations in imaging protocols.
- We investigate the effect of different knee MRI views for the detection of knee OA.
- We propose a multi-view learning approach that utilizes various models for training data obtained from diverse MRI views of the knee joint.
- We provide Gradient-weighted Class Activation Mapping to visually interpret and validate the decision-making process of our deep learning model in diagnosing knee OA from MRI scans.
- Results show that our proposed SSL approaches enhance the classification performance of various CNN models on multiple knee MRI scan acquisitions.

- Results indicate that combining multiview MRI for knee OA detection enhances performance by increasing the size of the dataset, enriching it with multiview features, and boosting classification accuracy.

4.2 Methodology

Our proposed methodology consists of three main parts: (i) Model Architecture, (ii) Semi-Supervised training, and (iii) Multi-view classification. Firstly, to ensure the quality and relevance of the data, light preprocessing of the scans is performed to remove noise and irrelevant features. Next, the proposed 3D-ResCNN-GAP model is developed and trained from scratch, allowing for full customization of the model to suit the specific task. The Semi-Supervised training step involves leveraging the pseudo-labeling approach to augment the training data. Finally, to further improve the model's performance, the proposed architecture is trained using a multi-view approach.

Each step of the proposed methodology is described in detail in the following subsections:

4.2.1 Proposed 3D-ResCNN-GAP model

We propose a new knee OA classification model that leverages three fundamental components: (i) stacked 3D convolutional blocks, (ii) GAP to aggregate local and global feature representations, and (iii) residual and skip connections, making the model wide and deep. These components are carefully designed and integrated to enhance the model's ability to learn complex patterns and capture informative features from high-dimensional data.

4.2.1.1 Stacked Convolutional Blocks

The 3D-ResCNN-GAP model employs a stacked architecture consisting of five successive CNN blocks, designed to extract features from the MRI slices at multiple levels.

The structure of the 3D-ResCNN-GAP model is illustrated in Figure 4.1. The 3D-ResCNN-GAP architecture consists of N stacked convolutional blocks, denoted by C_n, where $n \in 1,...,N$. Each C_n block contains a 3D convolutional layer with F feature maps, followed by L^2 regularization and the Rectified Linear Unit (ReLU) activation function. The output of the ReLU layer is then passed through a Batch Normalization (BN) layer and a 3D Max-Pooling (MaxPool) layer.

Let iC_n be the entry of each C_n block. The output, oC_n is defined as:

$$oC_n = \text{3D-MaxPool}\Big(\text{BN}\Big(\text{ReLU}\Big(\text{Conv3D}\big(iC_n\big)\Big)\Big)\Big) \tag{4.1}$$

The specific design of this architecture is described in further detail in the following section.

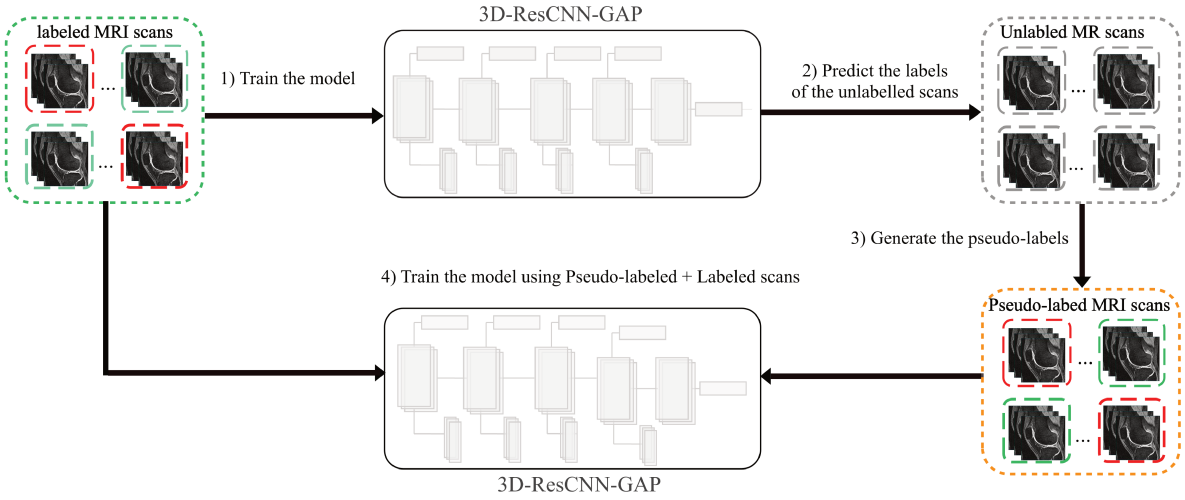


Figure 4.2: Pseudo-labeling approach. We trained our 3D-ResCNN-GAP model on the labeled data, then we used the trainable weights to predict the labels on batches of the unlabeled MRI scans; the generated labels are then assigned, and the pseudo-labeled and labeled data are merged to train the 3D-ResCNN-GAP model again.

4.2.1.2 Global Average Pooling (GAP)

The GAP modules [144] can reduce computational complexity and overfitting by reducing the number of parameters and regularizing the model. They also ensure the conservation of all features extracted from each block, preventing the loss or underrepresentation of important features. In order to ensure this, we extract from each 3D CNN Conv Block a feature vector, V_k using a GAP layer, as presented in Fig.4.1. The feature vector V_k of the n^{th} building block C_n is defined as:

$$V_k(C_n) = \frac{1}{N} \sum_{ij} z_{ij}^c \tag{4.2}$$

where z_{ij}^c represents the activation value of the c-th feature map at position (i, j). The subscript c refers to the channel or feature dimension of the input, while the subscripts i and j refer to the spatial dimensions. The final fusion feature vector of the extracted GAP modules, V_F is obtained as follows:

$$V_F = \sum_{p=1}^{5} (V_k(C_p)) \tag{4.3}$$

4.2.1.3 Residual connections

To improve our model's ability in feature extraction, we adopt residual connections in the corp of the 3D-ResCNN-GAP model. These connections, comprising two 3D convolutional layers and BN, enable the preservation of information from earlier layers and have been shown to improve

performance in image classification tasks [93]. The outputs of the convolutional and the residual blocks are concatenated and transferred to subsequent stacked convolutional blocks, as presented in Fig. 4.1.

Each residual block is composed of two 3D convolutional layers and a BN layer that help accelerate the training process by reducing internal covariate shift, thus improving the generalization performance of the network. The BN is defined as follows:

$$BN = \frac{x - E(x)}{\sqrt{(Var(x) + \varepsilon)}} * \gamma + \beta \tag{4.4}$$

BN normalizes the input to a layer using its mean and variance over the current mini-batch, and then scales and shifts the normalized output using learnable parameters γ and β. The input which corresponds to a mini-batch is denoted by x. $E(x)$ and $Var(x)$ represent the mean and variance, respectively, over the current mini-batch.

4.2.2 Pseudo-labels: Label augmentation

In this section, we introduce the semi-supervised learning (SSL) approach as a way of augmenting the training data and making use of the entirety of available OAI-MRI scans that have not been labeled using the MOAKS score for knee OA.

Our SSL approach involves training the proposed model on the data in two stages. First, the model is trained on labeled MRI scans according to the definition mentioned in [16]. Secondly, the trained 3D-ResCNN-GAP model is used to predict the labels and assign them to unlabeled MRI scans, as presented in Fig. 4.2.

To formulate this, we define *pseudo-labels* as labels assigned to unlabeled data. This is achieved by selecting the class that exhibits the highest predicted probability for each unlabeled sample:

Let X be a set of n MRI samples and x_i an element of X. X contains both labeled and unlabeled samples. Let X_L be the set of L labeled samples as OA or non-OA and X_U the set of the remaining non-labeled scans. Consequently, $X = X_L \cup X_U$. The corresponding labels for X_L are denoted by Y_L.

Let u be an unlabeled sample, and let $P(u)$ be the predicted probability distribution over the set of possible labels. We can assign a pseudo-label $l(u)$ to u based on a threshold value τ, such that:

$$l(u) = \begin{Bmatrix} 1 & \text{if } \operatorname{argmax}_\ell P_{i'}(u)) \geq \tau \\ 0 & \text{otherwise} \end{Bmatrix} \tag{4.5}$$

where, $\operatorname{argmax}_\ell P(u)$ returns the label with the highest predicted probability, and 0 is assigned when the predicted probability is below the threshold value τ = 0.5.

The predicted *pseudo-labels* are therefore assigned to their associated (X_U) scans.

The pseudo-labeled scans are then merged with the already labeled MRI scans, thus resulting in a larger dataset. The specifics of the size of each data subset are reported in Table 4.1. Notably, the SAG 3D-DESS acquisitions were augmented to 3830 labeled MRI scans, the SAG IW-TSE were augmented to 4297 labeled MRIs, and for the COR MPR scans, the number of labeled scans increased to 4089 labels.

All evaluated models (before and after pseudo-labeling) were trained to optimize the Binary Cross Entropy Loss function, BCE as:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \tag{4.6}$$

where N is the total number of training samples, y_i is the ground truth label for the i-th sample, and $\hat{y}_i$ is the predicted probability for the i-th sample.

4.2.3 Multi-view framework for knee OA classification

In this section, our aim is to use multiple views (sagittal and coronal) of the knee to classify OA and non-OA knees. Multi-view frameworks are motivated by the findings that multi-view images of a common object can furnish supplementary information, thereby enhancing the analysis and interpretation of the object.

Fig 4.3 illustrates the multi-view classification network utilized in the present chapter. Specifically, the framework employs two distinct MRI scans of the knee in the sagittal plane, namely the SAG 3D-DESS and SAG IW-TSE, as well as the coronal MPR scans. SAG-3D-DESS is a high-resolution sequence that provides high quality images of the knee joint shape and other tissues such as the cartilage from the sagittal view. This is important for the detection of early OA, as the cartilage is one of the first tissues affected by the disease. SAG-IW-TSE is a T2-weighted sequence that provides good contrast between the cartilage and other tissues in the knee. The combination of these two sagittal sequences in the multi-view framework will increase the feature detection capability of the framework. COR-MPR is not as sensitive to early OA as SAG-3D-DESS or SAG-IW-TSE, but it is useful for assessing the full extent of cartilage damage from the coronal view [4]. It is also useful for visualizing other structures in the knee, such as the menisci and ligaments.

This approach enables the representation of knee tissues from both sides, providing a more comprehensive view of the knee joint and potentially aiding in the identification of abnormalities or injuries.

The MRI scans of the knee joint are first preprocessed (see section 4.3.2) and then fed to the 3D-ResCNN-GAP model. The extracted features from each model-view M_f are concatenated to form a feature vector M_{conc} and then passed into a Multi-Layer Perceptron (MLP). The MLP

consists of linear dense layers, separated by ReLU activation functions and normalization layers. M_{conc} is obtained as:

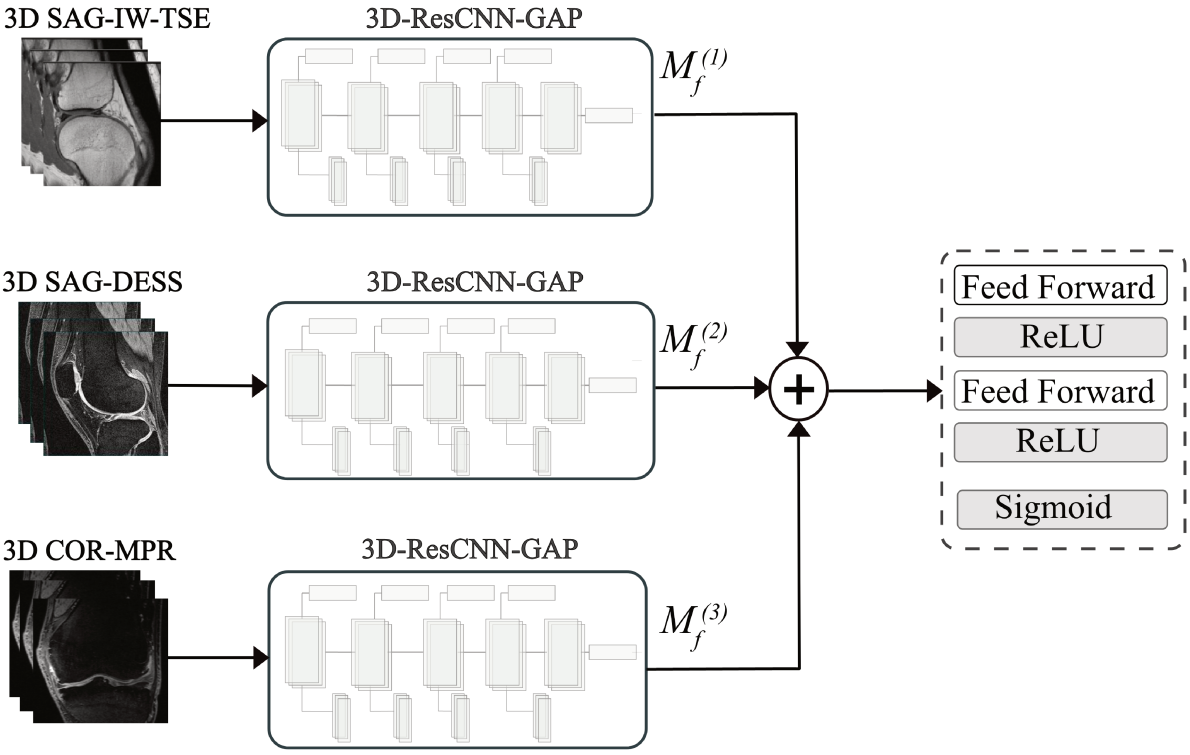


Figure 4.3: Proposed multi-view framework using the proposed 3D-ResCNN-GAP for knee OA detection. ⊕ denotes the concatenation operation.

$$M_{conc} = [M_f^{(1)} \oplus M_f^{(2)} \oplus M_f^{(3)}] \tag{4.7}$$

where ⊕ denotes the concatenation operation. M_{conc} has a length of $\sum_{i=1}^{3} F_i$, where F_i is the number of features extracted from the modality i.

4.3 Experiments

In this section, we compare the performance of several models by conducting experiments and evaluate their accuracy and AUC scores. First, the models were trained on labeled data from the OAI database [46], using different MRI sequences: Sagittal 3D DESS, Sagittal IW TSE, and Coronal MPR, and subsequently the models were trained with the added pseudo-labeled MRI samples.

4.3.1 Dataset

To train and validate the performance of different models across multiple experiments, we used the OAI dataset (previously described in section 2.4.1).

We compare the experimental results using two knee views (coronal and sagittal) from different MRI sequences (3D DESS, IW TSE, and MPR), which is important to gain a better understanding of the impacts of OA on the knee joint.

Table 4.1: Number of MRI samples before and after SSL

MRI acquisition	**Label**	**Before SSL**	**After SSL**
SAG 3D-DESS	Normal / OA	450 / 450	1680 / 2150
SAG IW TSE	Normal / OA	450 / 450	2015 / 2282
COR-MPR	Normal / OA	450 / 450	1880 / 2209

Table 4.2: Distribution of the knee OA patients from the OAI database that participated in this study. KL grades are from the radiographic data of the same patients.

	KL-0	KL-2	KL-3
MOAKS	0	308	142

4.3.2 Data preparation and processing

The OAI dataset contains only 450 patients graded with the MOAKS score as defined in [16], and some samples are missing certain acquisitions. This explains the samples' size difference in Table 4.1. The distribution of the patients is presented in Table 4.2. There were only OA patients with symptoms visible on X-rays according to the Kellgren-Lawrence (KL) scale [14] (KL-2 and KL-3) and none of them was healthy (KL-0). In order to balance our dataset, we included 450 MRI scans of healthy knees that were graded as KL-0 [14].

We used a specific preprocessing method for each type of MRI sequence, including zooming in on the knee joint and cropping the scans to remove irrelevant knee compartments, where the osteophytes and cartilage loss are not observed. The preprocessing method improves the accuracy of our analysis by facilitating better feature extraction and enabling us to focus on the relevant features of the knee joint.

Table 4.3 presents the parameters used to preprocess various MRI slices. These values were determined through multiple experiments to identify the optimal parameters for zooming, cropping, and slice removal.

Table 4.3: Preprocessing parameters for the MRI sequences

MRI acquisition	Slices Dim	Nb. of slices	New Dim	Removed slices
SAG 3D DESS	360x360	160	60x60	[130:160]
SAG IW TSE	444x444	45	60x60	[0:2],[34:45]
COR MPR	444x448	59	70x70	[57:59]

4.3.3 Knee OA detection using the sagittal views

The sagittal view is a cross-sectional view that shows the structures of the knee from the side. This view can provide information about the presence and extent of any abnormality such as the cartilage.

In this chapter, we explored the potential use of two sagittal MRI sequences, namely 3D-DESS and IW-TSE for the detection of knee OA based on the presence of osteophytes and cartilage condition as visible in each sagittal view.

One main difference between the sagittal 3D-DESS and IW-TSE is the manner in which the MR image is generated. 3D-DESS uses a steady state signal to create the image, while IW-TSE uses multiple echoes. The size of each MRI volume is presented in Table 4.3. The number of sequences as well as the size of slices is higher in the 3D-DESS.

In the first experiment, we trained each sagittal MRI scan separately. Secondly, we combined the two sagittal scans, as presented in Section 4.2.3. Table 6.2 shows the results obtained using both sagittal MRI sequences.

4.3.4 Knee OA detection using the coronal view

The coronal view is a cross-sectional view that allows the visualization of the femur, tibia, and patella. In the coronal view, it is possible to see the shape and structure of the cartilage and the presence of osteophytes. In this chapter, we used the 3D Coronal MPR sequence. The configuration and information related to this sequence are presented in Tables 4.1 and 4.3. Table 4.1 shows the number of MRI scans for each MRI sequence, and Table 4.3 shows the size and dimensions of the COR-MPR sequence.

4.3.5 Knee OA detection using both coronal and sagittal views

Multi-view models have been effectively used by the research community as they provide feature aggregation from different views rather than the use of a single view. In this section, we present a multi-view configuration at the core of our methodology to evaluate the combination of different MRI views (sagittal 3D-DESS, IW-TSE, and Coronal MPR) to detect knee OA. By that means, multi-view features (M_f) were aggregated to obtain a single feature vector, M_{conc} as defined by Eq. (4.7). The proposed multi-view framework is presented in Fig. 4.3, and as previously discussed in Sec. 4.2.3.

4.4 Results and Discussion

The objective of this chapter is to conceive and implement an automatic approach to accurately and consistently detect knee OA using 3D MRI data. At the core of this chapter, we introduce 3D-ResCNN-GAP; a deep CNN model that incorporates multiple deep learning mechanisms, including GAP, residuals, and feature aggregation modules. In this section, the experimental results are presented and discussed.

4.4.1 Results

As can be seen in Table 6.2, results show that the proposed 3D-ResCNN-GAP model achieves noteworthy performances, with an overall AUC of 79.01%, 81.48%, and 86.15% for SAG-3D-DESS, SAG-IW-TSE, and COR-MPR, respectively. The use of the multi-view features and SSL further enhanced the performance of the model, resulting in AUC values of 85.32%, 92.85%, and 92.45% respectively for (SAG-3D-DESS + SAG-IW-TSE), (SAG-IW-TSE + COR-MPR), (COR-MPR + SAG-3D-DESS). These results show that combining the multi-view knee MRIs and increasing the training data using our SSL approach boosted our model's performance to achieve an AUC of 93.20%.

Table 4.4: Results obtained with our 3D-ResCNN-GAP model before and after the SSL using the sag 3E-DESS, cor mrp, and sag iw-tse. the table also includes the results of aggregated MRI scans

	Before SSL		After SSL	
MRI acquisition	ACC(%)	AUC (%)	ACC (%)	AUC (%)
Sag 3D-DESS (1)	77.43	79.01	79.13	81.21
Sag IW TSE (2)	79.14	81.48	82.51	85.45
COR MPR (3)	83.49	86.15	87.49	91.45
(1)+(2)	83.88	89.02	84.14	85.32
(2)+(3)	85.56	90.77	87.31	92.85
(1)+(3)	85.66	**91.59**	89.73	92.45
(1)+(2)+(3)	85.22	91.04	90.43	**93.20**

4.4.2 Discussion

To our knowledge, most of the previous studies on knee OA detection were devoted to the use of plain radiographs (X-rays) [13, 133, 134, 147], while the utilization of 3D MRI is comparatively rare for a variety of reasons such as data complexity, and big data which eventually lead to computational costs [102]. This chapter expands upon previous research by examining the feasibility of using multiple MRI views of the knee to evaluate the effectiveness of deep learning for knee OA detection.

Table 4.5: Effects of different components on the performance of the 3D-ResCNN-GAP using the multi-view scans

Ablation setting		Results			
		Before SSL		After SSL	
GAP	Residuals	ACC (%)	AUC (%)	ACC (%)	AUC (%)
✗	✗	79.54	81.39	80.34	82.17
✓	✗	81.84	83.39	83.41	85.28
✗	✓	81.75	83.65	84.75	87.51
✓	✓	85.22	91.04	**90.43**	**93.20**

Table 4.6: Comparison with related work in the literature, using mr images or other biomarkers

Paper	Biomarker	Nb. 3D MRI samples	Results (%)
Chaudhari et al. [145]	osteophytes	124	90
Namiri et al. [100]	Bone	582	89
Guida et al. [76]	knee OA	440	91
Alexopoulos et al. [102]	knee OA	593	66.99
Pedoia et al. [75]	knee OA	4384	82.44
Chang et al. [146]	knee OA	710	85.3
3D-ResCNN-GAP	knee OA	3600	**93.20**

To the best of our knowledge, our proposed approach is the first to apply the reliable OA definition presented in [16] in a deep learning classifier. Some studies have utilized grades of structural OA for 2D X-rays [102], [76], but there is limited evidence of their applicability in assessing knee OA from 3D MR images. In contrast, this chapter relies on reliable MOAKS grades of OA approved by [4, 16] and shows interesting performance compared to these state-of-the-art models.

Using the proposed SSL approach, we were able to label over 3000 additional MR scans that were not previously annotated with MOAKS scores. The results obtained (Table 4.4) show that the proposed SSL technique to augment the size of the dataset improves the model's performance, increasing the AUC from 86.15% to 91.45% in coronal MPR and from 81.48% to 85.45% in sagittal IW-TSE.

When using the coronal MPR or sagittal IW-TSE sequences separately, we reached the highest AUCs for knee OA detection, with an AUC of 91.45 % and 85.45 %, respectively. Interestingly, although the sagittal 3D-DESS scans offer high-resolution images and more sequences, the performance of our model was higher in the sagittal IW-TSE, which has fewer slices and lower resolution.

Table 4.7: Effects of SSL on state-of-the-art models

Model	MRI scans	Before SSL		After SSL	
		ACC(%)	AUC (%)	ACC (%)	AUC (%)
3D InceptionV3	Sag 3D-DESS (1)	70.83	76.11	71.05	77.31
	Sag IW TSE (2)	63.56	75.1	62.14	72.44
	COR MPR (3)	77.33	81.14	77.19	81.32
	(1) + (2)	**81.25**	83.23	81.14	**84.12**
	(2)+(3)	79.44	82.82	81.85	83.54
	(1) + (3)	80.18	**83.89**	75.53	83.65
	(1) + (2)+ (3)	80.56	81.70	**82.41**	82.42
3D Densenet-121	Sag 3D-DESS (1)	70.13	75.15	69.41	73.87
	Sag IW TSE (2)	63.82	75.09	55.83	70.85
	COR MPR (3)	72.30	73.43	71.07	72.41
	(1) + (2)	70.03	79.97	70.03	79.97
	(2)+(3)	79.97	83.05	81.31	85.33
	(1) + (3)	**85.11**	**88.69**	**86.46**	**90.12**
	(1) + (2)+ (3)	83.71	87.03	84.84	89.41

4.4.3 Attention Maps for Decision-Making

Gradient-weighted Class Activation Mapping (Grad-CAM) [148] offers a visual window into deep learning models, enhancing the interpretability of decisions in medical imaging. The application of Grad-CAM in our study underscores the potential of explainable AI in the realm of medical diagnostics. By visualizing the areas of highest focus on MRI scans, particularly over the cartilage and meniscus regions, the heatmaps corroborate the model's alignment with clinical knowledge and diagnostic criteria for knee OA.

Attention maps presented in Figure 4.4 reveal that the model's attention is concentrated on the same anatomical structures, meniscus and cartilage regions, that clinicians examine when evaluating knee OA using MOAKS.

4.4.4 Effects of the SSL on state-of-the-art models

Using the experiments detailed in Section 4.3, we also investigated the impact of the SSL on the performance of InceptionV3 and DensNET-121, which were also trained from scratch on the same MRI scans. As an example, our model achieved an AUC of 86.15% with the COR-MRP slices (Table 6.2), while InceptionV3 and DensNET-121 achieved AUCs of 81.14%, and 73.43%, respectively (Table 4.7). These findings are also valid for other views, which confirm that the proposed processing pipeline and the building blocks of the modules significantly boosted the performance of the 3D-ResCNN-GAP model. In addition, as shown in Table 4.7, without the semi-supervised labeling step, our proposed 3D-ResCNN-GAP model attains better scores (Table 6.2) compared to these models.

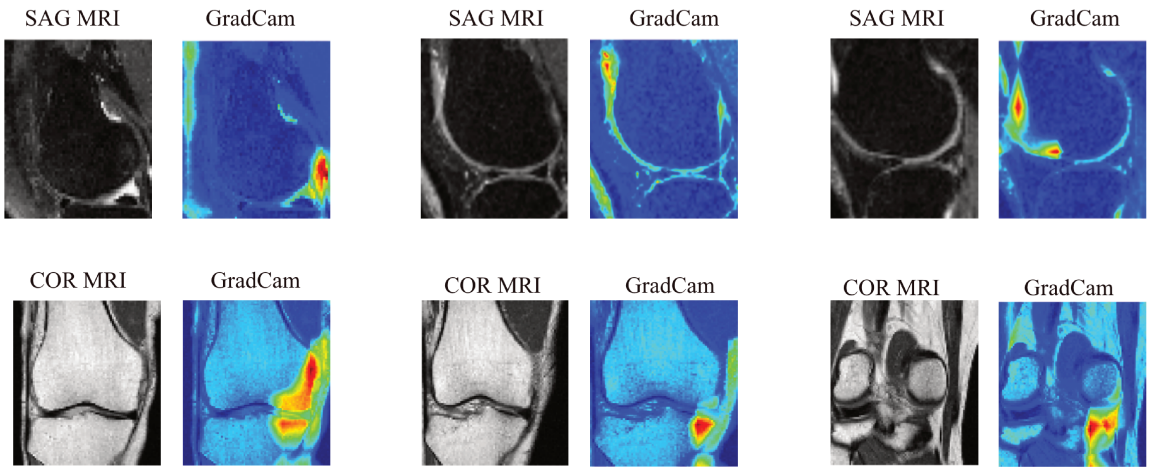


Figure 4.4: Obtained attention maps for Sagittal and Coronal MRI Scans. This visualization shows that the model focused its attention on the meniscus and cartilage regions, highlighted in red, aligning with key indicators used in MOAKS for diagnosing knee OA

.

Furthermore, the evaluation of the 3D-InceptionV3 and 3D-DensNET-121 shows that the multi-view network enhances their ability to detect knee OA more accurately for the models that achieved good classification performance before the pseudo-labeling. However, the results before the semi-supervised training were not sufficiently good to generalize and make accurate predictions on unlabeled scans, hence the performance did not increase after semi-supervised labeling.

To validate the effectiveness of our proposed model in diagnosing knee OA using multi-view MRI scans, we conducted a statistical analysis comparing its performance against the 3D-DenseNet-121 model which is the second performant model in our experiments. We hypothesized that our model would demonstrate a statistically significant improvement over the baseline.

Table 4.8: Comparison of AUC scores between our proposed model and 3D-DenseNet-12

Metric	**Our model**	**3D-DenseNet-121**	**p-value**
AUC	93.20	89.41	0.0011

As shown in table 4.8, the p-value obtained (0.0011) is below the significance level of 0.01. This indicates that the difference in the AUC scores between our model and the baseline model is statistically significant.

4.4.5 Ablation studies

We investigated the effectiveness of the 3D-ResCNN-GAP model's components (GAP and residual connection) on its performance, as well as the effect of aggregating the multi-view scans. Such experiments provide valuable insights into the importance of these components for the overall performance and enable the design of the model to be optimized. Table 6.2 shows that training the 3D-ResCNN-GAP in a multi-view framework improves its ability to accurately detect knee OA. The findings suggest that the combination of learned features from both sagittal and coronal scans provides more information, resulting in better classification performance.

In Table 4.5, we can see that when both GAP concatenation and Residuals are excluded, the model achieved the lowest AUC value of 82.17%. The inclusion of only GAP concatenation or Residuals led to a considerable improvement in the model's performance, with AUC values ranging from 85.28% to 87.51%, respectively. The highest performance was achieved when both GAP concatenation and residual units were incorporated into the model, resulting in an AUC of 93.20%. The same performance results were observed for the experiments before the semi-supervised training.

The ablative study suggests that both GAP concatenation and Residuals units play an important role in the performance of the 3D-ResCNN-GAP model and that their inclusion is critical for achieving optimal performance.

4.4.6 Limitation of the proposed study

In order to balance our dataset, we utilized knees from patients with KL-0 grade as proxies for healthy controls. It is noteworthy that a KL-0 grade does not necessarily correspond to a knee that would also be classified as completely healthy or devoid of OA changes according to the MOAKS scoring system. Unlike the KL grade, the MOAKS score provides a more detailed examination of the knee by assessing a range of MRI-visible features, including but not limited to cartilage defects, bone marrow lesions, and meniscal tears, which are not accounted for in the radiographic KL grading system. This suggests that our classification of 'healthy' controls may inadvertently include individuals with MRI-detectable signs of OA not visible on plain radiographs.

To conclude, with regard to the KL scoring system, our approach enables early prediction of OA (KL-0 vs. KL-2). If we consider only the MOAKS score, this assertion is rather difficult to support, as we do not have data with a zero MOAKS score corresponding to patients without non-OA.

4.5 Conclusion

Knee OA is a prevalent and debilitating joint disorder that affects millions of people worldwide, and the diagnosis is crucial for effective management and treatment. In this chapter, we have

proposed a new framework for the detection of knee OA using multi-view 3D MR scans. We have also introduced a semi-supervised multi-view framework and conceived a novel 3D deep CNN model for knee OA detection using multiple views from 3D MR images. We conducted a series of experiments to evaluate the effectiveness of the proposed model compared to SOTA models for the detection of knee OA. With an AUC of 93.20%, results show that our proposed methodology is more efficient than classical CNN models. Overall, our findings suggest that the integration of multiple MRI views of the knee enhances the performance of all the models evaluated. Also, the proposed semi-supervised learning approach enhances the training step, which also contributes to the overall performance of all the evaluated models. These findings highlight the potential of deep learning to revolutionize the diagnosis of knee OA and make it faster and more accurate.

In the following chapter, we explore an advanced deep-learning network designed to detect other knee injuries and also include additional MRI views, namely axial, coronal, and sagittal.

CHAPTER 5

M3NET: A 3D JOINT MULTIVIEW MULTI-LABEL MULTIMODAL NETWORK FOR KNEE INJURIES CLASSIFICATION

Knee injuries are among the principal causes of disability, affecting millions of individuals worldwide. The associated risk of developing knee OA further emphasizes the critical importance of diagnosis in clinical settings. Therefore, the development of an accurate and cost-effective system for identifying knee injuries is of critical importance. Nevertheless, training models for individual tasks is expensive, and single-view approaches often miss features visible in other views. In this chapter, we present our approach to address these limitations for efficient knee injuries analysis.

5.1 Introduction

Knee injury is a widespread condition affecting millions of individuals globally, causing severe pain, reduced mobility, and diminished quality of life. Injuries to the meniscus and cartilage are particularly common and can lead to prolonged rehabilitation and costly medical interventions [149]. Diagnosis and early treatment of knee injuries are crucial to prevent further damage, such as knee OA, and improve patients' lives [150].

In clinical practice, radiologists often use both X-ray and MRI to evaluate knee injuries and determine the extent of meniscus and cartilage damage. X-ray imaging is typically the first imaging modality used to evaluate knee injuries, and MRI is used to confirm the diagnosis and assess the severity of the injury [150].

Furthermore, clinical studies have shown a strong association between advanced meniscus and cartilage damage, and the two conditions can occur simultaneously [151]. On the contrary, cartilage damage may cause increased stress on the meniscus, causing it to wear down more

quickly and increasing the risk of meniscus damage. This clear evidence of the strong association between meniscus damage and cartilage damage requires clinicians to consider both conditions when evaluating and treating knee injuries.

Multiview and multi-label learning has emerged as an important research direction in recent years, with potential applications in a wide range of fields, including medical imaging [142, 152]. Multiview multi-label learning can be used to jointly analyze multiple images of the same anatomical structure, captured using different imaging modalities, and to perform multiple tasks, such as classification, segmentation, and disease diagnosis. By leveraging the complementary information captured by different imaging modalities, multiview multi-label learning can improve the accuracy and robustness of medical image analysis, as well as reduce the need for expert annotations.

Recent advances in deep learning have enabled the development of powerful multiview, multi-label image classification techniques for medical imaging [142]. These techniques can extract high-level features from multiple views of an image, such as CT [153], MRI [152], and X-rays [154], and use these features to jointly classify the image across multiple tasks, such as disease diagnosis and severity assessment. By sharing the same feature representation across tasks, these techniques can effectively reduce the number of parameters and computation required, while improving the generalization performance of the model.

In this chapter, we deal with two knee injuries that can potentially cause disability in patients, namely meniscus tears and cartilage degeneration. Due to the important impact of these two types of injuries on individuals and their potential risks of developing knee OA, multiple studies have developed methods to assist radiologists and clinicians in effectively diagnosing such cases. However, current methods have limitations. They often employ a single imaging modality, X-ray [155] or MRI [84, 156], where studies on MRI are comparatively fewer. In addition, these approaches often use a single model for each type of disease. This can potentially result in the omission of important features and information that might only be visible from certain views, which limits the generalizability of the findings. Another limitation is that some approaches operate on 2D slices or X-rays, potentially missing out on important feature correlations in the 3D space.

One key challenge in multi-view multi-label image classification for medical imaging is how to effectively combine features extracted from different modalities and tasks. Feature fusion techniques have been proposed to address this challenge. These techniques aim to combine the features learned from multiple views and tasks to maximize the overall classification accuracy [152].

To overcome some of these challenges, we introduce a novel model M3NET - Multiview Multimodal multi-label NETwork, capable of aggregating learned features from 3D MRI multiviews and 2D X-rays to classify the meniscus and cartilage tears simultaneously. The multiview framework consists of combining informative features from three 3D MRI views (axial, coronal, and

sagittal). The multi-label part consists of predicting the presence of full cartilage loss and torn meniscus simultaneously in the same network, and the multimodal part consists of learning radiological features from knee X-rays associated with cartilage and meniscus and combining them with the 3D MRI features. Further, the proposed M3NET framework introduces a comparative experiment to evaluate separate-feature fusion using feature concatenation, averaging, summation, and multimodal attention. Our major contributions are as follows:

- We develop the M3NET model, an innovative integration of multi-label learning, multi-view learning, and multi-modal analysis into a single network. M3NET excels in learning representative features from multiple perspectives, enabling the simultaneous detection of knee cartilage lesions and meniscus tears from MRI and X-ray scans.

- We leverage a customized 3D YOLOv8 model which is specifically designed for precise automatic detection of knee joint regions in MRI scans.

- We introduce a novel fusion architecture named separate-feature late fusion (SFLF) that effectively processes both X-ray and MRI data, utilizing multiple feature aggregation techniques for consistent classification across modalities.

- We conduct an extensive evaluation on 1519 MR scans from the OAI initiative, demonstrating our model's robust performance across axial, sagittal, and coronal views.

5.2 Background

5.2.1 Multiview Learning

Multiview Learning plays a crucial role in enhancing image analysis and interpretation, particularly within the realm of medical imaging [142, 152, 157]. The fundamental premise behind this approach lies in the intuitive notion that observing an object, such as the intricate knee joint, from multiple perspectives can provide a richer understanding of its complex structure and potential pathologies. This is analogous to a physician examining a patient from different angles to gain a comprehensive assessment of their condition.

In the context of knee MRI, the standard acquisition protocol involves capturing images from three orthogonal planes: axial, sagittal, and coronal. These views, each offering a unique vantage point, provide complementary information that contributes to a holistic representation of the knee joint. By integrating these diverse perspectives, we can effectively leverage the inherent richness of multiview data to enhance the accuracy and robustness of automated knee injury detection and classification.

The utilization of multiview learning in knee MRI analysis allows for the extraction of a richer set of features, encompassing both local and global information. These features can then be used

to train sophisticated deep learning models capable of automatically detecting and classifying knee injuries with higher accuracy compared to models trained on single-view data [157, 158].

Among the various multiview learning techniques, feature concatenation using deep learning has emerged as a powerful and versatile approach. This method leverages the ability of deep neural networks, particularly CNNs, to automatically learn hierarchical representations from raw image data. In the context of multiview learning, separate CNNs can be trained on each view to extract high-level features. These features, encapsulating the unique information captured by each perspective, are then concatenated into a single, comprehensive feature vector [159].

This concatenated feature vector serves as a rich and informative input for subsequent machine learning models, enabling them to learn complex relationships between the different views and ultimately improve the accuracy of knee injury detection and classification. The power of this approach lies in its ability to effectively combine the complementary information from each view, leading to a more holistic and accurate representation of the knee joint.

Furthermore, the flexibility of deep learning architectures allows for the incorporation of various advanced techniques, such as attention mechanisms and multi-scale feature extraction, to further enhance the performance of multiview feature concatenation. These techniques enable the model to selectively focus on relevant features and capture information at different levels of granularity, leading to improved accuracy and robustness.

5.2.2 Multimodal learning

Multimodal learning holds significant promise for enhancing the diagnostic process by integrating information from diverse imaging modalities [160, 161]. In knee injuries analysis, combining MRI and X-ray data offers a compelling approach. MRI excels at visualizing soft tissues, such as cartilage and ligaments, while X-ray provides valuable insights into bone morphology and joint space narrowing, key indicators of OA progression. Fusing these complementary modalities allows for a more comprehensive and accurate assessment of knee OA.

The rationale for combining MRI and X-ray data stems from their respective strengths and limitations. MRI, with its superior soft tissue contrast, facilitates detailed evaluation of cartilage degeneration, meniscal damage, and synovial inflammation, providing crucial information about the early stages of OA. However, MRI can be expensive, time-consuming, and its availability may be limited. X-ray, conversely, is widely accessible, cost-effective, and offers a readily interpretable view of bone structure and joint space narrowing, hallmarks of advanced OA. However, X-ray provides limited information about soft tissue structures and may not be sensitive to early changes associated with OA. By integrating MRI and X-ray data, we leverage the strengths of each modality while mitigating their limitations, potentially leading to earlier and more accurate diagnosis, as well as a more nuanced understanding of OA progression.

In our experimental setup, we will evaluate the effectiveness of incorporating and removing

multimodality in our proposed model, M3NET. By systematically adding and removing MRI and X-ray data, we aim to quantify the impact of multimodal fusion on the performance of our model for knee injuries analysis. This analysis will provide valuable insights into the benefits and limitations of multimodal learning in this context and guide the development of more robust and effective diagnostic tools. This systematic evaluation will allow us to assess the contribution of each modality to the overall performance of M3NET and identify potential scenarios where multimodal fusion offers significant advantages over single-modality analysis. Furthermore, it will help us understand the robustness of our model to missing data and inform strategies for handling such scenarios in clinical practice.

Several strategies have been proposed for multimodal learning in medical imaging, each offering distinct advantages and disadvantages depending on the specific application and data characteristics. These strategies can be broadly categorized as:

- **Early Fusion:** This approach involves combining the raw data from different modalities before feeding it into a model. This can be achieved through various techniques, such as concatenating the raw pixel values or creating a composite image by stacking the different modalities as channels. While early fusion can be straightforward to implement, it often requires careful consideration of the data heterogeneity and alignment issues inherent in multimodal data. Different modalities may have varying resolutions, noise levels, and represent different physical properties, making direct fusion of raw data challenging.

- **Late Fusion:** In contrast to early fusion, late fusion involves training separate models for each modality and then combining their predictions. This approach allows each modality to be processed independently, leveraging its specific characteristics and potentially leading to better performance for each individual modality. The predictions from each model are then combined, often through simple averaging or weighted averaging, to produce a final prediction. Late fusion offers greater flexibility in handling data heterogeneity and can be particularly effective when modalities have significantly different characteristics.

- **Intermediate Fusion:** This strategy combines features extracted from different modalities at an intermediate level within the model architecture. This allows for a more nuanced integration of information from different modalities, potentially capturing complex relationships between them. Intermediate fusion often involves designing specialized architectures that can effectively fuse features extracted from different modalities, such as using fusion layers or attention mechanisms. This approach can be more complex to implement but offers the potential for improved performance by leveraging the complementary information from different modalities.

Each of these fusion strategies has its own strengths and weaknesses, and the choice of the most appropriate strategy depends on the specific characteristics of the data and the desired

outcome. In this study, we focus on *late fusion* due to its inherent flexibility and robustness in handling the complexities of multimodal knee MRI data. Late fusion allows us to train specialized models for each view (axial, sagittal, coronal) and MRI sequence (e.g., T1-weighted, T2-weighted, proton density), effectively capturing the unique information provided by each modality. This approach is particularly appealing when dealing with data heterogeneity and potential misalignment between different views and sequences. By training separate models and fusing their predictions, we can leverage the strengths of each modality while mitigating the challenges associated with direct fusion of raw data. This strategy allows for a more modular and adaptable approach, enabling us to explore and optimize the performance of each individual model before combining their predictions for a more comprehensive and accurate assessment of knee injuries.

5.2.3 Multi-Label Learning

Multi-label learning in deep learning enables the prediction of multiple labels simultaneously for a single instance, contrasting with traditional single-label classification [162]. This is particularly relevant for knee injury analysis, where a single traumatic event can often affect multiple anatomical structures. For example, a patient might present with both a meniscus tear and cartilage damage. A multi-label model can identify both injuries concurrently, offering a more comprehensive understanding of the overall knee pathology.

The strength of multi-label learning lies in its ability to exploit correlations between labels. Certain knee injuries frequently co-occur, such as combined ACL and MCL tears. By learning these relationships, the model can refine its predictive accuracy and provide a more nuanced diagnosis. This is achieved through architectures that output multiple binary classifications, one for each potential injury. Deep learning models, particularly CNNs, are well-suited for this task due to their ability to extract intricate patterns and features from medical images.

A multi-label model can perform a comprehensive evaluation of the knee, encompassing:

- **Injury Classification:** Distinguishing between various types of tears (e.g., ACL, meniscus), fractures (e.g., tibial plateau, patella), and other conditions like cartilage damage (osteochondral defects), osteophytes, and bone marrow edema. For instance, the model can differentiate between a bucket-handle meniscus tear and a radial tear, or identify the presence of both cartilage damage and osteophytes in the same knee joint. This granular classification facilitates targeted interventions.

- **Injury Localization:** Pinpointing the precise location of each injury within the 3D space of the knee joint. For example, the model could identify the specific location of a meniscus tear (e.g., anterior horn, posterior horn, body) or highlight the region of cartilage damage on a 3D rendering. This spatial information is crucial for pre-operative planning, especially for minimally invasive procedures.

- **Severity Assessment:** Quantifying the extent of damage for each identified injury. For example, the model could classify a meniscus tear based on its size and displacement, grade the severity of cartilage damage using the Outerbridge classification, or assess the size and location of osteophytes. This information can aid in prognosis and treatment planning.

By integrating these tasks, multi-label learning provides a holistic assessment of knee injuries, surpassing the capabilities of single-task models. Additionally, multilabel models offer advantages in medical imaging analysis [163, 164]:

- **Improved Diagnostic Accuracy:** Identifying co-occurring injuries that might be missed by single-label models, leading to a more complete and accurate diagnosis. For example, identifying both a meniscus tear and cartilage damage, which might be overlooked if only focusing on the more prominent injury.
- **Enhanced Clinical Decision-Making:** Providing clinicians with a unified report detailing all identified injuries, their locations, and severities, facilitating informed treatment decisions. This is especially valuable in cases with multiple pathologies, such as a combination of meniscus tear, cartilage damage, and osteophytes.
- **Increased Efficiency:** Streamlining the analysis process by evaluating multiple aspects of knee injuries concurrently, saving time for radiologists and clinicians.

By harnessing the power of deep learning, multilabel learning holds immense potential to transform knee injury diagnosis and management, paving the way for more accurate, efficient, and personalized care.

Adopting multiview, multimodal, and multi-label learning approaches holds significant promise for enhancing the accuracy and efficiency of knee injury detection. By leveraging the complementary information offered by different views, modalities, and diagnostic tasks, these approaches can lead to more precise and streamlined diagnostic outcomes, ultimately improving patient care.

5.3 Proposed M3NET

This section introduces our proposed M3NET model, designed to leverage multimodality and multiview imaging for the detection of cartilage lesions and meniscus tears. We explore the efficacy of separate-feature late fusion (SFLF), as detailed in Fig. 5.1. These strategies employ various feature aggregation mechanisms, including linear fusion, averaging, concatenation, and multimodal attention, to enhance model performance. In the following sections, we will get into the various components of M3NET's methodology (multiview, multimodal, and multi-label) and explore our strategic approach to optimizing the training process for the entire network.

5.3.1 Multiview Process

In the multiview setting, we utilize three MRI views namely: axial MPR, Sagittal SAG-TSE, and coronal MPR. These views are integrated into a single network to enhance the feature representation of the knee joint from various views.

To formulate the multiview setting, we consider three custom 3D-ResNet encoders [93]. These encoders take three distinct views of knee MRI scans (axial, coronal, and sagittal) as inputs, allowing the network to capture diverse information related to the knee joint. The 3D-ResNet models are trained in parallel, and the weights are not shared between the models. This approach ensures that each model can focus on its specific view and extract unique insights.

Let $I_{axi}^{(i)}$, $I_{sag}^{(i)}$ and $I_{cor}^{(i)}$ denote the axial, sagittal and coronal views of the i^{th} scan, respectively. It should be noted that the input MRI scans have different shapes, hence we designed the M3NET to support multi-size MRI scans, so as to take as input multiview MRI scans in separate convolution nodes. Each model processes each MRI view independently and produces two output predictions $\hat{y}_m^{(i)}$ and $\hat{y}_c^{(i)}$, corresponding to the predicted labels for meniscus tears and cartilage loss, respectively. We represent the model as a function f_θ, where θ represents the learnable parameters of the network: $\hat{y}^{(i)} = f_\theta(I_{axi}^{(i)}, I_{sag}^{(i)}, I_{cor}^{(i)})$.

During training, the parameters θ are learned by minimizing a loss function that compares the predicted outputs $\hat{y}^{(i)}$ with the ground-truth labels $y_m^{(i)}$ and $y_c^{(i)}$. The objective function of the Multi-view setting in the M3NET is as follows:

$$M_v = \min_\theta \frac{1}{N} \sum_{i=1}^{N} \left[L\left(\hat{y}_m^{(i)}, y_m^{(i)}\right) + L\left(\hat{y}_c^{(i)}, y_c^{(i)}\right) \right], \tag{5.1}$$

where $\hat{y}_m^{(i)} = f_\theta(I_{axi}^{(i)}, I_{sag}^{(i)}, I_{cor}^{(i)})$ represents the predicted labels for meniscus tears from the extracted features f_θ using $I_{axi}, I_{sag}, I_{cor}$ and $\hat{y}_c^{(i)} = f_\theta(I_{axi}^{(i)}, I_{sag}^{(i)}, I_{cor}^{(i)})$ represents the predicted labels for cartilage loss. L is the binary cross entropy loss function:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \left[y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \right]. \tag{5.2}$$

5.3.2 Multimodal Process

In the core of our proposed M3NET, we integrate a multimodal framework that combines MRI and X-ray modalities and extracts features from three 3D MRI acquisitions, and 2D X-ray images of the same knee. The extracted features are fused to generate a joint representation that captures complementary information from all modalities. The objective function of multimodality in the M3NET is defined as follows:

$$M_m = \min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L\left(\sum_{j=1}^{2} F_{\theta j}(M_j^{(i)}), y^{(i)} \right), \tag{5.3}$$

where $F_\theta j$ represents the extracted feature from modality j, where $j \in [X-ray, MRI]$, and $M_j^{(i)}$ represents the input for sample i from modality j. The outer loss function L computes the loss between the aggregated feature extractions from all modalities and the true label $y^{(i)}$.

The fused features are then passed through a multi-layer perceptron (see Fig. 5.1, which is trained to predict the presence or absence of cartilage lesions and meniscus tears disease, as explained in the following section.

5.3.3 Multi-label Process

Let F_{axi}, F_{sag} and F_{cor} be the feature extractors for axial, sagittal, and coronal scans, respectively.

Given a training set $D = (I_{axi}^{(i)}, I_{sag}^{(i)}, I_{cor}^{(i)}, y_m^{(i)}, y_c^{(i)})_{i=1}^{N}$, where $y_c^{(i)}$ denotes the label of the i^{th} sample for cartilage loss, and $y_m^{(i)}$ is the label of the same i^{th} sample for meniscus tears.
Using the training set D, we represent the multi-label approach as follows:

- For the meniscus tear detection task, we train the M3NET using fused meniscus features, f_m with w_m parameters, which takes the fused features z_i as input and outputs a predicted probability, $\hat{y}_m$:

$$\hat{y}_m = f_m(z_i; w_m). \tag{5.4}$$

- For the task of cartilage lesion detection, we train the M3NET model using specific cartilage features, denoted as f_c, along with their corresponding parameters, w_c. The model inputs include the fused features z_i, which are integrated into the network to generate a predicted probability, $\hat{y}_c$, for the presence of cartilage lesions:

$$\hat{y}_c = f_c(z_i; w_c). \tag{5.5}$$

To train the proposed M3NET, we minimize the training loss over the training set for both tasks simultaneously using the following objective function:

$$M_t = min_{wm,wc} \frac{1}{N} \sum_{i=1}^{N} L_m(\hat{y_m}^{(i)}, Y_m^{(i)}) + L_c(\hat{y_c}^{(i)}, Y_c^{(i)}), \tag{5.6}$$

where L_m and L_c are the loss functions for the meniscus and cartilage tasks, respectively. And Y_m, Y_c are the true labels for meniscus and cartilage scans, respectively.

5.3.4 Separate-Feature Fusion Strategy in M3NET

M3NET leverages a novel separate-feature fusion strategy to harness the distinct and complementary information embedded within MRI and X-ray data for the robust identification of meniscus and cartilage pathologies. This approach capitalizes on the strengths of each modality while mitigating their inherent limitations.

In this chapter, we used our 3D-RES-GAP model (presented in the previous chapter, see section (4.2.1) as an encoder part of M3NET for the MRI input branches, while we adapted 3D-RES-GAP to support 2D data for the X-ray branch. This parallel processing paradigm allows for the extraction of modality-specific features optimally tailored for characterizing meniscus and cartilage abnormalities.

Each network branch, responsible for processing either MRI or X-ray data, is meticulously designed to extract two distinct feature sets: one specifically characterizing meniscus abnormalities and the other delineating cartilage lesions. This separation allows the model to learn specialized representations for each anatomical structure, potentially enhancing both performance and interpretability. By focusing on distinct feature sets, the model can better capture subtle nuances in imaging characteristics that might be indicative of specific pathologies.

To achieve this separate feature extraction, we utilize a dual-output layer architecture within each model. Instead of a single output layer producing a combined representation, two parallel output layers are employed, each focusing on a specific anatomical structure. This architecture enables the model to learn distinct feature representations while retaining the ability to capture interdependencies between the meniscus and cartilage through shared layers preceding the dual-output layer. This shared representation allows information to flow between the two feature extraction pathways, fostering a more holistic understanding of the knee joint.

The features extracted from the MRI and X-ray branches are subsequently fused using specialized layers that integrate the meniscus and cartilage feature sets across modalities. We rigorously evaluate several fusion techniques to determine the optimal approach for combining these multimodal multiview representations.

5.3.5 Fusion Methods

This section delves into the intricacies of feature fusion, a critical element in maximizing deep learning model performance when integrating data from multiple MRI views and X-ray images. We explore various fusion techniques to effectively combine these feature representations.

Let F_{cor}, F_{xray}, F_{sag}, and F_{ax} represent the feature vectors extracted from 3D Coronal MPR, 2D X-ray, 3D Sagittal TSE, and 3D Axial scans, respectively.

1. **Linear Fusion (Feature Sum)**

   M3NET leverages feature addition for linear fusion of multi-modal and multiview imaging data, capturing complementary information across different modalities. Given the feature

vectors from corresponding imaging techniques, the linearly fused feature vector F_{linear} is defined as:

$$F_{\text{linear}} = F_{\text{cor}} + F_{\text{xray}} + F_{\text{sag}} + F_{\text{ax}}, \tag{5.7}$$

where + denotes element-wise vector addition. This fusion strategy promotes feature synergy and is instrumental for subsequent meniscus and cartilage evaluation tasks within M3NET.

2. **Feature Concatenation**

   M3NET employs feature concatenation to integrate multi-view MRI and X-ray features for enhanced representation. The concatenated feature vector F_{concat} is formulated as:

$$F_{\text{concat}} = [F_{\text{cor}} \oplus F_{\text{xray}} \oplus F_{\text{sag}} \oplus F_{\text{ax}}], \tag{5.8}$$

   where $\oplus$ represents the concatenation operator, the concatenated vector F_{concat} serves as input for dedicated dense layers, facilitating simultaneous meniscus and cartilage pathology predictions within the unified framework of M3NET.

3. **Feature Averaging**

   The feature averaging strategy consolidates extracted feature sets from the respective imaging modalities into a single representation. The averaged feature vector F_{avg} is computed as the arithmetic mean across all modalities for each feature index, subsequently utilized for both meniscus and cartilage detection tasks. Formally, the fusion process for a set of M modalities is defined as:

$$F_{\text{avg}} = \frac{1}{M} \sum_{i=1}^{M} F_i, \tag{5.9}$$

   where F_i represents the feature set from the i-th modality. This unified feature vector encapsulates a comprehensive representation conducive for subsequent classification layers in M3NET.

4. **Multimodal Attention Fusion**

   Given the feature sets extracted from different modalities, denoted as F_{cor}, F_{xray}, F_{sag}, and F_{ax} for 3D coronal MPR, 2D X-ray, 3D sagittal TSE, and 3D axial MRI, respectively, we consider each feature set containing N features extracted by the 3D-ResNet models.

To compute attention scores, we introduce a trainable attention network parameterized by θ. The attention scores α_i for each feature set F_i, where $i \in \{\text{cor}, \text{xray}, \text{sag}, \text{ax}\}$, are calculated as:

$$\alpha_i = \frac{\exp(w_i^\top F_i)}{\sum_{j \in \{\text{cor,xray,sag,ax}\}} \exp(w_j^\top F_j)}, \tag{5.10}$$

where w_i are learned weight vectors signifying the importance of the features in F_i.

We perform attention-weighted feature fusion for meniscus and cartilage lesion detection tasks by computing:

$$F_k = \sum_{i \in \{\text{cor,xray,sag,ax}\}} \alpha_{i,k} \cdot F_{i,k}, \tag{5.11}$$

where $\alpha_{i,k}$ is the weighting factor for the features of sample i specific to the injury class k, and $F_{i,k}$ represents the features of sample i from the injury class feature $k \in$ {meniscus features, cartilage features}.

This process yields two attention-weighted composite feature representations, $F_{\text{meniscus\_fused}}$ and $F_{\text{cartilage\_fused}}$, for subsequent analysis tasks.

These specialized fusion layers are implemented as trainable layers within the network architecture, enabling the model to learn optimal weights for combining modality-specific features. The fused feature representation is then used for the final prediction of meniscus and cartilage pathologies.

We conduct an empirical assessment to gauge the impact of each fusion mechanism on the model's predictive accuracy, identifying the most effective fusion strategy for combining MRI and X-ray data in the context of meniscus and cartilage pathology detection. This evaluation sets the stage for a robust comparison with traditional late fusion strategies, where features are extracted and processed separately for each modality before being combined at the final prediction stage. By comparing our separate-feature fusion approach with late fusion, we can assess the potential benefits of early integration and joint feature learning across modalities.

5.4 Experiments

5.4.1 Dataset

This study utilizes the comprehensive dataset provided by the OAI database, which includes multiple imaging modalities. A total of 1,519 MRI scans were selected based on their annotations with MOAKS labels for cartilage loss. Concurrently, meniscus labels and corresponding X-rays

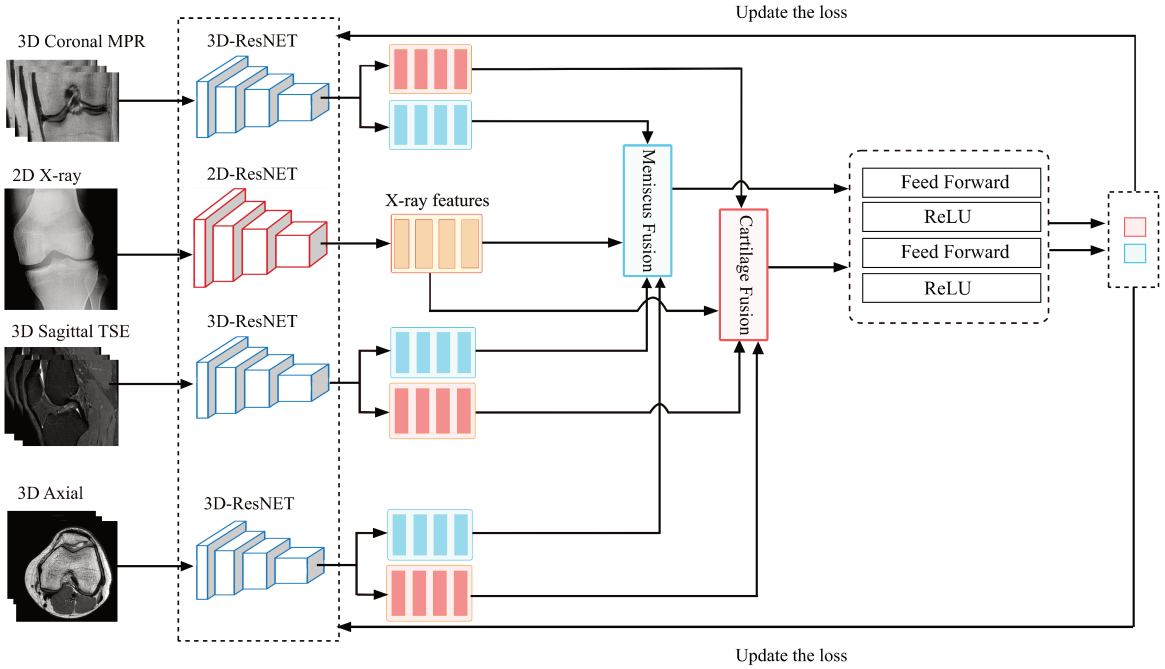


Figure 5.1: M3NET using separate-feature late fusion (SFLF). Here, we make sure that the features are separated for each injury. The blue rectangles refer to the features related to the meniscus, whereas the red rectangles refer to features related to cartilage loss.

for these patients were retrieved to augment the dataset and provide a fuller understanding of each subject's condition. The X-ray features, while included, were not assigned specific labels, thereby serving as supplementary data to the MRI scans' detailed labels.

The OAI database comprises 2,400 MRI scans annotated with MOAKS labels for meniscus injuries. In an effort to maintain consistency across data subsets, the selection was limited to the same 1,519 scans that had been labeled for cartilage loss. The process for assigning simplified MOAKS labels to these scans is described in the subsequent section.

Additionally, three-view sequence data were extracted for each selected MRI scan to provide comprehensive coverage of the anatomical regions of interest.

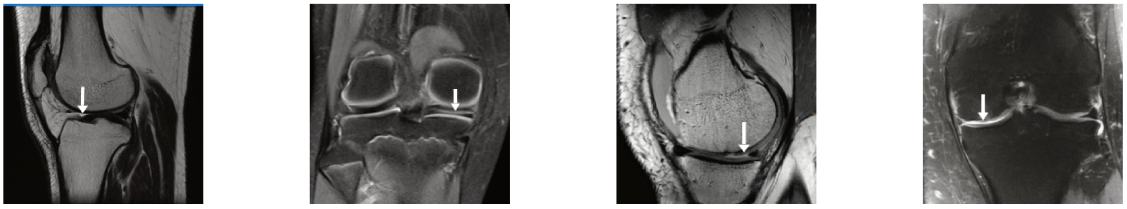
5.4.2 Data Preparation and Processing

MRI scans were processed to identify and extract ROI pertinent to both cartilage and meniscus injuries. A custom YOLOv8 model was employed for this purpose, with its training and configuration details provided in Section 3.1.4.2.

The MOAKS scoring system, typically detailed at the subregion level, was simplified for the purpose of this study. The focus shifted towards a more generalized assessment of injury severity, as detailed below:

- **Cartilage Loss Assessment in Knee MRI:** The primary concern in OA is the integrity of

articular cartilage. Traditionally, MOAKS scores are detailed, assessing the size and depth of cartilage damage in specific subregions. In contrast, our approach introduces a binary classification, where a knee MRI scan is classified as exhibiting full-thickness cartilage loss if the damaged area encompasses 75% or more of the cartilage. This threshold was selected based on its clinical relevance in indicating severe cartilage degradation.



a) torn meniscus in lateral side from sagittal view

b) torn meniscus from coronal view

c) cartilage loss in medial side from sagittal view

(d) cartilage loss from coronal view

Figure 5.2: Complete cartilage loss and a torn meniscus in the knee joint. Both coronal and sagittal scans show details information from the medial and lateral sides of the knee that cannot be obtained through the traditional X-ray imaging technique, which can be valuable for diagnosing and monitoring knee joint conditions and guiding treatment decisions.

- **Meniscus Damage Assessment in Knee MRI:** The meniscus plays a crucial role in the mechanical functionality of the knee, contributing to load distribution, shock absorption, and joint stability (meniscus pointed in white arrows in Fig. 5.2). It is well-documented that meniscus damage is closely linked to the progression of OA [24]. In our simplified scoring approach, an MRI scan is classified as indicative of a torn meniscus based on the presence of any graded abnormalities in the lateral or medial meniscus views.

5.5 Results

Table 5.1 presents the performance metrics, including AUC and ACC, for each fusion technique and fusion modality.

Table 5.1: Experimental results of M3NET using multiple fusion techniques.

Fusion	**Modality**	**Cartilage**		**Meniscus**	
		AUC	ACC	AUC	ACC
Separate-feature Fusion	Linear	**86.87**	**87.71**	**87.42**	**87.39**
	Concatenation	85.22	85.62	85.69	85.40
	Averaging	84.17	84.50	84.43	84.92
	MultiModal attention	85.43	85.29	85.23	85.09

In the Separate-Feature fusion experiments, the linear fusion modality consistently outperforms the other methods, achieving an AUC of 86.87 for cartilage and 87.42 for meniscus, with ACC values of 87.71 and 87.39, respectively. The Concatenation fusion modality, while robust, falls slightly behind the linear fusion, with AUC values of 85.22 for cartilage and 85.69 for meniscus, and ACC values of 85.62 and 85.40, respectively. Averaging modality within separate-feature fusion also trails behind the linear fusion, with AUC values of 84.17 for cartilage and 84.43 for meniscus, and ACC values of 84.50 and 84.92, respectively. The MultiModal Attention fusion modality delivers competitive performance but is generally slightly below the linear fusion.

5.5.1 Discussion

In this section, we discuss experimental results that evaluate the performance of various fusion techniques within the context of our M3NET framework. The primary objective of our study is to simultaneously classify meniscus and cartilage injuries by aggregating features extracted from multi-view 3D MRI scans and 2D X-rays. We have investigated the separate-feature fusion method, employing diverse modalities for feature aggregation. This approach was chosen to leverage the unique information captured by each imaging modality while maintaining a computationally manageable framework.

As illustrated in Table 5.1, a comparison of the fusion modalities reveals that separate-feature fusion techniques achieve reasonably good performance in terms of AUC and ACC for both cartilage lesion and meniscus tear detection. This finding supports the hypothesis that integrating information from multiple imaging modalities can enhance the diagnostic accuracy of knee injury classification.

The results demonstrate that the linear fusion modality consistently yields the most robust classification performance. This outcome underscores the efficacy of feature combination through element-wise addition, as it successfully harnesses the synergistic information present in the distinct knee scanning modalities. By simply adding the feature vectors, the linear fusion method effectively combines the complementary information captured by MRI and X-ray, leading to improved discrimination between healthy and injured tissues.

Similarly, the concatenation fusion method exhibits meritorious performance across various scenarios, affirming the capability of feature concatenation to effectively capture discriminative information from diverse sources. By concatenating the feature vectors, this method creates a richer representation of the knee joint, allowing the model to learn more complex relationships between the different imaging modalities and the presence of injury.

While the Averaging fusion modality demonstrates competitive results, it generally trails behind the linear fusion modality. This finding implies that the practice of combining features through averaging may not optimally exploit the complementary characteristics inherent in the different modalities or views. Averaging can potentially dilute the discriminative power of individual features, leading to a less effective representation of the underlying pathology.

The MultiModal Attention fusion modality, while designed to selectively attend to relevant features from different modalities, did not outperform the linear fusion method in this study. This suggests that the simpler linear fusion approach is sufficient to capture the relevant information for accurate classification in this context. Further investigation is needed to explore the potential benefits of attention mechanisms in more complex scenarios or with different network architectures.

Overall, the results of this study highlight the effectiveness of separate-feature fusion techniques, particularly the linear fusion modality, for integrating information from multi-view 3D MRI and 2D X-ray data for the simultaneous classification of meniscus and cartilage injuries. The superior performance of the linear fusion method suggests that a simple yet effective combination of features can lead to robust and accurate classification results. Future research will focus on exploring more sophisticated fusion techniques and incorporating larger and more diverse datasets to further improve the performance and generalizability of the M3NET framework.

When combining data from multiple viewpoints and modalities, there is a possibility of capturing overlapping or redundant information. This redundancy can arise from the inherent correlations between different views of the same anatomical structure or from the complementary nature of different imaging modalities that may capture similar underlying features. For instance, both MRI and X-ray can reveal information about bone structure and alignment, potentially leading to redundant features when integrated into a single model.

5.5.2 Ablation study

In this section, we conduct an ablation study to assess the influence of the different modalities, views and task variations on the M3NET framework's performance for both classification tasks. It is worth noting that these results in the MRI view and modality ablations are the averages of ACC and AUC across meniscus tear classification and cartilage lesion classification using linear fusion.

First, we examine the impact of the MRI and X-rays modalities on the M3NET framework. As can be seen in Table 5.2, results indicate that compared to X-ray scans, multi-view MRI tends to give superior performance due to the large data used for training. The type of modality (MRI vs. X-ray) impacts significantly the accuracy and the AUC, with MRI outperforming X-ray.

Investigating the MRI views shows that the Sagittal view scans (using the SAG IW TSE volumes) reach an ACC of 80.15 and an AUC of 80.46, underlining its effectiveness. Using the coronal view scans only excels with an ACC of 83.68 and an AUC of 82.19. While the axial view scans closely follow with an ACC of 83.54 and an AUC of 82.71. When all MRI views are combined, the framework achieves its best performance, resulting in an ACC of 83.87 and an AUC of 84.15. Different MRI views (Sagittal, Coronal, Axial) exhibit varying performance, and combining them enhances overall results. Training the M3NET framework to classify both cartilage and meniscus using multi-views yields the highest performance, with an ACC of 87.6 and an AUC of 87.15.

Focusing solely on cartilage classification with multi-views results in an ACC of 85.15 and an AUC of 86.83. While concentrating on meniscus classification alone, using multi-views achieves an ACC of 86.34 and an AUC of 86.54.

If there is a need to simultaneously classify both cartilage and meniscus conditions, configuring the M3NET framework for dual classification offers the highest overall performance. However, if the application demands specialized classification, the framework can be fine-tuned for either cartilage or meniscus classification, with both configurations providing solid accuracy and AUC results. These findings highlight the adaptability of the M3NET framework for different medical imaging tasks.

Table 5.2: Ablative study experiments on M3NET components

Ablation	**Modality**	**View**	**A**CC	AUC
Modality Ablation	MRI	multi-views	82.15	81.53
	X-ray	–	78.63	79.29
	MRI & X-ray	–	**86.87**	**87.71**
MRI View Ablation	N/A	Sagittal	80.15	80.46
	N/A	Coronal	83.68	82.19
	N/A	Axial	83.54	82.71
	N/A	multiview	**83.87**	**84.15**
Task Ablation	Both	multiview	**87.60**	**87.15**
	Cartilage Only	multiview	85.15	86.83
	Meniscus Only	multiview	86.34	86.54

5.5.3 Limitations

While integrating multiview, multimodal imaging data offers significant potential for enhancing diagnostic accuracy and providing a more comprehensive understanding of knee OA and related injuries, it also presents certain challenges and limitations. One notable concern is the potential for feature redundancy and its impact on model performance.

In the context of this study, while the employed separate-feature fusion approach aims to leverage the unique information captured by each modality, the potential for feature redundancy remains a consideration. Future research will focus on incorporating feature selection or dimensionality reduction techniques to further optimize the model and mitigate the potential impact of redundant features. Moreover, exploring more sophisticated fusion techniques that can intelligently select and combine relevant features from different modalities could lead to further improvements in model performance and generalizability.

5.6 Conclusion

In this chapter, we introduced an innovative learning model, M3NET for knee injuries diagnostics, which integrates multi-modal, multi-view, and multi-label learning paradigms into a single three-dimensional network. Utilizing a large annotated dataset, our method capitalized on the strengths of 3D-ResNet models for MRI analysis and 2D-ResNet models for X-ray evaluation, extracting relevant features necessary for the assessment of meniscus and cartilage health.

M3NET leverages a distinct feature fusion strategy to effectively utilize 1519 MRI scans and 1519 X-ray images from multiple views, all carefully annotated for accurate training and validation. A custom training of YoloV8 was employed for ROI detection, significantly enhancing the efficiency and accuracy of the feature extraction process.

The results from our extensive experimentation demonstrate that M3NET is not only viable but also excels in extracting and utilizing diagnostic features from complex imaging data. Its capability to adapt and identify critical features from both MRI and X-ray modalities makes it a valuable tool for the advancement of knee injury diagnostics.

In the following chapter, we explain how we effectively trained 2D deep learning models on 3D volumetric MRI for the detection of knee OA using multi-view MRI scans.

CHAPTER 6

MVMI: Multi-View Multi-Instance Network for the Detection of Knee Osteoarthritis

Traditional methods for diagnosing OA from 3D MRI often struggle to recognize this complex disease effectively due to limitations in capturing the critical, subtle details in individual MRI volumes. Also, the complexity of training large models on volume-based data is expensive and complex. In this chapter, we present a new effective approach to training simple 2D CNN models on multi-view 3D MRI scans, making the training faster, and more efficient.

6.1 Introduction

MRI is a reliable tool for assessing knee OA, surpassing traditional X-ray methods that primarily capture bone abnormalities [38, 39]. Unlike X-rays, MRI provides a comprehensive view of the entire knee joint, including cartilage, bone, synovitis, ligaments, and soft tissues. This detailed imaging capability allows for the detection of OA signs that are not visible in X-rays [4], offering a deeper understanding of the joint's structural integrity and the nuanced pathological changes associated with the condition.

MRI not only fills this gap by providing detailed and high-resolution images but also offers the unique advantage of non-invasive, multi-planar imaging capabilities. This enables a more thorough assessment of the knee joint from multiple views, which is crucial for a comprehensive analysis of the disease's progression. Nevertheless, the vast amount of data produced by MRI scans requires sophisticated processing techniques to extract diagnostically relevant information effectively, highlighting the importance of advanced computational models in enhancing the diagnostic process.

Despite the advantages of current deep learning models in MRI analysis, their implementation in knee OA detection presents specific limitations:

- **Volume-Based input**: When entire MRI volumes are fed into 3D CNNs, the models often face significant computational demands [165]. This can lead to practical constraints on the resolution and size of the data that can be processed, potentially reducing the clinical utility of the models. Moreover, the aggregation of volumetric data may result in a loss of critical localized details. For example, small but clinically significant anomalies might be averaged out or obscured in the larger volume context, which can lead to missed diagnoses or inaccuracies [166].

- **Single-View Constraint:** The prevalent reliance on a single MRI view in most existing models limits their diagnostic scope. Important pathological indicators of OA, which may be more clearly visible in different planes, are often overlooked [142, 157]. This single-view approach undermines the capacity of the models to provide a comprehensive assessment of the knee joint, as different views can reveal different aspects of the joint pathology. As a result, the diagnostic accuracy is compromised, particularly in cases where OA affects areas of the knee that are not optimally visualized from the sagittal plane.

To address the challenges and limitations of existing approaches, we propose the Multiview Multi-instance Network (MVMI). The proposed MVMI network is inspired by the efficiency of both multiview learning and multi-instance learning, particularly their abilities to enhance the interpretation of complex structures like the knee joint and its compartments. By incorporating both coronal and sagittal views of the knee, MVMI captures a holistic representation of the knee. The coronal view, with its frontal viewpoint, is important in assessing the knee joint's alignment and evaluating the condition of the menisci, as well as the lateral and medial compartments. In contrast, the sagittal view provides a detailed side perspective, crucial for examining the integrity of the anterior and posterior cruciate ligaments and assessing cartilage thickness. By synthesizing these distinct but complementary views, MVMI ensures a more complete analysis, where pathologies that might be obscured in one view are highlighted in another, thus enhancing the accuracy and reliability of knee OA diagnosis.

In the MVMI architecture, Multi-Instance Learning (MIL) complements our multiview approach by enhancing the analysis of MRI scans slice by slice. Similar to how radiologists review MRI scans, MIL treats each slice as an individual instance within a broader collection or "bag". This method not only mimics the sequential examination process used in clinical practice but also aligns seamlessly with the multiview strategy by allowing each view's unique characteristics to be analyzed in detail. By enabling a focused assessment of each slice, MIL reduces computational demands and ensures that critical, view-specific details are not overlooked. This integration of MIL into the multiview framework enhances the overall diagnostic precision of the MVMI model, making it a robust tool for identifying knee OA with high accuracy.

The overall objective of this paper is to introduce a pioneering model, MVMI that integrates multiview learning with multi-instance learning to address the complexities of diagnosing knee OA via MRI. This approach not only enhances diagnostic accuracy by leveraging diverse anatomical perspectives and focusing on crucial localized features but also improves computational efficiency. Our contributions extend to demonstrating MVMI's superior performance through rigorous validation on the OAI dataset, highlighting its potential as a new standard in knee injuries diagnostics.

Despite the advancements and good performance of current deep learning methods in diagnosing knee OA using MRI, these approaches still face significant limitations. These challenges largely stem from the traditional reliance on whole MRI volumes as input for model training and the frequent use of a single MRI view, which can impede both the efficacy and efficiency of the diagnostic process. Recent studies achieved better performance in terms of knee OA detection from multi-view MRI on the OAI database, Berrimi et al. [157] introduced a pseudo-labeling framework for detecting knee OA using multi-view MRI scans, and aggregates multiple MRI views using a concatenation layer obtaining an accuracy of 93.21%.

Using entire MRI volumes as input for 3D CNN models presents substantial computational challenges [165]. Processing these large volumes demands significant memory and computational power, hindering deployment in clinical settings where rapid diagnoses are critical. Additionally, analyzing entire volumes risks diluting crucial localized features [166]. Subtle pathological indicators, like small lesions, can become obscured within the aggregated data, compromising detection and accurate staging.

MIL offers a compelling solution to these challenges. In MIL, a 3D MRI volume is treated as a "bag" composed of multiple smaller 3D "instances" (e.g., patches or sub-volumes). Instead of training on the entire volume, the model learns from these individual instances, significantly reducing computational burden while preserving localized information. By focusing on smaller, information-rich regions within the larger volume, MIL enables efficient and sensitive analysis of 3D medical images, paving the way for more practical and accurate diagnostic models.

Furthermore, Most existing models rely on a single MRI view, typically the sagittal plane. While this view can provide valuable information about certain aspects of the knee joint, it may not capture other critical features visible only in the coronal or axial planes. This limitation restricts the model's ability to perform a comprehensive assessment of the knee joint. Pathological features that are more pronounced or exclusively visible in alternative views can be completely missed, potentially leading to incomplete or inaccurate diagnoses. For instance, certain types of meniscus tears or cartilage degradation may only be evident in the coronal view and not in the sagittal plane.

Multiview learning frameworks are fundamentally motivated by the understanding that images of the same object, captured from multiple perspectives, can provide complementary information that is crucial for a more comprehensive analysis [157]. In the context of knee MRI,

employing a multiview approach is particularly beneficial because it enables the model to capture and integrate diverse anatomical features that may not be visible or are less discernible from a single view.

In knee MRI, the complexity of the joint's anatomy necessitates detailed imaging from multiple planes to accurately assess the health of all structures involved. For instance, sagittal view is optimal for evaluating the anterior and posterior aspects of the knee, including the cruciate ligaments and the cartilage covering the femoral and tibial surfaces. Certain types of meniscal tears, particularly those in the posterior horns, are more clearly visible in the sagittal view. Although, the coronal view provides critical insights into the medial and lateral compartments of the knee. It is especially valuable for assessing the spacing between the femur and tibia, detecting meniscus extrusion, and evaluating the overall alignment of the knee. Some types of cartilage wear and osteophyte formation are best observed from this perspective.

As previously mentioned in 5.3.1 multiview learning not only enhances the model's ability to perform a thorough and accurate analysis by leveraging the unique strengths of each view, but it also mimics the holistic approach that a radiologist takes in evaluating MRI scans to diagnose knee OA.

6.2 Multi-View Multi-Instance network (MVMI)

In this section, we present the overall MVMI architecture, which is composed of three key components: (i) Feature extraction and encoding in which we develop a new custom and efficient CNN model that we train from scratch on the dataset employing complex and robust modules for the extraction of relevant knee OA features from the MRI slices. (ii) Multi-Instance setting in which we explain how we adapt the CNN model with a two-dimensional setting for a 3D MRI input. Doing so, we leverage the power of MIL frameworks to enable the model treating each 2D slice as a distinct instance within a larger "bag" of slices, optimizing the analysis for detailed and localized feature recognition. (iii) Lastly, to enhance the model's diagnostic capability, the MVMI framework is customized to support dual branches, each processing a different anatomical view of the 3D MRI scans.

6.2.1 Res-Squeez-Net

: For feature extraction, we propose Res-Squeez-Net, which integrates sophisticated residual blocks [93] and Squeeze-and-Excitation (SE) blocks [167], to efficiently harness and decode complex patterns within MRI slices, facilitating robust extraction of critical features indicative of knee OA:

1. Residual learning: Residual blocks in Res-Squeez-Net incorporate shortcut connections that facilitate identity mapping, ensuring that the signal can bypass one or more layers directly.

This architecture supports effective gradient propagation, which is critical for maintaining the performance of the network as depth increases.

- Direct Shortcut Connection: When the input and output feature maps of the residual block have the same dimensions, the input x is directly added to the function $C(x, W_i)$:

$$x_{out} = C(x, \mathbb{W}_i) + x \tag{6.1}$$

  $C(x, \mathbb{W}_i)$ applies transformations to the input x using the specified weights $\mathbb{W}_i$. The result of this transformation is then added directly to the original input x through a shortcut connection.

- Adaptive Shortcut Connection: If a mismatch in dimensions exists, particularly in the number of channels, a 1×1 convolution W_s is applied to match the dimensions:

$$x_{out} = C(x, \mathbb{W}_i) + \mathbb{W}_s x \tag{6.2}$$

  The weights $\mathbb{W}_s$ operate primarily to align the channel dimensions before the addition, ensuring that the input x can seamlessly integrate with the block's output.

2. Squeeze-and-Excitation (SE): The SE block acts as a mechanism to adaptively adjust the importance (or weights) of each channel in a feature map by enabling it to perform dynamic channel-wise feature recalibration [167]. Henceforth, it allows the network to focus more on informative features and suppress less useful ones. This is important for our use case since the clinical features of knee OA in the MRI slices are located only in some parts of the scans.

   - The SE block begins with the $squeeze$ operation, which is performed using Global Average Pooling (GAP). In this step, each channel of the input feature map is compressed into a single scalar. This scalar represents the spatial average of the feature activations across the entire channel. For a given feature map F of dimensions HxW (height and width), the output, z_c of the GAP for each channel is calculated as:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i, j) \tag{6.3}$$

   - The $excitation$ operation applies a fully connected neural network to each channel descriptor FCN with channels 1x1xK. This network consists of two layers: a dimension-reduction layer with a ReLU activation function, followed by a dimension-expansion layer with a sigmoid activation function. The output of the excitation phase is a set of scalar coefficients for each channel, calculated as:

$$s_c = \sigma(\mathbb{W}_1 \cdot \mathrm{ReLU}(\mathbb{W}_2 \cdot z_c)) \tag{6.4}$$

σ is the sigmoid function, $\mathbb{W}_1$ and $\mathbb{W}_2$ are respectively the weights of the first and second layers of the fully connected network.

- The final step of the SE block involves applying the learned channel-wise scaling factors s_c back to the original feature map. This scaling adjusts the amplitude of the input feature activations, enhancing those that are more relevant and suppressing those that are less important.

$$\tilde{y} = s_c \odot x'' \tag{6.5}$$

where $\tilde{y}$ is the output of the SE block after scaling, and s_c represents the vector of scaling factors derived from the excitation operation. The $\odot$ is the element-wise multiplication, and x'' is the input to the SE block, typically the output from the convolutional layers within the residual block. The purpose of adding scaling is to recalibrate the feature channels according to their importance, which the model learns during training. By adjusting the activations in this manner, the network can focus more on useful features while diminishing less useful ones.

6.2.2 Multi-View, Multi-Instance Modules:

Two essential steps of MVMI are employed to ensure robustness in the model training, (i) the creation of bag instances (slices) from the multi-view MRI volumes, and (ii) the aggregation of instance-level predictions to summarize the features across all instances in a bag, where general trends across slices can indicate important diagnostic information, as illustrated in Fig. 6.1.

1. **Bag instance creation:** In the multi-instance module within MVMI, the concept of Multi-Instance Learning (MIL) is crucial for effectively handling the 3D MRI data, which consists of numerous 2D slices. Here, each 3D MRI volume is treated as a "bag," and the 2D slices that make up the volume are considered "instances" within that bag. This decomposition process is essential because it allows the model to examine each slice individually while still considering the entire volume's structural and contextual information.

   For both coronal and sagittal view scans, the MRI volume V is decomposed into its constituent slices:

$$V_{\mathrm{Cor}} = \{s_1^{\mathrm{cor}}, s_2^{\mathrm{cor}}, \ldots, s_N^{\mathrm{cor}}\} \tag{6.6}$$

$$V_{\text{Sag}} = \{s_1^{\text{sag}}, s_2^{\text{sag}}, \ldots, s_M^{\text{sag}}\} \tag{6.7}$$

where N and M represent the number of slices in the coronal and sagittal views, respectively. These slices differ in number due to the varying acquisition protocols used in COR MPR and SAG IW-TSE imaging.

The MIL framework treats each bag of slices from a single MRI volume as a unit that must be processed together. During model training, it is critical that all slices in a bag are passed through the network simultaneously in both forward and backward passes. This is achieved by setting the batch size to 1, ensuring that the spatial and contextual integrity of the slices within the volume is maintained. This approach allows the model to capture the intricate relationships between slices, which is particularly important for identifying patterns and features that are distributed across multiple instances within the same volume.

2. **Instance-features aggregation**: After processing the individual instances within a bag, the next step involves aggregating the features extracted from each slice to form a comprehensive representation of the entire MRI volume. In the MVMI model, this aggregation is performed using average pooling, which plays a critical role in the MIL process.

   The purpose of average pooling in this context is to synthesize the feature embeddings obtained from each slice into a single, cohesive feature vector for each MRI view (either sagittal or coronal). By calculating the mean of the instance embeddings, the model produces a feature vector that captures the dominant characteristics of the pathology across all slices. This method ensures that the final representation is robust to variations within individual slices, such as noise or irrelevant artifacts, which might otherwise distort the diagnostic conclusions.

   The aggregation process is mathematically expressed as follows:

$$P_{avg} = \frac{1}{K} \sum_{i=1}^{K} p_i \tag{6.8}$$

   where P_{avg} denotes the aggregated feature vector for a specific view, and p_i represents the feature embedding of the i-th slice within that view. K is the total number of slices processed for the view. This resulting vector encapsulates the essential diagnostic information derived from the entire set of slices, allowing the model to make informed predictions based on the aggregated features.

   The design of this MIL module within MVMI ensures that the model can effectively learn from the complex, multi-slice structure of MRI data. By aggregating instance-level features

into a bag-level representation, the model is better equipped to detect and interpret subtle patterns indicative of pathological conditions, ultimately leading to improved diagnostic accuracy.

3. **Multi-View aggregation:** This module extracts features from both the coronal and sagittal views using the proposed Res-squeeze-net model. Detailed explanation of multiview concepts is explained in Section 5.3.1.

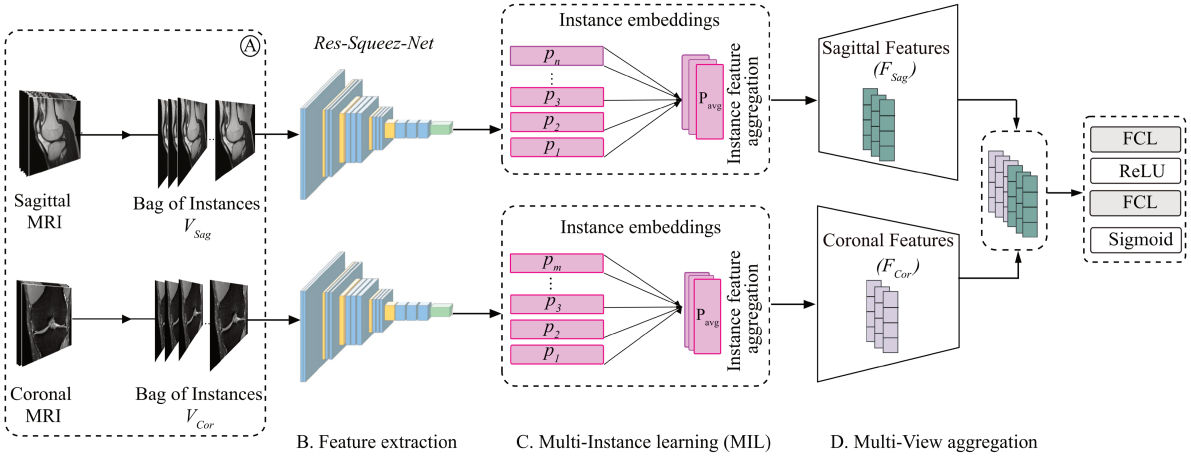


Figure 6.1: MVMI Architecture with Res-Squeez-Net as feature extraction module, MIL mechanism and Multi-view features aggregation.

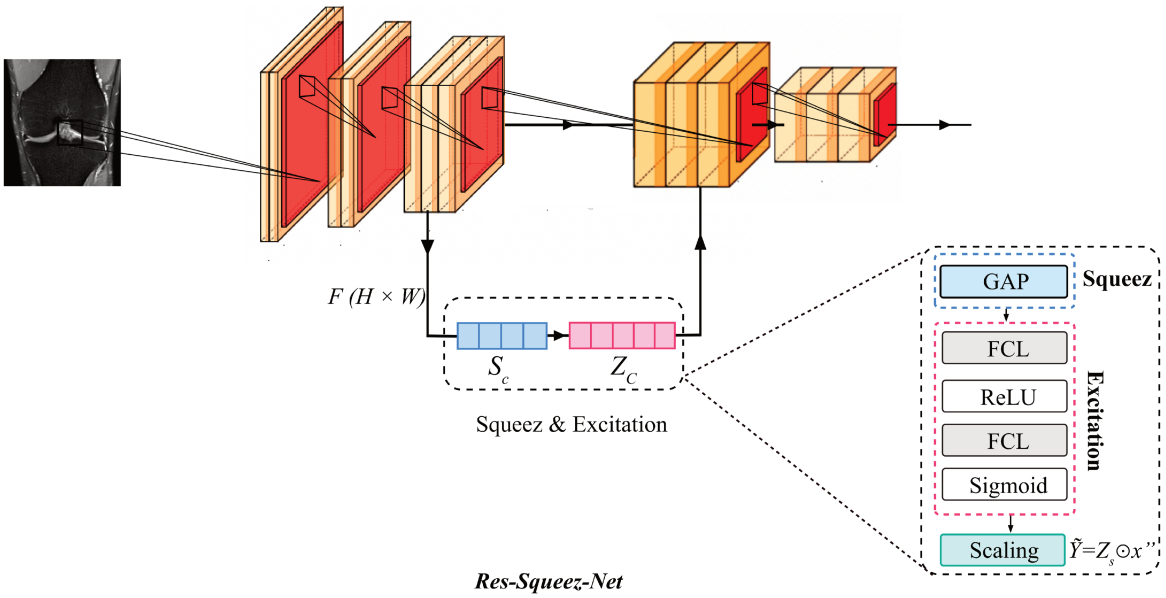


Figure 6.2: Proposed Res-Squeez-Net model. The figure delineates the Squeez and Excitation operations within convolutional layers.

In the MVMI model, the features are extracted from both the coronal and sagittal views

using the proposed Res-Squeeze-Net model and then fused to form a comprehensive feature set F that the model can use to determine the final class of the input scans. The feature fusion is achieved through a concatenation operation $\odot$ (see Fig. 6.1-D), which aligns and merges the feature vectors from each view into a single extended vector. The last is then used as input to the multi-layer-perceptron which is composed of two dense layers with ReLU activations.

Let F_{cor}, F_{sag} be the extracted features from the Res-Squeeze-Net model from the coronal and sagittal views, respectively. The concatenated feature vector F for multi-view features can be represented as:

$$F = [F_{cor} \odot F_{sag}] \tag{6.9}$$

where $\odot$ denotes the concatenation operation.

The MVMI network is trained with the objective of minimizing the Binary Cross-Entropy (BCE) loss function:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}\left[y^{(i)}\log(\hat{y}^{(i)}) + (1-y^{(i)})\log(1-\hat{y}^{(i)})\right] \tag{6.10}$$

where, y is the true label (0 or 1 for absence or presence of knee OA, respectively), and $\hat{y}$ is the predicted probability. This probability is derived by applying a sigmoid function to the neural network's output, which utilizes the concatenated features.

In the following section, we detail the experiments and ablation studies conducted to evaluate the efficiency of the MVMI network, utilizing the OAI database.

6.3 Experiments and results

6.3.1 Experimental Setup

In our experiments, we employed kernel regularization using the L^2 norm [168]. The training was conducted using the Adam optimizer for 250 epochs, starting with an initial learning rate of 3×10^{-4}. To enhance training efficiency and promote effective convergence, we implemented an *Exponential Decay* schedule for the learning rate. This approach allows for substantial initial learning rate values that facilitate rapid early progress, which then gradually decrease to enable finer adjustments as the training progresses.

The computational environment consisted of TensorFlow v2.12.0 running on an Nvidia TESLA A100 GPU, which is equipped with 40 GB of memory. This setup ensured robust handling of computational tasks. To ensure a rigorous evaluation of the model's performance, the dataset was split into 80% training and 20% validation.

6.3.2 Results

In this study, we evaluated the performance of our custom model, Res-Squeez-Net, for the detection of knee OA using the proposed MVMI network. The model was assessed in various configurations and compared with baseline models, namely Inception-V3 and DenseNet-121, using two different MRI scans: Coronal MPR (COR) and Sagittal IW-TSE (SAG).

Table 6.1: Results obtained using our proposed Res-Squeez-Net model with and without MV (Multi-View) and MIL (Multi-Instance) modules, using the COR MPR and SAG IW-TSE scans. The table also includes a comparison of our Res-Squeez-Net with other baseline models.

Model	MRI View		MIL		ACC (%)	AUC (%)
	COR	**SAG**	**instance-based**	**volume-based**		
Inceptionv-V3 [94]	✓	✗	✓	✗	83.62	92.14
	✓	✗	✗	✓	77.33	81.14
	✗	✓	✓	✗	66.10	81.80
	✗	✓	✗	✓	63.56	75.10
	✓	✓	✗	✓	79.44	82.82
	✓	✓	✓	✗	80.71	84.46
DenseNet-121 [96]	✓	✗	✓	✗	90.09	92.99
	✗	✓	✓	✗	78.53	88.61
	✗	✓	✗	✓	63.82	75.09
	✓	✗	✗	✓	72.30	73.43
	✓	✓	✓	✗	77.40	84.51
	✓	✓	✗	✓	79.97	83.05
ResSqueezNet	✓	✗	✓	✗	84.75	93.41
	✓	✗	✗	✓	79.10	87.23
	✗	✓	✓	✗	72.32	86.76
	✗	✓	✗	✓	74.01	83.25
	✓	✓	✗	✓	75.14	83.45
	✓	✓	✓	✗	**88.13**	**94.52**

As shown in Table 6.1, the various configurations of the Res-Squeez-Net, InceptionV3, and DenseNet-121 models were evaluated using COR and SAG MRI views, with different combinations of Multi-Instance Learning (MIL) and volume-based learning strategies.

The proposed Res-Squeez-Net model achieved its highest performance when both COR and SAG views were used with the instance-based MIL, resulting in an accuracy of 88.13% and an AUC of 94.52%. In single-view setups, the COR view with instance-based MIL reached an accuracy of 84.75% and an AUC of 93.41%. When using the SAG view alone with instance-based MIL, the performance dropped to 72.32% for accuracy and 86.76% for AUC. Volume-based

learning in dual views yielded lower results, with an accuracy of 75.14% and an AUC of 83.45% for both COR and SAG views combined.

InceptionV3, in contrast, demonstrated the best performance under a dual-view, instance-based MIL setup with an accuracy of 80.71% and an AUC of 84.46%. When this model employed a single COR view with instance-based MIL, it achieved an accuracy of 83.62% and an AUC of 92.14%. Performance declined when using the SAG view alone, recording an accuracy of 66.10% and an AUC of 81.80%. The volume-based learning configurations generally resulted in lower performance metrics across the board.

DenseNet-121 showed its best single-view performance with the COR view and instance-based MIL, where it achieved an accuracy of 90.09% and an AUC of 92.99%. This was the highest single-view performance across all models. However, in a dual-view, instance-based MIL configuration, accuracy was lower at 77.40% and AUC at 84.51%. The SAG view alone with instance-based MIL yielded an accuracy of 78.53% and an AUC of 88.61%.

Table 6.2: Comparative evaluation of the proposed MVMI network with several studies for knee OA detection using different MRI views

Model	View	AUC (%)
Guida et al. [76]	Sag	91.00
Chang et al. [146]	Sag	85.30
Pedoia et al. [75]	Sag	82.44
Alexopoulos et al. [74]	Sag	66.99
Berrimi et al. [157]	Multiview	93.20
MVMI (**ours**)	Multi-view	**94.52**

Table 6.2 presents the comparative evaluation of the proposed MVMI network alongside existing related works for the detection of knee OA using MRI scans. The MVMI model demonstrated superior performance, achieving the highest AUC of 94.52% among all compared models.

The comparison underscores the efficacy of the multiview approach over single-view methodologies. While single-view models, employing only sagittal views, show lower performance metrics (ranging from AUC of 66.99% to 91.00%), multiview models consistently deliver higher AUC. Notably, the model by Berrimi et al. [157], another multiview approach, also performs well, highlighting the advantages of leveraging multiple anatomical perspectives.

6.4 Discussion

The superior performance of the proposed MVMI network over other models can be attributed to its integration of multiview data and the effectiveness of its underlying CNN architecture, the Res-Squeeze-Net, which integrates robust components, namely, residual connections and SE modules. The multiview integration employed by the MVMI network harnesses the distinctive yet complementary perspectives provided by different anatomical planes of knee MRI scans.

This approach enables the model to capture a broader and more nuanced range of pathological features, which are often visible in one anatomical slice but not in others. In contrast, volume-based learning might dilute such critical localized information by averaging features across the entire volume.

Comparing the performance of Res-Squeez-Net with state-of-the-art models like , Inception-V3, and DenseNet-121 provided additional insights. Although DenseNet-121 performed exceptionally well in single-view, instance-based configurations, it did not exhibit the same level of performance in multiview settings as Res-Squeez-Net. This could indicate that while DenseNet-121 is highly capable of extracting effective features from a single slice, its adaptability to combine and synthesize information across multiple planes might be limited compared to Res-Squeez-Net. This observation aligns with the designed flexibility and integration capability embedded within the architecture of Res-Squeez-Net, which may be better suited for handling the complexities of multiview MRI data.

The differential performance across COR and SAG views is another aspect worth noting. The COR view generally resulted in higher accuracy and AUC than the SAG view, which could be due to the orientation and visibility of certain OA symptoms that are more pronounced or easier to detect in the coronal view.

Regarding the lower performance noted in SAG TSE acquisition (e.g., 72.32% accuracy and 86.76% AUC for Res-Squeez-Net with MIL), several factors may contribute to this outcome. The SAG TSE scan, consisting of 34 slices, provides a different anatomical perspective compared to the 50 slices of the COR MPR. The lower number of slices and potential differences in slice thickness and inter-slice spacing in the SAG TSE might lead to a less comprehensive representation of the knee joint. Moreover, the alignment and capture of knee structures in sagittal views can be more susceptible to patient movement and positioning errors, potentially complicating the learning process for the models.

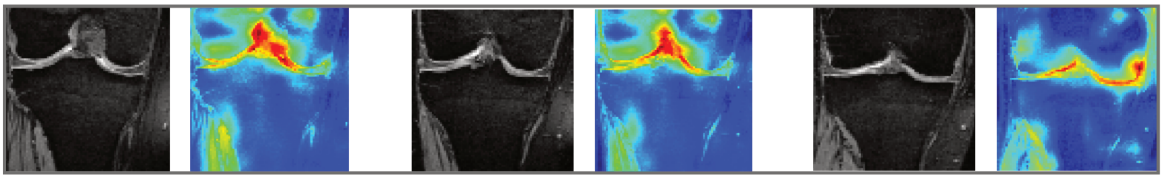
Moreover, the Res-Squeeze-Net architecture, a key component of the MVMI network, enhances this capability by efficiently extracting and encoding relevant features from each view. The network leverages residual connections to promote deeper learning without the risk of vanishing gradients, alongside Squeeze-and-Excitation blocks that dynamically recalibrate channel-wise feature responses based on their relevance. This enables the model to emphasize important features while suppressing less informative ones, optimizing the feature extraction process for each specific view.

Together with the MIL framework, which processes each slice as an independent instance within the broader context of its respective MRI volume, MVMI ensures that both local and global features are accurately captured and integrated. The model's ability to treat each slice independently but within the framework of its view-specific characteristics facilitates a more refined analysis, enhancing the detection capabilities for early and subtle signs of knee OA that

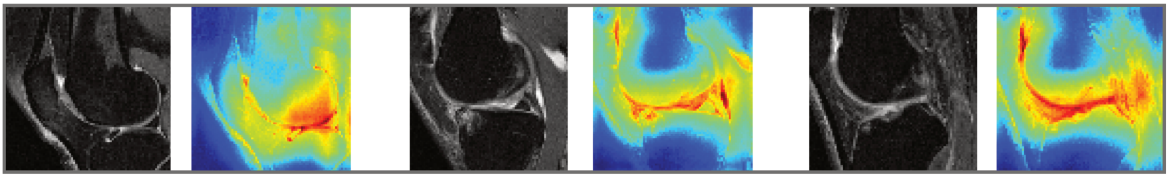
might be overlooked by volume-based models.

The obtained results affirm the significance of utilizing MVMI for knee OA diagnosis, especially in contexts requiring the synthesis of complex, multi-dimensional data to enhance classification accuracy.

Grad-cam To enhance the visualization of regions implicated in the detection of knee OA from MRI slices in both coronal and sagittal views, we employed Grad-CAM [148]. This method improves the medical decision making by effectively highlighting areas within the knee vulnerable to changes associated with OA.



Coronal MPR Slices with grad maps



Sagittal TSE Slices with grad maps

Figure 6.3: Obtained Grad-Cam maps using our proposed MVMI network for Coronal and Sagittal slices. The red regions reflect the area where the model is paying more attention to, and which has more relevant features.

In the sagittal MRI slices, the heatmaps illustrated in Fig. 6.3 emphasizes the posterior aspect of the patellar region and the surrounding soft tissues. This could indicate regions where there might be cartilage degeneration or inflammation, both of which are important in the diagnosis and assessment of knee OA. The heatmap shows increased intensity around the femoral condyles and tibial plateau, which are common sites for osteoarthritic changes, such as cartilage wear and osteophyte formation.

Additionally, in the coronal MPR slices, the heatmaps highlight the areas around the joint space, particularly around the menisci and the articular cartilage of the femur and tibia. These areas are critical in the diagnosis and monitoring of OA as they are prone to cartilage wear and degeneration.

The results from applying gradcam provide a clearer understanding of the model's focus areas, validating its ability to target relevant anatomical features in OA assessments.

6.5 Conclusion

In this chapter, we have introduced a Multiview Multi-instance Network (MVMI) for diagnosing knee OA. This approach utilizes multiview MRI data, combining features from both coronal and sagittal planes to thoroughly analyze the knee joint. The core feature extractor of the MVMI model is a custom Res-Squeeze-Net architecture, enhanced with residual and Squeeze-and-Excitation blocks for improved relevant feature extraction.

By employing Multi-Instance Learning (MIL), each MRI slice is treated as an individual instance, allowing the model to assess a range of pathological features with precision. The MVMI network was validated using the OAI database and achieved an AUC of 94.52%, demonstrating its capability. These results suggest that the MVMI model is a promising tool for the accurate diagnosis of knee OA, showing potential improvements over some current methods.

In the following chapter, we present a new 2D MRI dataset with its corresponding knee segmentation masks and our efforts towards building multi-task models to boost the classification accuracy of knee OA with the help of segmentation maps.

CHAPTER 7

A Multi-Task Approach for Automated Segmentation and Classification from MRI slices for Knee OA detection

Multi-task learning, which leverages shared information across related tasks, has demonstrated superior performance in various medical imaging applications. This chapter explores the potential of multi-task learning for enhancing knee OA classification accuracy using MRI. We first describe the creation of a large-scale dataset comprising 2D MRI scans from two extensive cohorts, meticulously annotated for both image segmentation and OA severity classification. Next, we present a novel multi-task learning framework designed to simultaneously perform segmentation of relevant knee structures and classify OA severity. This approach integrates segmentation as a critical component, enhancing the overall predictive performance by providing detailed anatomical insights.

7.1 Introduction

The accurate and timely diagnosis of knee OA is crucial for effective disease management and improving patient outcomes. MRI has emerged as a powerful tool for assessing knee OA due to its ability to visualize soft tissues and cartilage, providing valuable insights into the underlying pathology [9]. While most MRI-based knee OA evaluation studies have traditionally relied on 3D analysis methods, these approaches often face significant challenges, including the need for large amounts of labeled data and the computationally demanding nature of 3D image processing. Our previous work [157] addressed this challenge by developing a 3D analysis method for knee OA detection, achieving promising results with an accuracy of up to 93.20%. However, the limitations of 3D analysis, particularly in terms of computational burden and processing time, necessitate the exploration of more efficient and clinically practical alternatives.

This has led to a growing shift towards 2D analysis methods, which offer greater efficiency and potential for wider accessibility in clinical settings. However, relying solely on 2D analysis without incorporating segmentation models presents its own set of challenges. The absence of segmentation models could result in the loss of critical spatial information essential for accurate disease characterization and progression tracking. Segmentation models provide detailed anatomical maps that help in precisely isolating affected areas, enhancing both diagnostic accuracy and the efficacy of targeted treatments. Without these models, analyses might overlook subtle yet clinically significant anomalies, potentially leading to less effective patient management.

Therefore, this chapter proposes a novel multi-task learning approach that combines the efficiency of 2D analysis with the rich anatomical information provided by segmentation models. This approach leverages the inherent synergy between segmentation and classification tasks, allowing the model to learn a more comprehensive representation of the underlying OA pathology. By incorporating segmentation as an auxiliary task, we aim to improve the accuracy and robustness of OA classification while simultaneously providing valuable anatomical insights for clinical decision-making.

7.2 Multitask Learning for Knee Osteoarthritis Detection from MRI

Multitask learning (MTL) [169, 170] has emerged as a powerful paradigm in deep learning, particularly within the realm of medical image analysis. Its ability to learn shared representations across multiple related tasks simultaneously offers significant advantages over traditional single-task approaches, including improved performance, enhanced generalization, and increased data efficiency. While the pioneering application of MTL in medical imaging focused on Alzheimer's disease staging [171], a comprehensive review of the literature reveals a significant gap in its application to knee osteoarthritis (OA) detection from MRI data. This research gap underscores the need for further exploration of MTL's potential in improving the accuracy and efficiency of OA diagnosis and management.

In this context, we introduce a novel multitask learning framework designed to address this gap by accurately detecting knee OA from 2D MRI images. Our work aims to automate the diagnostic process, providing clinicians with a robust and reliable tool for identifying and assessing OA with greater accuracy and efficiency. Specifically, we propose an approach that integrates both classification and 2D segmentation tasks for OA prediction, a methodology not previously explored in the existing literature. This integrated approach leverages the inherent correlations between these tasks, allowing the model to learn a richer and more comprehensive representation of the underlying OA pathology.

The Synergy of Multitasking in Knee OA Analysis

The essence of multitask learning lies in its ability to exploit the synergistic relationships between related tasks. In the context of knee OA detection, we hypothesize that learning to segment the regions of cartilage degradation and bone spur formation (segmentation task) can provide valuable contextual information that enhances the model's ability to classify the presence or absence of OA (classification task). Conversely, the classification task can guide the segmentation process by providing a global understanding of the disease state. This mutual information exchange between tasks facilitates a more holistic understanding of the underlying pathology, leading to improved performance and generalization compared to training separate models for each task independently.

Mathematically, we define our tasks as follows:

- Classification Task (T1): Predicting the presence or absence of knee OA, denoted by $y_c \in \{0,1\}$, where $y_c = 1$ indicates the presence of OA and $y_c = 0$ indicates its absence.
- Segmentation Task (T2): Segmenting the regions of interest (ROIs) associated with OA, such as cartilage and bone spurs. This is represented as a pixel-wise classification problem, where each pixel p is assigned a label $y_s(p) \in \{0,1\}$, with $y_s(p) = 1$ indicating that pixel p belongs to an ROI.

Our multitask learning framework aims to learn a model parameterized by θ that minimizes the combined loss across both tasks:

$$\mathscr{L}(\theta) = \lambda_c \mathscr{L}_c(\theta;\mathscr{D}) + \lambda_s \mathscr{L}_s(\theta;\mathscr{D}) \tag{7.1}$$

where $\mathscr{L}_c(\theta;\mathscr{D})$ represents the classification loss function, typically binary cross-entropy, and $\mathscr{L}_s(\theta;\mathscr{D})$ represents the segmentation loss function, commonly Dice loss or cross-entropy. The parameters λ_c and λ_s are task-specific weighting parameters that control the relative contribution of each task to the overall loss. By jointly optimizing this combined loss function, the model learns to extract features that are not only discriminative for OA classification but also informative for segmenting the relevant anatomical structures.

Building upon this multitask learning paradigm, our work makes several key contributions to the field of automated knee OA diagnosis:

- Comprehensive Input Modality Analysis: We conduct a rigorous comparative analysis to investigate the influence of different input image modalities on knee OA detection performance. This analysis encompasses various MRI sequences, including T1-weighted, T2-weighted, and proton density-weighted images. By evaluating the performance of our MTL framework across these modalities, we gain valuable insights into the optimal choice of input data for maximizing diagnostic accuracy.

- Novel Multitask Learning Architecture: We develop a novel multitask learning framework specifically tailored for knee OA detection from 2D MRI images. This framework comprises a shared encoder network responsible for extracting common features from the input images, followed by task-specific decoder networks for classification and segmentation. The shared encoder leverages the synergistic relationship between the tasks, enabling the model to learn a more robust and generalizable representation of OA-related features.

- Optimized Hybrid Loss Function: We introduce a hybrid loss function, as defined in Equation 1, that combines a classification loss (e.g., binary cross-entropy) with a segmentation loss (e.g., Dice loss) to optimize the training process. This carefully crafted loss function balances the contributions of each task, ensuring that the model effectively learns both to classify the presence of OA and to accurately segment the affected regions within the knee joint.

This research aims to bridge the gap in the existing literature by exploring the potential of multitask learning for knee OA detection from 2D MRI images. The proposed framework offers a promising approach for improving diagnostic accuracy, efficiency, and clinical applicability, ultimately contributing to better patient care and management of knee OA.

7.3 Materials and Methods

7.3.1 Dataset

This chapter leverages the Osteoarthritis Initiative (OAI) database [oai], a publicly available, invaluable resource encompassing multimodal data from a cohort of 4,796 patients. A detailed discussion of the OAI database is provided in Section 2.4.1. For this study, we specifically focused on the 3D DESS MRI data, utilizing the accompanying femoral and tibial bone and cartilage segmentations meticulously generated by the Zuse Institute Berlin (ZIB) [131].

While the OAI provides volumetric 3D MRI data, this study focuses on a 2D analysis approach. This strategic decision was made to reduce computational complexity and facilitate the development of a more readily deployable model. Processing 3D volumes requires significant computational resources and can be time-consuming, limiting the practical applicability of such models in clinical settings. By focusing on 2D slices, we aim to develop a more efficient and accessible solution that can be readily integrated into existing clinical workflows.

To this end, we extracted 2D slices from the 3D DESS MRI volumes and meticulously aligned each slice with its corresponding segmentation mask provided by the ZIB dataset. This alignment ensured accurate correspondence between the image data and the anatomical labels. Following slice extraction and alignment, a crucial preprocessing step was implemented to enhance data quality and model performance. Empty slices, defined as those lacking discernible knee joint information and corresponding to entirely black segmentation masks, were identified and removed

from the dataset. These empty slices typically represent regions outside the anatomical area of interest and do not contribute meaningful information for OA classification.

This rigorous data acquisition and preprocessing pipeline yielded a final dataset comprising 105,919 MRI slices with their corresponding segmentation masks. This translates to a total of 211,838 knee images across 799 knees, providing a substantial and diverse dataset for model training and evaluation. Importantly, the dataset exhibits a balanced gender distribution, mitigating potential bias, and includes a near-equal proportion of normal cases (49.5%) and OA cases (50.5%). The OA cases were classified using the MOAKS grading system [4, 16], ensuring a robust and clinically relevant ground truth for model training and validation.

7.3.2 Model Architectures

In this study, two models, as described in Table 7.1, were employed. All models utilize a novel approach that integrates a reduced-depth VGG16-inspired encoder. This architectural choice optimizes computational efficiency without compromising the model's ability to extract crucial features from high-dimensional medical images. Unlike traditional deep learning models that often rely on deeper architectures, our approach utilizes only the first three convolutional blocks of the VGG16 architecture [95]. This strategy effectively captures essential image features while ensuring computational feasibility, making the model suitable for deployment in resource-constrained environments.

Table 7.1: Details of proposed models.

Model	Tasks	Total parameters	Memory sizes (MB)
A	classification	1,768,513	6.75
B	classification + segmentation	2,506,242	9.56

7.3.2.1 Model A: The Baseline Classification Model

Model A (Figure. 7.1) serves as the baseline method, utilizing the VGG16-inspired architecture as the backbone network for the classification task. The model starts with an input layer accepting MRI images of dimensions (384, 384, 3). The encoder begins with two convolutional layers, each containing 64 filters with a 3x3 kernel size, ReLU activation, and 'same' padding. Following these layers is a max-pooling operation with a 2x2 pool size, reducing the spatial dimensions from (384, 384) to (192, 192) while increasing the depth to 64. This hierarchical pattern continues through additional convolutional and pooling layers, progressively enhancing feature extraction for effective OA classification.

The extracted features are then flattened and passed through a series of fully connected layers, culminating in a final output layer with a sigmoid activation function. This output layer produces a probability score indicating the likelihood of OA presence in the input image. Model A

employs Binary Cross-Entropy (BCE) loss for the classification task, measuring the similarity between predicted probabilities and actual binary labels. This loss function is well-suited for binary classification problems and is commonly used in medical image analysis tasks.

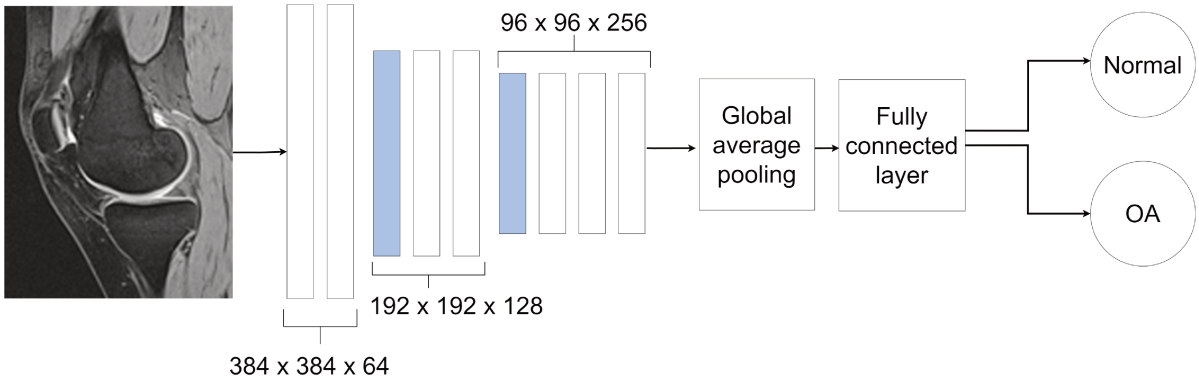


Figure 7.1: Architecture of the baseline method (Model A) that utilizes the VGG16 architecture as the backbone network for the classification task.

7.3.2.2 Model B: Integrating Segmentation for Enhanced Performance

Model B (Fig. 7.2) enhances the classification task by incorporating mask segmentation as an auxiliary task. The model architecture remains similar to Model A but includes additional layers dedicated to segmentation. After the last convolutional layer in the encoder, the model branches into a segmentation module that processes the feature maps to predict masks highlighting regions of interest.

Decoder Architecture and Upsampling The decoder part of the model takes the high-dimensional features from the encoder and upsamples them back to the original image size, using a series of decoder blocks. Each decoder block starts with an upsampling layer that doubles the spatial dimensions of the input feature maps. These upsampled features are then concatenated with the corresponding features from the encoder, creating a rich set of feature maps that combine high-resolution details from earlier layers with the semantic information from deeper layers. After concatenation, two convolutional layers with ReLU activation and 'same' padding are applied to refine the features. This process is repeated in multiple stages, gradually increasing the spatial resolution while reducing the depth, until the original input dimensions are restored.

Hybrid Loss Function for Multitask Optimization The model's loss function is a weighted combination of different loss components corresponding to its main and auxiliary tasks: classification and segmentation. The total loss function is represented by the equation:

$$\mathscr{L}_{\text{total}} = \lambda_1 \mathscr{L}_{\text{classification}} + \lambda_3 \mathscr{L}_{\text{segmentation}}$$

where λ_1 and λ_2 are the weights assigned to the classification loss ($\mathscr{L}_{\text{classification}}$) and segmentation loss ($\mathscr{L}_{\text{segmentation}}$), respectively. These weights allow for the balancing of different tasks during training to optimize the overall model performance.

The classification loss, $\mathscr{L}_{\text{classification}}$, also uses BCE loss, facilitating accurate binary classification of the data. The segmentation loss, $\mathscr{L}_{\text{segmentation}}$, utilizes Binary Cross-Entropy (BCE) loss, which measures the similarity between predicted probabilities and actual binary labels:

$$\mathscr{L}_{\text{BCE}}(y,p) = -[y\log(p) + (1-y)\log(1-p)]$$

This balanced approach, facilitated by the hybrid loss function, allows the model to optimize its performance across multiple tasks, enhancing its utility in medical imaging applications. By learning to segment the relevant anatomical structures, Model B gains a deeper understanding of the underlying OA pathology, leading to improved classification accuracy and a more comprehensive assessment of the disease.

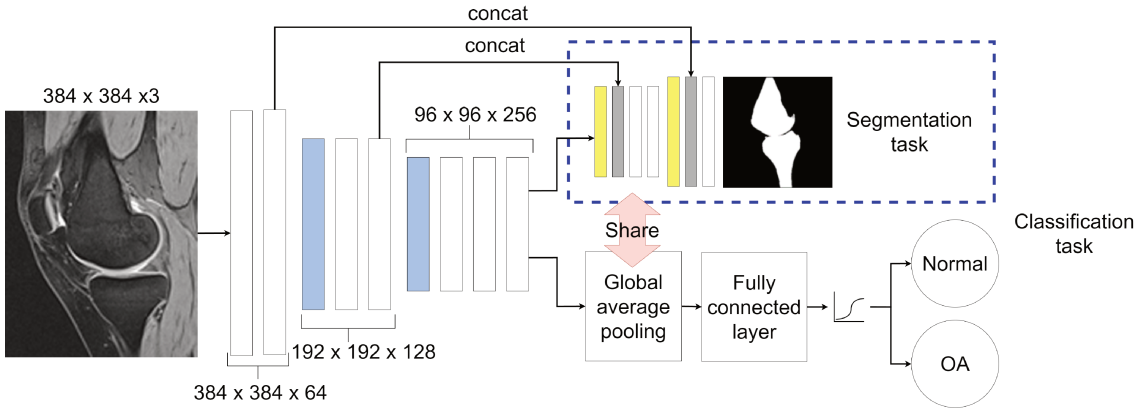


Figure 7.2: Architecture of the proposed method (Model B) that incorporates mask segmentation as an auxiliary task.

7.4 Experiments

In this study, two experiments were conducted. The study started with the selection of optimal input data, followed by the classification of OA.

7.4.1 Selection of Optimal Input Data

The first experiment was designed to identify the input image types that produced the best classification outcomes. This is crucial for ensuring that the subsequent OA classification task is based on the most informative and discriminative data representation. For this purpose, the single-task Model A was employed to evaluate different variations of input data, including raw MRI slices, binary mask images, and segmented MRI slices, as shown in Figure 7.3. By comparing

the classification performance across these different input types, we aimed to determine the most effective data format for subsequent analysis. This experiment allows us to assess the relative importance of raw image intensity values versus segmented anatomical information in the context of OA classification.

Figure 7.3: Illustration of input data types: (a) raw MRI slice, (b) binary mask image, and (c) segmented MRI slice (mask).

7.4.2 Classification of OA

The second experiment aimed to classify OA based on MOAKS labels, utilizing the optimal input data identified in the first experiment. This experiment involved comparing the classification performances of Model A and Model B. Both models were trained and tested using the selected input data, and their performance metrics were analyzed to determine which model provided superior accuracy and overall effectiveness in OA classification. This comparative analysis allows us to evaluate the benefits of incorporating segmentation as an auxiliary task in the context of multitask learning for knee OA detection. By comparing the performance of Model B (multitask) with Model A (single-task), we can assess the impact of the synergistic relationship between segmentation and classification on the overall diagnostic accuracy.

7.5 Results and Discussion

The results of both experiments highlight the prediction power of different MRI input images and the comparative effectiveness of single-task learning versus multitask learning in predicting knee OA. These findings contribute valuable insights into the development of accurate and efficient automated diagnostic tools for knee OA.

7.5.1 Selection of Optimal Input Data

The classification performance across three different types of input data was compared using Model A, as summarized in Table 7.2. The results demonstrate the model's sensitivity to the type of input data, with raw MRI slices providing the highest accuracy (90.84%) compared to binary mask images (76.60%) and segmented MRI slices (81.61

Table 7.2: Comparison of detection outcomes using single-task and multitask learning approaches.

Input data	Accuracy (%)	Loss
Raw MRI slices	90.84	0.2699
Binary mask images	76.60	0.6581
Segmented MRI slices	81.61	0.8181

Interpreting the Performance Differences The lower accuracy observed with binary mask images and segmented MRI slices can be attributed to the loss of potentially valuable information during the segmentation process. While segmentation helps highlight specific regions of interest, it also inherently discards certain details present in the raw MRI data. These discarded details, such as subtle variations in pixel intensity within the segmented regions, might contain valuable information for OA classification. Therefore, while segmentation can be beneficial in certain contexts, it is crucial to consider the potential trade-off between highlighting specific features and preserving the richness of the original data.

Implications for Model Development Based on these findings, we selected raw MRI slices as the optimal input data for the subsequent OA classification experiment. This choice ensures that the models are trained and evaluated on the most informative data representation, maximizing their potential for accurate and robust OA detection.

7.5.2 Classification of OA: A Comparative Analysis

This section delves into the classification performance of Osteoarthritis (OA) using both single-task learning (Model A) and multitask learning (Model B) approaches. The results are compared not only between these two models but also against findings reported in existing literature, providing a comprehensive evaluation of the proposed methodologies. Table 7.3 provides a concise summary of the key performance metrics.

Performance of Single-Task Learning (Model A) Model A, employing a single-task learning approach with 2D raw MRI slices as input, achieved a commendable accuracy of 90.84%. This result underscores the inherent capability of single-task learning strategies, even with simpler 2D image data, to effectively classify knee OA. The performance highlights the potential of

Table 7.3: Comparison of classification outcomes using single-task and multitask learning approaches.

Model	Model size (MB)	Input data	Approach	Accuracy (%)
3D-ResCNN-GAP [157]	-	3D MRI	Single-task	93.20
OA-MTL [172]	64.30	3D MRI	Multitask	82.50
A (our)	**6.75**	Raw MRI slices	Single-task	90.84
B (our)	9.56	Raw MRI slices	Multitask	**96.70**

leveraging 2D data for OA classification, particularly in scenarios where computational resources or 3D data acquisition might be limited. This finding aligns with the growing trend in medical image analysis to explore the efficacy of 2D approaches, especially when considering their reduced computational demands and potential for wider accessibility.

The Power of Multitask Learning (Model B) Model B, leveraging a multitask learning framework with the same 2D raw MRI slices as Model A, exhibited a significant improvement in accuracy, reaching 96.70%. Furthermore, it achieved a segmentation accuracy of 91.74%. This substantial leap in performance underscores the advantages of multitask learning in this context. By simultaneously learning the classification and segmentation tasks, the model effectively leverages shared representations and inter-task dependencies within the data. This leads to a more robust and generalized model capable of superior performance in both tasks. The ability to extract richer information from the data through multitask learning is particularly relevant in complex medical image analysis tasks like OA detection, where subtle features and contextual information play a crucial role in accurate diagnosis.

Benchmarking Against Existing Studies Comparing Model A and Model B with existing studies provides further insights into the efficacy of our proposed approaches. The single-task 3D-ResCNN-GAP model [157], utilizing 3D MRI data, achieved an accuracy of 93.20%. While Model A's performance (90.84%) with 2D data is slightly lower, it demonstrates competitive performance considering the reduced complexity associated with 2D data processing. Crucially, Model B's performance (96.70%) significantly surpasses that of 3D-ResCNN-GAP, indicating that multitask learning with 2D data can not only rival but even outperform single-task models trained on 3D MRI data. This suggests that the benefits derived from multitask learning, such as improved feature representation and generalization, outweigh the potential advantages of using 3D data in this particular application.

The OA-MTL model [172], another multitask learning approach employing 3D MRI data and a more complex architecture with five residual blocks, achieved an accuracy of 82.50%. This result, notably lower than both Model A and Model B, suggests that the effectiveness of multitask learning is not solely dependent on using 3D data or complex architectures. Factors such as input data quality, model design, and appropriate task selection play a crucial role in achieving

optimal performance. The superior performance of Model B, despite its smaller size compared to OA-MTL, further highlights the importance of model efficiency and design considerations in achieving high accuracy. This underscores the need for careful consideration of model complexity and the potential for overfitting when designing deep learning models for medical image analysis, particularly in the context of multitask learning.

The Advantages of Efficiency and Generalizability The findings presented in this section highlight the potential of leveraging multitask learning with 2D data for accurate and efficient OA classification. The success of Model B, coupled with its smaller model size, emphasizes the importance of model design and efficient utilization of 2D images in achieving high accuracy. This research suggests that a well-designed multitask learning model using 2D data can provide a practical and effective alternative to more complex 3D methods, paving the way for more accessible and efficient OA diagnosis. This contribution is particularly significant in the context of OA research, where accurate and timely diagnosis is crucial for effective disease management and improving patient outcomes. The enhanced generalizability of Model B, demonstrated by its superior performance compared to existing methods, suggests its potential for broader applicability across diverse datasets and clinical settings.

Future Directions and Considerations While the results presented in this chapter are promising, further research is warranted to explore the full potential of multitask learning for knee OA detection. Future work could investigate the impact of different loss function weightings, alternative segmentation techniques, and the integration of other relevant clinical data into the multitask framework. Additionally, exploring the applicability of this approach to other imaging modalities, such as X-ray, could further enhance its clinical utility and accessibility. By addressing these future research directions, we can continue to refine and improve the accuracy, efficiency, and clinical applicability of automated diagnostic tools for knee OA, ultimately contributing to better patient care and management of this prevalent and debilitating disease.

7.6 Conclusion

This chapter provides compelling evidence for the efficacy of multitask learning in the automated detection of knee OA from MRI slices. The developed multitask learning model, utilizing raw MRI slices as input, achieved a remarkable accuracy of 96.70%, significantly outperforming its single-task learning counterpart. This superior performance underscores the inherent advantages of multitask learning in leveraging shared representations and inter-task dependencies, leading to more robust and generalized models capable of capturing complex relationships within the data.

The findings of this study strongly suggest that multitask learning offers a more effective strategy for knee OA detection compared to traditional single-task approaches. The ability to

simultaneously learn both classification and segmentation tasks allows the model to extract richer information from the data, leading to improved accuracy and a more comprehensive understanding of the underlying pathology. This is particularly relevant in the context of OA diagnosis, where subtle changes in cartilage and bone structure can be challenging to detect using single-task approaches.

Furthermore, the high accuracy achieved with raw MRI slices highlights the potential of multitask learning to simplify the image pre-processing pipeline. By effectively learning from raw data, the model reduces the reliance on complex and often time-consuming pre-processing steps, potentially streamlining the clinical workflow and facilitating wider adoption of AI-based diagnostic tools. This is particularly crucial in resource-constrained settings where access to specialized image processing expertise may be limited.

CHAPTER 8

GENERAL CONCLUSIONS

This dissertation has explored the potential of deep learning for the automated detection of knee OA using MRI data, structured across five detailed studies, each addressing distinct challenges in knee OA detection. These studies focused on developing and evaluating deep learning frameworks based on multi-view, multi-modal, multi-task, multi-label, and multi-instance learning paradigms. The overarching goal was to enhance diagnostic accuracy and streamline the clinical workflow for knee OA detection.

The foundational chapters set the stage by discussing the motivation for applying deep learning techniques to knee OA assessment. We emphasized the limitations of conventional radiography, particularly its insufficiency in detecting knee OA. MRI was highlighted as a superior imaging modality due to its capability to visualize soft tissue structures such as cartilage, menisci, and ligaments, making it more effective for OA diagnosis. A comprehensive overview of the Osteoarthritis Initiative (OAI) dataset was provided, which served as the primary data source for this research. The specific MRI sequences used, including 3D-DESS, IW-TSE, COR MPR, and AXR MPR, were selected for their ability to offer complementary views of the knee joint, enabling a detailed assessment of OA-related structures. The MOAKS scoring system was also emphasized for its role in providing a consistent framework for grading OA severity.

Chapter 4 introduced a novel semi-supervised multi-view 3D CNN designed specifically for knee OA detection. By integrating multiple views from 3D MRI scans and employing pseudo-labeling to increase the training data, this approach achieved an AUC of 93.20%. This result surpasses current state-of-the-art methods relying solely on X-ray or single-view MRI data. The study demonstrated the potential of multi-view learning to improve diagnostic accuracy by capturing a more comprehensive representation of OA-related changes. However, the use of KL-0 grade knees as "healthy" controls was identified as a limitation, highlighting the challenges in

defining appropriate control groups in OA research.

In Chapter 5, we introduced M3NET, the first model to combine multi-modal (X-ray and MRI) and multi-view MRI data for the analysis of two different knee injuries within the same framework. M3NET combined custom 3D-ResNet models for MRI analysis with 2D-ResNet models for X-ray evaluation, achieving significant advancements over previous and single-task, single modality, and single-view approaches. This model not only improved diagnostic accuracy but also illustrated the feasibility of integrating different imaging modalities for a holistic assessment of knee injuries. The study emphasized the potential of multi-modal approaches, though it also recognized the need for more sophisticated fusion techniques to address potential feature redundancy.

Chapter 6 presented the Multiview Multi-instance Network (MVMI), which employed a custom Res-Squeeze-Net architecture combined with multi-instance learning for knee OA diagnosis. The model achieved an AUC of 94.52%, demonstrating the effectiveness of multi-instance learning in capturing subtle pathological features from individual MRI slices. By treating each slice as an individual instance, MVMI proved particularly effective in processing 3D MRI scans using pretrained 2D CNN models, outperforming traditional 3D models. This work represents a significant advancement in the application of multi-instance learning to medical image analysis.

In Chapter 7, we proposed a novel multitask learning framework for knee OA detection using two large datasets, each with 105.919 MR slices, marking the first time such an approach has been applied to knee OA analysis. The model was trained to perform both segmentation to extract relevant features from the input scans, and detect the appearance of knee OA, tasks simultaneously, achieving an accuracy of 96.70%. This 2D multitask model demonstrated significant improvements over single-task models by leveraging shared representations and capturing complex inter-task dependencies. The ability to train directly from raw MRI slices, without extensive pre-processing, highlighted the potential of this approach to simplify clinical workflows and facilitate the adoption of AI-based diagnostic tools in practice.

Limitations

Despite the promising results, some limitations of this thesis must be acknowledged. First, the reliance on the OAI dataset, while providing a well-characterized cohort, may introduce biases that limit the generalizability of the findings to other populations. The dataset primarily includes participants from specific demographic groups, and the models developed here may not perform as well across more diverse populations or in different clinical settings.

Second, the potential for feature redundancy in multi-modal and multiview learning models emerged as a potential challenge. While the integration of X-ray and multiview MRI data improved diagnostic performance, it also introduced the possibility of overlapping features that may not contribute additional value to the model's predictions. Addressing this redundancy is

crucial for optimizing model performance and interpretability in clinical settings.

Future Work

Future work will focus on addressing these limitations and further enhancing the applicability of deep learning models for knee OA detection. One key area of focus will be the validation of these models on external datasets that include more diverse demographics and varying disease characteristics. This effort will ensure the broader applicability and robustness of the models in different patient populations and clinical environments.

Additionally, research will explore advanced techniques for feature selection and dimensionality reduction to mitigate the challenges associated with feature redundancy. Refining the selection of meaningful features will improve both the interpretability and efficiency of these deep learning frameworks in clinical practice.

Future directions also include the exploration of emerging technologies such as multimodal vision-language models and advanced segmentation models like SAM and MED-SAM. These innovations hold the potential to further expand the capabilities of deep learning in medical image analysis, allowing for more precise and comprehensive assessments of knee OA and related conditions.

In conclusion, this dissertation has laid a strong foundation for the application of deep learning in knee OA detection using MRI data. The novel frameworks developed here not only advance the field of medical image analysis but also pave the way for future research aimed at improving the diagnosis, treatment, and management of knee OA. By continuing to harness the power of deep learning, future work can aspire to create more accurate, efficient, and personalized healthcare solutions for this prevalent and debilitating condition.

BIBLIOGRAPHY

[1] Frank W Roemer, C Kent Kwoh, Michael J Hannon, David J Hunter, Felix Eckstein, Tomoko Fujii, Robert M Boudreau, and Ali Guermazi.
What comes first? multitissue involvement leading to radiographic osteoarthritis: magnetic resonance imaging–based trajectory analysis over four years in the osteoarthritis initiative.
Arthritis & rheumatology, 67(8):2085–2096, 2015.

[2] Scott Wong, Lynne Steinbach, Jian Zhao, Christoph Stehling, C Benjamin Ma, and Thomas M Link.
Comparative study of imaging at 3.0 t versus 1.5 t of the knee.
Skeletal radiology, 38:761–769, 2009.

[3] John A Lynch, Frank W Roemer, Michael C Nevitt, David T Felson, Jingbo Niu, Charles B Eaton, and Ali Guermazi.
Comparison of bloks and worms scoring systems part i. cross sectional comparison of methods to assess cartilage morphology, meniscal damage and bone marrow lesions on knee mri: data from the osteoarthritis initiative.
Osteoarthritis and cartilage, 18(11):1393–1401, 2010.

[4] D.J. Hunter, A. Guermazi, G.H. Lo, A.J. Grainger, P.G. Conaghan, R.M. Boudreau, and F.W. Roemer.
Evolution of semi-quantitative whole joint assessment of knee oa: Moaks (mri osteoarthritis knee score).
Osteoarthritis and Cartilage, 19(8):990–1002, 2011.

[5] David G Disler.
Articular cartilage in the knee: current mr imaging techniques and applications in clinical practice and research. invited commentary.
Radiographics: a review publication of the Radiological Society of North America, Inc, 31(1):61–62, 2011.

[6] CK Kam, Daniel WY Chee, and Wilfred CG Peh.
Magnetic resonance imaging of cruciate ligament injuries of the knee.

Canadian Association of Radiologists Journal, 61(2):80–89, 2010.

[7] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I S'anchez.
A survey on deep learning in medical image analysis.
Medical image analysis, 42:60–88, 2017.

[8] Tribikram Dhar, Nilanjan Dey, Surekha Borra, and R Simon Sherratt.
Challenges of deep learning in medical image analysis—improving explainability and trust.
IEEE Transactions on Technology and Society, 4(1):68–75, 2023.

[9] Pauline Shan Qing Yeoh, Khin Wee Lai, Siew Li Goh, Khairunnisa Hasikin, Yan Chai Hum, Yee Kai Tee, and Samiappan Dhanalakshmi.
Emergence of deep learning in knee osteoarthritis diagnosis.
Computational intelligence and neuroscience, 2021(1):4931437, 2021.

[10] Michael T Hirschmann and Werner Müller.
Complex function of the knee joint: the current understanding of the knee.
Knee Surgery, Sports Traumatology, Arthroscopy, 23:2780–2788, 2015.

[11] Charlotte Beaudart, Nannan Li, Annelies Boonen, and Mickael Hiligsmann.
Burden of osteoarthritis in the netherlands: a scoping review.
Expert Review of Pharmacoeconomics & Outcomes Research, 23(10):1147–1167, 2023.

[12] Valia P Leifer, Jeffrey N Katz, and Elena Losina.
The burden of oa-health services and economics.
Osteoarthritis and cartilage, 30(1):10–16, 2022.

[13] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala.
Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach.
Sci. Rep., 8(1):1727, 2018.

[14] J. H. Kellgren and J. S. Lawrence.
Radiological assessment of osteo-arthrosis.
Ann. Rheum. Dis., 16(4):494–502, 1957.

[15] Alejandro Morales Martinez, Francesco Caliva, Io Flament, Felix Liu, Jinhee Lee, Peng Cao, Rutwik Shah, Sharmila Majumdar, and Valentina Pedoia.
Learning osteoarthritis imaging biomarkers from bone surface spherical encoding.
Magnetic resonance in medicine, 84(4):2190–2203, 2020.

[16] D. Hunter et al.
Definition of osteoarthritis on mri: results of a delphi exercise.
Osteoarthritis and Cartilage, 19:963–969, 2011.

[17] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei.
Imagenet: A large-scale hierarchical image database.
In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[18] William Micheo, Belmarie Rodríguez-Santiago, Fernando Sepulveda-Irizarry, and Brenda Castillo.
Knee injuries.
Essential Sports Medicine: A Clinical Guide for Students and Residents, pages 315–340, 2021.

[19] Jeffrey B Driban, Charles B Eaton, Grace H Lo, Robert J Ward, Bing Lu, and Timothy E McAlindon.
Association of knee injuries with accelerated knee osteoarthritis progression: data from the osteoarthritis initiative.
Arthritis care & research, 66(11):1673–1679, 2014.

[20] Armaghan Mahmoudian, L Stefan Lohmander, Ali Mobasheri, Martin Englund, and Frank P Luyten.
Early-stage symptomatic osteoarthritis of the knee—time for action.
Nature Reviews Rheumatology, 17(10):621–632, 2021.

[21] Marta Favero, Roberta Ramonda, Mary B Goldring, Steven R Goldring, and Leonardo Punzi.
Early knee osteoarthritis.
RMD open, 1(Suppl 1):e000062, 2015.

[22] Ali Hasanpour-Dehkordi, Fariba Kabiri, and Fatemeh Dris.
Comparing the effects of massage therapy and aromatherapy on knee pain, morning stiffness, daily life function, and quality of life in patients with knee osteoarthritis.
Complementary Medicine Research, 28(4):292–299, 2021.

[23] Patrick Omoumi, Charbel Mourad, Jean-Baptiste Ledoux, and Tom Hilbert.
Morphological assessment of cartilage and osteoarthritis in clinical practice and research: intermediate-weighted fat-suppressed sequences and beyond.
Skeletal Radiology, 52(11):2185–2198, 2023.

[24] Arthur A. De Smet.

How i diagnose meniscal tears on knee mri.
American Journal of Roentgenology, 199(3):481–499, 2012.
PMID: 22915388.

[25] M.D. Shari M. Ling and M.D. Joan M. Bathon.
Osteoarthritis: Pathophysiology.
`https://www.hopkinsarthritis.org/arthritis-info/osteoarthritis/oa-pathophysiology/`.
Accessed: 2024-07-31.

[26] AGCM Chen, C Gupte, K Akhtar, P Smith, and J Cobb.
The global economic cost of osteoarthritis: how the uk compares.
Arthritis, 2012(1):698709, 2012.

[27] Margaret Lethbridge-Çejku, Charles G Helmick, and Jennifer R Popovic.
Hospitalizations for arthritis and other rheumatic conditions: data from the 1997 national hospital discharge survey.
Medical care, 41(12):1367–1373, 2003.

[28] I Rosemont.
United states bone and joint decade. the burden of musculoskeletal diseases in the united states.
Book United States Bone and Joint Decade. The Burden of Musculoskeletal Diseases in the United States (Editor ed.ˆ eds.). City: American Academy of Orthopaedic Surgeons, 2008.

[29] Joseph A Buckwalter, Charles Saltzman, and Thomas Brown.
The impact of osteoarthritis: implications for research.
Clinical Orthopaedics and Related Research (1976-2007), 427:S6–S15, 2004.

[30] Harry Kotlarz, Candace L Gunnarsson, Hai Fang, and John A Rizzo.
Osteoarthritis and absenteeism costs: evidence from us national survey data.
Journal of occupational and environmental medicine, 52(3):263–268, 2010.

[31] Sanjay Gupta, GA Hawker, A Laporte, R Croxford, and Peter C Coyte.
The economic burden of disabling hip and knee osteoarthritis (oa) from the perspective of individuals living with this condition.
Rheumatology, 44(12):1531–1537, 2005.

[32] Claude Le Pen, Camille Reygrobellet, and Isabelle Gérentes.
Financial cost of osteoarthritis in france: the “coart” france study.
Joint Bone Spine, 72(6):567–570, 2005.

[33] Philippe Bertin, François Rannou, Laurent Grange, Jean-Noel Dachicourt, Pierre Bruel, Corinne Emery, Nathalie Grandfils, and Charles Taieb.
Annual cost of patients with osteoarthritis of the hip and knee in france.
Journal of Musculoskeletal Pain, 22, 07 2014.

[34] Xia Wang, Win Min Oo, and James M Linklater.
What is the role of imaging in the clinical diagnosis of osteoarthritis and disease management?
Rheumatology, 57(suppl_4):iv51–iv60, 2018.

[35] Anuradha Vashishtha and Anuja kumar Acharya.
An overview of medical imaging techniques for knee osteoarthritis disease.
Biomedical and Pharmacology Journal, 14(2):903–919, 2021.

[36] Ali Guermazi, Frank W Roemer, Deborah Burstein, and Daichi Hayashi.
Why radiography should no longer be considered a surrogate outcome measure for longitudinal assessment of cartilage in knee osteoarthritis.
Arthritis research & therapy, 13:1–11, 2011.

[37] Brooke V Wessman, Andrew K Moriarity, Vanda Ametlli, and David J Kastan.
Reducing barriers to timely mr imaging scheduling.
Radiographics, 34(7):2064–2070, 2014.

[38] M.A. Omer, S.S. Malik, A.A. Malik, M.N. Anjum, Az Riaz, and R. Ali.
Diagnostic accuracy of ultrasound in detecting meniscal tears taking magnetic resonance imaging as gold standard.
Biological and Clinical Sciences Research Journal, 2020(1), 2020.

[39] C.E. Quatman, C.M. Hettrich, L.C. Schmitt, and K.P. Spindler.
The clinical utility and diagnostic performance of magnetic resonance imaging for identification of early and advanced knee osteoarthritis: a systematic review.
The American journal of sports medicine, 39(7):1557–1568, 2011.

[40] Huiyu Zhao, Hongqiu Li, Shuo Liang, Xinyue Wang, and Feng Yang.
T2 mapping for knee cartilage degeneration in young patients with mild symptoms.
BMC Medical Imaging, 22(1):72, 2022.

[41] Richard Kijowski.
3d mri of articular cartilage.
In *Seminars in Musculoskeletal Radiology*, volume 25, pages 397–408. Thieme Medical Publishers, Inc., 2021.

[42] BAM Snoeker, M Ishijima, J Kumm, F Zhang, AT Turkiewicz, and M Englund.
Are structural abnormalities on knee mri associated with osteophyte development? data from the osteoarthritis initiative.
Osteoarthritis and Cartilage, 29(12):1701–1708, 2021.

[43] Pieter Van Dyck, Filip M Vanhoenacker, Valérie Lambrecht, Kristien Wouters, Jan L Gielen, Lieven Dossche, and Paul M Parizel.
Prospective comparison of 1.5 and 3.0-t mri for evaluating the knee menisci and acl.
JBJS, 95(10):916–924, 2013.

[44] Richard Kijowski, Donna G Blankenbaker, Kirkland W Davis, Kazuhiko Shinki, Lee D Kaplan, and Arthur A De Smet.
Comparison of 1.5-and 3.0-t mr imaging for evaluating the articular cartilage of the knee joint.
Radiology, 250(3):839–848, 2009.

[45] Michael T Modic, Will Pflanze, David HI Feiglin, and George Belhobek.
Magnetic resonance imaging of musculoskeletal infections.
Radiologic Clinics of North America, 24(2):247–258, 1986.

[46] C. G. Peterfy, E. Schneider, and M. Nevitt.
The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee.
Osteoarthritis and Cartilage, 16(12):1433–1441, 2008.

[47] N. Bien et al.
Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of mrnet.
PLoS Medicine, 15(11):e1002699, 2018.

[48] Ivan Štajduhar, Mihaela Mamula, Damir Miletić, and Goezde Uenal.
Semi-automated detection of anterior cruciate ligament injury from mri.
Computer methods and programs in biomedicine, 140:151–164, 2017.

[49] Neil A Segal, Michael C Nevitt, K Douglas Gross, Jean Hietpas, Natalie A Glass, Cora E Lewis, and James C Torner.
The multicenter osteoarthritis study (most): opportunities for rehabilitation research.
PM & R: the journal of injury, function, and rehabilitation, 5(8), 2013.

[50] Ari V"a"ar"al"a, Victor Casula, Arttu Peuna, Egor Panfilov, Ali Mobasheri, Marianne Haapea, Eveliina Lammentausta, and Miika T Nieminen.
Predicting osteoarthritis onset and progression with 3d texture analysis of cartilage mri dess: 6-year data from osteoarthritis initiative.

Journal of Orthopaedic Research®, 40(11):2597–2608, 2022.

[51] Dean Inglis, Andy Kin On Wong, Felix Eckstein, Jonathan D Adachi, and Karen A Beattie.
Multiplanar reconstruction recovers morphological cartilage assessment reproducibility from maloriented coronal mri scans.
Magnetic Resonance in Medicine, 65(3):790–795, 2011.

[52] Andrea Wenger, Wolfgang Wirth, Martin Hudelmaier, Iris Noebauer-Huhmann, Siegfried Trattnig, Katja Bloecker, Richard B Frobell, C Kent Kwoh, Felix Eckstein, and Martin Englund.
Meniscus body position, size, and shape in persons with and persons without radiographic knee osteoarthritis: quantitative analyses of knee magnetic resonance images from the osteoarthritis initiative.
Arthritis & Rheumatism, 65(7):1804–1811, 2013.

[53] Andreas Hecker, Rainer J Egli, Emanuel F Liechti, Christiane S Leibold, and Frank M Klenke.
Multiplanar reformation improves identification of the anterolateral ligament with mri of the knee.
Scientific reports, 11(1):13216, 2021.

[54] Tsuyoshi Ohishi, Masaaki Takahashi, Masashi Abe, Takuya Tsuchikawa, Masashi Mori, and Akira Nagano.
The use of axial reconstructed images from three-dimensional mri datasets for morphological diagnosis of meniscal tears of the knee.
Archives of Orthopaedic and Trauma Surgery, 125:622–627, 2005.

[55] S Guilbert, V Chassaing, C Radier, C Hulet, F R'emy, J Chouteau, F Chotel, P Boisrenoult, A Sebilo, P Ferrua, et al.
Axial mri index of patellar engagement: a new method to assess patellar instability.
Orthopaedics & Traumatology: Surgery & Research, 99(8):S399–S405, 2013.

[56] Frank W Roemer, David J Hunter, Michel D Crema, C Kent Kwoh, Elena Ochoa-Albiztegui, and Ali Guermazi.
An illustrative overview of semi-quantitative mri scoring of knee osteoarthritis: lessons learned from longitudinal observational studies.
Osteoarthritis and cartilage, 24(2):274–289, 2016.

[57] CG Peterfy, A Guermazi, S Zaim, PFJ Tirman, Y Miaux, D White, M Kothari, Y Lu, K Fye, S Zhao, et al.
Whole-organ magnetic resonance imaging score (worms) of the knee in osteoarthritis.
Osteoarthritis and cartilage, 12(3):177–190, 2004.

[58] David J Hunter, Grace H Lo, D Gale, Andrew J Grainger, Ali Guermazi, and Philip G Conaghan.
The reliability of a new scoring system for knee osteoarthritis mri and the validity of bone marrow lesion assessment: Bloks (boston–leeds osteoarthritis knee score).
Annals of the rheumatic diseases, 67(2):206–211, 2008.

[59] Michel D Crema, Frank W Roemer, David T Felson, Martin Englund, Ke Wang, Mohamed Jarraya, Michael C Nevitt, Monica D Marra, James C Torner, Cora E Lewis, et al.
Factors associated with meniscal extrusion in knees with or at risk for osteoarthritis: the multicenter osteoarthritis study.
Radiology, 264(2):494–503, 2012.

[60] Matteo Dunnhofer, Niki Martinel, and Christian Micheloni.
Deep convolutional feature details for better knee disorder diagnoses in magnetic resonance images.
Computerized Medical Imaging and Graphics, 102:102142, 2022.

[61] Dinggang Shen, Guorong Wu, and Heung-Il Suk.
Deep learning in medical image analysis.
Annual review of biomedical engineering, 19(1):221–248, 2017.

[62] Andre Esteva, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean, and Richard Socher.
Deep learning-enabled medical computer vision.
NPJ digital medicine, 4(1):5, 2021.

[63] Fouzia Altaf, Syed MS Islam, Naveed Akhtar, and Naeem Khalid Janjua.
Going deep in medical image analysis: concepts, methods, challenges, and future directions.
IEEE Access, 7:99540–99572, 2019.

[64] F'elix Renard, Soulaimane Guedria, Noel De Palma, and Nicolas Vuillerme.
Variability and reproducibility in deep learning for medical image segmentation.
Scientific Reports, 10(1):13724, 2020.

[65] Junghwan Lee, Cong Liu, Junyoung Kim, Zhehuan Chen, Yingcheng Sun, James R Rogers, Wendy K Chung, and Chunhua Weng.
Deep learning for rare disease: A scoping review.
Journal of Biomedical Informatics, 135:104227, 2022.

[66] Gabriel Chartrand, Phillip M Cheng, Eugene Vorontsov, Michal Drozdzal, Simon Turcotte, Christopher J Pal, Samuel Kadoury, and An Tang.
Deep learning: a primer for radiologists.
Radiographics, 37(7):2113–2131, 2017.

[67] Johann Li, Guangming Zhu, Cong Hua, Mingtao Feng, BasheerBennamoun, Ping Li, Xiaoyuan Lu, Juan Song, Peiyi Shen, Xu Xu, Lin Mei, Liang Zhang, Syed Afaq Ali Shah, and Mohammed Bennamoun.
A systematic collection of medical image datasets for deep learning, 2021.

[68] A Omar Adil Deheyab, Mohammed Hasan Alwan, Islam khalid Abdul Rezzaqe, Omar Abdulkareem Mahmood, Yousif I Hammadi, Ali Noori Kareem, and Maha Ibrahim.
An overview of challenges in medical image processing.
In *Proceedings of the 6th International Conference on Future Networks & Distributed Systems*, pages 511–516, 2022.

[69] Tianming Song, Xiaoyang Yu, Shuang Yu, Zhe Ren, and Yawei Qu.
Feature extraction processing method of medical image fusion based on neural network algorithm.
Complexity, 2021(1):7523513, 2021.

[70] Marcin Kociołek, Michał Strzelecki, and Rafał Obuchowicz.
Does image normalization and intensity resolution impact texture classification?
Computerized Medical Imaging and Graphics, 81:101716, 2020.

[71] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton.
Deep learning.
nature, 521(7553):436–444, 2015.

[72] Yann LeCun, L'eon Bottou, Yoshua Bengio, and Patrick Haffner.
Gradient-based learning applied to document recognition.
Proceedings of the IEEE, 86(11):2278–2324, 1998.

[73] Ian Goodfellow, Yoshua Bengio, and Aaron Courville.
Deep learning.
MIT press, 2016.

[74] A. Alexopoulos, J. Hirvasniemi, S. Klein, C Donkervoort, E.H.G. Oei, and N. Tümer.
Early detection of knee osteoarthritis using deep learning on knee mri.
Osteoarthritis Imaging, 3:100112, 2023.
17th International Workshop on Osteoarthritis Imaging, Lausanne, Switzerland, June 28-30, 2023.

[75] V. Pedoia, J. Lee, B. Norman, T. M. Link, and S. Majumdar.
Diagnosing osteoarthritis from t2 maps using deep learning: an analysis of the entire osteoarthritis initiative baseline cohort.
Osteoarthritis Cartilage, 27(7):1002–1010, 2019.

[76] C. Guida, M. Zhang, and J. Shan.
Knee osteoarthritis classification using 3d cnn and mri.
Appl. Sci. (Basel), 11(11):5196, 2021.

[77] Zhe Wang, Aladine Chetouani, and Rachid Jennane.
Key-exchange convolutional auto-encoder for data augmentation in early knee osteoarthritis classification, 2023.

[78] Guoyong Jiang, Jianguang Ding, and Chenglu Ge.
Deep learning-based ct imaging to evaluate the therapeutic effects of acupuncture and moxibustion therapy on knee osteoarthritis.
Computational and mathematical methods in medicine, 2022(1):1135196, 2022.

[79] Marc Moreno L'opez, Joshua M Frederick, and Jonathan Ventura.
Evaluation of mri denoising methods using unsupervised learning.
Frontiers in Artificial Intelligence, 4:642731, 2021.

[80] Zhe Guo, Xiang Li, Heng Huang, Ning Guo, and Quanzheng Li.
Deep learning-based image segmentation on multimodal medical imaging.
IEEE Transactions on Radiation and Plasma Medical Sciences, 3(2):162–169, 2019.

[81] J Van der Veen, S Willems, S Deschuymer, D Robben, W Crijns, F Maes, and S Nuyts.
Benefits of deep learning for delineation of organs at risk in head and neck cancer.
Radiotherapy and Oncology, 138:68–74, 2019.

[82] Jon N Marsh, Matthew K Matlock, Satoru Kudose, Ta-Chiang Liu, Thaddeus S Stappenbeck, Joseph P Gaut, and S Joshua Swamidass.
Deep learning global glomerulosclerosis in transplant kidney frozen sections.
IEEE transactions on medical imaging, 37(12):2718–2728, 2018.

[83] Chayakrit Krittanawong, Kipp W Johnson, Robert S Rosenson, Zhen Wang, Mehmet Aydar, Usman Baber, James K Min, WH Wilson Tang, Jonathan L Halperin, and Sanjiv M Narayan.
Deep learning for cardiovascular medicine: a practical primer.
European heart journal, 40(25):2058–2073, 2019.

[84] A. Tack, A. Shestakov, D. Lüdke, and S. Zachow.
A multi-task deep learning method for detection of meniscal tears in mri data from the osteoarthritis initiative database.
Frontiers in Bioengineering and Biotechnology, 9:747217, 2021.

[85] Andreas Kofler, Christian Wald, Christoph Kolbitsch, CV Tycowicz, and Felix Ambellan.

Joint reconstruction and segmentation in undersampled 3d knee mri combining shape knowledge and deep learning.
Physics in Medicine & Biology, 69(9):095022, 2024.

[86] Qizheng Wang, Meiyi Yao, Xinhang Song, Yandong Liu, Xiaoying Xing, Yongye Chen, Fangbo Zhao, Ke Liu, Xiaoguang Cheng, Shuqiang Jiang, et al.
Automated segmentation and classification of knee synovitis based on mri using deep learning.
Academic Radiology, 31(4):1518–1527, 2024.

[87] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby.
An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

[88] Ilya O Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, et al.
Mlp-mixer: An all-mlp architecture for vision.
Advances in neural information processing systems, 34:24261–24272, 2021.

[89] Satya P Singh, Lipo Wang, Sukrit Gupta, Haveesh Goli, Parasuraman Padmanabhan, and Bal'azs Guly'as.
3d deep learning on medical images: a review.
Sensors, 20(18):5097, 2020.

[90] Mohammad Amin Morid, Alireza Borjali, and Guilherme Del Fiol.
A scoping review of transfer learning research on medical image analysis using imagenet.
Computers in biology and medicine, 128:104115, 2021.

[91] Lucas Beyer, Olivier J. Hénaff, Alexander Kolesnikov, Xiaohua Zhai, and Aäron van den Oord.
Are we done with imagenet?, 2020.

[92] Bhakti Baheti, Sarthak Pati, Bjoern Menze, and Spyridon Bakas.
Leveraging 2d deep learning imagenet-trained models for native 3d medical image analysis.
In *International MICCAI Brainlesion Workshop*, pages 68–79. Springer, 2022.

[93] K. He, X. Zhang, S. Ren, and J. Sun.
Deep residual learning for image recognition.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[94] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna.
Rethinking the inception architecture for computer vision.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.

[95] K. Simonyan and A. Zisserman.
Very deep convolutional networks for large-scale image recognition.
arXiv preprint arXiv:1409.1556, 2014.

[96] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger.
Densely connected convolutional networks.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[97] Mingxing Tan and Quoc Le.
Efficientnet: Rethinking model scaling for convolutional neural networks.
In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[98] Jonathan K Riek, A Murat Tekalp, Warren E Smith, and Edmund Kwok.
Out-of-plane motion compensation in multislice spin-echo mri.
IEEE transactions on medical imaging, 14(3):464–470, 1995.

[99] Rizwan Qureshi, MOHAMMED GAMAL RAGAB, SAID JADID ABDULKADER, ALAWI ALQUSHAIB, EBRAHIM HAMID SUMIEA, Hitham Alhussian, et al.
A comprehensive systematic review of yolo for medical object detection (2018 to 2023).
Authorea Preprints, 2023.

[100] N. K. Namiri et al.
Deep learning for large scale mri-based morphological phenotyping of osteoarthritis.
Sci. Rep., 11(1):10915, 2021.

[101] E. Panfilov, A. Tiulpin, M. T. Nieminen, S. Saarakkala, and V. Casula.
Deep learning-based segmentation of knee mri for fully automatic subregional morphological assessment of cartilage tissues: Data from the osteoarthritis initiative.
J. Orthop. Res., 40(5):1113–1124, 2022.

[102] A. Alexopoulos, J. Hirvasniemi, and N. Tümer.
Early detection of knee osteoarthritis using deep learning on knee magnetic resonance images.
arXiv [physics.med-ph], 2022.

[103] A. A. Tolpadi, J. J. Lee, V. Pedoia, and S. Majumdar.
Deep learning predicts total knee replacement from magnetic resonance images.

Sci. Rep., 10(1):6371, 2020.

[104] Y. Hu, J. Tang, S. Zhao, and Y. Li.
Deep learning-based multimodal 3 t mri for the diagnosis of knee osteoarthritis.
Computational and Mathematical Methods in Medicine, 2022, 2022.

[105] D. Azcona, K. McGuinness, and A. F. Smeaton.
A comparative study of existing and new deep learning methods for detecting knee injuries using the mrnet dataset.
In *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pages 149–155, 2020.

[106] Benjamin Fritz, Giuseppe Marbach, Francesco Civardi, Sandro F Fucentese, and Christian WA Pfirrmann.
Deep convolutional neural network-based detection of meniscus tears: comparison with radiologists and surgery as standard of reference.
Skeletal radiology, 49:1207–1217, 2020.

[107] Yaodong Du, Rania Almajalid, Juan Shan, and Ming Zhang.
A novel method to predict knee osteoarthritis progression on mri using machine learning methods.
IEEE Transactions on NanoBioscience, 17(3):228–236, 2018.

[108] Pauline Shan Qing Yeoh, Khin Wee Lai, Siew Li Goh, Khairunnisa Hasikin, Xiang Wu, and Pei Li.
Transfer learning-assisted 3d deep learning models for knee osteoarthritis detection: Data from the osteoarthritis initiative.
Frontiers in Bioengineering and Biotechnology, 11:1164655, 2023.

[109] Jean-Baptiste Schiratti, Rémy Dubois, Paul Herent, David Cahané, Jocelyn Dachary, Thomas Clozel, Gilles Wainrib, Florence Keime-Guibert, Agnes Lalande, Maria Pueyo, et al.
A deep learning method for predicting knee osteoarthritis radiographic progression from mri.
Arthritis Research & Therapy, 23:1–10, 2021.

[110] Fang Liu, Zhaoye Zhou, Alexey Samsonov, Donna Blankenbaker, Will Larison, Andrew Kanarek, Kevin Lian, Shivkumar Kambhampati, and Richard Kijowski.
Deep learning approach for evaluating knee mr images: Achieving high diagnostic performance for cartilage lesion detection.
Radiology, 289(1):160–169, 2018.
PMID: 30063195.

[111] A. Tiulpin et al.
Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data.
Scientific Reports, 9(1):20038, 2019.

[112] Adam J Schwartz, Henry D Clarke, Mark J Spangehl, Joshua S Bingham, David A Etzioni, and Matthew R Neville.
Can a convolutional neural network classify knee osteoarthritis on plain radiographs as accurately as fellowship-trained knee arthroplasty surgeons?
The Journal of Arthroplasty, 35(9):2423–2428, 2020.

[113] B Guan, F Liu, AH Mizaian, S Demhri, T Neogi, A Guermazi, and R Kijowski.
Deep learning approach to predict radiographic knee osteoarthritis progression.
Osteoarthritis and Cartilage, 27:S395–S396, 2019.

[114] A Tiulpin, S Klein, S Bierma-Zeinstra, J Thevenot, J van Meurs, E Oei, and S Saarakkala.
Deep learning predicts knee osteoarthritis progression from plain radiographs.
Osteoarthritis and Cartilage, 27:S397–S398, 2019.

[115] Bochen Guan, Fang Liu, Arya Haj-Mirzaian, Shadpour Demehri, Alexey Samsonov, Tuhina Neogi, Ali Guermazi, and Richard Kijowski.
Deep learning risk assessment models for predicting progression of radiographic medial joint space loss over a 48-month follow-up period.
Osteoarthritis and cartilage, 28(4):428–437, 2020.

[116] Zhe Wang, Aladine Chetouani, and Rachid Jennane.
Transformer with selective shuffled position embedding for early detection of knee osteoarthritis.
In *Proceedings of the 20th International Conference on Content-based Multimedia Indexing*, pages 261–264, 2023.

[117] Bin Liu, Jianxu Luo, and Huan Huang.
Toward automatic quantification of knee osteoarthritis severity using improved faster r-cnn.
International journal of computer assisted radiology and surgery, 15:457–466, 2020.

[118] Yassine Nasser, Mohammed El Hassouni, Didier Hans, and Rachid Jennane.
A discriminative shape-texture convolutional neural network for early diagnosis of knee osteoarthritis from x-ray images.
Physical and Engineering Sciences in Medicine, 46(2):827–837, 2023.

[119] Adam J Schwartz, Henry D Clarke, Mark J Spangehl, Joshua S Bingham, David A Etzioni, and Matthew R Neville.

Can a convolutional neural network classify knee osteoarthritis on plain radiographs as accurately as fellowship-trained knee arthroplasty surgeons?
The Journal of Arthroplasty, 35(9):2423–2428, 2020.

[120] Shi Yan, Taghi Ramazanian, Elham Sagheb, Walter K Kremers, Vipin Chaudhary, Michael Taunton, Hilal Maradit Kremers, and Ahmad P Tafti.
Give me a knee radiograph, i will tell you where the knee joint area is: a deep convolutional neural network adventure.
arXiv preprint arXiv:2202.05382, 2022.

[121] Pingjun Chen, Linlin Gao, Xiaoshuang Shi, Kyle Allen, and Lin Yang.
Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss.
Computerized Medical Imaging and Graphics, 75:84–92, 2019.

[122] Berk Norman, Valentina Pedoia, and Sharmila Majumdar.
Use of 2d u-net convolutional neural networks for automated cartilage and meniscus segmentation of knee mr imaging data to determine relaxometry and morphometry.
Radiology, 288(1):177–185, 2018.

[123] Liping Si, Kai Xuan, Jingyu Zhong, Jiayu Huo, Yue Xing, Jia Geng, Yangfan Hu, Huan Zhang, Qian Wang, and Weiwu Yao.
Knee cartilage thickness differs alongside ages: a 3-t magnetic resonance research upon 2,481 subjects via deep learning.
Frontiers in medicine, 7:600049, 2021.

[124] Wolfgang Wirth, Felix Eckstein, Jana Kemnitz, Christian Frederik Baumgartner, Ender Konukoglu, David Fuerst, and Akshay Sanjay Chaudhari.
Accuracy and longitudinal reproducibility of quantitative femorotibial cartilage measures derived from automated u-net-based segmentation of two different mri contrasts: data from the osteoarthritis initiative healthy reference cohort.
Magnetic Resonance Materials in Physics, Biology and Medicine, 34:337–354, 2021.

[125] Gayatri Kompella, Maria Antico, Fumio Sasazawa, S Jeevakala, Keerthi Ram, Davide Fontanarosa, Ajay K Pandey, and Mohanasankar Sivaprakasam.
Segmentation of femoral cartilage from knee ultrasound images using mask r-cnn.
In *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 966–969. IEEE, 2019.

[126] Fang Liu, Zhaoye Zhou, Alexey Samsonov, Donna Blankenbaker, Will Larison, Andrew Kanarek, Kevin Lian, Shivkumar Kambhampati, and Richard Kijowski.

Learning from highly confident samples for automatic knee osteoarthritis severity assessment: Data from the osteoarthritis initiative.
IEEE Journal of Biomedical and Health Informatics, 26(3):1239–1250, March 2022.

[135] A. Tack, A. Shestakov, D. Ludke, and S. Zachow.
A multi-task deep learning method for detection of meniscal tears in mri data from the osteoarthritis initiative database.
Frontiers In Bioengineering And Biotechnology, 9, 2021.

[136] L. C. Ribas, R. Riad, R. Jennane, and O. M. Bruno.
A complex network based approach for knee osteoarthritis detection: Data from the osteoarthritis initiative.
Biomed. Signal Process. Control, 71:103133, 2022.

[137] S. Gaj, M. Yang, K. Nakamura, and X. Li.
Automated cartilage and meniscus segmentation of knee mri with conditional generative adversarial networks.
Magn. Reson. Med., 84(1):437–449, 2020.

[138] Jian-Qing Zheng, Ngee Han Lim, and Bartłomiej W. Papież.
Accurate volume alignment of arbitrarily oriented tibiae based on a mutual attention network for osteoarthritis analysis.
Computerized Medical Imaging and Graphics, 106:102204, 2023.

[139] H. H. Nguyen, S. Saarakkala, M. B. Blaschko, and A. Tiulpin.
semixup: In- and out-of-manifold regularization for deep semi-supervised knee osteoarthritis severity grading from plain radiographs.
IEEE Trans. Med. Imaging, 39(12):4346–4356, 2020.

[140] Abdelbasset Brahim, Rachid Jennane, Rabia Riad, Thomas Janvier, Laila Khedher, Hechmi Toumi, and Eric Lespessailles.
A decision support tool for early detection of knee osteoarthritis using x-ray imaging and machine learning: Data from the osteoarthritis initiative.
Computerized Medical Imaging and Graphics, 73:11–18, 2019.

[141] Charis Ntakolia, Christos Kokkotis, Serafeim Moustakidis, and Dimitrios Tsaopoulos.
Identification of most important features based on a fuzzy ensemble technique: Evaluation on joint space narrowing progression in knee osteoarthritis patients.
International Journal of Medical Informatics, 156:104614, 2021.

[142] L. Sun, J. Wang, Z. Hu, Y. Xu, and Z. Cui.
Multi-view convolutional neural networks for mammographic image classification.
IEEE Access, 7:126273–126282, 2019.

[143] Q. Liu, L. Yu, L. Luo, Q. Dou, and P. A. Heng.
Semi-supervised medical image classification with relation-driven self-ensembling model.
IEEE Trans. Med. Imaging, 39(11):3429–3440, 2020.

[144] Min Lin, Qiang Chen, and Shuicheng Yan.
Network in network.
arXiv preprint arXiv:1312.4400, 2013.

[145] A. S. Chaudhari et al.
Utility of deep learning super-resolution in the context of osteoarthritis mri biomarkers.
J. Magn. Reson. Imaging, 51(3):768–779, 2020.

[146] G. H. Chang, D. T. Felson, S. Qiu, A. Guermazi, T. D. Capellini, and V. B. Kolachalama.
Assessment of knee pain from mr imaging using a convolutional siamese network.
European Radiology, 30(6):3538–3548, 2020.

[147] P. Chen, L. Gao, X. Shi, K. Allen, and L. Yang.
Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss.
Comput. Med. Imaging Graph., 75:84–92, 2019.

[148] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra.
Grad-cam: Visual explanations from deep networks via gradient-based localization.
International Journal of Computer Vision, 128(2):336–359, October 2019.

[149] J. C. Mora, R. Przkora, and Y. Cruz-Almeida.
Knee osteoarthritis: pathophysiology and current treatment modalities.
J. Pain Res., 11:2189–2196, 2018.

[150] Y. H. Aghdam, A. Moradi, L. G. Großterlinden, M. S. Jafari, J. T. Heverhagen, and K. Daneshvar.
Accuracy of magnetic resonance imaging in assessing knee cartilage changes over time in patients with osteoarthritis: A systematic review.
Journal of Orthopaedic Surgery and Research, 15(1):316, 2020.

[151] D. J. Hunter et al.
The association of meniscal pathologic changes with cartilage loss in symptomatic knee osteoarthritis.
Arthritis & Rheumatism, 54(3):795–801, 2006.

[152] T. Zhou, H. Fu, G. Chen, J. Shen, and L. Shao.
Hi-net: Hybrid-fusion network for multi-modal mr image synthesis.

IEEE Transactions on Medical Imaging, 39(9):2772–2781, 2020.

[153] Penghua Zhai, Huaiwei Cong, Enwei Zhu, Gangming Zhao, Yizhou Yu, and Jinpeng Li.
Mvcnet: multiview contrastive network for unsupervised representation learning for 3-d ct lesions.
IEEE Transactions on Neural Networks and Learning Systems, 2022.

[154] Maram Mahmoud A Monshi, Josiah Poon, and Vera Chung.
Convolutional neural network to detect thorax diseases from multi-view chest x-rays.
In *Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part IV 26*, pages 148–158. Springer, 2019.

[155] J. C. Teo, I. M. Khairuddin, M. A. M. Razman, A. P. P. A. Majeed, and W. H. M. Isa.
Automated detection of knee cartilage region in x-ray image.
MEKATRONIKA, 4(1):104–109, 2022.

[156] B. Rizk et al.
Meniscal lesion detection and characterization in adult knee mri: a deep learning model approach with external validation.
Physica Medica, 83:64–71, 2021.

[157] M. Berrimi, D. Hans, and R. Jennane.
A semi-supervised multiview-mri network for the detection of knee osteoarthritis.
Computerized Medical Imaging and Graphics, 114:102371, 2024.

[158] Chunyu Li, Ming Deng, Xiaoli Zhong, Jinxia Ren, Xiaohui Chen, Jun Chen, Feng Xiao, and Haibo Xu.
Multi-view radiomics and deep learning modeling for prostate cancer detection based on multi-parametric mri.
Frontiers in Oncology, 13:1198899, 2023.

[159] Jin Liu, Yi Pan, Fang-Xiang Wu, and Jianxin Wang.
Enhancing the feature representation of multi-modal mri data by combining multi-view information for mci classification.
Neurocomputing, 400:322–332, 2020.

[160] Zhe Guo, Xiang Li, Heng Huang, Ning Guo, and Quanzheng Li.
Deep learning-based image segmentation on multimodal medical imaging.
IEEE Transactions on Radiation and Plasma Medical Sciences, 3(2):162–169, 2019.

[161] Aiham Taleb, Christoph Lippert, Tassilo Klein, and Moin Nabi.
Multimodal self-supervised learning for medical image analysis.

In *International conference on information processing in medical imaging*, pages 661–673. Springer, 2021.

[162] Eman A Abdel Maksoud, Sherif Barakat, and Mohammed Elmogy.
Medical images analysis based on multilabel classification.
In *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*, pages 209–245. Elsevier, 2019.

[163] Constantinos Loukas and Nicholas P Sgouros.
Multi-instance multi-label learning for surgical image annotation.
The International Journal of Medical Robotics and Computer Assisted Surgery, 16(2):e2058, 2020.

[164] Manuel Alejandro Rodríguez, Hasan AlMarzouqi, and Panos Liatsis.
Multi-label retinal disease classification using transformers.
IEEE Journal of Biomedical and Health Informatics, 27(6):2739–2750, 2022.

[165] H. Lee, Y.-S. Kim, M. Kim, and Y. Lee.
Low-cost network scheduling of 3d-cnn processing for embedded action recognition.
IEEE Access, 9:83901–83912, 2021.

[166] Z. Shao et al.
Transmil: Transformer based correlated multiple instance learning for whole slide image classification.
Advances in Neural Information Processing Systems, 34:2136–2147, 2021.

[167] J. Hu, L. Shen, and G. Sun.
Squeeze-and-excitation networks.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[168] C. Cortes, M. Mohri, and A. Rostamizadeh.
L2 regularization for learning kernels.
arXiv preprint arXiv:1205.2653, 2012.

[169] Y. Zhang and Q. Yang.
A survey on multi-task learning.
IEEE Transactions on Knowledge and Data Engineering, 34(12):5586–5609, 2021.

[170] Y. Zhao, X. Wang, T. Che, G. Bao, and S. Li.
Multi-task deep learning for medical image computing and analysis: A review.
Computers in Biology and Medicine, 153:106496, 2023.

[171] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, and J. Li.

A robust deep model for improved classification of ad/mci patients.
IEEE Journal of Biomedical and Health Informatics, 19(5):1610–1616, 2015.

[172] P. S. Q. Yeoh, S. L. Goh, K. Hasikin, X. Wu, and K. W. Lai.
3d efficient multi-task neural network for knee osteoarthritis diagnosis using mri scans: Data from the osteoarthritis initiative.
IEEE Access, 11:135323–135333, 2023.