University of Thessaly

School of Science

Department of Computer Science
and Biomedical Informatics

Harnessing Deep Learning for Precise ACL Tear Diagnosis and Knee Abnormality Detection in Medical Imaging

DOCTORAL DISSERTATION

Υποψήφιος Διδάκτωρ: Siouras Athanasios

Ημερομηνία: 29 Νοεμβρίου 2024

Members of the examination committee

Full name	University/Research centre	Information
Dr. Sotirios Tasoulis	Assistant Professor, Department of Computer Science and Biomedical Informatics, University of Thessaly, Greece	Supervisor
Dr. Dimitrios E. Tsaopoulos	Researcher A’, Bio-Economy and Agri-Technology (IBO)/ Centre for Research and Technology – Hellas (CERTH)	Three-member committee
Dr. Michael Hantes	Professor, Department of Orthopaedic Surgery, University of Thessaly, Larissa, Greece	Three-member committee
Dr. Archontis Giannakidis	Senior Lecturer, Department of Data Science with the School of Science and Technology, Nottingham trent University	Seven-member committee
Dr. Marianna Vlychou	Professor, Department of Radiology, University of Thessaly, Larissa, Greece	Seven-member committee
Dr. Giannis Giakas	Professor, Department of Physical Education & Sport Science, University of Thessaly, Greece	Seven-member committee
Dr. Vassilis Plagianakos	Professor, Department of Computer Science and Biomedical Informatics, University of Thessaly, Greece	Seven-member committee

Declaration

Athanasios Siouras, hereby declare that this thesis has not been previously submitted in this University or any other University for the award of any degree, diploma, associateship, fellowship or other similar titles of recognition.

Thesis title: Harnessing Deep Learning for Precise ACL Tear Diagnosis and Knee Abnormality Detection in Medical Imaging.

Τίτλος: Αξιοποίηση της βαθιάς μάθησης για την ακριβή διάγνωση της ρήξης του πρόσθιου χιαστού συνδέσμου και την ανίχνευση ανωμαλιών της άρθρωσης του γονάτου στην ιατρική απεικόνιση

Pages: 113

Words: 39175

Figures: 19

Tables: 16

Plagiarism Disclaimer

I declare that this thesis is based on a total of four studies/papers of which three (Chapter 2, Chapter 3, and Chapter 4) have been published and one have been accepted for publication. These papers are referred to and cited in the thesis chapters (with specification in the intro to each chapter of the study/paper referenced).

List of Publications

[1] Siouras A, Moustakidis S, Giannakidis A, Chalatsis G, Liampas I, Vlychou M, Hantes M, Tasoulis S, Tsaopoulos D. Knee Injury Detection Using Deep Learning on MRI Studies: A Systematic Review. Diagnostics. 2022; 12(2):537. https://doi.org/10.3390/diagnostics12020537

[2] A. Siouras et al., "Automated Recognition of healthy Anterior Cruciate Ligament in Sagittal MR images using Lightweight Deep Learning," 2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA), Corfu, Greece, 2022, pp. 1-8, doi:10.1109/IISA56318.2022.9904387.

[3] Siouras, A., Moustakidis, S., Chalatsis, G., Bohoran, T. A., Hantes, M., Vlychou, M., M., Tasoulis, S., Giannakidis, A. and Tsaopoulos, D. (2024). Economical hybrid novelty detection leveraging global aleatoric semantic uncertainty for enhanced MRI-based ACL tear diagnosis. Computerized Medical Imaging and Graphics, 102424.

[4] Graft Index: An Artificial Intelligence-Based Index for the Assessment of Graft Integrity in ACL Reconstructed Knees

Accepted for Publication:

"Graft Index: An Artificial Intelligence-Based Index for the Assessment of Graft Integrity in ACL Reconstructed Knees."

Georgios Chalatsis, Athanasios Siouras, Vasileios Mitrousias, Ilias Chantes, Serafeim Moustakidis, Dimitris Tsaopoulos, Marianna Vlychou, Sotiris Tasoulis, Michael Hantes

Journal of Bone and Joint Surgery (JBJS), Manuscript ID: JBJS-D-24-00427R3.

[5] Advanced Dual-Phase Deep Learning Approach for ACL Tear Identification in Sagittal MRI Scans, Authorblock**:** A. Siouras, S. Moustakidis, G. Chalatsis, T. A. Bohoran, M. Hantes, A. Giannakidis, S. Tasoulis, D. Tsaopoulos, M. Vlychou; Poster No.: C-23348 , Congress: ECR24

Acknowledgments

I would like to express my sincere gratitude to my supervisors, Dr. Dimitrios Tsaopoulos, Dr. Sotiris Tasoulis, and Dr. Michael Hantes, for their invaluable guidance and support throughout this journey. Their expertise, patience, and encouragement have been instrumental in helping me navigate the complexities of this research. Without their mentorship, this thesis would not have been possible.

I am deeply thankful to Dr. Archontis Giannakidis for his meticulous assistance as an external professor. His insightful feedback and unwavering support significantly contributed to the development and refinement of this work.

I would also like to express my deep appreciation to Dr. George Chalatsis. Without his support, none of this would have happened. He helped me start this PhD and, as a medical expert, provided both mental and technical assistance that was invaluable throughout this journey.

A special thank you goes to Dr. Serafeim Moustakidis, who has been a constant source of inspiration from beginning to end. His helpful ideas and our brainstorming sessions greatly enriched my work.

I extend my gratitude to all who have been part of my PhD journey. Their encouragement, collaboration, and shared experiences have made this path more meaningful and fulfilling.

This work was supported by funding from the European Community's H2020 Programme, under grant agreement Nr. 777159 (OActive project), and by the National Operational Program Competitiveness, Entrepreneurship and Innovation, under the call Research – Create – Innovate (SafeACL project, code: T1EDK-04234).

Finally, I want to thank my family for their unwavering support. To my wife, whose understanding has been my anchor; to my children, who have been a constant source of joy and motivation; and to my friends, who provided the encouragement that kept me focused on this goal. Without their belief in me, this achievement would not have been possible.

Thanos Siouras

Contents

List of Figures

List of Tables

Abstract

The accurate diagnosis of anterior cruciate ligament (ACL) rupture is crucial for the effective treatment and rehabilitation of knee injuries, which are particularly prevalent among athletes. Recent advancements in deep learning (DL) have significantly improved the detection of ACL injuries using MRI studies. This dissertation aims to explore and enhance the application of DL techniques for the precise identification and evaluation of ACL injuries in medical imaging.

A systematic review was conducted to summarize existing DL methodologies for detecting knee injuries, highlighting the strengths and limitations of current approaches. Building upon this foundation, a lightweight DL pipeline using YOLOv5-Nano was developed to identify healthy ACLs in sagittal MRI slices, achieving high accuracy and efficiency with minimal computational resources. Additionally, the thesis introduces the Thessaly Graft Index (TGI), an AI-based index for assessing graft healing post-ACL reconstruction. The TGI demonstrated a strong correlation with clinical and functional outcomes, proving its potential as a reliable tool for postoperative evaluation.

To address class imbalance and enhance diagnostic precision, an innovative hybrid DL framework was formulated, reconfiguring ACL tear classification as a novelty detection problem. This approach leveraged aleatoric semantic uncertainty and demonstrated superior performance across multiple datasets, including cross-database testing with public and clinical MRI data.

The dissertation's findings underscore the potential of DL to match human-level performance in MRI-based knee injury diagnosis. Key contributions include the development of resource-efficient models, the introduction of novel diagnostic indices, and the demonstration of robust generalization capabilities across diverse datasets. These advancements pave the way for the widespread clinical application of DL in the accurate and cost-effective diagnosis of ACL injuries, promising significant improvements in patient outcomes and healthcare efficiency.

Περίληψη

Η ακριβής διάγνωση της ρήξης του πρόσθιου χιαστού συνδέσμου (ACL) είναι ζωτικής σημασίας για την αποτελεσματική θεραπεία και αποκατάσταση των τραυματισμών στο γόνατο, οι οποίοι είναι ιδιαίτερα συχνοί στους αθλητές. Οι πρόσφατες εξελίξεις στην τεχνολογία της βαθιάς μάθησης έχουν βελτιώσει σημαντικά την ανίχνευση τραυματισμών του ACL μέσω μελετών μαγνητικής τομογραφίας (MRI). Η παρούσα διδακτορική διατριβή έχει ως στόχο να διερευνήσει και να βελτιώσει την εφαρμογή τεχνικών βαθιάς μάθησης για την ακριβή αναγνώριση και αξιολόγηση των τραυματισμών του ACL στην ιατρική απεικόνιση.

Πραγματοποιήθηκε συστηματική ανασκόπηση για τη σύνοψη των υφιστάμενων μεθοδολογιών βαθιάς μάθησης που αφορούν την ανίχνευση τραυματισμών στο γόνατο, αναδεικνύοντας τα πλεονεκτήματα και τους περιορισμούς των τρεχουσών προσεγγίσεων. Βάσει αυτής της θεμελίωσης, αναπτύχθηκε μια ελαφριά διαδικασία βαθιάς μάθησης με χρήση του YOLOv5-Nano για την αναγνώριση υγιών πρόσθιων χιαστών συνδέσμων σε οβελιαίες τομές MRI, επιτυγχάνοντας υψηλή ακρίβεια και αποδοτικότητα με ελάχιστους υπολογιστικούς πόρους. Επιπλέον, η διατριβή εισάγει τον Δείκτη Μοσχεύματος Θεσσαλίας (Thessaly Graft Index - TGI), έναν δείκτη που βασίζεται στην τεχνητή νοημοσύνη για την αξιολόγηση της επούλωσης μοσχεύματος μετά από ανακατασκευή του ACL. Ο TGI επέδειξε ισχυρή συσχέτιση με τα κλινικά και λειτουργικά αποτελέσματα, αποδεικνύοντας τη δυνατότητά του ως αξιόπιστο εργαλείο για μετεγχειρητική αξιολόγηση.

Για την αντιμετώπιση της ανισορροπίας των κατηγοριών και τη βελτίωση της διαγνωστικής ακρίβειας, διαμορφώθηκε ένα καινοτόμο υβριδικό πλαίσιο βαθιάς μάθησης, επαναπροσδιορίζοντας την ταξινόμηση της ρήξης του ACL ως πρόβλημα ανίχνευσης νεότητας (novelty detection). Αυτή η προσέγγιση αξιοποίησε την αληθοκρατική σημασιολογική αβεβαιότητα και επέδειξε ανώτερη απόδοση σε πολλαπλά σύνολα δεδομένων, συμπεριλαμβανομένων δοκιμών σε διασταυρούμενες βάσεις δεδομένων με δημόσια και κλινικά δεδομένα MRI.

Τα ευρήματα της διατριβής υπογραμμίζουν τη δυνατότητα της βαθιάς μάθησης να επιτύχει επιδόσεις αντίστοιχες με αυτές του ανθρώπου στη διάγνωση τραυματισμών στο γόνατο μέσω MRI. Οι βασικές συνεισφορές περιλαμβάνουν την ανάπτυξη αποδοτικών μοντέλων, την εισαγωγή νέων διαγνωστικών δεικτών και την απόδειξη ισχυρών ικανοτήτων γενίκευσης σε ποικίλα σύνολα δεδομένων. Αυτές οι εξελίξεις ανοίγουν το δρόμο για την ευρεία κλινική εφαρμογή της βαθιάς μάθησης στην ακριβή και οικονομικά αποδοτική διάγνωση τραυματισμών του ACL, υποσχόμενες σημαντικές βελτιώσεις στα αποτελέσματα των ασθενών και στην αποδοτικότητα της υγειονομικής περίθαλψης.

Chapter 1: Introduction

Knee injuries are not just a setback for athletes; they are experienced by millions of people globally, impacting daily activities and overall quality of life. Among these, Anterior Cruciate Ligament (ACL) injuries stand out due to their frequency and the significant challenges posed in diagnosis and treatment. Traditional methods often rely heavily on a clinician's expertise, which can lead to variability in diagnosis and delays in treatment planning.

In an era where technology is rapidly evolving, a promising avenue exists to enhance how these injuries are addressed. This chapter is devoted to exploring the pivotal role of Deep Learning (DL) techniques in transforming ACL injury diagnosis and management. The prevalence of knee injuries and the substantial socio-economic burdens they impose are examined first. By analyzing the limitations of current diagnostic practices, the pressing need for more accurate and efficient solutions is highlighted.

To build a solid foundation for this exploration, key concepts and terminology related to DL and medical imaging are introduced. The ways in which neural networks and advanced algorithms are revolutionizing the analysis of complex medical images like MRI scans are discussed. A focus is placed on how DL models—particularly Convolutional Neural Networks (CNNs) and object detection frameworks—are applied to identify and classify ACL injuries and to assess graft ealing following reconstruction.

The accuracy of ACL diagnosis is more than a clinical concern; it is essential for effective treatment planning, preventing long-term complications, and improving patient outcomes. With this understanding, the objectives and scope of this dissertation are outlined. The aims include conducting a systematic literature review, developing lightweight DL models, proposing a new hybrid diagnostic framework, and introducing the Tibial Graft Index (TGI).

Finally, the structure of the dissertation is presented, previewing the upcoming chapters that detail the methodologies employed, the findings obtained, and the implications of the research. This introduction sets the stage for a comprehensive journey into how DL can enhance ACL injury diagnosis and make significant contributions to orthopedic research and patient care.

1.1. Background and Motivation

Knee injuries represent a significant portion of sport-related severe injuries, with ACL ruptures accounting for more than half of these cases. In the United States alone, ACL injuries affect over 200,000 individuals annually, leading to substantial healthcare costs exceeding $7 billion. These injuries not only impact the physical well-being of individuals but also impose a heavy socio-economic burden due to direct medical expenses and indirect costs such as lost wages and reduced productivity.

The high prevalence of knee injuries, particularly among young and athletic populations, underscores the importance of developing accurate, cost-effective, and reliable diagnostic methods. Early diagnosis and appropriate treatment of ACL injuries can prevent the onset of long-term complications, including knee osteoarthritis (OA), which is a common consequence of such injuries. Magnetic resonance imaging (MRI) has become the most widely used non-invasive technique for diagnosing knee injuries due to its ability to provide detailed images of the knee joint's internal structures.

Despite the effectiveness of MRI, the accurate interpretation of these images remains challenging. The diagnostic process heavily relies on the expertise and experience of clinicians, leading to variability in diagnoses due to human factors such as subjectivity, fatigue, and diagnostic uncertainties. The potential for misinterpretation is further compounded by clinical-diagnostic discrepancies between radiologists and orthopedic surgeons, which can hinder optimal patient management.

Given these challenges, there is a growing interest in utilizing artificial intelligence (AI) and machine learning (ML) technologies to enhance the diagnostic process. AI, and more specifically deep DL, has shown significant promise in medical image analysis by automatically learning complex patterns and relationships within the data. DL algorithms, particularly convolutional neural networks (CNNs), have been successfully applied in various medical imaging tasks, including the detection of skin cancer, diabetic retinopathy, lung nodules, and breast cancer.

The application of DL to knee injury diagnosis has gained considerable attention due to its potential to improve diagnostic accuracy and efficiency. Several studies have demonstrated that DL models can achieve performance levels comparable to, or even surpassing, those of experienced radiologists. For instance, customized CNN architectures have been developed to detect ACL injuries with high accuracy, providing a robust and reliable tool for clinical use.

In the context of ACL injuries, DL models have been employed to identify and classify the presence of ACL tears in MRI scans. These models can automatically detect and localize the ACL within the images, reducing the reliance on manual interpretation and potentially lowering the risk of diagnostic errors. Additionally, DL techniques have been explored for their ability to assess the maturation of ACL grafts post-reconstruction, providing valuable insights into the healing process and guiding rehabilitation protocols.

One of the key challenges in applying DL to medical imaging is the need for models that are both accurate and computationally efficient. Traditional DL models often require significant computational resources, which can limit their practicality in clinical settings. To address this, lightweight DL models such as YOLOv5-Nano have been developed. These models maintain high detection accuracy while being optimized for use on devices with limited computational power, making them suitable for deployment in a variety of healthcare environments.

Another critical aspect of DL applications in ACL injury diagnosis is the handling of data variability and class imbalance. Novelty detection frameworks, which identify deviations from normal patterns, have been employed to tackle the challenges associated with classifying ACL tears. By leveraging uncertainty measures, these models can provide robust and reliable classifications, even in the presence of noisy or incomplete data.

In addition to detecting ACL injuries, this dissertation introduces the TGI, an AI-based index designed to evaluate the maturation of ACL grafts post-reconstruction. The TGI uses MRI images to provide a quantitative measure of graft healing, which can be correlated with clinical outcomes and patient-reported measures. This tool aims to enhance postoperative monitoring and decision-making, ultimately improving patient recovery and outcomes.

Overall, the integration of DL into the diagnostic process for ACL injuries offers numerous benefits, including improved accuracy, reduced analysis time, and enhanced consistency in image interpretation. By addressing the limitations of traditional diagnostic methods and leveraging the strengths of DL algorithms, this dissertation seeks to advance the field of knee injury diagnosis and contribute to better patient care.

The following sections will delve into the specifics of these advancements, starting with a systematic review of existing DL methodologies for knee injury detection, followed by the development and validation of new models for ACL injury diagnosis and graft assessment. Each chapter builds upon the previous one, providing a comprehensive exploration of the potential of DL in improving the diagnosis and treatment of ACL injuries.

1.2. Definitions and Terminology

To provide a comprehensive overview of the key definitions and terminology used in this dissertation, including the various models and techniques referenced across the four papers, a table format can indeed be more effective. Here are detailed Table 1, Table 2, and Table 3 summarizing the critical terms and their descriptions:

Term/Model	Description
Deep Learning	A subset of machine learning involving neural networks with many layers that learn to perform tasks by example, typically without task-specific programming.
YOLO (You Only Look Once)	A family of DL models for object detection, known for their speed and accuracy. Variants like YOLOv5-Nano are used for detecting ACL in MRI scans in this dissertation.
YOLOv5-Nano	A lightweight version of the YOLOv5 model, designed for scenarios with limited computational resources. Maintains high detection accuracy, suitable for mobile and embedded applications.
Novelty Detection	A machine learning task that identifies new or unknown data not seen during training. Used to detect abnormalities in MRI images indicating ACL injuries.
Aleatoric Uncertainty	Uncertainty arising from inherent noise in the data. Addressed in the dissertation to improve the reliability of ACL injury detection models.
Spearman's Correlation Coefficient	A non-parametric measure of rank correlation assessing the strength and direction of association between two ranked variables. Used to evaluate the relationship between TGI scores and PROMs.
MiniROCKET	A lightweight, efficient time series classification model used to determine the

	presence of ACL tears by analyzing MRI data sequences.
Hybrid Deep Learning Framework	A novel approach combining YOLOv5-Nano for object detection and MiniROCKET for time series classification, enhancing ACL tear diagnosis by addressing class imbalance and uncertainty.

Table 1: AI/ML Definitions

Term/Model	Description
Magnetic Resonance Imaging (MRI)	A non-invasive imaging technology that produces detailed three-dimensional anatomical images, commonly used for disease detection, diagnosis, and treatment monitoring.
KT-1000 Arthrometer	A device measuring anterior translation of the tibia relative to the femur, assessing ACL stability.
Patient-Reported Outcome Measures (PROMs)	Standardized questionnaires completed by patients to measure their functional status and well-being, including KOOS, IKDC, Lysholm, and TAS, used to correlate clinical outcomes with model predictions.
ACL	A key ligament in the knee joint that stabilizes the knee. ACL injuries, particularly tears, are common in athletes and can significantly impact mobility and performance.

Table 2: Health Definitions

Term/Model	Description
Thessaly Graft Index (TGI)	An AI-based index developed to evaluate the maturation of ACL grafts post-reconstruction, providing a quantitative measure of graft healing by comparing MRI images with healthy ACLs.
Spearman's Correlation Coefficient	A non-parametric measure of rank correlation assessing the strength and direction of association between two ranked variables. Used to evaluate the relationship between TGI scores and PROMs.

Table 3: Other Relevant Terms

These definitions and descriptions provide a foundational understanding of the concepts and models that underpin the research presented in this dissertation. The following sections will build on this foundation, detailing the methodologies, results, and implications of the studies conducted.

1.3. Overview of Deep Learning in Medical Imaging

The integration of DL into medical imaging has significantly advanced the analysis and interpretation of complex medical images, such as MRI scans. DL, a branch of artificial intelligence, employs neural networks with multiple layers to learn from large datasets, enabling the identification of intricate patterns that can assist clinicians in diagnosis and treatment planning.

In medical imaging, DL algorithms, particularly Convolutional Neural Networks (CNNs)—have been instrumental due to their ability to effectively process image data. CNNs automatically learn hierarchical feature representations from raw images, making them well-suited for detecting and classifying abnormalities within medical scans.

In the context of musculoskeletal injuries, specifically ACL injuries, DL models have been developed to enhance the accuracy and efficiency of diagnosis from MRI scans. Traditional diagnostic methods rely heavily on the expertise of radiologists, which can be time-consuming and subject to variability. DL models offer consistent and rapid analysis, potentially reducing diagnostic errors and improving patient outcomes.

Applications in ACL Injury Detection

- **Object Detection**: Models like YOLOv5-Nano have been adapted to identify the presence of ACLs in sagittal MRI slices. By detecting ACLs in specific images, these models aid in the preliminary identification of healthy ligaments and focus clinical evaluation on relevant slices, streamlining the diagnostic process.
- **Classification**: Novel frameworks incorporating methods like novelty detection and uncertainty estimation have been applied to classify ACL tears. By handling variability and ambiguity in medical images, these models improve the robustness and reliability of classification results.
- **Postoperative Assessment**: The development of indices such as the TGI utilizes DL to assess ACL graft maturation from MRI images. The TGI provides a quantitative measure of graft healing, aiding clinicians in monitoring patient recovery and making informed decisions about rehabilitation.

Benefits and Challenges

The use of DL in medical imaging offers several benefits:

- **Improved Diagnostic Accuracy**: Enhanced ability to detect subtle abnormalities that may be missed by human observers.
- **Efficiency**: Reduced analysis time through automated image interpretation.
- **Consistency**: Standardized evaluations that minimize inter-observer variability.

However, challenges remain:

- **Data Requirements**: DL models require large, high-quality datasets, which can be difficult to obtain in medical contexts.

- **Generalizability**: Ensuring models perform well across different imaging protocols and equipment.
- **Explainability**: Providing interpretable results that clinicians can trust and understand.

As DL techniques continue to evolve, their integration into clinical practice is expected to grow. Ongoing research aims to address current challenges by developing models that are both accurate and generalizable, with a focus on explainability to facilitate adoption in healthcare settings. This progress holds promise for enhancing patient care through more precise and timely diagnoses.

1.4. Importance of Accurate ACL Diagnosis

Accurate diagnosis of ACL injuries is critical for effective treatment and rehabilitation, significantly impacting patient outcomes and long-term knee health. Misdiagnosis can lead to prolonged recovery periods, inappropriate treatment, and an increased risk of further injury or complications such as osteoarthritis.

According to recent epidemiological data, ACL injuries affect an estimated 200,000 individuals annually in the United States alone, with incidence rates ranging from 68.6 to 81 per 100,000 person-years [1], [2]. These injuries are particularly prevalent among young, physically active populations, especially athletes involved in high-risk sports such as soccer, basketball, and skiing [3]. Female athletes are at a higher risk compared to males, with studies indicating they are two to eight times more likely to sustain an ACL injury due to anatomical and hormonal differences [4], [5]. Globally, the incidence of ACL injuries has been on the rise, partly attributed to increased participation in competitive sports and improved diagnostic awareness. This growing prevalence underscores the critical need for accurate diagnostic methods to ensure timely and effective treatment, ultimately improving patient outcomes and reducing the risk of long-term complications.

Traditional diagnostic methods for ACL injuries primarily involve clinical examinations and MRI scans interpreted by radiologists. Clinical assessments, such as the Lachman test or pivot shift test, rely heavily on the clinician's expertise and experience, which can result in variability and potential inaccuracies in diagnosis. While MRI scans provide detailed images of the knee joint, their interpretation depends on the radiologist's skill and can be subjective, leading to possible human error.

DL models offer a promising solution to these challenges by providing consistent and objective analysis of MRI scans. These models assist in detecting ACL injuries with high accuracy, reducing reliance on human interpretation and potentially lowering the incidence of diagnostic errors. For instance, as detailed in the paper "Knee Injury Detection Using Deep Learning on MRI Studies: A Systematic Review," the application of DL techniques has shown prediction accuracies ranging from 72.5% to 100%, demonstrating the potential of these models to achieve human-level performance in ACL injury diagnosis.

The development of lightweight and efficient DL models, such as the YOLOv5-Nano used in the study "Automated Recognition of Healthy ACL in Sagittal MR Images Using Lightweight Deep Learning," further emphasizes the importance of accurate diagnosis. These models not only achieve high detection performance but also operate efficiently on devices with limited computational resources, making advanced diagnostic tools more accessible in diverse clinical settings.

Moreover, accurate assessment of ACL graft maturation post-reconstruction is essential for determining the success of surgery and guiding rehabilitation protocols. The TGI, introduced in the study "Thessaly Graft Index: An Artificial Intelligence-Based Index for the Assessment of Graft Healing in ACL Reconstructed Knees—A Pilot Study," provides a quantitative measure of graft healing by comparing postoperative MRI images with those of healthy ACLs. This AI-based index aids in monitoring the recovery process, ensuring that patients are progressing as expected and facilitating timely interventions if necessary.

Accurate diagnosis is also crucial in managing the overall healthcare costs associated with ACL injuries. Early and precise identification can prevent progression to more severe conditions, reducing the need for extensive treatments and long-term rehabilitation. Additionally, the use of DL models can streamline the diagnostic process, allowing for faster decision-making and more efficient use of medical resources.

In summary, the importance of accurate ACL diagnosis cannot be overstated. It plays a vital role in ensuring effective treatment, reducing recovery times, and preventing further complications. By leveraging DL techniques, this dissertation aims to enhance the diagnostic process, providing reliable and objective tools that support clinicians in delivering optimal care for patients with ACL injuries. The subsequent chapters will delve into the methodologies, results, and implications of the studies conducted, all aimed at improving ACL injury diagnosis and treatment.

1.5. Objectives and Scope of the Dissertation

The primary objective of this dissertation is to explore and enhance the application of DL techniques for the accurate diagnosis of anterior ACL injuries and the assessment of ACL graft maturation using MRI studies. By addressing both the detection of ACL injuries and the postoperative evaluation of ACL grafts, this work aims to provide a comprehensive approach to improving patient care through advanced imaging analysis.

Systematic Literature Review

Initially, a comprehensive systematic literature review was conducted to analyze existing DL methods for knee injury detection, focusing on ACL tears as well as meniscus and cartilage injuries. This review followed the PRISMA guidelines and included searches across multiple databases such as PubMed, Cochrane Library, EMBASE, and Google Scholar. The review identified the current state of DL models in MRI-based knee injury diagnosis, highlighting their potential to match human-level performance. However, it also uncovered limitations in existing approaches, including data imbalance, lack of model generalizability across different centers, verification bias, and ground-truth subjectivity. The review emphasized the need for explainability and lightweight models to enable widespread clinical adoption.

Development of a Lightweight Deep Learning Pipeline

Building upon these insights, a robust and lightweight DL pipeline was developed for identifying healthy ACLs in 3D MRI data of healthy knees. The goal was to locate the sagittal slices where the ACL is present, facilitating further clinical evaluation. An advanced pipeline was implemented using the YOLOv5-Nano object detection network, which was compared against other models. The YOLOv5-Nano model demonstrated superior performance while maintaining a minimal model size, offering a viable solution for robust object detection in healthcare systems, especially on devices with limited computational resources.

Novel Hybrid Deep Learning Framework for ACL Tear Diagnosis

To address class imbalance and improve ACL tear diagnosis, an innovative hybrid DL framework was developed, reformulating the ACL tear classification task as a novelty detection problem. This approach introduced a highly discriminative novelty score by leveraging aleatoric semantic uncertainty from the YOLOv5-Nano model's class scores. By accounting for tissue continuity through the use of global probability vectors across entire sagittal sequences, the framework enhances robustness. The second module incorporates the MINIROCKET time-series classification model to determine the presence of an ACL tear. Cross-database testing demonstrated that the proposed method consistently outperformed existing approaches, achieving better generalization and environmental sustainability.

Introduction of the TGI

Recognizing the need for standardized assessment of ACL graft maturation post-reconstruction, the TGI was introduced. This AI-based index evaluates graft healing by computing the probability of accurately detecting a healthy ACL on a percentage scale. In a pilot study involving patients treated with hamstring tendon autografts, preoperative and postoperative MRIs were analyzed using the TGI. The index was found to be in total agreement with independent radiological assessments and showed correlations with patient-reported outcome measures (PROMs), indicating that the TGI is a valuable tool for accurately recognizing ACL graft ruptures and assessing graft maturation.

Integration and Clinical Implications

By integrating the findings from the systematic review, model development, and validation studies, this dissertation provides a cohesive analysis of the role of DL in ACL injury diagnosis and treatment. The implications for clinical practice are significant, offering potential improvements in diagnostic accuracy, efficiency, environmental sustainability, and patient outcomes. The scope of this work encompasses the entire process of developing and validating deep learning models for ACL injury diagnosis, from reviewing existing methodologies to creating and testing new models.

Ultimately, this dissertation aims to contribute significantly to the field of medical imaging and orthopedic research, demonstrating the potential of DL techniques to enhance clinical practice and improve patient outcomes. The following chapters will detail the methodologies, results, and implications of each study conducted, building towards a unified understanding of DL applications in this critical area of orthopedic research.

1.6. Structure of the Dissertation

This dissertation is organized into six main chapters, each addressing different aspects of DL applications in the diagnosis and assessment of ACL injuries. The structure is designed to provide a logical progression from a general introduction to detailed studies, followed by a comprehensive conclusion. The chapters are outlined as follows:

Chapter 1: Introduction

This chapter provides an overview of the dissertation, including the background and motivation for the study, definitions and terminology, an overview of DL in medical imaging, the importance of accurate ACL diagnosis, the objectives and scope of the dissertation, and the structure of the dissertation.

Chapter 2: Systematic Review of Knee Injury Detection Using Deep Learning

This chapter presents a systematic review of existing DL methodologies for detecting knee injuries, focusing on ACL tears. It includes the methods used for the review, the results of the analysis, and a discussion of the strengths and limitations of current approaches. The chapter concludes with a summary of key findings and potential areas for future research.

Chapter 3: Automated Recognition of Healthy ACL in Sagittal MR Images Using Lightweight Deep Learning

This chapter details the development of a lightweight DL pipeline using YOLOv5-Nano for identifying healthy ACLs in sagittal MRI slices. It covers the data collection, preprocessing, model architecture, training and validation processes, and the results obtained. A discussion of the findings and their implications for clinical practice is also included.

Chapter 4: Economical Hybrid Novelty Detection Leveraging Global Aleatoric Semantic Uncertainty for Enhanced MRI-Based ACL Tear Diagnosis

This chapter introduces a novel hybrid DL framework for ACL tear diagnosis, reconfiguring the task as a novelty detection problem. It describes the methods used, including data collection, preprocessing, and the integration of global aleatoric semantic uncertainty. The chapter presents the results of model validation, including cross-database testing, and discusses the robustness and reliability of the framework.

Chapter 5: Thessaly Graft Index: An Artificial Intelligence-Based Index for the Assessment of Graft Healing in ACL Reconstructed Knees - A Pilot Study

This chapter focuses on the development and validation of the TGI, an AI-based index for evaluating ACL graft maturation post-reconstruction. It outlines the study design, participant details, surgical techniques, rehabilitation protocols, and the methods used for clinical and MRI evaluation. The chapter presents the results of the TGI validation and its correlation with patient-reported outcome measures (PROMs) and clinical evaluations.

Chapter 6: General Conclusion and Future Work

The final chapter summarizes the key findings from the previous chapters, highlighting the contributions of the dissertation to the field of ACL injury diagnosis and treatment. It discusses the limitations of the studies conducted and provides recommendations for future research. The chapter concludes with final thoughts on the potential impact of DL techniques in clinical practice.

Each chapter builds on the previous one, providing a comprehensive and cohesive narrative that guides the reader through the research process and findings. This structure ensures that the dissertation is accessible and informative, offering valuable insights into the application of DL in medical imaging for ACL injury diagnosis.

Chapter 2: Systematic Review of Knee Injury Detection Using Deep Learning

Published as:

Siouras A, Moustakidis S, Giannakidis A, Chalatsis G, Liampas I, Vlychou M, Hantes M, Tasoulis S, Tsaopoulos D. Knee Injury Detection Using Deep Learning on MRI Studies: A Systematic Review. *Diagnostics*. 2022; 12(2):537. https://doi.org/10.3390/diagnostics12020537

This chapter provides a comprehensive systematic review of the current DL methodologies employed in the detection of knee injuries using magnetic resonance imaging (MRI), with a particular focus on ACL tears, meniscus injuries, and cartilage lesions. Recognizing that knee injuries constitute a significant portion of sports-related severe injuries and impose substantial socio-economic burdens due to healthcare costs and long-term disabilities, the chapter underscores the critical need for accurate and cost-effective diagnostic procedures.

The chapter begins by highlighting the limitations of traditional diagnostic methods, which often rely heavily on clinician expertise and are subject to human factors such as subjectivity and fatigue, leading to variability and potential errors in diagnosis. It then introduces the growing interest in artificial intelligence (AI) and DL as tools to enhance medical image interpretation. By automating the analysis of complex MRI data, DL algorithms, particularly convolutional neural networks (CNNs), offer the potential to improve diagnostic accuracy, efficiency, and consistency.

To ensure clarity, the chapter provides essential definitions and terminology related to machine learning (ML) and DL, outlining the differences between traditional ML pipelines and DL architectures. It explains key concepts such as supervised learning, feature extraction, CNNs, transfer learning, and various evaluation metrics used to assess model performance.

The systematic review adheres to the PRISMA guidelines, detailing the methodology for literature search, study selection, data extraction, and the types of outcome measures considered. A total of 22 relevant articles published between January 2013 and November 2021 are analyzed. These studies are categorized based on their application domains: detection of ACL injuries, meniscus tears, cartilage lesions, and combined knee injuries.

The chapter critically examines the methodologies and findings of these studies, comparing traditional ML approaches with DL techniques. It discusses how DL models have increasingly outperformed traditional methods, achieving high accuracy levels in detecting knee injuries. Particular attention is given to the use of transfer learning, custom-made DL architectures, data augmentation strategies, and localization techniques that focus on regions of interest within MRI scans.

Limitations identified across the reviewed studies are highlighted, including issues such as data imbalance, lack of multi-center validation, verification bias, and the subjectivity of ground truth annotations. The chapter emphasizes the need for more generalizable models that can perform reliably across different imaging protocols and patient populations. It also underscores the importance of developing explainable and lightweight AI models to facilitate clinical adoption and gain trust from medical professionals.

By synthesizing the current state of research, this chapter lays the groundwork for the subsequent development of novel DL models and frameworks presented in later chapters. It emphasizes the potential of AI to revolutionize knee injury diagnosis while acknowledging the challenges that must be addressed to integrate these technologies effectively into clinical practice.

2.1. Introduction

2.1.1. Backdrop

Knee injuries account for the largest percentage of sport-related severe injuries (i.e., injuries that cause more than 21 days of missed sport participation)[6]–[9]. ACL ruptures represent more than 50% of the cases, affecting 200,000 individuals in United States each year [6], [10]–[12]. Knee cartilage lesions affect around 900,000 individuals in United States every year, resulting in over 200,000 surgical procedures [10]–[13]. Menisci injuries is the second most common knee impairment with an incidence of 12-14% [14] and a prevalence of 60-70 cases per 100,000 in United Kingdom [7]. ACL injuries alone account for an expenditure of more than $7 billion in United States [15]. Both short and long-term pain, disability, and negatively affected health-related quality of life have all been strongly associated with knee injuries [16]–[18]. In regard to young and athletic individuals, the more time they spend engaging in occupational and/or recreational activities, the higher predisposition to knee injuries they have, which, in turn, contributes to a higher likelihood of developing osteoarthritis (OA) [19]. On average, half of the individuals, that have an injury that involved ACL and/or meniscal tear, develop radiographically confirmed knee OA, ten to 20 years' post-injury [20], [21]. Another two possible consequences of knee injuries are: (i) structural muscle injuries of the lower limb [22]; and (ii) tendinopathies [23]. All the above reflect the direct and indirect (lost wages, productivity, and disability) socio-economic burden conferred on the society by knee injuries. The high prevalence of knee injuries in the general population and the socio-economic impact have created the necessity of developing accurate and cost-effective procedures that can detect and quantify the severity of knee injuries. Early diagnosis and, consequently, treatment of ligament rupture, menisci tear and/or cartilage lesion can prevent early onset of knee OA [6].

Arthroscopy is considered the "gold-standard" for the diagnosis of intra-articular knee pathologies but is limited by potential complications and its invasive nature [24]. Therefore, magnetic resonance imaging (MRI) is the most widely used non-invasive imaging technique for diagnosing knee injuries [25], [26]. However, the MRI-based diagnosis of knee injuries can be a very challenging procedure, with the experience of clinicians playing a critical role in image interpretation. Human-based image interpretation pitfalls such as subjectivity, distraction, and fatigue, as well as diagnostic uncertainties often lead to erratic diagnoses hindering the optimal management of knee injuries [27], [28]. Moreover, clinical-diagnostic discrepancies among non-musculoskeletal radiologists and orthopedic surgeons are commonly encountered in everyday clinical practice [16].

Due to the above-listed factors, as well as the exponentially increasing number of clinical examinations, the idea of using computers for improving the challenging task of image interpretation of medical examinations has been recently adopted by the scientific community [29]. Imaging data proliferation, algorithmic advances, and recent technological advances in fast computing have already resulted in a strong push towards the utilization of artificial intelligence

(AI) algorithms in medical image analysis. The term AI broadly refers to any method that enables computers to mimic human intelligence [30]. DL in particular is a class of machine learning (ML) algorithms that is currently driving the AI boom [31]. Numerous applications of DL in medical image analysis have been reported including skin cancer classification, diabetic retinopathy detection, lung nodule detection, and mammography cancer detection among others [32]. The aforementioned AI-empowered solutions are expected to revolutionize medical sectors by improving the accuracy and productivity of different diagnostic and therapeutic measures in clinical practice [25].

Drawing attention to the diagnosis of knee injuries, several early DL studies have exhibited better performance than traditional ML techniques while in some cases they have proved to be even superior to radiologists [31]. However, the previously published review studies in the MRI field were either focused on other application domains (e.g. fracture detection [33]) or limited to the performance of the proposed networks without paying attention to their specifics (learning methodology, processing stages, technical limitations etc.) [34]. In the light of the advancements of AI technology and the increasing number of studies in this field, in this paper we conducted a systematic literature review, covering all DL-oriented techniques that have been employed in the diagnostic process of knee injuries. The purpose of this systematic review is to identify all recent studies that investigate the use of DL technology in the MRI-based diagnosis of the injured knee. It was decided that the primary focus would be on studies that examine at least one of the following pathologies: (i) injuries of knee ligaments, (ii) meniscus tears, and (iii) cartilage lesion. The main methodological features and findings of retrieved articles are presented in the clinical context, while future implications from the usage of AI in the knee injury domain are discussed.

2.1.2. Machine learning in a nutshell: Definitions and terminology

To enhance the understanding of the readers and for the sake of completeness, this section quickly presents the relevant terminology and definitions with respect to ML and DL algorithms used in the studies involved in the present review. ML is a branch of AI that focuses on the development of algorithms that automatically learn to make accurate predictions by relying on experience (data) rather than on hard-coded instructions.

Supervised ML systems (Figure 1) operate in two phases: the learning phase (training) and the testing one. In a traditional ML pipeline, a feature extraction/selection stage (also referred to as feature engineering) is firstly implemented to extract or identify the most informative features [20]. These features can be extracted from the input images employing various algorithms including grey-level co-occurrence matrix (GLCM), first- and second-order statistics and shape/edge features among others [35]. Next, a ML model is fit to the extracted features and the optimal model parameters are obtained. During the testing phase, the trained model is shown previously unseen samples (represented as images or features extracted from images) which are then classified. As opposed to traditional programming, where the rules are manually crafted by a programmer, a supervised ML algorithm automatically formulates rules from the data.

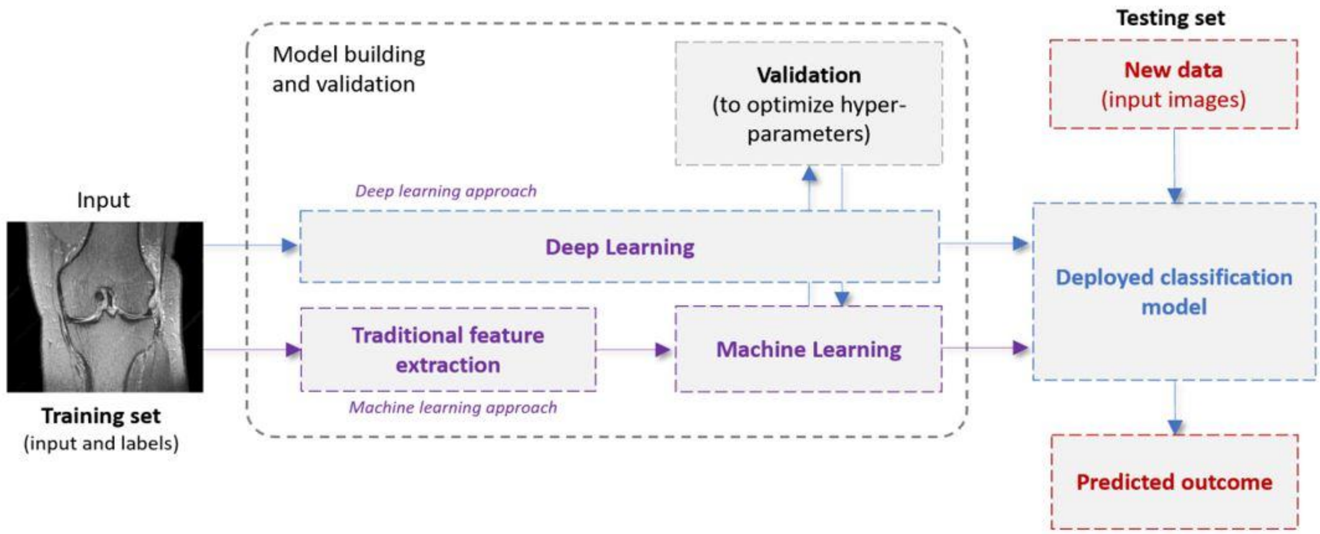


Figure 1: Examples of typical machine learning and deep learning pipelines.

DL [36] is a subfield of ML that sets an alternative architectural paradigm by shifting the process of extracting features from images to the underlying learning mechanism. The most informative features for the task at hand are extracted by the algorithm itself. The mainstream DL architecture for computer vision applications is the convolutional neural network (CNN). A CNN typically consists of multiple building blocks (layers such as convolutional, pooling, and fully connected) that automatically extract increasingly abstract spatial hierarchies of features. The CNN training is carried out via a backpropagation algorithm. The huge popularity of CNNs is attributed to certain characteristics they possess, such as the weight sharing and spatial invariance.

Transfer learning is a common strategy where a network, that was pre-trained on a big dataset, is partly re-used to provide decisions on a problem with a different dataset. The main idea behind transfer learning is that generic features learned on a large dataset could be useful and applicable to other domain tasks with potentially a limited amount of accessible data. Numerous pre-trained networks are currently available such as DenseNet **[37]**, AlexNet **[38]** and VGG **[39]**. When employing DL with transfer learning for feature extraction, the pre-trained network is treated as an arbitrary feature extractor: the input image propagates through multiple layers until a pre-specified layer, the outputs of which are considered as the finally extracted features. Table 1 provides a brief presentation of the main ML and DL algorithms that were reported in the papers of this review.

Category	**Models**	**Description**
Feature extraction	Histogram of oriented gradient (HOG) [40]	This is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image.
	Generalized search tree (GIST) [35]	GIST descriptor represents holistic spatial scene properties (spatial envelope) of an image. It summarizes gradient information on different spatial scales and orientations by splitting the image into a grid of cells on several scales and convolving each cell using a Gabor filter bank from different perspectives.
	Gray-level co-occurrence matrix (GLCM) [41]	GLCM is a way of extracting second-order statistical texture features. In particular, the texture of an image is estimated by calculating how often pairs of pixels with specific values and a certain spatial relationship occur.

Traditional Machine Learning	k-nearest neighbor (K-NN) [42]	KNN algorithm is a simple, easy-to-implement supervised ML algorithm that can be used to solve both classification and regression problems. It works by (i) finding the distances between a query and all the examples in the data, (ii) selecting the K nearest neighbors of the query, and (iii) voting for the most frequent label (in the case of classification) or averaging the labels (in the case of regression).
	Support vector machines (SVMs) [43]	SVMs is a supervised method that identifies a hyperplane that best divides the data into two classes. To separate the two clouds of data points, there are many possible hyperplanes that could be chosen. The objective of the SVM algorithm is to find a slab that has the maximum thickness, i.e., the maximum distance between data points of the different classes.
	Shallow artificial neural networks (ANNs) [44]	The ANN vaguely simulates the way the human brain analyzes and processes information. They consist of sequential layers: input, hidden and output layers. The hidden layer processes and transmits the input information to the output layer.
Deep Learning	Convolutional neural networks (CNNs) [45]	This is a class of DL algorithms commonly used in computer vision and pattern recognition. CNNs are a specific type of neural networks that are generally composed of the following layers: (i) input layer, (ii) convolution layers, (iii) pooling layers and (iv) fully connected layers. The convolution layers use filters that perform convolution operations as they are scanning the input with respect to its dimensions. Pooling is a down-sampling operation, which is typically applied after a convolution layer. The fully connected layers operate on a flattened input where each input is connected to all neurons in the next layer and are usually found towards the end of CNN architectures to optimize objectives such as class scores.
	Region based convolutional neural networks (R-CNNs) [46]	The method of detecting and classifying objects in an image is known as object detection. R-CNN (regions with convolutional neural networks) is a DL technique that blends rectangular area proposals with convolutional neural network functionality. The R-CNN algorithm is a two-stage detection method.
	Deep residual networks [47]	A residual neural network (ResNet) is an ANN variant that uses residual mapping and shortcut connections to tackle the problem of vanishing and exploding gradients that is characteristic of deep CNNs. As a consequence of this, deep residual networks achieve better performance when compared to plain very deep networks, whereas their training is easier as well. Typical ResNet models are implemented with double- or triple-layer skips that contain nonlinearities such as rectified linear unit (ReLUs) and batch normalization in between.
	3D-CNNs [48]	A 3D CNN is simply the 3D generalization of 2D CNNs. It takes as input a 3D volume or a sequence of 2D frames (e.g., slices in an MRI scan). Then kernels move through 3 dimensions of data producing 3D activation maps. Overall, they learn powerful representations of volumetric data.
	Computer Vision Transformers [49]	When data is modelized as a sequence of embeddings, the Transformer model is a basic yet scalable technique that can be used for any type of data. Even without typical convolutional pipelines, transformers can be utilized to provide SOTA results in Computer Vision. It is a DL network that extracts inherent properties of the interest domain via the self-attention technique.
Procedure	Training	The standard procedure involves a dataset of paired images and labels (x, y) for training and testing, an optimizer (e.g., stochastic gradient descent, Adam [43]), and a loss function to update the model parameters. The aim of the training is to find the optimal values for the network parameters so that the loss function is minimized.
	Data augmentation	Data augmentation is a strategy that artificially generates more training samples to increase the diversity of the training data. This can be done via applying affine transformations (e.g., rotation, scaling), flipping or cropping to original labeled samples.
	Dropout	Dropout is a regularization method that randomly drops some units from the neural network during training, encouraging the network to learn a sparse representation. It is used to reduce overfitting.

	Loss function	The metric to assess the discrepancy between model predictions and labels is called loss function. The gradients of the loss function are used to update the weights of the neural networks.
	Transfer learning	This aims to transfer knowledge from one task to another different but related target task. This is often achieved by reusing the weights of a pre-trained model, to initialize the weights in a new model for the target task. Transfer learning can help to decrease the training time and achieve lower generalization error.

Table 4: Brief presentation of the feature extraction techniques as well as the ml and dl models, and the main procedures that were reported in the papers of our survey.

2.2. Methods

2.2.1. Reporting

The present (systematic) review was performed in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines [50]. Each step of the review process (literature search, study selection, and data extraction) was independently performed by 2 authors (A.S., S.M.). Discrepancies were resolved by a 3rd author (D.T.). The present study was not registered in a database prior to its conduction.

2.2.2. Literature Search

A structured literature search was conducted in the following databases: (a) MEDLINE (through PubMed), (b) CENTRAL (through Cochrane Library), and (c) EMBASE (through Elsevier). Articles cited by the retrieved papers, as well as articles citing the retrieved papers (using Google Scholar), were also identified through a supplementary manual search. Grey literature was examined based on conference abstracts, English abstracts (from articles not published in English), and the OpenGrey database. The potential eligibility of the articles was initially decided based on their title and abstract. Full texts were investigated to verify whether the initial qualifiers fulfill the inclusion criteria. The structured search strategy per database is quoted in Table 16 (Appendix A).

2.2.3. Eligibility Criteria

The inclusion criteria were as follows: (i) papers were published between the 1st of January 2013 (the dawn of the DL era) and the 15th of November 2021 (date of final literature search)); (ii) MRI images were used for the evaluation of knee injuries; (iii) knee injury detection was conducted via AI-based algorithms, including both traditional ML and DL techniques; (iv) the performance of the AI-based algorithms was compared to clinical, human-based evaluations.

Papers were excluded according to the following criteria: (i) articles published before 2013; (ii) papers investigating OA or other bone pathology not directly linked with knee injuries; (iii) studies performed in animals; (iv) non-original research articles, such as protocols, reviews, meta-analysis, etc.; (v) articles not written in English (however, English abstracts were assessed as part of the grey literature); (vi) book chapters, editorials and commentaries.

2.2.4. Data Extraction

Extracted data were placed into a custom Microsoft Excel spreadsheet using a standardized table. The following information was included for each of the articles: first author, publication year, database, description of data and models, learning algorithm including pre-processing, size of training and test samples, validation method, and obtained results.

2.2.5. Statistical Analysis

Multiple evaluation criteria were employed to assess the predictive capacity of the proposed learning algorithms. The most common evaluation metric considered in this review study was accuracy, along with the receiver operator characteristic curves (ROCs) that visualize the performance of a classification model at various likelihood ratio thresholds. These curves plot two factors: true positive rate (sensitivity = TP/(TP+FN)) versus the false positive rate (specificity=FP/(FP+TN)), where TP, FN, FP and TN denote true positives, false negatives, false positives and true negatives, respectively. The quantitative output of this curve is the AUC which could be interpreted as an aggregated measure of performance across all possible classification thresholds.

2.2.6. Quality Assessment

Quality assessment was performed using a modified methodologic index for nonrandomized studies (MINORS) [51]. A seven-item checklist was considered, including information with respect to the following items: disclosure, study aim, input feature, ground truth label determination, dataset distribution, performance metric, and explanation of the applied AI models. Data were extracted and recorded using standardized forms (Microsoft Excel spreadsheet). To resolve conflicts over article selection, quality assessment, and data extraction, both observers (A.S., S.M.) convened a consensus meeting. The items were scored with 0 (not reported), 1 (reported but inadequate), or 2 (reported and adequate). The average modified MINORS score among all studies was 9.82 _ 1.99. It should be mentioned that the range of the score per item was between 0 and 44.

As shown in Figure 2, all the reported studies (22) clearly stated the study aim, input features, and the performance achieved using appropriate metrics. A clear distribution and description of the dataset were reported in twenty studies (90.09%). Fifteen studies (68.18%) clearly described how they established the ground truth (AI’s reference standards), whereas the others were subjected to AI models that were inadequately trained. The most prevalent causes of quality point loss were failures to describe ground truth assignments. Last, but not least, more than half of the studies (54.54%) failed to disclose a conflict of interest declaration.

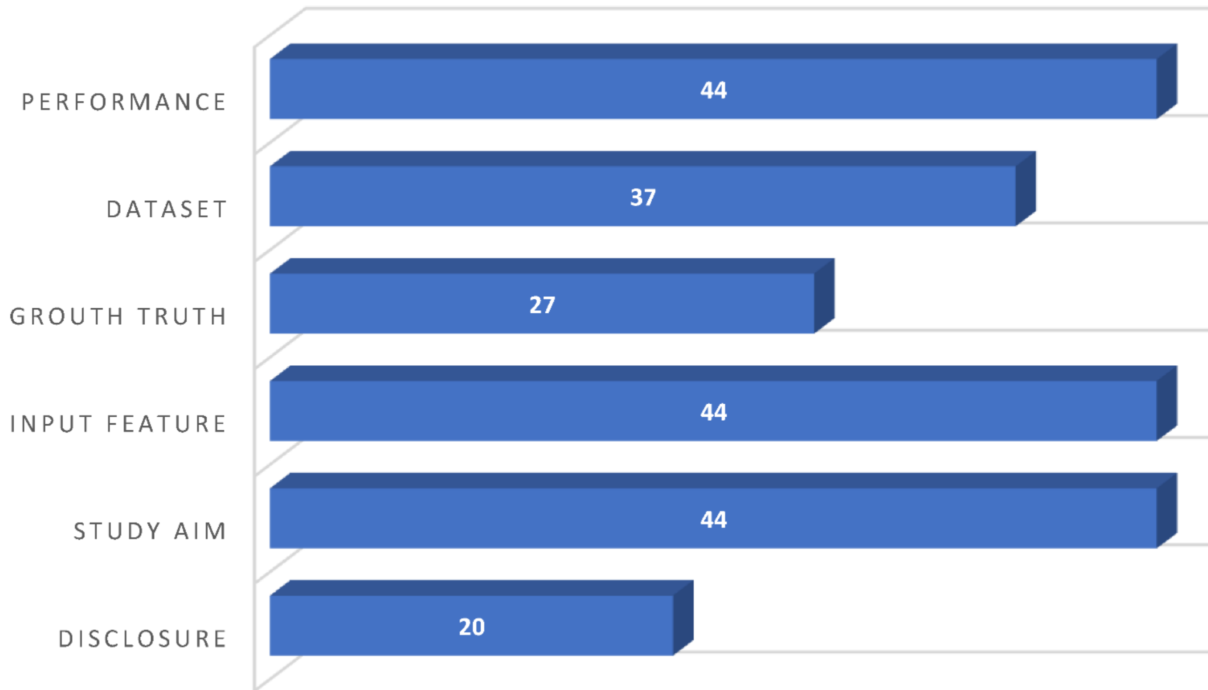


Figure 2. Quality assessment outcomes using the MINORS tool.

2.3. Results

In total, 407 studies were retrieved: 172 from MEDLINE (through PubMed), 170 from EMBASE (through ELSEVIER), 24 from CENTRAL, 40 from the structured search in Google Scholar, and 1 conference abstract (grey literature). Fifty-nine papers were selected after applying the proposed inclusion/exclusion criteria. Thirty-seven studies were further excluded due to irrelevant content (for example, those focusing only on segmentation or other scientific fields). Taking everything into consideration, 22 articles were finally included in the present systematic review. A flow chart of the literature search design is presented at Figure 3.

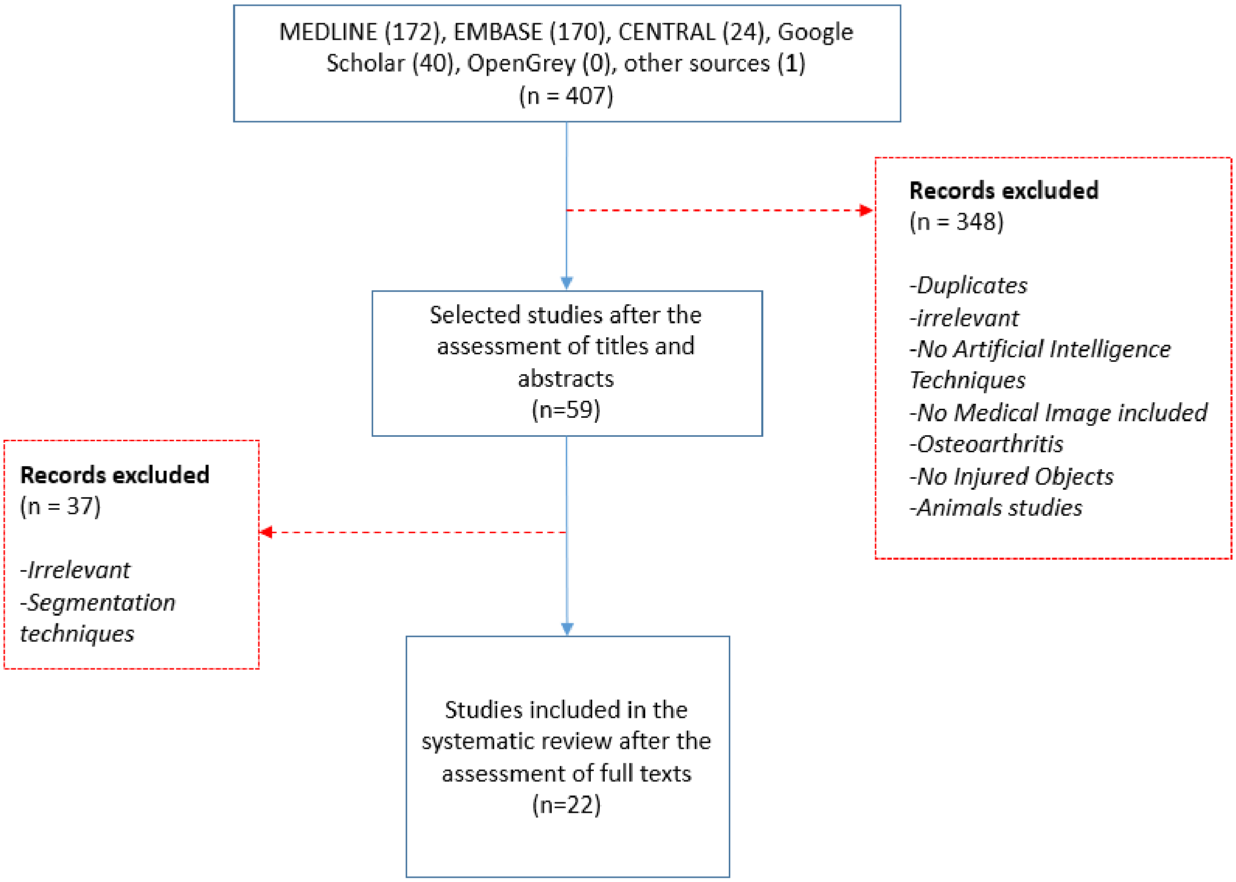


Figure 3: Flow chart presenting the design of the literature search.

The retrieved articles were categorized into the following application domains: (i) detection of ACL injuries alone (10 studies), (ii) detection of meniscus tears alone (7 studies), (iii) detection of cartilage lesions (1 study) and (iv) combined ACL and meniscus tears plus other knee injuries (4 studies). The main results of each study have been quoted, while the individual study validity has been determined based on its methodological strengths and weaknesses. Important methodological features of the retrieved articles have been commented.

The identified studies focusing on ACL and meniscus tears detection have been grouped into three categories: (i) those employing traditional ML pipelines, (ii) DL studies in which transfer learning is reported and (iii) papers that propose the use of custom-made DL architectures.

2.3.1. ACL injury detection

2.3.1.1. Machine Learning

Štajduhar and colleagues [52] utilized two different feature extraction techniques: histogram oriented gradient (HOG) [40] and generalized search tree (GIST) [35]. These feature extraction techniques were subsequently paired with two commonly used ML models: support vector machines (SVMs) [43] and random forests [53]. They found that the best performing ML model was the one that combined HOG with linear-kernel SVM, producing an AUC of 0.89 for differentiating between ACL injured and healthy subjects, and an area under curve (AUC) of 0.94 for detecting only completely ruptured ACL. Abdulah et al. [54] described a diagnostic system consisting of image pre-processing, feature extraction based on segment-derived spatial descriptors (perimeter, area, and shape), and finally classification. They compared k-nearest neighbor (K-NN) with back propagation artificial neural network (BP-ANN) for ACL tear classification. BP-ANN achieved a classification accuracy of 94.44%, whereas K-NN reached an accuracy of 87.33%. Another study [55] tested an SVM algorithm on a dataset that comprised 100 non-injured ACLs, 100 partially-torn ACLs, and 100 completely torn ACLs. All datasets underwent pre-processing. Features were extracted using shape descriptors such as objects' contour circularity, as-pect ratio, angle, and number of sides. It was reported that the SVM model had an accuracy of 100% for classifying ACL MRI samples as normal, partial-tear, or complete-tear. The authors also sought to compare the diagnostic capability of their AI model with that of two medical experts on a subset of 10 samples. No statistically significant differences be-tween the AI model and the radiologists were found.

2.3.1.2. Deep Learning with Transfer Learning

Bien et al. [32] used transfer learning in order to train a fully automated CNN for classifying MRI series and they combined predictions from 3 series per exam using logistic regression. The accuracy and the AUC of the model for detecting ACL tears were 86.7% and 0.965, respectively. These results were juxtaposed with the assessments by three musculoskeletal (MSK) radiologists on a testing set of 120 knee MR images. Radiologists achieved significantly higher sensitivities for tear diagnosis than the AI model (AUC: 0.91 vs. 0.76, p-value=0.002). The accuracy achieved by the radiologists (92%) was higher than the one achieved by the AI model (86.7%). Azcona et al. [56] proposed and evaluated the performance of four architectures: (i) deep residual network with transfer learning, (ii) custom deep residual network using fixed number of slices, (iii) multi-plane deep residual network, and (iv) multi-plane multi-objective deep residual network. They found that transfer learning combined with a carefully tuned data augmentation strategy were the crucial factors in achieving best performance. The authors modified the last layer to out-put a probability instead of a one-hot softmax vector for a number of classes and they also used transfer learning with pre-trained weights from ImageNet. By using the aforementioned DL architectures and data augmentation strategies for ACL detection, they achieved an AUC of 0.96 on the validation data.

2.3.1.3. Custom-made deep learning networks

Another study [13] evaluated three customized CNN models with variations in the input fields of view (i.e., full slice, cropped slice, dynamic patch-based sampling) as well as in dimensionality (single slice, three slices, five slices) for the detection of complete ACL tears. The importance of limiting the input field-of-view to the intercondylar region for high algorithm performance was demonstrated. The incremental value of contextual in-formation of adjacent image slices in improving network classification accuracy was also exhibited. The model that utilized dynamic sampling had an accuracy of 96.7% and AUC of 0.97. Liu et al. [57] trained multiple CNNs and applied them to a test set comprising 50 MR images of ACL tears with normal thickness and 50 MR images with intact ACLs. The best model they came up with for detecting the presence or absence of a full thickness ACL tear produced an AUC of 0.98 (95% CI: 0.93-1.00, p-value < 0.001). However, there was no statistically significant difference in diagnostic performance between the AI model (AUC: 0.98, 95% CI: 0.93- 1.00) and the clinical radiologist performance: Radiologist AUC: 0.98 (95% CI: 0.95-1.00); Fellow AUC: 0.98 (95% CI: 0.95- 1.00); Resident 1 AUC: 0.93 (95% CI: 0.88-0.98); Resident 2 0.97 (95% CI: 0.94-1.00); Resident 3 0.98 (95% CI: 0.95-1.00).

Namiri et al. [58] employed two CNN types for classification of ACL injuries: the first one involved three-dimensional (3D) kernels, whereas the second one made use of two dimensional (2D) filters. The overall accuracies using the 3D CNN and the 2D CNN were 89% (225 of 254) and 92% (233 of 254), respectively (p-value= .27), whereas both CNNs had a weighted Cohen k of 0.83. The 2D CNN and 3D CNN performed similarly in classifying intact ACLs (2D CNN: sensitivity of 93% and specificity of 90%; 3D CNN: sensitivity of 89% and specificity of 88%). Classification of full tears by both networks was also comparable (2D CNN: sensitivity of 82% and specificity of 94%; 3D CNN: sensitivity of 76% and specificity of 100%). The 2D CNN classified all reconstructed ACLs correctly. A separate study [11] proposed to perform CNN-based classification by relying on the architecture of 3D DenseNet [37]. They compared this DL approach with other two variants, namely VGG16 [37] and ResNet [47]. The accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the proposed customized architecture were calculated respectively. The average AUCs were 0.95, 0.86, 0.96 for ResNet, VGG16, and their proposed network, respectively. The diagnostic accuracies achieved by the proposed model, the residents, and the senior radiologists were 95.7%, 81.4%, and 89.9%, respectively.

Germann et al. [29] trained a deep convolutional neural network (DCNN) on 512 MR images of ACL tears from different patients (ACL tears were present in 45.7% and absent in 54.3% of the subjects). The network had a pre-processing step that involved selection, rescaling, and cropping of coronal and sagittal fluid-sensitive views. Next, the coronal and sagittal MRI scans were processed independently in parallel and were then concatenated before being processed by one dense layer fat-suppressed MRI scans. Finally, a soft-max layer extracted the confidence level for the ACL tear. Three fellowship-trained full-time academic MSK radiologists independently evaluated the MRI examinations for full-thickness ACL tears. ACL tears were present in 45.7% and absent in 54.3% of the subjects. The DCNN had a sensitivity of 96.1%, which was not significantly different from that of the readers (97.5%–97.9%; all p-values ≥ 0.118). However, the sensitivity of the DCNN (93.1%) was significantly lower than that of the readers (99.6%-100%, all p-values < 0.001), and a similar trend was observed in the AUC values (DCCN: 0.94, readers: 0.99-0.99, all p < 0.001). Finally, a related study [59] used a customized 14-layer Res-Net-based CNN with six different directions by using class balancing and data augmentation. The proposed ResNet-14 achieved AUC values of 0.98, 0.97, and 0.99 for detecting healthy tears, partial tear, and fully ruptured tear, respectively. Jeon et al. [60] proposed a 3D deep neural network model for diagnosing ACL tears from a knee MRI test that is both interpretable and lightweight. They used squeeze modules and fewer convolutional filters to represent the homogeneity of the features, as well as attention modules and Gaussian positional encoding to strengthen the searching of local features. Their model outperformed the prior SOTA on the Chiba and Stanford knee datasets, achieving average ROC and AUCs of 0.983 and 0.980, respectively. Recently, Dai et al. introduced TransMed [61] as a multi-modal medical picture categorization system. It combines the benefits of CNN and transformer to efficiently extract low-level characteristics from pictures and construct long-range relationships between modalities. The accuracy and the AUC of the model for detecting ACL tears were 94.9% and 0.98, respectively. These results were of higher accuracy than the MRNet technique. Astuto et al. [62] made use of 3D CNNs which were de-signed to identify and grade ACL injuries in MRI investigations. The reported binary le-sion sensitivity for ACL tissue is 88%. The specificity of the results is 89 %. The AUC is 0.90.

2.3.2. Meniscus Tear

2.3.2.1. Machine Learning

Fu et al. [63] compared the performance of two SVM models in detecting meniscus tears. One model was trained on selected MR features (from a pool of 180 spatial and tex-tural features using GLCM), while the other model implemented the SVM model without any feature selection. The SVM model without feature selection produced an AUC of 0.73, while their model with feature selection yielded an AUC value of 0.91. Zarandi et al. [64] performed MR image segmentation followed by the application of a perceptron neural network (PNN) for classifying meniscal tears. The model accomplished a 90% classifica-tion accuracy (meniscus tear versus no meniscus tear) on a testing dataset of 50 MRI stud-ies. Precision (%) was also reported for five different settings of meniscus tear including: (1) medial anterior horn and posterior horn normal (88.82%), (2) lateral anterior horn and posterior horn normal (92.13%), (3) medial anterior horn normal and posterior horn torn (84.24%), (4) lateral anterior horn normal and posterior horn torn (91.96%) and (5) lateral anterior horn torn and posterior horn normal (87.64%).

2.3.2.2. Deep Learning with Transfer Learning

Another group [32] utilized MRNet as the primary building block of their prediction system, that is a CNN mapping a 3D MRI series to a probability. The input to MRNet had dimensions: s × 3 × 256 × 256, where s was the number of images in the MRI series (3 is the number of color channels). In diagnosing meniscus tear, this group reported an accuracy of 72.5% (95% CI: 0.639-0.797) and an AUC of 0.85 (95% CI: 0.78-0.91). Furthermore, they compared the performance of the proposed model with unassisted MSK radiologists for detecting meniscus tear (intact, degenerative changes without tear, postsurgical changes without tear). When compared to the MSK radiologists in the study, the AI model had a statistically significant lower specificity (AUC: 0.88, 95% CI: 0.85-0.91 versus AUC: 0.741, 95% CI: 0.62-0.84; p-value=0.003) and accuracy (0.85, 95% CI: 0.82-0.87 versus 0.725, 95% CI: 0.64-0.80, p-value=0.015). The sensitivity was also shown to be lower for the AI model (0.82, 95% CI: 0.78-0.85) compared to MSK radiologists (0.71, 95% CI: 0.59-0.81; p-value = 0.504), although this was not statistically significant. Azcona and colleagues [56] lever-aged the baseline MRNet architecture and replaced the AlexNet feature extractor with more modern residual architectures such as Resnet18, Resnet50 and Resnet152. They applied a series of transformations including horizontal flips and photometric augmenta-tions (with respect to random contrast, gamma, and brightness). They reported an AUC performance of 0.91 on the validation data by using ResNet18.

2.3.2.3. Custom-made deep learning networks

Couteaux et al. [65] used a region-based convolutional neural network (R-CNN) model for tear detection and localization (anterior or posterior). The anterior meniscus was classified as torn when at least one network had detected a torn anterior meniscus and the posterior meniscus was classified as torn when the strict majority of the networks had detected a torn posterior meniscus. A weighted AUC score of 0.91 was achieved by the proposed network on a test set of 700 MRIs. Another paper [66] also used an R-CNN trained on a dataset of 700 MRI images to perform three tasks, namely detection of meniscus tear presence, position, and orientation. Their AI model produced an AUC of 0.94 on the task of detecting the presence of a meniscal tear, 0.92 for detecting the position of the two meniscal horns, and 0.83 for detecting the orientation of the tear. The overall combined AUC was 0.90.

Another group [67] created a DL model that combined meniscus segmentation and a 3D CNN for accomplishing both detection and severity staging of meniscus lesions. The segmentation task for both cartilage and meniscus was implemented using 2D U-Net [68]. The model was first built to recognize the presence of a lesion (including intrasubstance abnormalities), and subsequently to quantify the lesion severity. This model produced a lesion detection AUC performance of 0.89 on the test dataset and accuracies of 80.74%, 78.02%, and 75.00% for determining severe, mild-moderate and no lesions, respectively. Comparisons were made between the model and experts. The authors also sought to de-termine the inter-rater variability between three MSK radiologists (expert 1: >20 years of experience, expert 2: 10 years of experience and expert 3: <1 year of experience) for as-sessing meniscus lesion severity on selected cases. They restored an average agreement among the three experts of 86.27% for no meniscus lesions, 66.48% for mild-moderate le-sions, and 74.66% for severe lesions, while the best model obtained accuracies of 80.74% for no meniscus lesions, 78.02% for mild-moderate lesions, and 75.00% for severe lesions.

Fritz et al. [29] proposed that deep CNN-based meniscus tear detection be performed in a fully automated manner with a similar specificity but a lower sensitivity in comparison with the MSK radiologists. The AUC of the deep CNN employed was 0.88, 0.78, and 0.96 for the detection of medial, lateral, and overall meniscus tear, respectively. The sensitivity, specificity, and accuracy for medial meniscus tear detection were 93%, 91%, and 92%, respectively, for readers 1; 96%, 86%, and 92%, respectively, for reader 2; and 84%, 88%, and 86%, respectively, for the DCNN. The sensitivity, specificity, and accuracy for lateral meniscus tear detection were 71%, 95%, and 89%, respectively, for reader 1; 67%, 99%, and 91%, respectively, for reader 2; and 58%, 92%, and 84%, respectively, for the DCNN. The sensitivity for medial meniscus tears was significantly different between reader 2 and the DCNN (p-value = 0.039), but no significant differences were witnessed in all other comparisons (all p-value ≥ 0.092). Rizk et al. [69] used a 3D CNN architecture that incorporated meniscal localization and lesion classification. They achieved AUC values of 0.93 and 0.84 for medial and lateral meniscal tear detection, respectively, and 0.91 and 0.95 for medial and lateral meniscal tear migration detection, respectively. The combined medial and lateral meniscal tear detection models were externally validated and yielded an AUC of 0.83 without additional training and 0.89 after fine-tuning. Moreover, Dai et al. utilized TransMed [61] achieving accuracy and AUC 94.9% and 0.98, re-spectively, for detecting meniscus tears, improving over the MRNet technique. 3D CNNs were built by Astuto et al. [62] to identify and grade meniscus tear in MRI examinations. The reported binary lesion sensitivity and specificity values were 85% for both, whereas the AUC was 0.93. Lastly, Dai et al. used TransMed to also identify meniscus tears in the MRNet dataset. The group reported an AUC of 0.95 and an accuracy of 85.3%.

2.3.3. Cartilage Lesion and other abnormalities

Liu et al. [57] developed a fully automated DL–based cartilage lesion detection system by combining CNN-based semantic segmentation and disease classification. Segmentation was implemented via the use of a VGG-16-based encoder network consisting of a combination of 2D convolution layers, rectified-linear activations, batch normalization layers, and max-pooling layers to achieve image feature extraction and data compression at the same time. The classification CNN in the proposed pipeline was also based on the 2D VGG16. Their pipeline achieved AUC in the range of 0.91-0.92, indicating high overall diagnostic accuracy for detecting cartilage lesions. Also, there was good intra-observer agreement between two individual evaluations, with a k-statistic of 0.76. As previously indicated, Astuto et al. [55] used 3D CNNs to also detect cartilage lesions. The sensitivity and specificity of binary lesions were found to be 85% and 89%, respectively, whereas the AUC was 0.93. Finally, three of the reported papers [32], [56], [61] attempted to detect other knee abnormalities such as osteoarthritis, effusion, iliotibial band syndrome, posterior cruciate ligament tear, fracture, contusion, plica, and medial collateral ligament sprain. MRNet [32], a ResNet18 [70], and TransMed network were employed to implement the classification tasks achieving AUC values of 0.94, 0.94, and 0.976, respectively.

No.	Author	Year	AI Model used	Pretrained CNN	MRI(T)	Localization technique	Validation	Performance (Accuracy/AUC)	Application domain
1	Awan et al. [59]	2021	CNN	ResNet-14	1.5T	They applied normal approach to localize based upon region of interest (ROI)	5-fold cross-validation	92%/ (healthy tear=0.98, partial tear=0.97 and fully ruptured tear=0.99)	ACL tear
2	Jeon et al. [60]	2021	3D CNN	VGGNet, AlexNet, and SqueezeNet	3T & 1.5T	Custom localization technique	5-fold cross-validation	N/A / 0.983 and 0.980 on the Chiba and Stanford knee datasets, respectively	ACL tear
3	Rizk et al. [69]	2021	3D CNN	CNN-based localization model	1T (54%) – 1.5T(9.7%) – 3T (36.3%)	Custom localization technique	ten-fold cross validation	Meidal = N/A/ 0.93, Lateral = N/A/0.84	Meniscus tear
4	Dai et al [61]	2021	TransMed	N/A	3T & 1.5T	N/A	120 exams	ACL tear = 94.9%/0.98, Abnormality = 91.8% / 0.976, Meniscus tear = 85.3%/0.95	ACL tear - Meniscus tear - Abnormalities
5	Astuto et al. [62]	2021	3D CNN	N/A	3T	V-Net	Hold out (15% of sample)	N/A / from 0.83 to 0.93	ACL tear - Meniscus tear – Cartilage Lession
6	Fritz et al. [29]	2020	DCNN	N/A	1.5T (64%) – 3T (36%)	To visually localize the tear, the software computes the class activation map (CAM) of the last convolution layer in the CNN and maps it to an axial knee image	Hold out (10% of sample)	Medial= (86%/0.88), Lateral= (84%/0.78), Overall= (N/A/0.96)	Meniscus tear
7	Namiri et al. [58]	2020	CNN	N/A	3T	three-dimensional V-Net	Hold out (10% of sample)	3D-model= (89%/sensitivity of 89% and specificity of 88%), 2D-model= (92%/sensitivity of 93% and specificity of 90%)	ACL tear
8	Zhang et al. [11]	2020	CNN	3D DenseNet, VGG16, ResNet	1.5T (74%) – 3T (26%)	-	Hold out (20% of sample)	Custom= (95.7%/0.96), ResNet= (NA/0.95), VGG16= (NA/0.86)	ACL tear
9	Germann et al. [29]	2020	DCNN	N/A	1.5T – 3T	They cropped manually	Out of the 5802 MRI studies, 4802 were used for training, 500 for validation, and 500 for initial testing	N/A/0.94	ACL tear
10	Azcona et al. [56]	2020	CNN	MRNet, ResNet18, Resnet50 and	3T (56.6%) – 1.5t (43.4%)	-	N/A	NA/0.96 – N/A/0.91 – N/A/0.94	ACL tear - Meniscus tear - Abnormalities

				ResNet152, ImageNet					
11	Chang et al. [13]	2019	CNN	ResNet	1.5T – 3T	The object localization CNN was implemented as a fully convolutional network based on U-net architecture	5-fold-cross-validation	96.7%/0.97	ACL tear
12	Liu et al. [71]	2019	CNN	LeNet-5, DenseNet, VGG16, AlexNet	N/A	They used object detection technique YOLO	50 subjects test set (14% of the sample)	N/A/0.98	ACL tear
13	Couteaux et al. [65]	2019	CNN	ResNet-101, ConvNet,R-CNN	N/A	To localize both menisci and identify tears in each meniscus, they used the Mask R-CNN framework	54 cases and the model with the highest validation accuracy was selected	N/A/0.90	Meniscus tear
14	Pedoia et al. [67]	2019	2d U-Net, CNN	N/A	3T	-	Hold out (20% of sample)	Sensitivity of 89.81% and specificity of 81.98%	Meniscus tear
15	Roblot et al. [66]	2019	CNN	AlexNet, MRNet	N/A	They used object detection technique Fast RCNN & Faster RCNN	The algorithm was thus used on a test dataset composed of 700 images for external validation	72.5%/0.85	Meniscus tear
16	Nicholas Bien et al. [32]	2018	CNN	AlexNET, MRNet	3T (56.6%) – 1.5T (43.4%)	-	120 exams	86.7%/0.97 – 72.5%/0.85 – N/A/0.94	ACL tear - Meniscus tear - Abnormalities
17	Liu et al. [57]	2018	CNN	VGG16	3T	-	fellowship trained musculoskeletal radiologist (R.K., with 15 years of clinical experience)	N/A/0.92	Cartilage lesion
18	Stajduhar et al. [52]	2017	HOG+linSVM,HOG+RF,GIST+rbfSVM,GIST+RF	N/A	1.5T	Manual extraction of a rectangular ROI	10- fold cross validation	(Injury detection problem, complete rupture) = (N/A/0.89, N/A/0.94), (N/A/0.88,N/A/0.94), (N/A/0.889, N/A/0.91), (N/A/0.88,N/A/0.90) respectively with the models	ACL tear
19	Mazlan et al. [55]	2017	SVM	N/A	N/A	They use cropping technique	Hold out (10% of sample)	100%/N/A	ACL tear
20	Zarandi et al. [64]	2016	IT2FCM, PNN	N/A	N/A	-	Hold out (20% of sample)	0 and 1 mode: 90%/N/A Binary mode: 78%/N/A	Meniscus tear
21	Fu et al. [63]	2013	SVM	N/A	N/A	Active Contours without Edges method. This method combines Active Contours with Level Sets and is called ACLS	5-Fold cross validation	SVM model: N/A/0.73 SFFS+SVM: N/A/0.91	Meniscus tear

22	Abdullah et al. [54]	2013	BP ANN, K-NN	N/A	N/A	-	5-fold and 6-fold	BP ANN: k-NN: 87.83%/N/A	94.44%/N/A	ACL tear

Table 5: Results of studies.

2.4. Discussion

The present literature review (Table 5) outlined the recent application of traditional ML and DL models to the diagnosis of the most common knee injuries using MRI as the main data source. Figure 4 shows an increasing trend in adopting ML based studies in this application area with most of the papers being published from 2017 onwards (whilst the first ML-based paper on the field was published in 2013). Medical imaging and specifically MRI has to be seen as one of the most instructive assets in the field of knee injuries diagnosis. The proliferation of MRI data has facilitated the effective training of ML and DL networks towards the development of: (i) novel methodologies that could enhance the medical experts' domain knowledge and understanding of MRI and (ii) new data-driven tools that could enable a more reliable, fast, and fully automated detection of knee injuries.

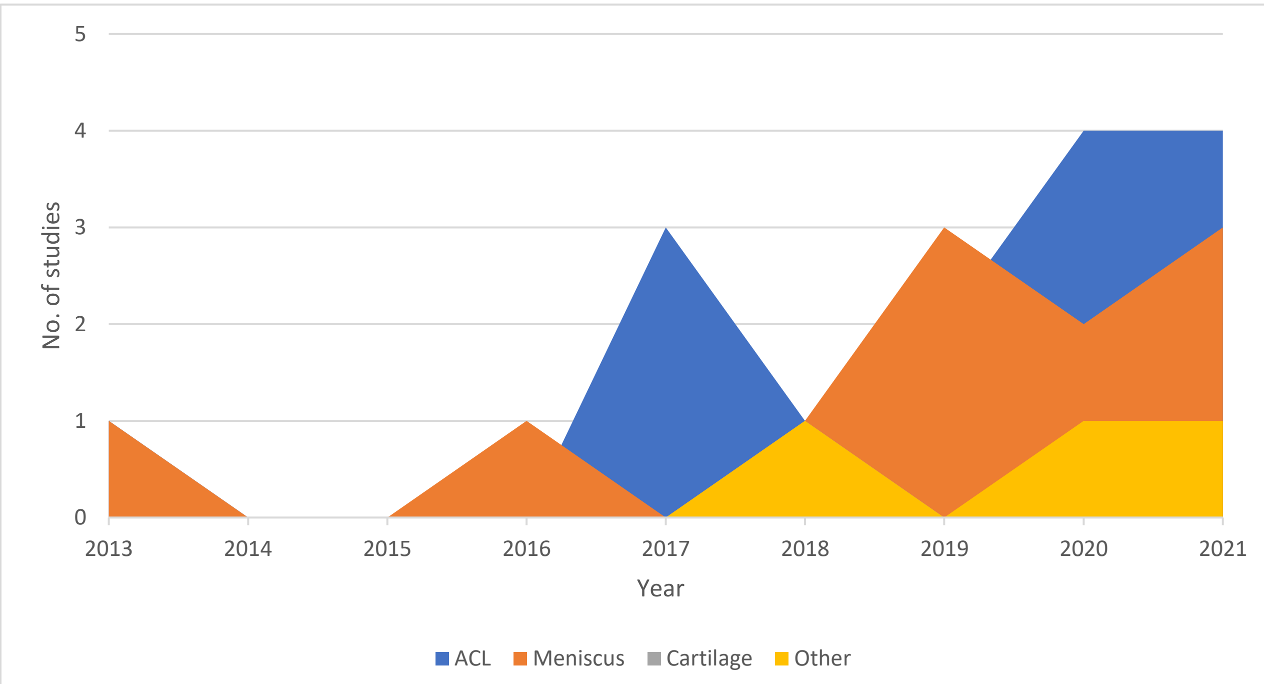


Figure 4: Temporal evolution chart depicting the number of ML papers per category published each year since 2013.

Although there is no clear acceptance of a "gold-standard" methodological pipeline for diagnosing knee abnormalities using MRI data, it was observed that a number of processing steps were commonly employed in the majority of the reported studies. Figure 5 visualizes a DL pipeline that was adopted by most of the papers including a pre-processing step, localization (optionally) by identifying regions of interest and finally a CNN-based classification step. Data augmentation was employed by a significant number of papers in the detection of ACL injuries [11], [32], [56], [58]–[62] also in papers where meniscus injuries were investigated [32], [56], [61], [62], [66], [67] and finally in studies focusing on cartilage lesion abnormalities [32], [56]. In particular, the available MRI images were modified (via a number of image transformations such as random rotations, shifting, flipping, and addition of noise) to expand the training dataset and thus help to improve the performance and ability of the employed DL models to generalize. Localization was employed in papers from all three subcategories: (i) ACL studies [11], [13], [29], [52], [58]–[60], [62], [71] (ii) meniscus injuries detection studies [13], [29], [62]–[67], [69] and for diagnosing lesion abnormalities [67]. Segmentation or objection detection algorithms were applied in the aforementioned studies to extract areas of interest enabling the application of CNN-based models on focused and more relevant parts of the initially available images. Given that the region of interest (ROI) may appear in slightly different positions within an image and may have different aspect ratios or sizes, identifying ROIs with an automatic manner has been proven to be a crucial processing step.

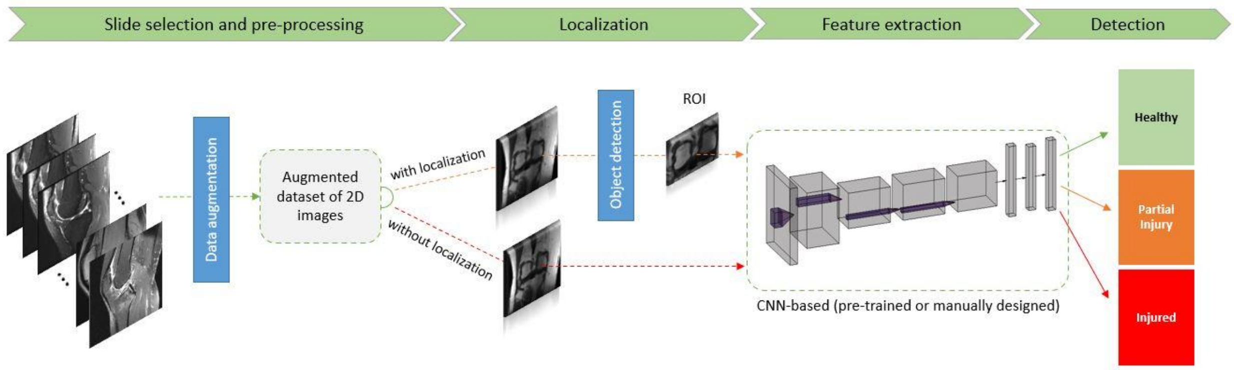


Figure 5: A typical DL pipeline for ACL detection.

CNN-inspired networks were identified as the dominant approach in the task of extracting informative features from either ROIs or entire MRIs and finally classifying them as normal (healthy) or abnormal (indicating either partial or complete tears). Transfer learning was preferred in most of the cases allowing the training of big and powerful deep architectures, even if the amount of available data was limited. As networks require a lot of information to be trained from scratch, this technique essentially 'steals' knowledge from already pre-trained large networks. Specifically, ResNet variants were used in five papers [11], [13], [56], [59], [65] of this review, whereas VGG [39], AlexNet [38], and MRNet [32] were used three times [11], [32], [56], [57], [66], [71]. Other pre-trained networks that were used at least once in this survey are: DenseNet [37], Le-Net [72], ImageNet [38], and R-CNN [46]. In five [52], [54], [55], [64], [73] out of the 22 studies of the present survey, more traditional ML pipelines were applied including a separate feature engineering step (where features were manually extracted from images). SVM classification was the preferred classifier in most of the cases.

Despite the excellent capability of CNNs to come up with valuable image representations, these models lack the capacity for capturing long-range relationships. To deal with this limitation, recent research studies [49], [74] have proposed to employ Transform-er-based architectures for various image recognition tasks. The Transformer [75] is a neural network architecture which relies on global self-attention mechanisms, and it was initially designed for sequence-to-sequence prediction. Papers that used this architectural paradigm have indeed achieved state-of the-art results [76], [77] in many natural language processing (NLP) tasks. Dai et al. [61] were the first to employ a Transformer-based archi-tecture for the MRI-based knee injury detection task. In particular, their hybrid (Trans-former and CNN) model was used to extract features that pick up the long-range depend-encies between MRI and other modalities.

The present review study demonstrated that the prediction accuracy of the DL models for the ACL and meniscus tears detection ranged from 72.5% to 100%. However, certain limitations have been identified in all studies that are included in this literature review. The lack of multi-center data has been recognized as a limitation in three papers [21], [32], [71] leading to the development of biased DL detection systems which have only been tested on knee MRIs carried out at a single institution. The results of these studies relate to knee examinations using specific MR acquisition protocols for knee joint assessment. In general, classification models, trained using data acquired by a specific MRI acquisition protocol, are unsuccessful or underperform when applied to data that was obtained differently. One way to tackle this lack of ability to generalize is by using DL models that learn MRI acquisition-invariant contrast-agnostic representations [78], [79]. The effect of data imbalance has been also highlighted in some cases [11], [58], [59] where the sample of patients was not properly balanced among all gradings leading the algorithms to pay more attention to the majority class (typically the class of healthy subjects). Applying down-sampling in the majority class has been proved to be an unreliable approach which led to a biased result in the case of the fully ruptured classes [59]. Verification bias was also identified [29], [71] mainly because subjects involved in the studies underwent arthroscopic knee surgery leading to increased sensitivity and decreased specificity for both the detection system and the clinical radiologists. Moreover, it should be stressed out that the grades used for the training of the detection algorithms are typically dependent on subjective assessment by a limited number of radiologists (one in some cases [58]). In most of the studies, only two categories (normal versus tears) were discriminated and the need of considering additional categories was highlighted [11] to allow more detailed classifications to happen. Overall, it was stressed that the diagnostic performance of the combined use of clinical radiologist and machine interpretation of the MRI examinations has not been evaluated [71].

Future studies should try to train and test the accuracy of AI prediction models for the detection of ACL and meniscal lesion based on the arthroscopic images and compare the outcome with that of direct non-arthroscopic assessments. Arthroscopy is a surrogate "gold-standard" for the validation of non-invasive assessments such as MRI as it provides highly magnified and direct views of articular cartilage with non-destructive interactive assessments of its structure and functional properties.

Radiological imaging data of the knee continues to grow at a disproportionate rate vastly outnumbering the trained MSK radiologists. The workload has also increased dramatically leading to inevitable errors in the decision-making process. Despite the identified limitations, AI systems have potential to relieve physician burnout, utilize clinicians in fields at which they have not been specialized (MSK MRI) and reduce the cost of knee injury diagnosis for the public health system. In addition to flagging abnormal cases, if an AI algorithm could rapidly identify negative exams (increased sensitivity and negative predictive value), then, a substantial amount of time and other resources could be made free. Such a concept would be really useful in countries without easy access to medical expertise.

Advances in medical imaging, in terms of quality, sensitivity and resolution have enabled the discrimination between the smallest differences in the various knee tissue densities. These differences sometimes are difficult to be recognized even by a trained, specialized eye. Expert's diagnostic capacity used to be superior but now we see this has been balanced out. As it was recently reported [29], deep CNN performance has reached performance levels like fellowship-trained full-time academic MSK radiologists in several tasks including detection and segmentation. Despite this, AI can provide several new tools to the field of radiology imaging of the knee and medicine in general. The major hope for automated intelligent systems in the knee injury diagnosis is to increase accuracy, efficiency and productivity in order to streamline patient care and outcomes. Newest, high-performance DL models should surpass the performance of traditional systems, meet the requirements for clinical utility and become more user-friendly for the MSK clinician. Furthermore, there is the possibility of better training of young MSK radiologists with the help of AI.

MRI data of the knee complemented by massive amounts of associated multi-dimensional data such as omics and electronic health records are only expected to grow. To fully exploit the full potential of this wealth of data, new paradigms should arise involving processes and workflows suitable for multi-institutional collaboration. Moreover, addressing the need for trustworthy detection systems of knee injuries, a medical diagnosis algorithm should meet a number of requirements (e.g. transparency, interpretability, explainability, and easiness of use) in order to gain trust from clinicians. AI explainability and lightweight DL are key enablers for the wide use of such systems in the everyday clinical practice. Exploiting the intersection and merits of traditional ML and DL methods, AI analytics are expected to revolutionize knee medical informatics enabling informed and accurate diagnoses needed by precision medicine.

Notwithstanding the huge potential of AI to improve the medical domain, the DL-based methods have yet to achieve significant deployment in clinical environments. This mainly ensues as a result of: (i) the intrinsic black-box nature of the DL algorithms, and (ii) the high computational cost. Explainable AI aims at building trust in the AI algorithms by providing medical experts with a diagnostic rationale behind the AI decision processes. The goal of the lightweight DL field is to develop models that have shallower architecture and are also faster and more data-efficient, while retaining the high-performance standards. Jeon et al. [60] were the first to get to grips with the clinical deployment of the MRI-based knee injury diagnosis. To this end, they proposed to use post-inference visualisation tools (such as CAM and Grad-CAM), and they also incorporated attention modules, Gaussian positional encoding, squeeze modules, and fewer convolutional filters.

2.5. Understanding YOLO and YOLOv5

See in Appendix B

2.6. Understanding MiniRocket: A Very Fast Transform for Time Series Classification

See in Appendix C

Chapter 3: Automated Recognition of Healthy Anterior Cruciate Ligament in Sagittal MR Images Using Lightweight Deep Learning

Published as:

Siouras, A., Moustakidis, S., Giannakidis, A., Chalatsis, G., Malizos, K. N., Hantes, M., ... & Tsaopoulos, D. (2022, July). Automated recognition of healthy anterior cruciate ligament in sagittal MR images using lightweight deep learning. In 2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA) (pp. 1-8). IEEE.

3.1. Introduction

ACL ruptures account for more than half of all knee injuries, impacting million people all over the world each year [7]–[12], [80]. For instance, in the United States, ACL injuries alone cost more than $7 billion in spending. In the case of young and athletic people, the more time they spend conducting professional and/or leisure activities, the greater their risk of knee injuries, which adds to a higher risk of developing Knee osteoarthritis (OA) [19]. Hence, the early, cost-effective, and reliable diagnosis of ACL injuries could help reduce health-care system discharge.

Arthroscopy is the “gold standard” for diagnosing intra-articular knee diseases, although it is restricted by its invasive nature and associated consequences [24]. In contrast with the aforementioned approach, the most extensively utilized non-invasive imaging approach for assessing ACL injuries is magnetic resonance imaging (MRI) [25], [81]. The accurate interpretation of the results obtained from this noninvasive procedure requires the use of trained and experienced clinicians. Because the human aspect plays such a big role in ACL diagnosis, differing perspectives on the diagnosis are frequently noted [27], [28]. As a result, techniques for assessing ACL injury must be developed that are both subjective and simple to use.

Recently, machine learning approaches have shown potential for are meeting this need. DL algorithms are particularly suited for modeling the complex relationships between medical images and their interpretations because they can automatically learn complicated nonlinear. Object detection (OD) is an area of computer vision [82], that refers to the process of both classifying and locating objects in an image or video. A systematic assessment of recent literature has yielded numerous research studies aimed at detecting ACL injury [83]. In particular, DL approaches have been found to reach even higher rates of classification performance than humans. Awan et al. [59] employed class balance and data augmentation to create a customized 14-layer ResNet-based CNN [84] with six different paths. For detecting a healthy tear, partial tear, and totally ruptured tear, the suggested ResNet-14 attained AUC values of 0.98, 0.97, and 0.99, respectively. Jeon et al. [60] proposed a 3D deep-neural network model for

diagnosing ACL tears from a knee MRI test. Their model is both interpretable and lightweight. To reflect the homogeneity of the features, they used squeeze modules and fewer convolutional filters, as well as attention modules and Gaussian positional encoding to improve the search for local features. On the Chiba and Stanford knee [32] datasets, their model beat the previous state-of-the-art (SOTA), with average ROC and AUC values of 0.983 and 0.980, respectively. TransMed was developed by Dai et al. [61] as a multi-modal medical image classification system. It combines the advantages of CNN and transformer architectures to extract low-level features from images quickly and build long-range correlations between modalities. In MRI examinations, Astuto et al. [62] used 3D CNNs, which were developed to diagnose and grade ACL damage. For ACL tissue, the reported binary lesion sensitivity was 88%. The results had an 89% specificity. The AUC for this study was 0.90.

Despite the enormous potential of DL to enhance the health field, the related approaches have yet to see widespread use in clinical practices. This is mostly due to the DL algorithms' high computational cost. The lightweight DL field's objective is to create models with a simpler architecture, as well as models that are lighter and more data-efficient while maintaining high performance criteria. Jeon et al. [60] was the first to grasp how to use an MRI-based knee injury diagnosis in a clinical setting. They suggested using attention modules, Gaussian positional encoding, squeeze modules, and less convolutional filters, as well as using post-inference visualization tools (such as CAM and Grad-CAM) [85].

This paper focuses on the first-ever full exploration of the latest OD networks in the task of Region of interest (ROI) detection, for ACL using only information from the sagittal plane. The objective of the current study is to provide a robust and reliable tool for the selection of the most informative MRI slices which could be used for further analysis from clinicians. To reach our goals, we have designed, implemented, and tested an advanced pipeline that employs the latest OD network~~s~~ (YOLOv5 nano). The performance of our pipeline is compared with those of three other state-of -the-art methods.

The rest of the paper is organised as follows. In section II, we outline the dataset used and the proposed methodologies for object detection. In Section II, I the results are presented and discussed, while in section IV the conclusions are provided.

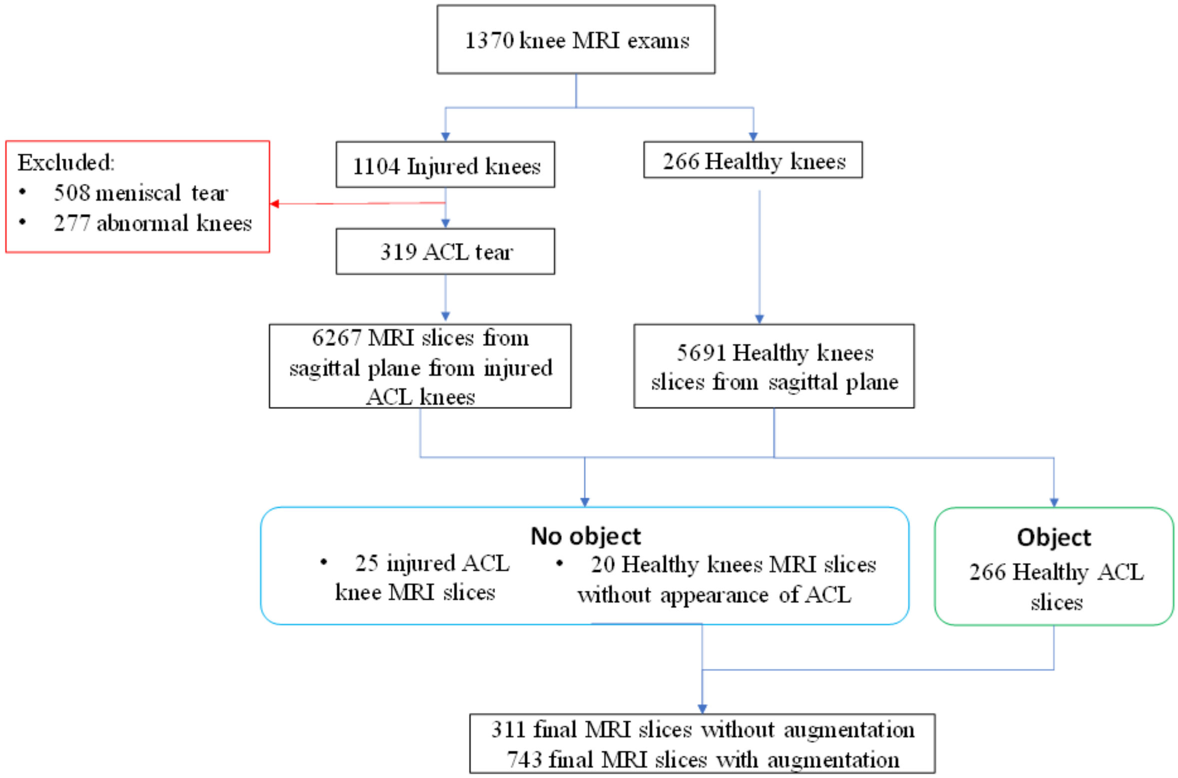


Figure 6: Study design

3.2. Methods

The most important aspect of any medical image analysis pipeline is the collection of a sufficient number of useful and representative images. Insufficient data collection can lead to inadequate object detection network training, which results in low accuracy or lack of generalization. To address the aforementioned issues, data augmentation and transfer learning approaches were used. Our object detection pipelines include: (i) data pre-processing and annotation, (ii) data augmentation, and (iii) learning models with pre-trained weights.

3.2.1. Dataset Description

The MRNet Dataset consists of 1,370 knee MRI tests performed at the Stanford University Medical Center. In this study, a subset of this database was employed. We only considered knees that were entirely healthy (n=266) for training and testing. A sample (n=319) of knees with ACL rupture was also investigated to evaluate the capacity of our methodology on injured knees. Only the sagittal plane was used in this study. The number of slices in each 3D MRI dataset ranges from 17 to 61.

3.2.2. Data preprocessing and Annotation

The effort of organizing, preparing, and annotating the training data has an impact on the accuracy of the deployed OD model, hence data preprocessing is crucial. All healthy and injured sagittal plane slices were initially converted to jpeg image format, leading to the generation of 5691 and 6267 images, respectively. A healthy subset of the MRNet examinations was manually annotated by a senior orthopedic/radiologist (GH) to identify healthy ACL slices on 266 MRI examinations at the sagittal plane and then the boundary boxes were annotated by the ML engineer (AS) in a specialized tool. To execute data annotation in the healthy ACL slices, the Roboflow [86] (Roboflow, n.d.) labeling graphic user interface (GUI) was employed. The images were splitted into two categories: (i) images with a healthy ACL as an object, and (ii) some extra slices were included that are either neighboring to slices with an ACL (but no ACL) or slices with an injured ACL as a no-object (Figure 6) . When there are no objects in an image, no bounding boxes must be recorded. This isn't always an issue; in fact, it could be desirable in order to teach a model that things aren't constantly in slice. These additions were specified (not as a new class) so that the model recognizes that this slice contains no objects. As a result, we placed no object in these aforementioned slices to boost our model shows the proposed study design that led to the selection of 266 slices of healthy ACL knees, and another 45 slices which are either ACL damaged or healthy but without the ACL.

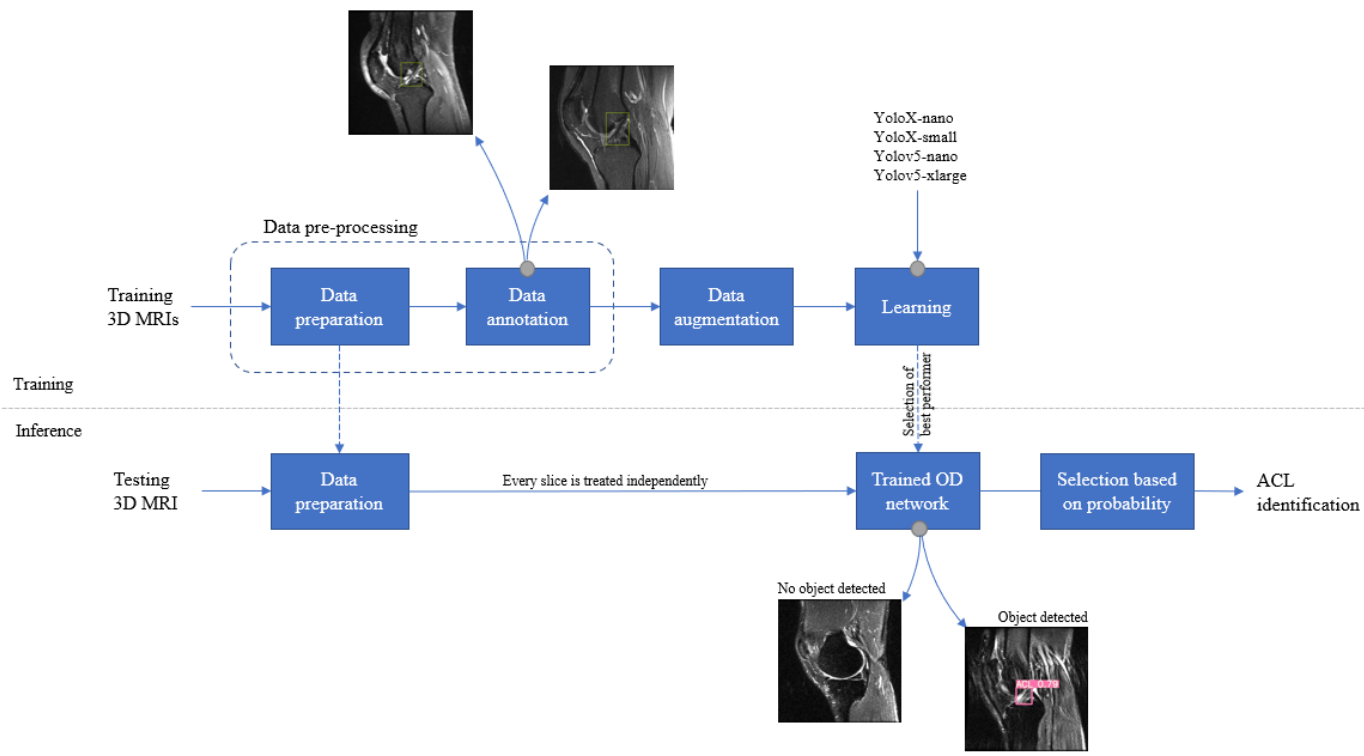


Figure 7: Proposed methodology for training and inference

3.2.3. Data Augmentation

Image augmentation is commonly used to increase the object detection network's generalizability. Data augmentation allows to increase the variety of the dataset that is eventually available for training without having to collect additional data. Image enhancement and transformation methods such as image brightness (between -32% and +32%), hue (between -39° and +39°), exposure (between -30% and +30%), blurring (up to 1.75px), and random noise (up to

7% of pixels) addition were used to multiply the obtained dataset. Exporting the annotations in the suitable format is the final step before OD training. After the data augmentation, the images produced are 743 (Figure 3). Each image is 256 by 256 pixels in size.

3.2.4. Learning Model

Our proposed method (Figure 7) is based on the most recent developments in object detecting technology. Transfer learning was used to solve the problem of data availability. In the objective of detecting healthy ACL in the MRNet dataset using a lightweight model, a SoTA OD network, the YOLOv5 nano was employed. This had been trained on extremely large datasets, despite the fact that the vast datasets did not contain the items of our interest. For comparison purposes, three other models, which are also leaders of the YOLO family (YOLOv5-xlarge, YOLOX-small, YOLOX-nano) were analyzed Eight tests were carried out, since each one of the OD networks was trained twice, with and without augmentation. In MRNet, each of the four OD networks was trained individually, and the best performers were chosen. Below is a brief overview of the analyzed OD networks.

YOLOv5 Pytorch Framework (Jocher, n.d.) [87]**:** This is the first release of YOLO models developed in PyTorch rather than PJ Reddie's Darknet. In terms of support and deployment, YOLOv5 is originally implemented in PyTorch, and thus benefits from the current PyTorch ecosystem. Because it is the most generally recognized research framework, iterating on YOLOv5 may be easier for the research community. Finally, YOLOv5 is lightning fast. The following versions were analyzed:

- The xlarge version which has 140.7 million parameters, release version six, FLOPs=209.8 and pretrained checkpoint in COCO dataset with mAP val 0.5:0.95=55.8.
- The nano version which has 1.9 million parameters, FLOPs=4.5, and pretrained checkpoint in COCO dataset with mAP val 0.5:0.95=28.0.

YOLOX [88] is a single-stage object detector with a DarkNet53 backbone that modifies YOLOv3 in numerous ways. The head of YOLO is replaced with a decoupled version. It uses a 1×1 conv layer to decrease the feature channel to 256 for each level of FPN feature, then adds two parallel branches with two 3×3 conv layers each for classification and regression tasks. The following versions were used:

- The small version which has 9.0 million parameters, release version six, FLOPs=26.8 and pretrained checkpoint in COCO dataset with mAP val 0.5:0.98=40.5.
- The nano version which has 0.91 million parameters, FLOPs=1.08, and pretrained checkpoint in COCO dataset with mAP val 0.5:0.95=25.8.

3.2.5. Validation of OD models

The suggested object detection approaches were evaluated using a holdout mechanism. The data was split into three stratified groups (training, validation, and testing), each comprising 70%, 20%, and 10% of the total. The training subset was used to train the detection networks, the validation subset was used to optimize the networks, and the testing subset was used to determine the final performance. Intersection over Union (IoU) is a typical method for determining whether or not one item has been detected. The following formula is used to determine IoU:

$$IoU(A,B) = \frac{A \cap B}{A \cup B} \quad (1)$$

where A represents the set of suggested object pixels and B represents the set of genuine object pixels.

The following metrics are then calculated for each class:

(1) True Positive (TP(c)): a proposal was made for class c and there actually was an object of class c.

(2) False Positive (FP(c)): a proposal was made for class c, but there is no object of class c.

The average precision (AP) for class c may be computed using the formula:

$$AP(c) = (\mathrm{TP(c)})/\ ((\mathrm{TP(c)} + \mathrm{FP(c)}) \quad (2)$$

and the mean average precision (mAP) can be derived using the formula shown below.

$$mAP = \frac{1}{|Classes|}\sum_{c \in Classes} AP(c) \quad (3)$$

The performance metric used to compare the performance of the various object detection techniques was mean average precision (mAP). An IoU threshold of 0.5 was explored to test the detection capabilities of the different techniques.

3.2.6. Inference

The best performing OD model was selected as the one achieving the best detection accuracy (measured by *mAP@0.5*) and computational efficiency (determined by the number of inference FLOPs(G), inference time (s) and the model size (MB)). During inference, all sagittal slices of the testing 3D MRI sample were extracted and the trained best performing OD model was applied on each one, separately. The slice that better represent a healthy ACL was selected as the one with the maximum probability.

3.3. Results and discussion

The OD findings after applying all of the competing DL networks on the MRNet dataset are presented below. Table 6 shows the results of the eight OD models that were evaluated for their suitability in detecting healthy ACL in the dataset (mAP@0.5 and mAP@.5:.95). The table also includes information on the model version, the number of images, and whether augmentation was used. Each experiment used the same framework and batch size, which were Pytotrch and 8 respectively.

The following conclusions can be drawn from Table 6: (i) All the OD models had a mAP score of 0.90 or above, highlighting the capacity of all competing models to identify healthy ACL objects. (ii) YOLOv5-nano had the highest overall mAP@0.5 performance (0.9727) achieved on augmented data. YOLOv5-nano also accomplished the second highest mAP@0.5 (0.9494) without augmentation;-(iii) Overall, data augmentation had a positive effect on the detection efficacy (mAP@0.5) of both YOLOX and YOLOv5.

Table 7 cites the computational efficiency of our models in three different aspects: inference floating point operations per second (FLOPs) utilization, inference time per scan (s), and model size (MB). When executing an application, FLOP(G) is the number of floating-point computations (such as additions and multiplications) completed per second. Only 1.08 GFLOPS are required by YOLOX-nano, however its accuracy was moderate (mAP@0.5 of 0.9395 with augmentation). YOLOv5-nano was found to be the overall best performer given that it accomplished the optimal trade-off between accuracy (best in mAP@0.5) and computational efficiency (second best inference FLOP and time and the smallest model size of the four competitors). This indicates that our suggested model (YOLOv5-nano) is lightweight given that it takes 0.009 seconds (in NVIDIA GeForce 2070 GPU with 6GB RAM) to provide a decision per slice. Overall, only 51.2 seconds were needed to scan all 266 MRIs which included 5691 slices.

exp.	Objection Detection results				
	Model	***Augmentation***	***No. of images***	***Model version***	***mAP@0.5***
1	YOLOv5	No	311	xlarge	0.9045
2		Yes	743		0.9169
3		No	311	nano	0.9494
4		Yes	743		0.9727
5	YOLOX	No	311	small	0.9110
6		Yes	743		0.9231
7		No	311	nano	0.9108
8		Yes	743		0.9395

Table 6: Performance of the proposed object detection approaches.

Model (Version)	***Inference FLOPs(G)***	***Inference time (s)***	***Model size (MB)***
YOLOv5-xlarge	209.8	0.10	175
YOLOv5-nano	4.5	0.009	3.7
YOLOX-small	26.8	0.04	71.9
YOLOX-nano	1.08	0.006	7.7

Table 7: Computational cost, speed, and size. The table displays the computational efficiency of our models in three different aspects: Floating point operations per second (FLOPs) utilization, and inference time per scan (s).

Figure 8 visualises detected objects of healthy ACL as identified by the four competing OD models in representative MRI slices from our dataset. All four models were able to properly identify the location of the healthy ACL. YOLOv5-xlarge identified the ACL objects with relative low probabilities (76%-82%), whereas the rest three OD models provided decisions with higher confidences ranging from 80%-90%.

The last part of our experimentation focused on the performance of the proposed methodology at inference. Figure 8 shows the performance of YOLOv5-nano for all the sagittal slices of a healthy subject (blue line). Specifically, the y-axis depicts the probability of detection for each one of the sagittal slices (0 means that there were no object(s) detected at the specific slice). ACL was detected at slices 19-24 with the highest probability (0.91) at slice 23. This means that ACL becomes partially visible at slice 19 and then its visibility progressively increases up to slice 23 that is finally selected by the algorithm as the best choice (the most informative slice). At slice 24, the ACL becomes less visible and then it disappears after slice 25.

As it has already been mentioned in Section II, the proposed methodology was trained using sagittal slices from only healthy subjects. Figure 9 also depicts the performance of the trained YOLOv5-nano models on sagittal slices obtained from a subject with an injured ACL (red line). The model misclassified the ruptured ACL as healthy, however the observed confidence levels (probabilities) were significantly lower (0.5-0.7) compared to the ones obtained on the healthy knee. This observation can be attributed to the fact that a normal ACL is characterized by tight, continuous fibers that extend from the tibial plateau to the medial side of the lateral femoral condyle. On the other hand, the primary sign of ACL tear is fibers discontinuity. In acute or subacute injury, thickening and oedema of the ACL are found. The ACL may appear lax and concave in appearance. In chronic cases, the fibers can be completely absorbed.

(a) YOLOx-nano

(b) YOLOx-small

(c) YOLOv5-nano

(d) YOLOv5-xlarge

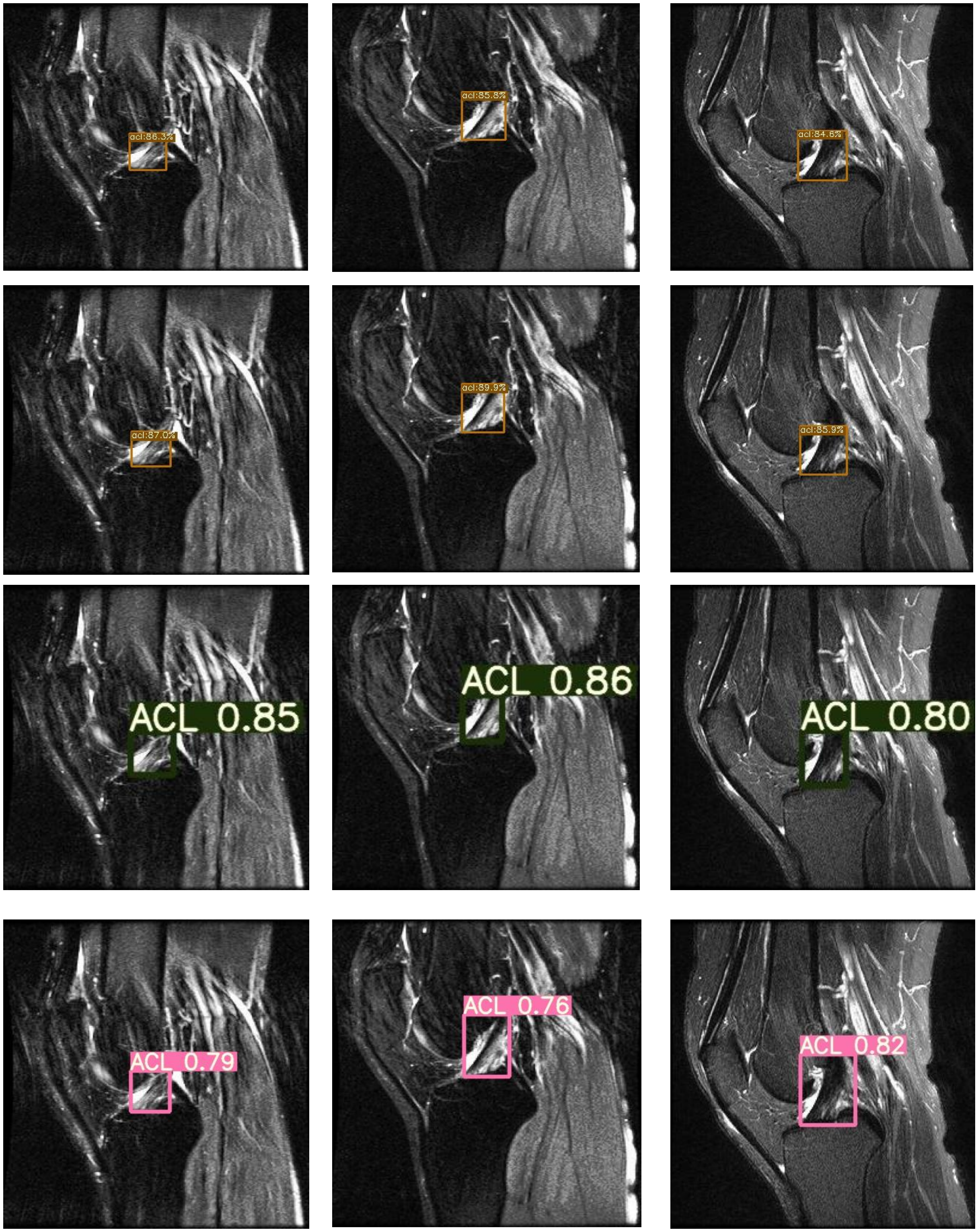


Figure 8: Visual examples of the object detection results obtained by: (a) YOLOX-nano, (b) YOLOX-small, (c) YOLOv5-nano and (d) YOLOv5-xlarrge.

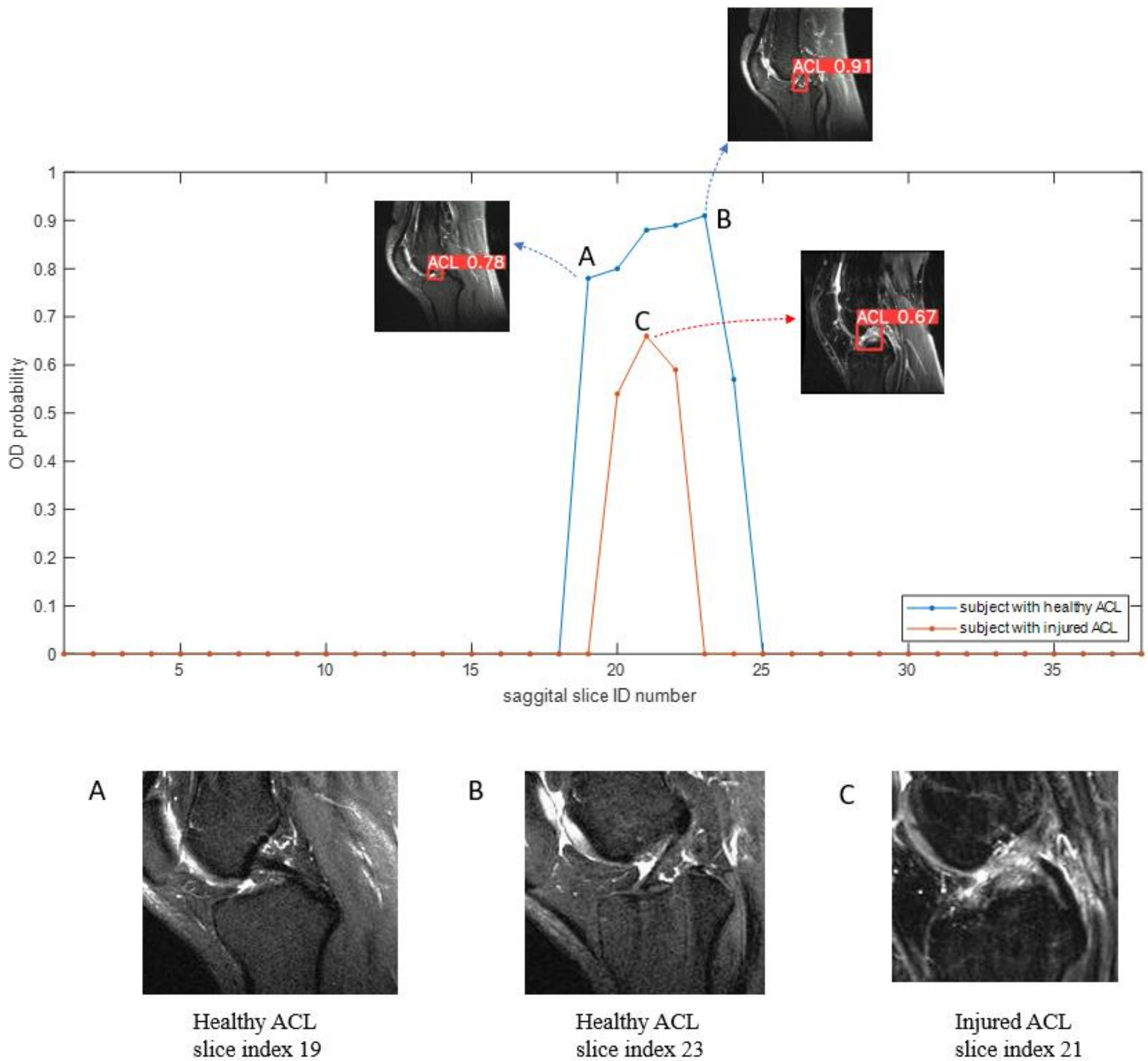


Figure 9: Inference results of the proposed methodology in all sagittal MRI slices of two subjects: one healthy (blue line) and one injured (red line). The confidence probability for each recognized object is visualized with respect to the slide ID number, whereas A, B, and C are three slices of interest: A corresponds to a slice where a healthy ACL was identified with a 0.78 probability, B corresponds to another slide that is the slide with the highest probability observed for the specific healthy subject. C corresponds to an injured ACL that was detected with moderate probability.

3.4. Conclusion

A lightweight DL based OD pipeline (using YOLOv5-nano) was proposed in this paper to locate the healthy ACL in a series of sagittal MRI slices. For comparison purposes, three additional models were analyzed. The proposed methodology accurately recognized the majority of healthy ACL objects and achieved the best performance while having the smallest size. A limitation of this research is that it was tested only on the MRNET dataset, so a more extended cross-dataset validation would be beneficial to provide more robust results. Increasing the size of the training and testing samples are also within our future plans. Another limitation of the proposed model is its inability to distinguish healthy from injured ACL. However, the observed differences on the OD probabilities between subjects with healthy and injured ACL could be further exploited to build new classification techniques that will use the extracted probabilities as inputs. The proposed methodology is a big step towards constructing a network architecture that detects injuries, and decisive object detection that can be successfully implemented in medical picture analysis. Due to the success of nano models, it is conceivable to deduce that lightweight models might be used as a tool for decision-making in therapeutic contexts.

3.5. Acknowledgement

This work has been Co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE (project code: T1EDK-04234)

Chapter 4: Economical Hybrid Novelty Detection Leveraging Global Aleatoric Semantic Uncertainty for Enhanced MRI-Based ACL Tear Diagnosis

Published as:

Siouras, A., Moustakidis, S., Chalatsis, G., Bohoran, T. A., Hantes, M., Vlychou, M., ... & Tsaopoulos, D. (2024). Economical hybrid novelty detection leveraging global aleatoric semantic uncertainty for enhanced MRI-based ACL tear diagnosis. *Computerized Medical Imaging and Graphics*, *117*, 102424.

4.1. Introduction

4.1.1. Background

The economic and societal impacts of ACL injuries are far-reaching, as they represent over 50% of all knee-related injuries and affect each year millions of individuals globally [2]. Only in the United States, the financial burden of ACL injuries exceeds $7 billion annually [15]. Young and athletic populations are particularly susceptible to these injuries, as engaging in professional or recreational activities can raise the risk of a knee injury [89]. This, in turn, contributes to a heightened likelihood of developing knee osteoarthritis [90], [91]. Therefore, prioritizing cost-effective and reliable diagnostic methods for ACL injuries is vital to alleviate the strain on healthcare systems and promote the well-being of affected individuals [15].

Although arthroscopy is considered the "gold standard" for diagnosing intra-articular knee pathology, it has limitations due to its invasive nature and potential complications [24], [92], [93]. Recently, magnetic resonance imaging (MRI) has emerged as the most widely adopted non-invasive imaging technique for evaluating ACL injuries [25]. However, the accurate image interpretation of knee MRI data necessitates the costly expertise of trained clinicians. In addition, given the substantial human subjectivity element, variations in ACL injuries diagnosis are not uncommon [27], [28]. Consequently, there is a pressing need for the development of MRI-based assessment methods that are both robust and user-friendly in order to enhance the diagnostic process for ACL injuries [29], [94].

In recent years, DL techniques have demonstrated promising capabilities in addressing the challenges associated with medical image interpretation by automatically learning complex, nonlinear patterns [83], [95], [96]. One notable area of computer vision that has benefitted greatly by DL is object detection (OD), which encompasses the simultaneous localization and classification of objects within images or videos. The DL-powered OD has the potential to significantly enhance the analysis and interpretation of knee MRI datasets, and further advancements in the field of ACL tear diagnosis.

4.1.2. Clinical background

MRI has become an invaluable noninvasive tool for assessing the integrity and healing of the ACL graft after reconstruction. Increasingly, MRI signal intensity measurements are utilized in clinical studies to evaluate injuries. Despite this, the heterogeneity in MRI acquisition and interpretation methodologies complicates the comparison of signal intensity between scans and different studies. The efficacy of using MRI in this context is quantified by a model that rates the normalcy of the ACL graft in terms of integrity, contour, direction, and thickness, with scores up to 100%. Interpretation of MR images is influenced by the radiologist's experience, study protocol parameters like signal scaling factors, voxel volume, pulse sequence weighting, and patient positioning. However, MRI is still valuable for monitoring the healing of the tendon-bone interface after ACL reconstruction, correlating with functional recovery of the knee joint, and assisting decisions regarding return to sports. Advances in artificial intelligence and DL further enhance the utility of MRI in this field. High-resolution imaging, such as with a 3T MR scanner, and specialized protocols like quantitative MRI with ultrashort echo time T2* and T2* mapping, offer improved evaluations of ACL injury. However, these advanced imaging protocols require longer scanning times and are not feasible for routine use in busy radiology departments. Recent studies have shown that even conventional imaging protocols can provide adequate assessments of ACL graft maturation.

4.1.3. Related Work

The topic of detecting ACL injuries from MRI has recently received considerable attention by the DL community [14]. Awan et al. [59] utilized class balancing and data augmentation to develop a customized 14-layer ResNet-based convolutional neural network (CNN) featuring six distinct paths. Jeon et al. [60] proposed an interpretable lightweight 3D deep-neural network model, outperforming previous state-of-the-art (SOTA) models on the Chiba and Stanford knee datasets [32]. Astuto et al. [62] employed 3D CNNs to diagnose and grade ACL damage. Dunnhofer et al. [97] developed MRPyrNet, a novel CNN architecture which focuses on small, localized regions. By integrating MRPyrNet into existing diagnostic pipelines, the researchers significantly improved diagnostic capabilities, particularly for ACL and meniscal tears, due to the architecture's ability to exploit relevant appearance features.

All the above studies exemplify the potential of DL techniques to revolutionize the detection and assessment of ACL injuries in clinical practice. These papers have nevertheless used an ordinary classification framework for the detection of ACL injuries. Yet training a network on datasets with severely imbalanced classes poses a serious challenge to the classification-based predictive modeling, since the algorithm is more prone to errors in the minority than the majority class. This can be particularly hazardous in the clinical context where the minority class is typically the most valuable. In addition, existing studies derived their decision by relying on a single slice, which had typically been obtained in a manual fashion by the clinicians. Nonetheless, the investigated tissue realistically spans across multiple slices. Moreover, the network architecture in previous studies was designed and evaluated using MRI data from a single database. Therefore, the knowledge about the ability of these algorithms to generalize is limited. Finally, previous papers disregarded the compute demand by the DL model. Yet, this is a significant matter since these computations necessitate mind-boggling amounts of generated power for fuelling them [98].

4.1.4. Our contribution

In this paper, in order to enhance ACL tear diagnosis by tackling the class imbalance issue, presented in most publicly available knee MRI datasets, we propose to reformulate this task as a novelty detection problem [99]. In brief, our framework first attempts to accurately model "normality" by training an OD model (YOLOv5-nano [87]) exclusively on a dataset of healthy knees, devoid of any instance that we would like the framework to detect. To obtain a discriminative novelty score, we leverage the aleatoric semantic uncertainty as this is naturally and implicitly modeled in the metadata (class scores) outputted by the OD model. Given that the ACL tissue spans across multiple slices, we propose to make use of the global class scores obtained when the OD model is applied to the whole image sequence. Therefore, the novelty score is a vector of generated probabilities which indicate the likelihood that the object inside the bounding box belongs to the “healthy knee” class, expecting alternate (lower) values when instances diverge from the training dataset. As the final step of the proposed hybrid framework, we utilize a timeseries classification model (MINIROCKET [100]) for determining whether a knee has an ACL tear or not.

As well as representing the SOTA, both machine learning modules of our pipeline are also known to be highly resource-efficient, fostering green DL research and aligning with efforts to democratize it. Notably, such economical strategies are more relevant to being adopted by lightweight edge devices which in turn could boost their use.

The optimal visualization of ACL necessitates having a sagittal scanning plane [101]. This is also backed by a growing literature demonstrating that sagittal plane factors are responsible for ACL injury mechanisms [102]. Consequently, only data from the sagittal plane is used in this study.

Another critical contribution of this study is that we implement both single- and cross-database testing to better evaluate the generalization capabilities of our approach, involving two public databases namely MRNet [32] and KneeMRI [52], and a third validation-only dataset sourced from the Department of Orthopaedic Surgery and Musculoskeletal Trauma at the University General Hospital of Larissa (UGHL) in Greece. Our hybrid (two-phase) novel detection framework is rigorously compared to two SOTA methods [60], [97] with regard to accuracy, resource efficiency and environmental cost.

4.2. Methods

4.2.1. Datasets and annotation

In this study, we utilized two public datasets (MRNet [32] and KneeMRI [52]) and one validation-only dataset sourced from UGHL in Greece.

MRNet [32] was created between 2001 and 2012 for the development and evaluation of artificial intelligence (AI) algorithms in medical imaging. It contains 1,370 knee examinations of 1,104 patients, along with labels indicating the presence or absence of abnormalities in four different categories: ACL tear, meniscus tear, abnormal cartilage, and abnormal bone marrow. A subset of this database was used, that is 266 healthy knees and 319 knees with ACL rupture. The number of slices in each 3D MRI scan ranges from 17 to 61, and the size of each MRI slice is 256×256 pixels.

The KneeMRI [52] dataset was retrospectively collected using proton density weighted fat suppression and a Siemens Avanto 1.5T MRI scanner at the Clinical Hospital Centre Rijeka, Croatia from 2006 until 2014. The dataset comprises 917 left- or right-knee 12-bit grayscale volumes. Following radiologist reports, the authors labeled each scan according to ACL disorder: non-damaged (690 scans), partially injured (172 scans), and totally ruptured (55 scans). Each MRI examination comprises 21–45 slices and the spatial resolution of each slice is 320×320 or 290×300 pixels.

At UGHL, MRI examinations of 129 ACL ruptured knees were conducted from 2018 onwards using a 3.0T MRI scanner (Signa HDxt 3.0T, GE Healthcare) with a quadrature knee coil, field of view (FOV) of 18x18 pixels, thickness of 1.0 (in millimeters), and a resolution of 254×256 pixels. The examinations were performed preoperatively on knees that were slated for surgery. The MRI acquisition protocol included the capture of sagittal and axial T1-weighted as well as T2-weighted images, supplemented by sagittal and coronal proton density weighted Turbo Spin Echo (TSE) images with fat saturation. A musculoskeletal radiologist with two decades of experience (M.V.) reviewed the MRI scans.

In the two public datasets each MRI slice comes pre-annotated. Regarding the UGHL dataset, there was only a single overarching annotation for the entire MRI scan.

4.2.2. Data preprocessing

Image augmentation was employed to enhance the generalization capacity of our pipeline by modifying image brightness (within a range from -32% to +32%), altering hue (ranging from -39° to +39°), adjusting exposure (between -30% and +30%), incorporating blurring (up to 1.75 pixels), and adding random noise (affecting up to 7% of pixels). The concluding step prior to initiating OD training involved exporting the annotations in an appropriate format.

4.2.3. Hybrid network architecture

The proposed hierarchical hybrid architecture (Figure 1) is composed of two key elements.

At first, we incorporate the latest advancements in OD technology. Specifically, we apply YOLOv5-nano [87], a SOTA OD network and part of the "You Only Look Once" series, to the complete sagittal MRI scan. The reason why we opted for the nano version of YOLOv5 for our task is due to its lightweight nature. When compared with the other available versions of YOLOv5, the nano variant trades off detection accuracy for computational efficiency. In brief, the architecture of YOLOv5 is divided into three essential parts (**Error! Reference source not found.**):

- Backbone: The core CNN that processes input images and extracts features at various levels.
- Neck: Layers that mix and combine features from the backbone, preparing them for accurate predictions.
- Head: This part uses the combined features to predict object bounding boxes and their classes.

This structure allows YOLOv5 to efficiently detect objects (by predicting bounding boxes) and classify them in a single unified framework.

In this study, solely healthy ACLs are fed into YOLOv5-nano, in line with our aim to build a novelty-detection framework. To obtain an appropriate novelty score, we propose to capitalize on the metadata outputted by the OD model. In more detail, we exploit the aleatoric semantic uncertainty which is naturally and implicitly modeled in the class probability score, every time our framework is fed with an observation that does not belong to the trained distribution. These class scores are typically estimated by employing a combination of techniques such as softmax activation function, class-specific filters in the final layer of the neural network, and class detection thresholds. Taking into account that the ACL tissue spans across multiple MRI slices, we propose to make use of the global class probabilities obtained from the whole sagittal image sequence. As a result, the novelty score is a vector of generated probabilities which indicate the likelihood that the object inside the bounding box belongs to the “healthy knee” class (rather than background), and the expectation is that these will take different (lower) values every time the input instance diverges from the training population.

In the second phase of the proposed hybrid framework, we utilize a timeseries classification model (MINIROCKET [100]) for determining whether a knee has an ACL injury or not by applying the probability vector to it. MINIROCKET (MINImally RandOm Convolutional KErnel Transform) marked a significant advancement in the field of time series classification due to its exceptional computational efficiency and speed, particularly on larger datasets. It offers the advantage of mostly deterministic operation, with an option for full determinism. This ensures both consistent and reliable results. In a nutshell, MINIROCKET's architecture is characterized by its use of a small, fixed set of convolutional kernels, each with a length of 9 and weights restricted to -1 and 2. This simplification not only maintains a similar level of accuracy compared to more complex models but also significantly reduces computational complexity. The algorithm is tailored to the length of the input time series through its dilation process and uses alternated padding for each kernel/dilation combination, adding to the transformation's consistency. In terms of feature handling, MINIROCKET simplifies its approach by focusing primarily on the Proportion of Positive Values (PPV) pooling, resulting in a total of about 10,000 features. This feature reduction effectively balances accuracy with computational efficiency, making the algorithm versatile for various applications. MINIROCKET's speed is further enhanced through several optimization techniques. These include computing PPV for a kernel and its inverse simultaneously, reusing convolution outputs for multiple feature computations, and performing most of the computations for all kernels at once for each dilation. While it brings in several advancements, MINIROCKET maintains linear scalability in relation to the number of features and the length of input time series. Although it requires slightly more memory than traditional methods due to the temporary storage of additional vectors during transformation, this increase is generally minimal. Utilizing the features it generates, MINIROCKET is capable of training linear classifiers, such as ridge regression or logistic regression, making it highly suitable for real-time and large-scale time series classification tasks where a balance between high accuracy and low computational expense is crucial. The ridge classifier was used in this study.

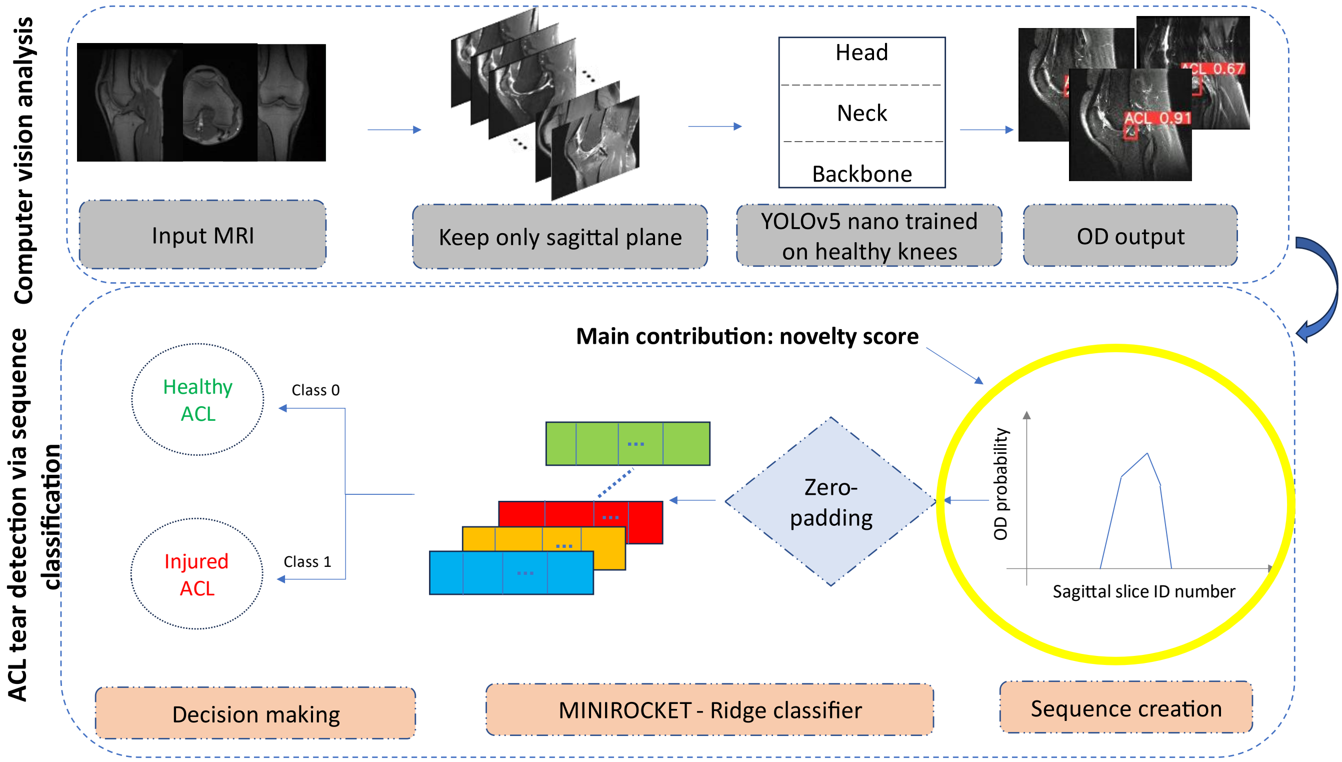


Figure 10: An illustration of the proposed hybrid DL pipeline.

4.2.4. Implementation and training

To better assess the generalization capabilities of our framework, we performed a total of six experiments, also allowing a cross-database testing. The two public databases (MRNet and KneeMRI) were utilized for training and evaluation purposes, with the third database (UGHL) being employed for external validation. The model evaluation protocols that we employed for the public databases adhere to the existing literature to ensure rigorous benchmarking and facilitate reliable comparisons across studies. In particular, all papers that discuss MRNet have undergone testing on its validation set. Likewise, every publication referring to KneeMRI has adopted a 5-fold cross-validation technique.

Therefore, the six experiments we carried out were:

- Exp1: Training on MRNet (Train) and testing on MRNet (Validation)
- Exp2: Training on MRNet (Train) and testing on the entire KneeMRI dataset
- Exp3: Training on MRNet (Train) and testing on the entire UGHL database
- Exp4: Training on KneeMRI and testing on KneeMRI (cross-validated)
- Exp5: Training on KneeMRI and testing on the entire MRNet.
- Exp6: Training on KneeMRI and testing on the entire UGHL.

We trained the OD model using exclusively healthy knees. Transfer learning was employed to address the limited dataset size. Zero padding was performed to the class probability vector outputted by YOLOv5-nano to create uniform data structures, essential for the legit application of our subsequent Ridge Classifier.

The pseudocode outlining the training and inference protocolss for the proposed hybrid architecture is provided below Table 8.

Training protocol
DATASET: $D\ (X)$ D represents the dataset MRNet or KneeMRI. **Splitting the dataset:** Partition D into training and validation sets using a split function S: $(X_{train}, X_{val}) = S(D)$ where $\|X_{train}\| = 0.8\ \|D\|, \|X_{val}\| = 0.2\ \|D\|$ **Data preprocessing or modifications** Keeping Only Sagittal Plane Data: Extract the sagittal plane data from X_{tr-} using function $F_{S.}$ Mathematically, this can be represented as: $X_{trs} = F_S(X_{tr}); F_S{:}X_{train} \rightarrow X_{trs}$ • Extract healthy data from X_{trs} using a function F_H: $X_{trhs} = F_H\ (X_{trs}); F_H{:}\ X_{trs} \rightarrow X_{trhs}$ **Training phase for the first model:** Train YOLOv5-nano model 1 (M_1) on X_{trhs} **Data Preparation for the second model:** $Predictions_{train}\ (P_{tr})$ = PREDICT using M_1 on X_{trs} P'_{tr} = EXTRACT probabilities from P_{tr} per slice and zero padding. **Training phase for the second model:** Train MINIROCKET (Ridge Classifier) model 2 (M_2) on P'_{tr}
Inference protocol
Initialize new MRI (Y) Keep sagittal slices from Y → Y_s FOR each slice i in Y_s: APPLY M_1 to slide i EXTRACT probability P_i from the output of M_1 PERFORM zero padding on P_i to match Model M_2 input dimensions → Vector (V) FOR V: APPLY Model M_2 EXTRACT final decision L_i RETURN L_i as final decisions for Y

Table 8: Pseudocode of the hybrid architecture training and inference protocols.

We trained the YOLOv5-nano model for 150 epochs with images resized to 640x640 pixels. For the RidgeClassifier, the default hyperparameters were employed: In this configuration, the alpha parameter, which controls the strength of regularization, was set to 1.0. The maximum number of iterations for the algorithm to converge, was not limited (max_iter= None). The parameter indicating whether to calculate the intercept for this model was enabled (fit_intercept=True). The tolerance for algorithm stopping was set to 0.0001. The solver, which is the algorithm used for optimization was set to automatically choose the best method ('auto'). Finally, the random_state parameter, used to seed the random number generator for reproducibility was not set (None), allowing the algorithm to use a random seed. Our framework was implemented[1] using Python version 3.10 on a system with operating system Ubuntu Linux 22.04 and an NVIDIA RTX A6000 (48GB) graphics card.

4.2.5. Model evaluation.

The ACL tear classification task involved evaluation metrics such as Receiver Operating Characteristic (ROC) curve and the respective Area Under the Curve score (AUC), accuracy, specificity, and sensitivity. Additionally, we employed the carbon dioxide equivalent emissions (CO_2eq) (g) produced and energy (kWh) spent during training, equivalent distance traveled by car (km), training time (h:min:s), and average inference time (ms) in order to provide a comprehensive assessment of the system's resource efficiency and generated carbon emissions. All evaluation metrics of the proposed methodology were juxtaposed with those yielded by two SOTA methods [18], [21] for comparison purposes.

4.3. Results

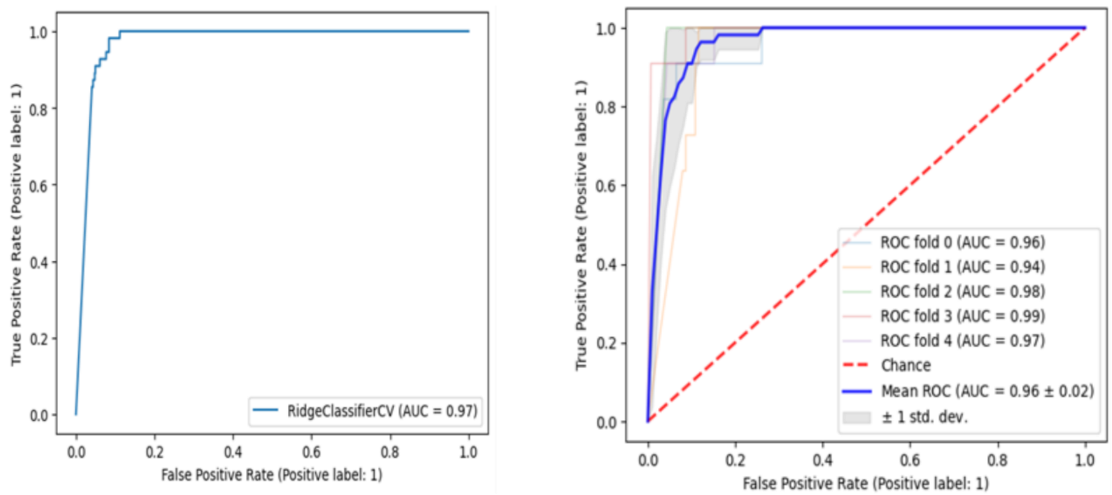
4.3.1. Accuracy evaluation

Error! Reference source not found. outlines the comparison results between the proposed f ramework and the SOTA methods in terms of ACL tear classification accuracy in single-database experiments (Exp1, Exp4). Our method consistently outperformed SOTA approaches on the KneeMRI dataset and achieved better accuracy and sensitivity on the MRNet dataset. Figure 2 shows the ROC curves and the corresponding AUC scores of our classification models for the same experiments. As shown, the achieved AUC scores are outstanding. Figure 3 illustrates the bounding boxes and the respective class scores outputted by the OD component of our framework in Exp1 and Exp4, when this is fed with representative ACL sagittal slices of both healthy and injured knees. The overall probability vector obtained when applying the whole sagittal image sequence for representative cases is shown in Figure 4. It is obvious that the class scores are different (lower) every time our OD model is fed with a scan that diverges from the training dataset. Apparently, probability scores exceeding 0.80 are suggestive of an intact ACL.

Approach	Experiment	Database	Accura cy	AUC	Sensitivi ty	Specifici ty
Dunnhofer et al. (2022) [97]	Exp1	MRNet	0.886	**0.976**	0.815	**0.944**
	Exp4	KneeMRI	0.834	0.914	0.806	0.843
Jeon et al. (2021) [60]	Exp1	MRNet	0.911	0.963	0.944	0.901
	Exp4	KneeMRI	0.826	0.851	0.801	0.874
Proposed	Exp1	MRNet	**94.94**	96.00	**96.30**	92.00
	Exp4	KneeMRI	**96.00**	**96**	**91.66**	**96.82**

Table 9: Comparison between the proposed and other SOTA approaches in terms of ACL tear classification performance in single database experiments. Boldface indicates best performance.

[1] https://github.com/ThanosUTH/ACL-Tear-Diagnosis-



(a) (b)

Figure 11: ROC curves and corresponding AUC scores illustrating the classification performance of the proposed models in Exp1 (a) and Exp4 (b).

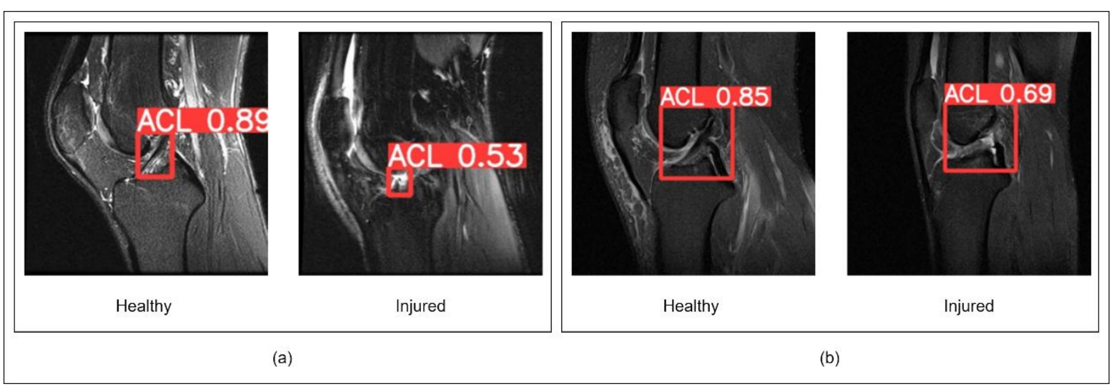


Figure 12: Indicative results (i.e. bounding boxes and class scores) of the OD component of our models in Exp1 (a), Exp4 (b), when these are fed with typical ACL sagittal slices of both healthy and injured knees.

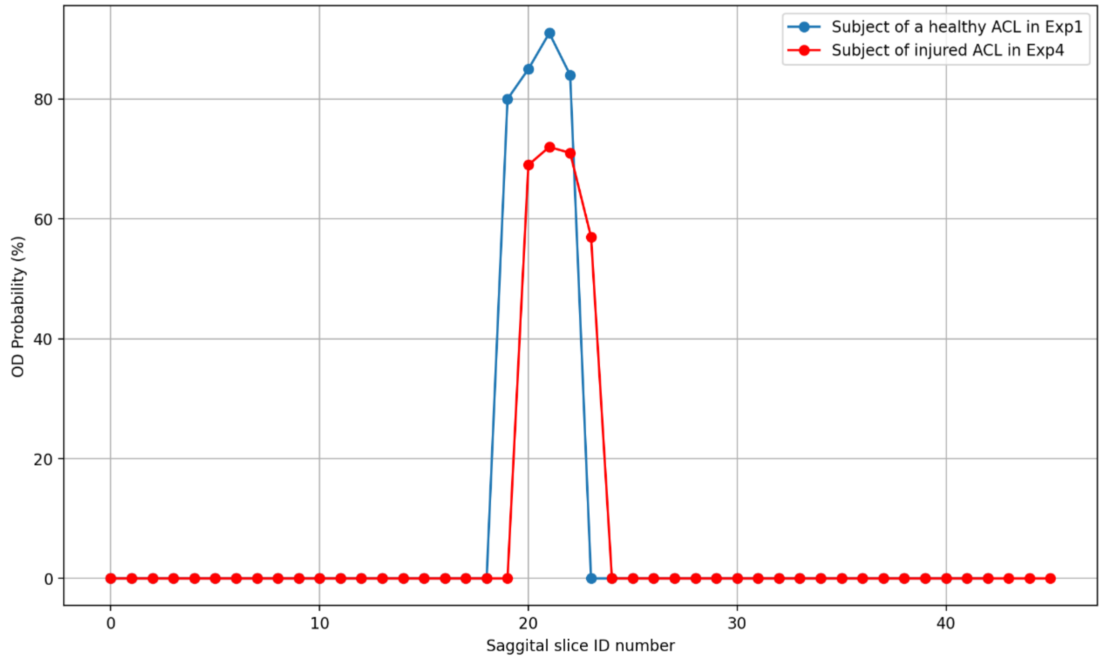


Figure 13: Indicative probability vectors obtained by the OD component of our framework in Exp1 and Exp4, when the whole sagittal plane image sequence is applied to it.

4.3.2. Resource efficiency evaluation

Error! Reference source not found. lists the generated CO_2eq emissions, the power spent, and t he equivalent distance a car could cover during the final training of 150 epochs for the proposed and SOTA methods. Also given are the training time and the average inference time. The proposed framework was found to be at least 2.1 times more eco-friendly and consume at least 2.6 times less fuel in comparison to the SOTA methods. Moreover, in terms of computational time, 1.5 times shorter training phase was achieved, while inference speed was at least 1.4 times higher.

Model	CO_2eq (g)	Energy (kWh)	Equivalent distance traveled by car (km)	Training time (h:min:s)	Average inference time (ms)
Jeon et al. (2021) [60]	51.921046	0.43277567	0.431238	0:55:00	2
Dunnhofer et al. (2022) [97]	347.505372	0.908115	2.886257	1:35:37	6
Proposed	**20.264802**	**0.208915**	**0.168312**	**0:36:06**	**1.4**

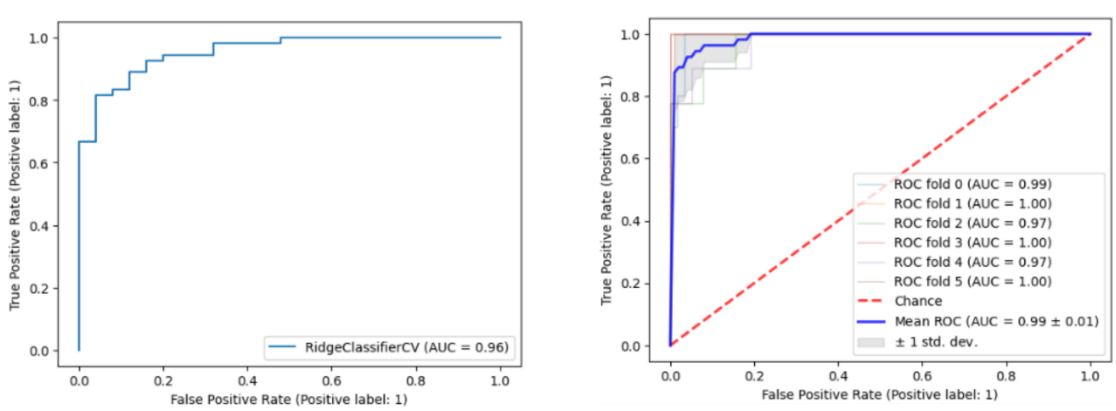
Table 10: The resource efficiency comparison of the proposed method with the SOTA approaches. Evaluation was based on carbon emissions, energy usage, and equivalent car travel distance during the final training. Additionally, total training and average inference times were recorded. All experiments were carried out utilizing an NVIDIA RTX A6000 (48GB) GPU workstation. Boldface indicates best performance.

4.3.3. Generalization capabilities evaluation

Error! Reference source not found. outlines the results of the proposed framework in terms of A CL tear classification accuracy in cross-database settings (Exp2, Exp3, Exp5, Exp6). Figure 5 shows the ROC curves and the corresponding AUC scores of our classification models for cross-database experiments with testing databases that involved more than one class (Exp2, Exp5). Figure 6 illustrates the bounding boxes and the respective class scores outputted by the OD component of our framework in Exp2, Exp3, Exp5, and Exp6. It can be seen that the proposed framework exhibited excellent generalization capabilities, holding performance across different databases, especially when it was trained on the KneeMRI dataset.

Experiment	Training Set	Testing Set	Accuracy	AUC	Sensitivity	Specificity
Exp2	MRNet	KneeMRI	88.99	97.00	98.10	88.12
Exp3	MRNet	UGHL	83.67	-	-	-
Exp5	KneeMRI	MRNet	91.14	96.00	96.30	80.00
Exp6	KneeMRI	UGHL	94.00	-	-	-

Table 11: ACL tear classification evaluation of the proposed framework in cross-database experiments. *AUC, Sensitivity and Specificity are absent in Exp3 and Exp6 because UGHL involves only samples from one class (ACL ruptured knees).*



(a) (b)

Figure 14: ROC curves and corresponding AUC scores illustrating the classification performance of the proposed models in Exp2 (a) and Exp5 (b).

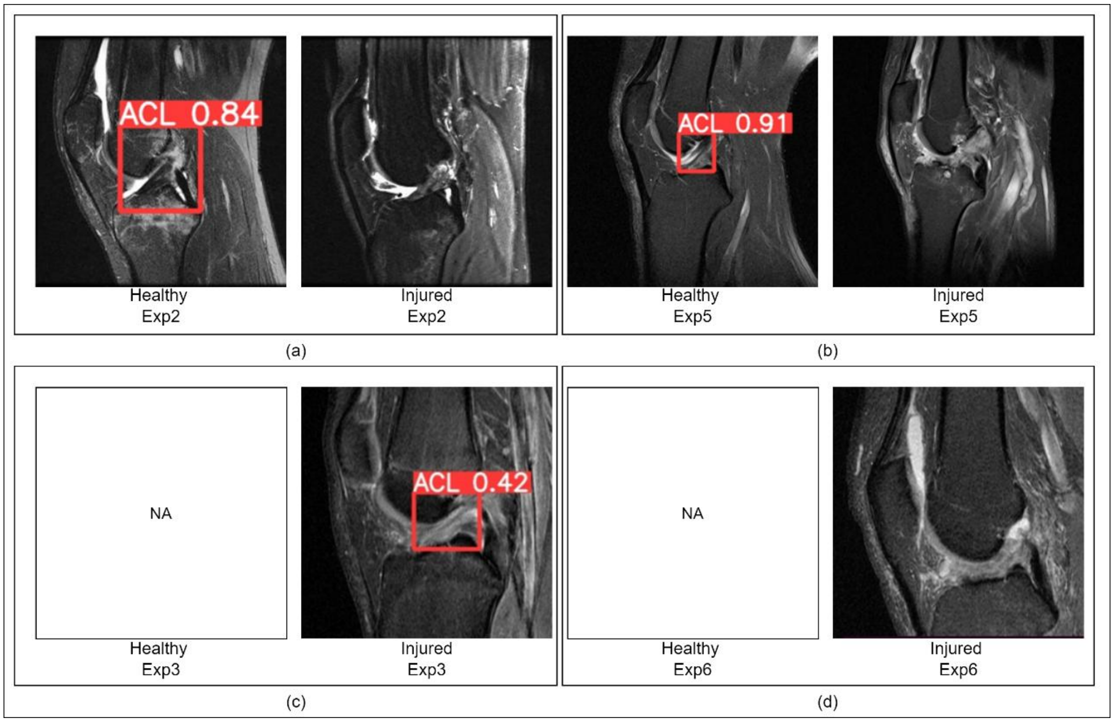


Figure 15: A visual presentation of results in cross-database experiments, showcasing sagittal planes, boundary boxes, and class scores of both healthy and injured knees. (a) Exp2, (b) Exp5, (c) Exp3 and (d) Exp6. Note: The absence of a boundary box in a graphic indicates a healthy knee class score less than 0.4, signaling potential injury. There were no healthy knees in the testing set of Exp3 and Exp6.

4.4. Discussion

This research successfully implemented an innovative hybrid DL framework that revised the sagittal MRI-based ACL tear classification task as a novelty detection problem to address the class imbalance challenge presented in public datasets. To obtain a fit-for-purpose novelty score, we capitalized on the aleatoric semantic uncertainty as this is modelled in the class scores outputted by the YOLOv5-nano OD model. Taking into consideration that the investigated tissue typically spans across multiple MRI slices, we proposed the utilization of the global scores (probability vector) obtained when the OD model is applied to the whole sagittal plane image sequence. The second part of the proposed hybrid framework comprised the MINIROCKET timeseries classification model to determine whether a knee has an ACL tear or not. To investigate our approach's ability to generalize, we performed both single- and cross-database testing involving two public databases (KneeMRI and MRNet) and a validation-only database from UGHL in Greece.

Our method was invariably superior to the SOTA approaches [18,21] on the KneeMRI dataset and achieved better accuracy and sensitivity on the MRNet dataset. The cross-database experiments we carried out verified the robustness and excellent generalization capabilities of our method, especially when the model had been trained on KneeMRI. In addition, the presented pipeline generated at least 2.1 times less carbon emissions and consumed at least 2.6 times less energy, when compared with the SOTA approaches. It is worth noting here that the numbers reported in Table 3 pertain to the final training only. However, a typical DL development process requires hundreds of training runs due to experimenting with multiple models and hyperparameter tuning. Therefore, this aspect is crucial and the concerns over the environmental footprint of DL research and its impact on accelerating global climate change have been growing [22]. Our results go along with efforts to propose “greener” DL approaches and democratize DL research [16].

In this paper, we innovatively framed the MRI-based ACL tear diagnosis as a novelty detection problem by leveraging the global aleatoric semantic uncertainty obtained by the YOLOv5-nano OD model. However other approaches do exist in the novelty detection field. For instance, a recent study [29] explored novelty detection by generating synthetic near-distribution anomalous data to bridge the gap in detecting subtle differences between normal and anomalous samples, an approach that demonstrates the flexibility of novelty detection techniques in handling near-distribution challenges. Another research [30] employed diffusion models for novelty detection, focusing on mitigating background bias in out-of-distribution samples through a method named Projection Regret (PR), which illustrates the potential of generative models in enhancing detection sensitivity. The varied applications of novelty detection across different tasks highlight the breadth of the evolving landscape of this field and the adaptability of these techniques to specific challenges, such as class imbalance and subtle anomaly distinction.

The integration of aleatoric semantic uncertainty-based scores into a novelty detection framework is a breakthrough contribution of this study. The above attribute, when combined with the use of lightweight OD and timeseries classification models, have the potential to revolutionize injury detection from MRI data by setting a new precedent in diagnostic precision, speed and environmental sustainability.

4.5. Limitations

One notable limitation of our study pertains to the fact that some MRI images within the datasets, especially those taken before 2005, lack the quality synonymous with modern magnetic resonance imagers. This discrepancy in image quality could impact the performance of the algorithm. In fact, this is possibly the reason why the model that had been trained on KneeMRI dataset generalized better (than the model that had been trained on MRNet dataset) on the UGHL dataset. As such, a worthwhile avenue for future exploration would be to conduct studies using contemporary MRI scanners, which offer superior image quality. Additionally, a challenge we encountered was the presence of MRI scans with a limited number of sagittal slices. This limitation is attributed to the large intervals at which the imager captures the data (or else, the low out-of-plane spatial resolution), necessitating the use of padding in our approach. In subsequent studies, this factor should be considered, ensuring that data is collected at optimal intervals to maximize the number of image slices and reduce the need for such adjustments.

4.6. Conclusion

In conclusion, this study introduced a hybrid novelty detection DL pipeline for ACL tear detection from MRI sagittal plane slices. The proposed framework achieved superior accuracy, consumed notably less fuel, and generated strikingly less carbon emissions, when compared against two SOTA approaches. The generalization capabilities of our pipeline were verified by running cross-database experiments. The integration of aleatoric uncertainty-based scores that are rapidly acquired into novelty detection DL frameworks holds promise as it not only sets a new standard in the MRI-based diagnostic speed and accuracy, but it also aligns with the increasing need for environmentally sustainable and democratic computational practices. The presented resource-efficient pipeline offers potential for widespread clinical application. The research paves the way for future advancements in the DL-powered MRI-based injury detection, emphasizing environmental sustainability alongside technological innovation.

Chapter 5: Thessaly Graft Index: An Artificial Intelligence-Based Index for the Assessment of Graft Healing in ACL Reconstructed Knees - A Pilot Study

Accepted for Publication:

"Graft Index: An Artificial Intelligence-Based Index for the Assessment of Graft Integrity in ACL Reconstructed Knees."

Georgios Chalatsis, Athanasios Siouras, Vasileios Mitrousias, Ilias Chantes, Serafeim Moustakidis, Dimitris Tsaopoulos, Marianna Vlychou, Sotiris Tasoulis, Michael Hantes

5.1. Introduction

The minimal postoperative period needed for the ACL graft in humans to fully mature is unclear [103]–[106]. Despite recent studies that suggest patients return to sport (RTS) at least 9 months post ACL reconstruction (ACLR) [107]–[109], objective guidelines, that ensure a secure return to full athletic participation after ACL reconstruction are lacking [110], [111]. Human histological 10 and magnetic resonance imaging (MRI) studies indicate that ACLR graft re-modelling is an ongoing process beyond one-year post-operation [112], [113].

Given its invasive nature, the histologic examination of the ACL to evaluate graft ligamentization in humans is difficult and costly to perform [114]. As a result, MRI has proven to be a valuable noninvasive tool to assess the graft healing process after ACL reconstruction [115], [116]. MRI signal intensity (SI) measurements have been used more and more in clinical studies to assess ACLR graft maturity [113], [117]. However, the vast heterogeneity of MRI acquisition and interpretation methodologies used to evaluate the ACL graft, hinders comparisons of SI between successive scans and separate investigations [118].

Furthermore, recent studies have tried to correlate MRI-based ACL graft maturity with clinical and functional outcomes [119]–[126]. Few subjective scores and functional criteria are recommended for clinical use in order to determine the time frame of return to sports (RTS) [127]–[130]. Despite the fact that they have been thoroughly investigated and illustrated in the literature, these scores and tests lack of objective data and consensus regarding the stability and healing status of the reconstructed graft to guide the timing of RTS [131].

In light of the advancements of AI technology and the growing number of studies in this field, different ML and DL algorithms have been employed in the diagnostic process of knee injuries [25], [34], [83] and in the assessment of functional outcomes post knee surgery [132]. Expert's diagnostic accuracy used to be superior, but recent studies report performance levels akin to fellowship-trained, full-time, academic musculoskeletal radiologists in several tasks, including detection and segmentation of ACL injury [29].

In a 2022 systematic review about AI in knee injuries [25], 14 studies were included presenting DL and ML models that could detect ACL injury with various accuracy ratings. However, despite the evolution of AI, no algorithm has been developed, until now, with the ability to assess the ACL graft maturity. More importantly, no algorithm has been developed to correlate the graft maturity with clinical and functional outcomes. In this way, the purpose of this study was to create an AI-based index (TGI) for the evaluation of graft maturation in ACL reconstructed knees and to investigate the correlation of the TGI with patients' clinical and functional outcomes. The hypothesis of this study is that the proposed AI model can successfully assess the ACL graft maturation, one year post-operatively, and that patients with higher TGI scores will present also better patient reported outcome measures (PROMs).

5.2. Methods

5.2.1. Study Design and Participants

The study included a cohort of 28 patients who had undergone ACLR using a hamstring tendon autograft after an isolated ACL tear. MRI examination was performed before surgery and 1 year postoperatively. Simultaneously, PROMs were used to assess clinical and functional outcomes before surgery and 1 year postoperatively. The main inclusion criterion was an isolated ACL reconstruction using a hamstring autograft. Exclusion criteria included open physes, concomitant meniscal repair/meniscectomy, chondral injuries, other ligament reconstruction, bilateral ACL rupture, and history of inflammatory disease or septic arthritis of the injured knee. The recruitment period lasted from September 2021 to May 2022. Twenty-eight patients were eligible to participate during this period, 3 patients refused to participate in the study and 1 patient was lost on follow-up. Finally, 24 patients were enrolled in the study. Ethical approval for the study protocol was obtained by the institutional review board of the General University Hospital of Larissa. The research was performed in accordance with the 2008 Declaration of Helsinki Ethical Principles for Medical Research involving Human Subjects. Written informed consent form was signed by all participants.

5.2.2. Surgical Technique

All patients underwent an anatomic ACL reconstruction using an ipsilateral hamstring tendon autograft reconstruction. All procedures were performed by the same senior surgeon (MH). The femoral tunnel was drilled at the center of the native ACL femoral footprint, through the anteromedial portal. The tibial tunnel was drilled outside-in from the tibial cortex, aiming at the middle of the ACL tibial footprint. The semitendinosus and gracilis tendons were folded to create a four-stranded (quadruple) autograft. A button was used for femoral fixation, while a bio-absorbable interference screw was used for tibial fixation. Post fixation was also performed using a bicortical screw.

5.2.3. Rehabilitation

All patients were instructed to perform basic exercises at home to regain range of motion and preserve quadriceps and hamstrings strength. The postoperative rehabilitation protocol emphasized in early weight-bearing progression and gradual acquisition of full range of motion (ROM), as tolerated. As soon as complete ROM was attained, patients followed a meticulously designed progressive exercise program for increasing strength and proprioception. Depending on the type of sport and knee function, return to athletic activities was personalized, but in general, it was advised no sooner than 9 months after surgery.

5.2.4. Preoperative and Postoperative Clinical evaluation and PROM's

Patients' knees were assessed clinically for their side-to-side stability, using the KT-1000 arthrometer preoperatively, at 3 months, at 6 months and during the final follow-up, 12 months postoperatively. PROMs were collected preoperatively and also at 1 year follow-up by two independent medical assistants. All PROMs have been previously validated and adapted in the Greek language [133]–[135]. The PROMS that were used, were the International Knee Documentation Committee Subjective Knee Function (IKDC-SKF) score [133], the Knee Injury and Osteoarthritis Outcome Score (KOOS) [136], the Lysholm knee score [134] and the Tegner Activity Scale (TAS) [135]. The minimum clinically important difference cutoff values for KOOS were based on the recent study of Ingelsrud et al. [137].

5.2.5. MRI Evaluation

5.2.5.1. The University Hospital of Larisa (UHL) Database

MRI examination was performed preoperatively and postoperatively using a 3T MRI scanner (Signa HDxt 3.0T, GE Healthcare) and a quadrature knee coil. The MRI protocol included axial, sagittal and coronal intermediate proton-weighted images with fat saturation, coronal T1-weighted fast spin echo images and 3D spoiled gradient-echo (SPGR) images with fat saturation. The MRI scans were reviewed by a musculoskeletal radiologist with 20 years of experience (MV). The native ACL was characterized either as normal or not normal/ruptured and the ACL graft either as normal/healed or not normal/ruptured. All MRI scans were also assessed by the DL model. The model's algorithm detected the integrity of the graft, compared it with a healthy native ACL presenting a binary result (healthy or ruptured native ACL and ACL graft) and a quantitative score on a percentile scale (%). The quantitative score was representing the model's estimation on how close to maturation (or to a native ACL) was the ACL graft.

5.2.5.2. The KneeMRI Dataset

The KneeMRI dataset, encompassing MRI scans of the knee, was compiled retrospectively by sourcing proton density weighted fat suppressed images gathered from 2006 to 2014 at the Clinical Hospital Centre Rijeka, Croatia. Data acquisition was achieved via a Siemens Avanto 1.5T MR scanner. This dataset is composed of 917 volumes in 12-bit grayscale, with each volume representing either the left or right knee. Each individual record within the dataset has been systematically classified into one of three categories: healthy, partially ruptured, or fully ruptured. This classification was executed following the assessment of radiologist reports, with each scan assigned a label in accordance with the state of the ACL: healthy (encompassing 690 scans), partially injured (consisting of 172 scans), and fully ruptured (comprising 55 scans). The spatial resolution of each MRI slice falls within the range of 320 × 320 to 290 × 300 pixels, with the number of slices per volume varying between 21 and 45.

5.2.6. Deep Learning Model

MRI scans were employed as training material for an advanced DL network, the YOLOv5 Nano version [87]. The training set was extracted from the Knee MRI database [52], consisting of 690 MRI images of healthy knees and an additional 227 images of knees displaying ACL injuries. This

DL model was meticulously designed to compute the probability of accurately detecting a healthy ACL.

Initially, slices were selected exclusively from the sagittal plane, showcasing only a healthy ACL, by a radiologist. Annotations for each slice were subsequently provided, contributed by the originator of the database. To boost the model's generalizability and adaptability to a variety of image conditions, augmentation techniques were employed. Involved adjustments to the image brightness.

Following the enhancement and curation of the dataset, the YOLOv5 Nano version model underwent training. Post-training, the model was applied to the complete dataset, which incorporated both healthy and injured knees. The output generated by the model encompassed images with bounding boxes, which marked the presence and location of the ACL, accompanied by a probability value. This value represented the model's confidence level in identifying a healthy ACL. The premise of the model functioned on the expectation of high probability values (thus high confidence) for healthy knees and conversely, low confidence for injured knees.

The model's efficacy was objectively represented as a percentage, symbolizing graft integrity. To evaluate the accuracy and utility of this model-derived index, an external patient database from the University of Thessaly, with 120 ACL injured knees, was employed. Finally, the model analyzed pre-operatively and one-year post-operative MRI scans of 24 patients undergoing ACL reconstruction from September 2021 to May 2022.

The DL model yielded two types of outcomes. The first was a binary classification, designating the graft as either normal or not normal. The second was a model-derived index, named the TGI, indicative of the probability of an ACL graft closely resembling a native or healthy ACL. The aim of this methodological approach was to establish a more quantifiable and accurate measure for evaluating ACL integrity and the effectiveness of reconstruction procedures, leveraging the power of DL.

5.2.7. Statistical Analysis

The reliability of the TGI as an evaluation tool for ACL reconstruction outcomes was validated. This index was correlated with the all PROMs, namely the KOOS, the IKDC, the Lysholm, the TAS and also with the KT-1000 measurements. The correlation between the TGI and these scores was determined using the Spearman’s correlation coefficient, a non-parametric measure used to assess statistical dependence between the rankings of two variables. In addition to our analyses, we utilized the Mann-Whitney U test, a non-parametric method, to examine the statistical significance of differences in PROMs before and after surgery. This provided a robust evaluation of the impact of ACL reconstruction outcomes.

5.3. Results

In total, 24 patients (24 knees) were enrolled in the study. There were 15 male and 9 female patients. The mean age of the population was 31.39 years. The basic demographic status of the recruited patients can be seen in Table 12. Side to side knee stability was statistically significantly improved from pre-operative to post-operative measurements using the KT-1000 arthrometer ($p < 0.001$) from 6.31 ± 1.76 to 1.21 ± 1.45, respectively. Two patients presented more than 3 mm side to side difference on KT-1000, between the ACL reconstructed and the healthy knee (4mm and 5 mm). Post-operative measurements of KT-1000 were statistically significantly improved ($p < 0.001$) in comparison to the pre-operative ones, from 11.68 ± 1.88 to 6.76 ± 1.66. All PROMs statistically significantly improved postoperatively. Results are presented in Table 13. Moreover, all improvements in KOOS successfully reached the minimum clinically important cutoff values described in recent literature [137].

Variable	**Results**
No of patients	24
Sex	
Male	15 (62.5%)
Female	9 (37.5%)
Mean age in years (SD)	31.39 (10.26)
Mean height in cm (SD)	168.62 (9.01)
Mean weight in kg (SD)	73.96 (14.55)

Table 12: Basic demographic characteristics of the recruited patients on follow-up.

PROMs and Index		**Mean Scores Preoperatively**	**SD**	**Mean Scores 1 year Follow-up**	**SD**	**p-value**
IKDC		61.36	16.66	85.76	17.97	**0.001**
Lysholm		71.00	19.95	92.30	11.36	**0.001**
KOOS	*Overall*	67.36	24.19	89.39	13.24	**0.011**
	Pain	76.36	19.67	92.84	13.03	**0.010**
	Symptoms	76.29	21.33	91.47	13.06	**0.017**
	ADL	83.43	15.92	95.85	9.10	**0.012**
	Sports & Recreation	41.50	28.65	82	23.51	**0.034**
	QoL	51.00	20.43	82.76	21.25	**0.006**

Tegner Activity Scale	2.18	1.03	6.73	1.54	**< 0.001**
KT-1000 of injured knee	11.68	1.88	6.76	1.66	**< 0.001**
KT-1000 side to side differences	6.31	1.76	1.21	1.45	**< 0.001**
Thessaly Graft Index	64.21	8.96	82.37	3.53	**< 0.001**

*Significant p values (<0.05) are indicated in bold.

Table 13: Results of PROMs, KT-1000 measurements and TGI pre-operatively and 1 year postoperatively. (IKDC: International Knee Documentation Committee, KOOS: Knee Injury and Osteoarthritis Outcome Score, TAS: Tegner Activity Scale, ADL: Activities of Daily Living, QoL: Quality of Life, SD: Standard Deviation).

5.3.1. MRI evaluation and Deep learning algorithm

In the results obtained from the Knee MRI dataset, the ML model displayed an average performance measure of 99.5% mean average precision (mAP), underscoring its capability in discerning healthy knees. To categorize a graft as healthy, a minimum threshold on the percentile scale was established from the KneeMRI and the University of Thessaly Databases at 79.21% and an average 15% increase in TGI score between pre-operative and post-operative images. This threshold served as a key criterion for differentiating between healthy and injured knees.

The average pre-operative and twelve months post-operative TGI scores were 64.21 ± 8.96 and 82.37 ± 3.53, respectively ($p < 0.001$) (Table 13). Pre-operative and post-operative images can be seen in Figure 16Figure 17,Figure 18, andFigure 19. Twenty-two grafts were characterized as healed and two as re-ruptured with post-op TGI scores 71% and 42% respectively, indicating a partial ACL graft rupture and a complete rupture (Figure 18Figure 19). The radiologist's assessment was in total agreement with the TGI scores for both pre-operative and post-operative MRI.

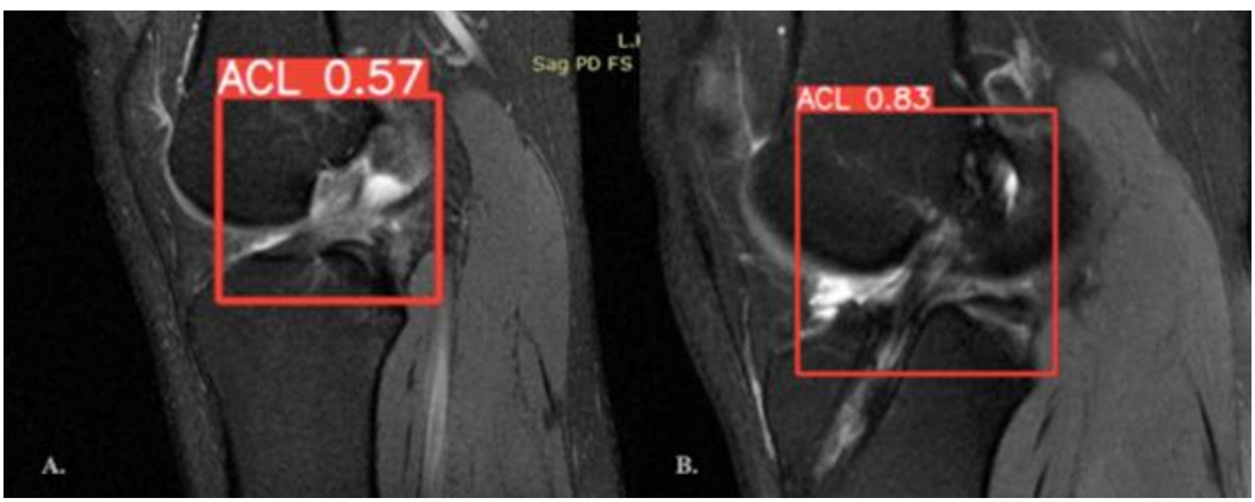


Figure 16: Sagittal T2-weighted images of a 33-year-old male patient with A) ACL rupture and TGI score of 57 on the preoperative MRI. B) ACL graft on the 1-year post operation MRI with TGI score of 83. IKDC, Lysholm, KOOS and TAS

score after ACL reconstruction were 78.2, 84, 87.2, and 4 respectively. (TGI: Thessaly Graft Index, IKDC: International Knee Documentation Committee, KOOS: Knee Injury and Osteoarthritis Outcome Score, TAS: Tegner Activity Scale)

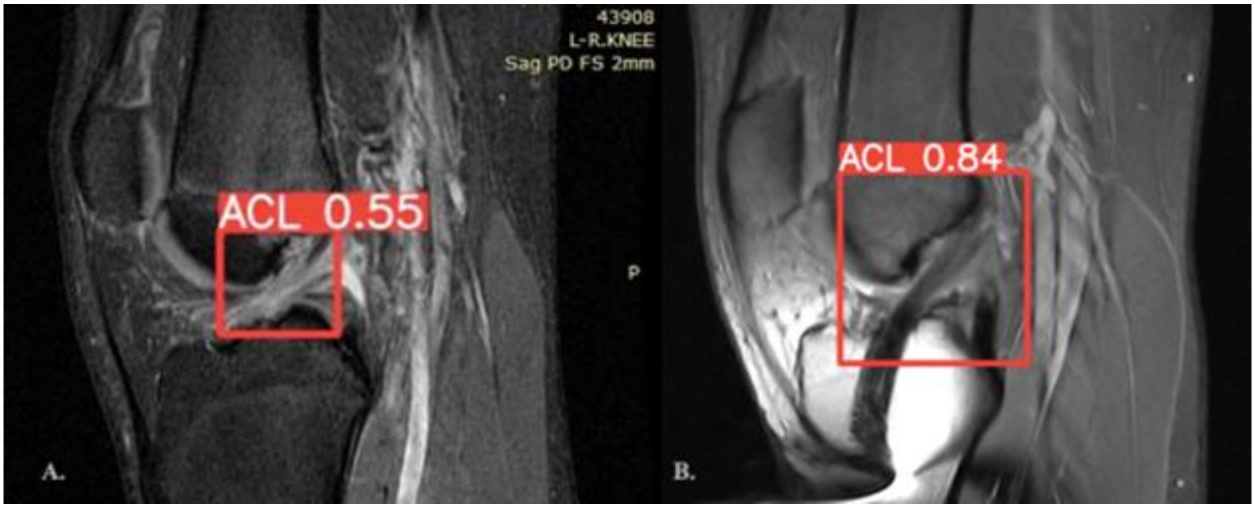


Figure 17: Sagittal T2-weighted images of a 17-year-old male patient with A) ACL rupture and TGI score of 55 on the preoperative MRI. B) ACL graft on the 1-year post operation MRI with TGI score of 84. IKDC, Lysholm, KOOS and TAS score after ACL reconstruction were 98,8, 100, 94, and 7 respectively. (TGI: Thessaly Graft Index, IKDC: International Knee Documentation Committee, KOOS: Knee Injury and Osteoarthritis Outcome Score, TAS: Tegner Activity Scale)

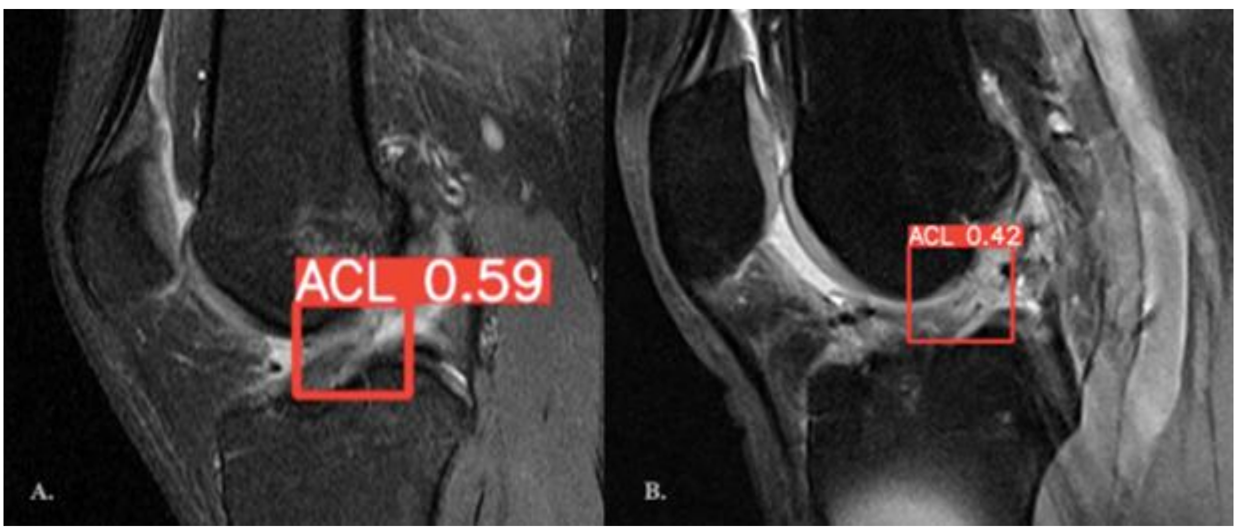


Figure 18: Sagittal T2-weighted images of a 25-year-old female patient with A) ACL rupture and TGI score of 59 on the preoperative MRI. B) ACL graft rupture and TGI score of 42 on the postoperative MRI. IKDC, Lysholm, KOOS and TAS score after ACL reconstruction were 36.8, 47, 66, and 4 respectively. (TGI: Thessaly Graft Index, IKDC: International Knee Documentation Committee, KOOS: Knee Injury and Osteoarthritis Outcome Score, TAS: Tegner Activity Scale)

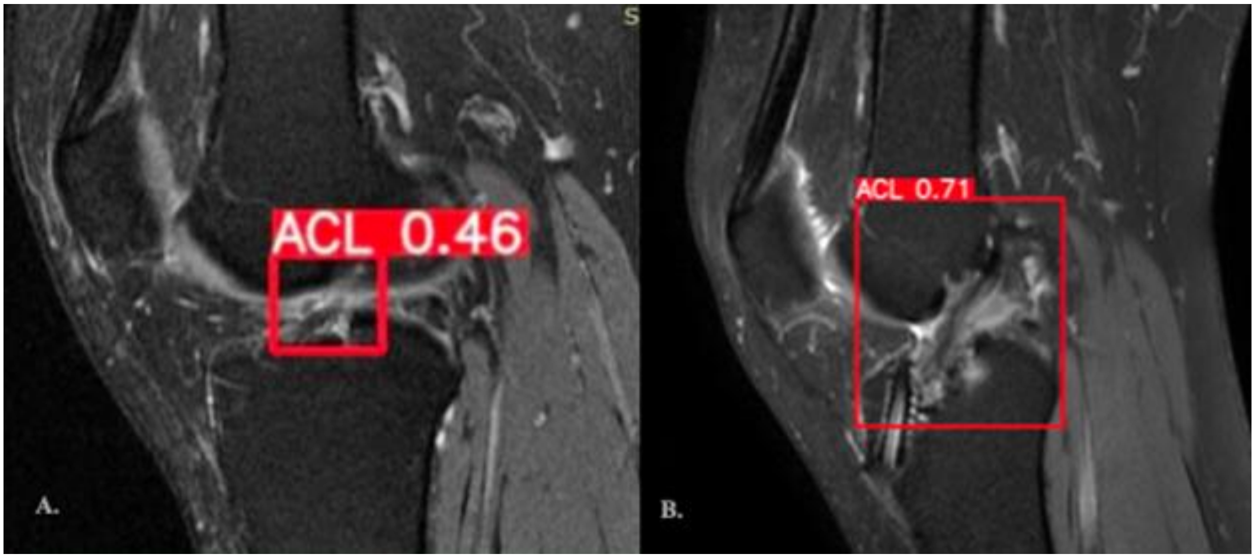


Figure 19: Sagittal T2-weighted images of a 25-year-old female patient with A) ACL rupture and TGI score of 46 on the preoperative MRI. B) ACL graft rupture and TGI score of 71 on the postoperative MRI. IKDC, Lysholm, KOOS and TAS score after ACL reconstruction were 58.6, 85, 72, and 4 respectively. (TGI: Thessaly Graft Index, IKDC: International Knee Documentation Committee, KOOS: Knee Injury and Osteoarthritis Outcome Score, TAS: Tegner Activity Scale).

5.3.2. PROMs and KT-1000 correlation with TGI

All PROMs were significantly correlated with the TGI. The correlation scores between PROMs and the TGI ranged from moderate to very good. The TAS and KOOS Sports and QoL sub-categories presented the higher correlation with 0.668, 0.663 and 0.621 Spearman's rho values. All correlations can be found in Table 14. Postoperative TGI and KT-1000 values were significantly correlated 0.561 ($p<0.001$), however, no correlation was noted between the TGI and the side-to-side postoperative differences ($p=0.108$) (Table 15).

	IKDC	LYSHOLM	KOOS Total	KOOS Pain	KOOS Symptoms	KOOS ADL	KOOS Sports	KOOS QoL	TAS
Spearman's Rho	0.516	0.521	0.594	0.594	0.504	0.597	0.663	0.621	0.668
p-value	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.01**	**<0.001**	**<0.001**	**<0.001**	**<0.001**

**Significant p values (<0.05) are indicated in bold.*

Table 14: Correlation of the TGI with the PROMs.

	KT-1000 post-operative	Postoperative side to side difference with KT-1000
Spearman's rho	**0.561**	0.380
p-value	**<0.001**	0.108

*Significant p values (<0.05) are indicated in bold.

Table 15: Correlation of the TGI with the KT-1000 postoperative values and knee side to side differences.

5.4. Discussion

The most important finding of this study is that the TGI is the first AI tool that could, via MRI, successfully assess healing of the ACL graft. Secondarily, the TGI achieved moderate to very good correlation with the PROMs one year after ACLR. The graft's healing rate was presented as a sensitive and easily reproducible method on a percentile scale. Very good correlation of the TGI was found with the TAS, KOOS QoL and KOOS Sports, followed by the KOOS Total, KOOS ADL, KOOS Pain, KOOS Symptoms subscale, the IKDC, the Lysholm score and the KT-1000 postoperative values. The TGI results were successfully validated by the musculoskeletal radiologist's diagnosis. The TGI tool successfully identified two ACL grafts with abnormal signal. The preoperative mean scores of the ACL were 64.21 ± 8.96 and the postoperative 82.37 ± 3.53. This incremental change in the postoperative index highlights the model's capability to detect improvements after ACL reconstruction procedures. The findings, therefore, underline the potential utility of this model in the clinical context, particularly in assisting with the assessment of the effectiveness of ACL reconstruction and the ongoing monitoring of patient recovery.

Although the role of MRI in monitoring and predicting outcomes after ACL reconstruction has been a challenge, limited studies until now are available in the literature and the results remain doubtful. On Van Dyck's et al. [118]m systematic review in 2019, only 7 of the 34 studies examined the correlation between MRI findings and clinical outcomes, while the T2-weighted graft signal was not considered a reliable factor on predicting clinical and functional outcomes after ACLR, at both early and long term follow-up. The role of MRI can be proved vital on the effort of researchers to determine the perfect timing for a patient after an ACLR to return to sports. Although evidence indicates that metabolic activity continues at 1 year 11 and shows that healing can go on beyond 2 years 43,44, Putnis et al. [113], noticed that in 42 patients there was no statistically significant change between signal intensity ratio of hamstring tendon autografts on MRI, 1 and 2 years, post ACLR. They also tried but failed to find any correlation between the MRI results, the PROMs and the patient's ability to return to sports. Similarly, Hong Li et al. [120] examined 38 patients after ACLR (21 with autograft and 17 with allograft) with MRI at 3, 6 and 12 months and tried to correlate the results with functional and physical examination, but no significant association between graft signal/noise quotient value and the IKDC, Lysholm and TAS scores was noted at each time point. Lutz et al. [123] also found poor correlation between ACL autograft signal intensity and the same PROMs in 17 male patients who underwent ACL reconstruction with hamstring autograft at 6 weeks, 3, 6, 12, and 24 months postoperatively, using 3T MRI. Along these results, Lin et al. [122]

studied the MRI of 54 patients (27 patients in the all-inside reconstruction group and 27 patients in the standard reconstruction group) after ACLR to measure the signal/noise quotient and assess the graft. The correlation between the graft maturity and the IKDC, TAS and Lysholm scores was not significant. Finally, Zhang et al. 24 enrolled 45 patients in two groups, using hamstrings graft (either free or with insertion preserved) but again no significant association was found between graft maturity (expressed with signal/noise quotient) and clinical scores.

On the other hand, recent studies point that correlation between MRI and clinical outcomes after ACLR exists but suggest that needs more research to clear this out. Biercevicz et al. [138] noticed that combined parameters of graft volume and median graft SI derived from T1-weighted 3D GRE MRI had the ability to predict clinical or in vivo outcomes in patients at 3 and 5 years post ACL reconstruction. More specifically in a multiple regression model the MRI parameters were correlated with 1-legged hop at both 3 and 5 years, demonstrating a possible way to assess graft strength in vivo. They also showed positive correlation with better patient reported outcomes in 4 out of 5 KOOS sub-scales. Li et al. [120] found that the MRI-based healing degree of the tendon bone interface was positively correlated with the functional scores at 12 months after ACLR. On a 250-patient case-control study, Putnis et al. [125] correlated the high graft signal with highest PROMs but also with higher risk of subsequent graft rupture. Finally, Flannery et al. [126] reported on their recent study that MRI measures of cross-sectional area, volume and estimated failure load were predictive of a positive outcome trajectory from 6 to 24 months after ACLR.

Although the interpretation of MR images relies on various factors, including the experience of the radiologist and parameters related to study protocol, i.e signal scaling factors, voxel volume, pulse sequence weighting, and patient positioning in the scanner which can complicate comparison between successive scans and between patients [139], [140], we believe that it can be still used to monitor the healing of tendon bone interface after ACLR, can be correlated with the knee joint functional recovery and could be also a criterion for return to sports. Latest research of the advancing technology of artificial intelligence and DL indicates that it can be facilitated through this path. In our study MRI examination was performed by a 3T MR scanner, providing a higher field strength and better imaging quality. Dedicated imaging protocols such as quantitative MRI ultrashort echo time T2* (UTE-T2*) and T2* mapping have been reported as alternative imaging protocols that outperform in evaluation of ACL graft maturation [141]. However, an advanced imaging protocol required longer scanning times and cannot be performed on routine basis in a busy radiology department. Besides, there is recently published data that rely on images with conventional protocols and produce acceptable suggestions regarding ACL graft maturation [123], [141], [142].

5.5. Limitation

The main limitation of this study is the size of the sample. Although it is a novel technique it may be underpowered to detect better correlations between PROMs and outcomes of TGI via MRI. While the sample size can be a common limiting factor which could be addressed with future studies (multi-center or not) reporting our primary results from this study is still important. The decision for return to sports and to previous level of activity for a patient after ACLR is still on debate and a tool that can provide an extra, individualized, aid on this direction can be crucial.

5.6. Conclusion

The TGI is an AI tool able to accurately recognize an ACL graft rupture. Moreover, in this pilot study, the TGI was found able to assess maturation of the ACL graft, using a sensitive and easily reproducible method and correlating with the KT-1000 postoperative values and the KOOS, IKDC, TAS, and Lysholm scores. Further development of the TGI can be an important step aiding in decision making regarding patient recovery and return to sports.

Chapter 6: General Conclusion and Future Work

6.1. Summary of Findings

This dissertation set out to enhance the application of DL techniques for the accurate diagnosis of ACL injuries and the assessment of ACL graft maturation using MRI studies. The research was structured around four primary studies: a systematic literature review of existing DL methods for knee injury detection; the development of a lightweight DL pipeline for identifying healthy ACLs in MRI data; the creation of a novel hybrid DL framework for ACL tear diagnosis; and the introduction and validation of the TGI for assessing ACL graft maturation post-reconstruction.

The initial phase involved conducting a comprehensive systematic literature review following the PRISMA guidelines. This review aimed to analyze existing DL approaches for knee injury detection, focusing on ACL tears as well as meniscus and cartilage injuries. Databases such as PubMed, Cochrane Library, EMBASE, and Google Scholar were meticulously searched. The studies reviewed employed various DL architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models.

Findings from the literature review indicated that DL models demonstrated prediction accuracies ranging from 72.5% to 100% for knee injury identification, with models utilizing transfer learning and pre-trained networks often achieving higher accuracies. However, several limitations were identified in the existing approaches. Many studies faced class imbalance issues, with fewer pathological cases compared to healthy ones, which affected model training and performance. Additionally, a lack of cross-dataset validation raised concerns about the generalizability of models across different imaging centers and patient populations. Some studies did not use independent test sets, leading to potential overestimation of model performance due to verification bias. Furthermore, variability in expert annotations introduced inconsistencies in the ground truth used for training and evaluation, highlighting the issue of ground-truth subjectivity. The review emphasized the need for developing models that are both computationally efficient and interpretable to facilitate clinical adoption.

Building upon these insights, a robust and lightweight DL pipeline was developed to identify healthy ACLs in 3D MRI data. The primary objective was to locate sagittal slices where the ACL is present, thereby aiding clinicians in focused evaluations. The pipeline employed the YOLOv5-Nano object detection network due to its balance of performance and computational efficiency. YOLOv5-Nano, being the smallest variant in the YOLOv5 family, is optimized for deployment on devices with limited resources. The model was trained and tested on augmented MRI datasets containing sagittal knee images with annotated ACL regions. Its performance was compared against other models, including YOLOv5-Xlarge, YOLOX-Small, and YOLOX-Nano.

The results demonstrated that YOLOv5-Nano achieved the highest mean Average Precision at 0.5 Intersection over Union (mAP@0.5) of 97.27% on augmented data. With a model size of 3.7 MB, YOLOv5-Nano was significantly smaller than its counterparts, facilitating faster inference times and lower computational demands. While YOLOv5-Xlarge and YOLOX-Small achieved competitive accuracies, they had larger model sizes and slower inference times. YOLOX-Nano, although similar in size, did not match the detection accuracy of YOLOv5-Nano. The lightweight nature of YOLOv5-Nano makes it suitable for integration into clinical workflows, including deployment on portable devices and real-time applications.

To address the challenges of class imbalance and improve diagnostic accuracy for ACL tears, a novel hybrid DL framework was developed. This framework reformulated the ACL tear classification task as a novelty detection problem and incorporated aleatoric semantic uncertainty. The YOLOv5-Nano model was utilized to generate class scores with associated uncertainties, and a highly discriminative novelty score was computed based on the aleatoric uncertainty. This approach enhanced the model's ability to distinguish between normal and abnormal cases. To account for tissue continuity in MRI sequences, global scores were calculated by aggregating the probability vectors across the entire sagittal sequence. The MINIROCKET algorithm, a fast transform for time series classification, was employed as the second module, processing the sequence of global scores to classify the knee as having an ACL tear or not.

The framework was trained and tested on two public datasets, KneeMRI and MRNet, and an independent validation was conducted using MRI data from the University General Hospital of Larissa, Greece. The results showed that on the KneeMRI dataset, the framework achieved an accuracy of 94.5%, outperforming state-of-the-art models with a statistically significant margin (p-value < 0.05). On the MRNet dataset, it achieved better accuracy and sensitivity compared to existing approaches. The model demonstrated robust performance across different datasets, indicating good generalizability. When trained on the KneeMRI dataset, the model maintained high accuracy on external validation data. Moreover, the framework consumed 2.6 times less energy and generated 2.1 times less carbon emissions compared to state-of-the-art models, highlighting its environmental sustainability. The integration of aleatoric uncertainty and novelty detection enhanced diagnostic precision, and the use of lightweight models like YOLOv5-Nano and MINIROCKET contributed to resource efficiency, making the framework suitable for widespread clinical application.

Recognizing the need for standardized assessment of ACL graft maturation post-reconstruction, the Thessaly Graft Index (TGI) was introduced and validated. This AI-based index quantifies graft healing by computing the probability of accurately detecting a healthy ACL on a percentage scale. In a pilot study involving 24 patients with isolated ACL injuries treated with hamstring tendon autografts, MRI scans were performed preoperatively and one year postoperatively. Clinical and functional evaluations included the KT-1000 arthrometer measurements and patient-reported outcome measures (PROMs) such as the Knee Injury and Osteoarthritis Outcome Score (KOOS), the International Knee Documentation Committee (IKDC) score, the Lysholm score, and the Tegner Activity Scale (TAS). The previously developed YOLOv5-Nano model was adapted to compute the TGI scores by analyzing the MRI scans and outputting a percentage representing graft health.

The findings indicated that the mean preoperative TGI score was 64.21%, while the mean postoperative TGI score was 82.37%, reflecting a mean increase of 15% between preoperative and postoperative images. A TGI score of 79.21% or higher categorized a graft as healthy on postoperative MRI. Out of the 24 grafts, 22 were characterized as healed, and 2 as re-ruptured with postoperative TGI scores of 71% and 42%, respectively. The TGI scores were in total agreement with evaluations by an independent radiologist. Additionally, TGI showed moderate to strong correlations with PROMs, including TAS (r = 0.668), IKDC (r = 0.516), Lysholm score (r = 0.521), and KOOS total (r = 0.594), supporting its validity as a tool for monitoring postoperative recovery. The TGI provides an objective, quantitative measure for assessing ACL graft maturation, and its correlation with PROMs and clinical assessments reinforces its potential clinical utility.

Collectively, the studies conducted in this dissertation contribute significantly to advancing the use of DL in ACL injury diagnosis and treatment. By integrating object detection, uncertainty modeling, novelty detection, and time-series classification, a robust framework for ACL injury diagnosis was established. The lightweight and efficient models facilitate real-time analysis, aiding clinicians in making prompt and accurate decisions. The resource-efficient nature of the models aligns with the growing emphasis on sustainable practices in healthcare technology. Furthermore, the models' generalizability and low computational requirements make them suitable for deployment in various clinical settings, including those with limited resources.

Overall, the developed models demonstrated high accuracy and sensitivity in detecting ACL injuries and assessing graft maturation. The use of lightweight architectures like YOLOv5-Nano and MINIROCKET ensured efficient processing without compromising performance. Cross-dataset testing confirmed the models' robustness and applicability to different patient populations and imaging protocols. The tools developed have practical applications in clinical practice, potentially improving patient outcomes through better diagnosis and monitoring. The findings of this dissertation highlight the significant potential of DL techniques in enhancing the diagnosis and treatment of ACL injuries. The integration of advanced algorithms with clinical expertise offers a promising pathway toward improved patient care in orthopedic medicine.

6.2. Contributions to the Field

This dissertation makes significant contributions to the field of orthopedic medicine and medical imaging by enhancing the diagnosis and treatment of ACL injuries through the application of advanced DL techniques. The work addresses critical challenges faced by clinicians in accurately diagnosing ACL injuries and monitoring postoperative recovery, ultimately aiming to improve patient outcomes.

One of the primary contributions is the development of a lightweight DL pipeline using the YOLOv5-Nano object detection network for identifying healthy ACLs in sagittal MRI slices. This model assists clinicians by rapidly locating the specific MRI slices where the ACL is present, thereby focusing their attention on the most relevant images for evaluation. The efficiency and high detection accuracy of the YOLOv5-Nano model enable it to function effectively on devices with limited computational resources, such as portable imaging systems. By streamlining the diagnostic workflow, this tool reduces the time clinicians spend sifting through extensive MRI data, allowing for quicker decision-making and potentially earlier intervention for patients.

In addressing the challenge of accurately diagnosing ACL tears, especially in the presence of class imbalance within datasets, the dissertation introduces a novel hybrid DL framework. This framework reformulates the ACL tear classification task as a novelty detection problem, incorporating aleatoric semantic uncertainty derived from the YOLOv5-Nano model's class scores. By accounting for uncertainty and tissue continuity across the entire sagittal MRI sequence, the model enhances robustness and reliability in detecting ACL tears. The integration of the MINIROCKET time-series classification algorithm further refines the diagnostic process by analyzing patterns across multiple images. For clinicians, this means a more dependable diagnostic tool that can distinguish between healthy and injured ligaments with greater precision, even when dealing with heterogeneous patient data or varying imaging conditions.

The introduction and validation of the TGI represent a significant advancement in postoperative assessment of ACL reconstruction. The TGI provides an objective, quantitative measure of ACL graft maturation by calculating the probability of accurately detecting a healthy ACL on a percentage scale from MRI images. This index correlates with established clinical evaluations and patient-reported outcome measures (PROMs), such as the Knee Injury and Osteoarthritis Outcome Score (KOOS), International Knee Documentation Committee (IKDC) score, Lysholm score, and Tegner Activity Scale (TAS). For orthopedic surgeons and rehabilitation specialists, the TGI serves as a valuable tool for monitoring patient recovery, assessing graft healing progress, and making informed decisions regarding rehabilitation protocols and the timing of returning to physical activities or sports. By offering a standardized assessment, the TGI reduces reliance on subjective interpretations of MRI scans and enhances the consistency of postoperative evaluations.

Clinically, the contributions of this dissertation offer several benefits to medical professionals:

1. **Improved Diagnostic Accuracy**: The developed models demonstrate high accuracy in detecting ACL injuries and assessing graft maturation, aiding clinicians in making precise diagnoses. Accurate identification of ACL tears is crucial for determining the appropriate treatment plan, whether it be surgical intervention or conservative management.

2. **Enhanced Efficiency**: By automating the detection of ACL presence in MRI slices and classifying tears with high reliability, the diagnostic process becomes more efficient. Clinicians can allocate their time more effectively, focusing on patient consultation and treatment planning rather than labor-intensive image analysis.

3. **Resource Optimization**: The lightweight nature of the models allows for deployment in various clinical settings, including those with limited computational infrastructure. This broadens the accessibility of advanced diagnostic tools to healthcare facilities that may not have extensive technological resources.

4. **Consistency in Evaluation**: The use of objective indices like the TGI minimizes inter-observer variability and subjectivity in interpreting MRI scans. Consistent evaluations contribute to better patient management and facilitate comparative studies across different clinical practices.

5. **Environmental Sustainability**: The resource-efficient models consume less energy and generate fewer carbon emissions compared to more complex architectures. This aligns with the healthcare industry's growing commitment to sustainability and responsible use of resources.

From a patient care perspective, these contributions have the potential to improve outcomes by enabling earlier and more accurate diagnoses, tailoring treatment plans to individual needs, and closely monitoring recovery progress. Patients benefit from reduced diagnostic delays, more personalized rehabilitation strategies, and potentially shorter recovery times. Additionally, the objective assessments provided by the models can enhance patient-clinician communication, as patients can be presented with clear, quantifiable information regarding their condition and progress.

The integration of these DL tools into clinical practice could transform the standard approach to diagnosing and treating ACL injuries. By providing clinicians with advanced analytical capabilities that are both efficient and accessible, the work supports a shift towards more data-driven and patient-centered care. The methodologies developed in this dissertation also have the potential to be adapted to other musculoskeletal injuries or imaging modalities, extending their impact beyond ACL injuries.

In summary, this dissertation contributes to the field by bridging the gap between advanced DL technologies and practical clinical applications in orthopedic medicine. It offers innovative solutions to longstanding challenges in ACL injury diagnosis and postoperative assessment, enhancing the capabilities of clinicians to deliver high-quality care. Through the development of efficient, accurate, and clinically relevant models and indices, the work paves the way for improved diagnostic processes, better patient outcomes, and a more sustainable approach to healthcare technology.

6.3. Limitations of the Study

While this dissertation advances the application of DL techniques in the diagnosis and assessment of ACL injuries, several limitations must be acknowledged.

One significant limitation pertains to the datasets used in the studies. The models were primarily trained and tested on specific public datasets, such as MRNet and KneeMRI, and a validation-only dataset from the University General Hospital of Larissa (UGHL), Greece. Although these datasets provided valuable information, they may not fully represent the diversity of MRI images encountered in different clinical settings. For instance, some MRI images, especially those taken before 2005, lacked the quality associated with modern magnetic resonance imaging scanners. This discrepancy in image quality could affect the performance and generalizability of the algorithms. The model trained on the KneeMRI dataset, which contained higher-quality images, generalized better on the UGHL dataset than the model trained on the MRNet dataset. Future research should include contemporary MRI scans acquired with advanced imaging technology to enhance the robustness and applicability of the models across various imaging conditions.

Another limitation involves the size and composition of the datasets. The presence of class imbalance, with a disproportionate number of healthy versus injured cases, may have influenced the model training and evaluation processes. While the hybrid DL framework addressed class imbalance by reformulating the classification task as a novelty detection problem, the overall effectiveness of this approach needs further validation with larger and more balanced datasets. Additionally, some MRI scans contained a limited number of sagittal slices due to large intervals at which the images were captured, resulting in low out-of-plane spatial resolution. This necessitated the use of padding in the models to standardize input sizes, which could introduce artifacts or affect the model's ability to learn relevant features. Ensuring that future datasets have sufficient spatial resolution and adequate numbers of image slices would enhance model performance and reliability.

The studies also faced limitations related to cross-dataset validation. Although the models demonstrated promising results on the datasets used, testing was limited to specific data sources. A more extensive cross-dataset validation involving a broader range of imaging centers and patient populations would provide stronger evidence of the models' generalizability and robustness. Such validation is crucial for clinical adoption, as it ensures that the models can perform reliably across different settings and imaging protocols.

In the development of the lightweight DL pipeline using YOLOv5-Nano, the model was primarily focused on detecting the presence of the ACL in MRI slices but did not distinguish between healthy and injured ligaments. While the observed differences in object detection probabilities between subjects with healthy and injured ACLs suggest potential for classification, the model's inability to explicitly differentiate between the two limits its immediate clinical utility. Future work should aim to extend the model's capabilities to include classification tasks, potentially by leveraging the probability outputs as inputs for additional classification algorithms.

Regarding the TGI, the main limitation lies in the sample size of the pilot study. With only 24 patients included, the study may be underpowered to detect stronger correlations between the TGI and patient-reported outcome measures (PROMs). While the initial results are promising, larger-scale studies, potentially multi-center collaborations, are necessary to validate the TGI's effectiveness fully. Additionally, the decision for patients to return to sports and previous activity levels after ACL reconstruction remains a complex issue. While the TGI offers an individualized assessment tool that could aid in this decision-making process, further research is needed to establish standardized thresholds and protocols for its use in clinical practice.

Lastly, the environmental sustainability assessment, while highlighting the reduced energy consumption and carbon emissions of the proposed models compared to state-of-the-art approaches, did not account for all potential environmental impacts. Factors such as the energy sources used, the lifecycle emissions of the hardware, and the broader ecological footprint of deploying these models in clinical settings were not fully explored. A more comprehensive environmental impact analysis would strengthen the case for the models' sustainability benefits.

In summary, while the studies presented in this dissertation offer valuable advancements in the application of DL for ACL injury diagnosis and assessment, they are subject to limitations related to dataset quality and diversity, sample size, model capabilities, and validation scope. Addressing these limitations in future research will be essential to enhance the models' clinical applicability, generalizability, and overall impact on patient care.

6.4. Recommendations for Future Research

Building upon the findings and acknowledging the limitations of this study, several avenues for future research emerge that could further advance the application of DL in the diagnosis and assessment of ACL injuries.

One primary recommendation is to expand the scope and quality of the datasets used for training and validating the models. The current research utilized MRI images from datasets that included scans taken with older imaging technology, some dating back to before 2005. These images may lack the resolution and clarity provided by modern MRI scanners. Future studies should incorporate contemporary MRI scans acquired with the latest imaging technology to enhance image quality. High-resolution images from recent scanners can improve the models' ability to detect subtle anatomical details and pathological changes, potentially increasing diagnostic accuracy. Additionally, utilizing data from updated scanners would ensure that the models remain relevant and effective in current clinical settings.

Expanding the datasets to include a larger and more diverse population is also crucial. A more extensive cross-dataset validation involving multiple imaging centers, different MRI machines, and diverse patient demographics would enhance the models' generalizability. This diversity would help the models learn a broader range of anatomical variations and pathological presentations, reducing the risk of overfitting to a specific dataset or imaging protocol. Collaborations with multiple institutions could facilitate the collection of such comprehensive datasets.

Another area for future research is the extension of these DL techniques to other types of musculoskeletal injuries and conditions. While this dissertation focused on ACL injuries, the methodologies developed—such as the lightweight object detection models and the hybrid classification framework—could be adapted to detect and assess injuries of other ligaments, tendons, and cartilaginous structures within the knee or other joints. For instance, meniscal tears, cartilage defects, and injuries to the posterior cruciate ligament (PCL) or collateral ligaments could be potential targets for similar diagnostic models. Expanding the application of these techniques would contribute to a more holistic approach to musculoskeletal imaging analysis.

In the rapidly evolving landscape of artificial intelligence, the emergence of large language models (LLMs) presents new opportunities for integrating multimodal systems into medical imaging analysis. Future research could explore the development of advanced multimodal models that combine imaging data with textual information, such as radiology reports, clinical notes, and patient histories. By incorporating LLMs, it would be possible to create intelligent chatbots or virtual assistants capable of generating detailed diagnostic reports based on the analysis of MRI images. These systems could provide clinicians with comprehensive insights by correlating imaging findings with clinical data, enhancing decision-making processes.

Such multimodal systems could facilitate interactive platforms where clinicians can query the model for specific information or clarification regarding a diagnosis. For example, a chatbot could explain the rationale behind a particular diagnostic suggestion, increasing the transparency and interpretability of the model's outputs. This integration would align with the growing emphasis on explainable AI in healthcare, where understanding the reasoning behind a model's decision is essential for clinical trust and adoption.

Moreover, integrating these advanced models into clinical workflows could provide real-time support during patient consultations or surgical planning. With the capability to process imaging data quickly and generate immediate reports, clinicians could offer more timely diagnoses and personalized treatment plans. Future research should investigate the feasibility and effectiveness of deploying such systems in real-world clinical environments, considering factors like user interface design, data security, and compliance with healthcare regulations.

Addressing the limitations related to the models' current capabilities, future work should focus on enhancing the models to not only detect the presence of the ACL in MRI slices but also to classify the ligament as healthy or injured explicitly. This could involve developing new classification algorithms that utilize the probability outputs from the object detection models as inputs. Additionally, improving the models' ability to handle variability in image quality and scanning protocols remains an important area of research. Techniques such as domain adaptation and data augmentation could be employed to make the models more robust to different imaging conditions.

Considering the environmental impact of deploying AI models, future studies should conduct comprehensive environmental assessments, including factors such as the energy sources used, the lifecycle emissions of the hardware, and the broader ecological footprint of implementing these technologies in healthcare settings. Exploring ways to optimize models for energy efficiency without compromising performance would contribute to sustainable practices in medical AI applications.

In summary, future research should aim to:

- **Utilize Contemporary MRI Scans**: Incorporate high-quality images from modern MRI scanners to improve model performance and relevance in current clinical practice.
- **Expand and Diversify Datasets**: Include larger, more diverse datasets from multiple centers to enhance model generalizability and robustness.
- **Extend Applications to Other Injuries**: Adapt the developed methodologies to detect and assess other musculoskeletal injuries, promoting comprehensive musculoskeletal care.
- **Integrate Multimodal Systems with LLMs**: Develop advanced models that combine imaging data with textual information, creating intelligent systems capable of generating detailed diagnostic reports and assisting clinicians interactively.
- **Enhance Model Capabilities**: Improve models to explicitly classify ligaments as healthy or injured and to handle variability in imaging conditions effectively.
- **Assess Environmental Impact**: Conduct thorough environmental impact analyses and optimize models for energy efficiency to support sustainable AI practices in healthcare.

By pursuing these recommendations, future research can build upon the foundations laid in this dissertation, further advancing the integration of DL into medical imaging. Such advancements hold the promise of improving diagnostic accuracy, enhancing patient care, and supporting clinicians with powerful, intelligent tools in the evolving landscape of healthcare technology.

6.5. Summary of findings

This dissertation has explored the application of DL techniques in the diagnosis and assessment of ACL injuries using MRI studies. Through a systematic approach that included a comprehensive literature review, the development of lightweight DL models, the creation of a novel hybrid diagnostic framework, and the introduction of the TGI, this work has demonstrated the significant potential of artificial intelligence to enhance orthopedic diagnostics.

The developed models address critical challenges faced by clinicians, such as the need for efficient, accurate, and accessible tools for diagnosing ACL injuries and monitoring postoperative recovery. By leveraging models like YOLOv5-Nano and MINIROCKET, the research provided solutions that are both technically robust and practical for real-world clinical settings. The TGI offers an objective measure for assessing graft maturation, contributing to improved patient management and outcomes.

While limitations were acknowledged, including dataset constraints and the necessity for broader validation, the findings lay a foundation for future advancements. The recommendations for future research emphasize the importance of incorporating contemporary imaging data, expanding the application of these techniques to other musculoskeletal injuries, and exploring the integration of multimodal systems with large language models.

In conclusion, this dissertation contributes to bridging the gap between advanced DL methodologies and their practical application in clinical practice. By enhancing diagnostic accuracy and efficiency, it supports the overarching goal of improving patient care in orthopedic medicine. The continued evolution of artificial intelligence presents exciting opportunities, and further research in this direction holds promise for significant impacts on healthcare delivery and patient well-being.

References

[1] N. A. Mall *et al.*, "Incidence and trends of anterior cruciate ligament reconstruction in the United States," *Am. J. Sports Med.*, vol. 42, no. 10, pp. 2363–2370, 2014, doi: 10.1177/0363546514542796.

[2] T. L. Sanders *et al.*, "Incidence of anterior cruciate ligament tears and reconstruction: A 21-year population-based study," *Am. J. Sports Med.*, vol. 44, no. 6, pp. 1502–1507, 2016, doi: 10.1177/0363546516629944.

[3] E. A. Arendt, J. Agel, and R. Dick, "Anterior Cruciate Ligament Injury Patterns among Collegiate Men and Women," *J. Athl. Train.*, vol. 34, no. 2, pp. 86–92, 1999.

[4] T. E. Hewett *et al.*, "Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: A prospective study," *Am. J. Sports Med.*, vol. 33, no. 4, pp. 492–501, 2005, doi: 10.1177/0363546504269591.

[5] et al. Herzog MM, Marshall SW, Lund JL, "Rising Rates of ACL Reconstruction in Young Females in the United States.," *J. Bone Jt. Surgery.*, pp. 1769–1776, 2017.

[6] V. Musahl and J. Karlsson, "Anterior Cruciate Ligament Tear," *N. Engl. J. Med.*, vol. 380, no. 24, pp. 2341–2348, 2019, doi: 10.1056/nejmcp1805931.

[7] I. Ahmed *et al.*, "Meniscal tear outcome Study (METRO Study): A study protocol for a multicentre prospective cohort study exploring the factors which affect outcomes in patients with a meniscal tear," *BMJ Open*, vol. 10, no. 7, pp. 1–7, 2020, doi: 10.1136/bmjopen-2020-038681.

[8] C. J. Darrow, C. L. Collins, E. E. Yard, and R. D. Comstock, "Epidemiology of severe injuries among United States high school athletes 2005-2007," *Am. J. Sports Med.*, vol. 37, no. 9, pp. 1798–1805, 2009, doi: 10.1177/0363546509333015.

[9] B. E. Gage, N. M. McIlvain, C. L. Collins, S. K. Fields, and R. D. Comstock, “Epidemiology of 6.6 million knee injuries presenting to United States emergency departments from 1999 through 2008,” *Acad. Emerg. Med.*, vol. 19, no. 4, pp. 378–385, 2012, doi: 10.1111/j.1553-2712.2012.01315.x.

[10] G. Merkely, J. Ackermann, and C. Lattermann, “Articular Cartilage Defects: Incidence, Diagnosis, and Natural History,” *Oper. Tech. Sports Med.*, vol. 26, no. 3, pp. 156–161, 2018, doi: 10.1053/j.otsm.2018.06.008.

[11] L. Zhang, M. Li, Y. Zhou, G. Lu, and Q. Zhou, “Deep Learning Approach for Anterior Cruciate Ligament Lesion Detection: Evaluation of Diagnostic Performance Using Arthroscopy as the Reference Standard,” *J. Magn. Reson. Imaging*, pp. 1–8, 2020, doi: 10.1002/jmri.27266.

[12] C. C. Kaeding, B. Léger-St-Jean, and R. A. Magnussen, “Epidemiology and Diagnosis of Anterior Cruciate Ligament Injuries,” *Clin. Sports Med.*, vol. 36, no. 1, pp. 1–8, 2017, doi: 10.1016/j.csm.2016.08.001.

[13] P. D. Chang, T. T. Wong, and M. J. Rasiej, “Deep Learning for Detection of Complete Anterior Cruciate Ligament Tear,” *J. Digit. Imaging*, vol. 32, no. 6, pp. 980–986, 2019, doi: 10.1007/s10278-019-00193-4.

[14] D. S. Logerstedt *et al.*, “Knee pain and mobility impairments: meniscal and articular cartilage lesions.,” *J. Orthop. Sports Phys. Ther.*, vol. 40, no. 6, pp. 1–58, 2010, doi: 10.2519/jospt.2010.0304.

[15] R. C. Mather *et al.*, “Societal and economic impact of anterior cruciate ligament tears,” *J. Bone Jt. Surg.*, vol. 95, no. 19, pp. 1751–1759, 2013, doi: 10.2106/JBJS.L.01705.

[16] K. L. Cameron, B. S. Thompson, K. Y. Peck, B. D. Owens, S. W. Marshall, and S. J. Svoboda, “Normative values for the KOOS and WOMAC in a young athletic population: History of knee ligament injury is associated with lower scores,” *Am. J. Sports Med.*, vol. 41, no. 3, pp. 582–589, 2013, doi: 10.1177/0363546512472330.

[17] C. M. Celano, M. E. Freedman, E. E. Beale, F. Gomez-Bernal, and J. C. Huffman, “A Positive Psychology Intervention to Promote Health Behaviors in Heart Failure: A Proof-of-Concept Trial,” *J. Nerv. Ment. Dis.*, vol. 206, no. 10, pp. 800–808, 2018, doi: 10.1097/NMD.0000000000000883.

[18] K. C. Lam, S. S. Thomas, A. R. S. Valier, T. C. V. McLeod, and R. C. Bay, “Previous knee injury and health-related quality of life in collegiate athletes,” *J. Athl. Train.*, vol. 52, no. 6, pp. 534–540, 2017, doi: 10.4085/1062-6050-50.5.01.

[19] M. Pouly, T. Koller, P. Gottfrois, and S. Lionetti, “Artificial intelligence in image analysis—fundamentals and new developments,” *Hautarzt*, vol. 71, no. 9, pp. 660–668, 2020, doi: 10.1007/s00105-020-04663-7.

[20] E. R. Garwood, R. Tai, G. Joshi, and G. J. Watts V, “The Use of Artificial Intelligence in the Evaluation of Knee Pathology,” *Semin. Musculoskelet. Radiol.*, vol. 24, no. 1, pp. 21–29, 2020, doi: 10.1055/s-0039-3400264.

[21] B. Fritz, G. Marbach, F. Civardi, S. F. Fucentese, and C. W. A. Pfirrmann, “Deep convolutional neural network-based detection of meniscus tears: comparison with radiologists and surgery as standard of reference,” *Skeletal Radiol.*, vol. 49, no. 8, pp. 1207–1217, 2020, doi: 10.1007/s00256-020-03410-2.

[22] S. Palermi *et al.*, “Indirect structural muscle injuries of lower limb: Rehabilitation and therapeutic exercise,” *J. Funct. Morphol. Kinesiol.*, vol. 6, no. 3, 2021, doi: 10.3390/jfmk6030075.

[23] F. SIRICO, S. PALERMI, B. MASSA, and B. CORRADO, “Tendinopathies of the hip and pelvis in athletes: A narrative review,” *J. Hum. Sport Exerc.*, vol. 15, no. September, pp. S748–S762, 2020, doi: 10.14198/jhse.2020.15.Proc3.25.

[24] I. Hetsroni, S. Lyman, H. Do, G. Mann, and R. G. Marx, “Symptomatic pulmonary embolism after outpatient arthroscopic procedures of the knee: The incidence and risk factors in 418 323 arthroscopies,” *J. Bone Jt. Surg. - Ser. B*, vol. 93 B, no. 1, pp. 47–51, 2011, doi: 10.1302/0301-620X.93B1.25498.

[25] H. O. Alanazi, A. H. Abdullah, and K. N. Qureshi, “A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care,” *J. Med. Syst.*, vol. 41, no. 4, 2017, doi: 10.1007/s10916-017-0715-6.

[26] M. J. Prickett, W.D.;Ward, S.I.; Matava, “Magnetic resonance imaging of the knee,” *Sport. Med.*, vol. 31, 997–10, 2001.

[27] W. Krampla, M. Roesel, K. Svoboda, A. Nachbagauer, M. Gschwantler, and W. Hruby, “MRI of the knee: How do field strength and radiologist’s experience influence diagnostic accuracy and interobserver correlation in assessing chondral and meniscal lesions and the integrity of the anterior cruciate ligament?,” *Eur. Radiol.*, vol. 19, no. 6, pp. 1519–1528, 2009, doi: 10.1007/s00330-009-1298-5.

[28] R. Mohankumar, L. M. White, and A. Naraghi, “Pitfalls and pearls in MRI of the knee,” *Am. J. Roentgenol.*, vol. 203, no. 3, pp. 516–530, 2014, doi: 10.2214/AJR.14.12969.

[29] C. Germann *et al.*, “Deep Convolutional Neural Network-Based Diagnosis of Anterior Cruciate Ligament Tears: Performance Comparison of Homogenous Versus Heterogeneous Knee MRI Cohorts with Different Pulse Sequence Protocols and 1.5-T and 3-T Magnetic Field Strengths,” *Invest. Radiol.*, vol. 55, no. 8, pp. 499–506, 2020, doi: 10.1097/RLI.0000000000000664.

[30] S. Gyftopoulos, D. Lin, F. Knoll, A. M. Doshi, T. C. Rodrigues, and M. P. Recht, “Artificial intelligence in musculoskeletal imaging: Current status and future directions,” *Am. J. Roentgenol.*, vol. 213, no. 3, pp. 506–513, 2019, doi: 10.2214/AJR.19.21117.

[31] G. W.-I. S. Dinggang Shen, “Deep Learning in Medical Image Analysis,” *Annu. Rev. Biomed. Eng.*, vol. 19, 2017.

[32] N. Bien *et al.*, “Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet,” *PLoS Med.*, vol. 15, no. 11, pp. 1–19, 2018, doi: 10.1371/journal.pmed.1002699.

[33] D. W. G. Langerhuizen *et al.*, “What Are the Applications and Limitations of Artificial Intelligence for Fracture Detection and Classification in Orthopaedic Trauma Imaging? A Systematic Review,” *Clin. Orthop. Relat. Res.*, vol. 477, no. 11, pp. 2482–2491, 2019, doi: 10.1097/CORR.0000000000000848.

[34] K. N. Kunze *et al.*, “Diagnostic Performance of Artificial Intelligence for Detection of Anterior Cruciate Ligament and Meniscus Tears: A Systematic Review,” *Arthrosc. - J. Arthrosc. Relat. Surg.*, 2020, doi: 10.1016/j.arthro.2020.09.012.

[35] A. Hellerstein, J.M.; Naughton, J.F.; Pfeffer, “Generalized Search Trees for Database

Systems.," *Proc. 21st VLDB Conf. Zurich, Switz.*, 1995.

[36] F. Liu, "SUSAN: segment unannotated image structure using adversarial network," *Magn. Reson. Med.*, vol. 81, no. 5, pp. 3330–3345, 2019, doi: 10.1002/mrm.27627.

[37] K. Q. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, "Densely connected convolutional networks.," *Proc. IEEE Conf. Comput. Vis. Pattern Recognition, Honolulu, HI, USA*, pp. 4700–4708, 2017.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[40] M. Freeman, W.T.; Roth, "Orientation histograms for hand gesture recognition.," *Proc. Int. Work. Autom. Face Gesture Recognition, Zurich, Switz.*, pp. 296–301, 1995.

[41] K. Sebastian, V.B.; Unnikrishnan, A.; Balakrishnan, "Gray level co-occurrence matrices: Generalisation and some new features.," *arXiv*, no. 1205.4831, 2012.

[42] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 1883, 2009.

[43] B. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, "Support vector machines.," *IEEE Intell. Syst. Appl.*, vol. 13, pp. 18–28, 1998.

[44] M. H. Hassoun, "Fundamentals of Artificial Neural Networks," *MIT Press Cambridge, MA, USA*, 1995.

[45] R. O'Shea, K.; Nash, "An introduction to convolutional neural networks.," *arXiv:1511.08458*, 2015.

[46] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[48] B. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation.," *Med. Image Anal.*, no. 36, pp. 61–78, 2017.

[49] G. . Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.;Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold and S. J. A. P. A. Gelly, "An image is worth 16_16 words: Transformers for image recognition at scale.," *arXiv:2010.11929*, 2020.

[50] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *BMJ*, vol. 339, no. 7716, pp. 332–336, 2009, doi: 10.1136/bmj.b2535.

[51] J. J. A. J. O. S. Slim, K.; Nini, E.; Forestier, D.; Kwiatkowski, F.; Panis, Y.; Chipponi, "Methodological index for non-randomized studies (MINORS): Development and validation of a new instrument.," *ANZ J. Surg.*, vol. 73, pp. 712–716, 2003.

[52] I. Štajduhar, M. Mamula, D. Miletić, and G. Ünal, “Semi-automated detection of anterior cruciate ligament injury from MRI,” *Comput. Methods Programs Biomed.*, vol. 140, pp. 151–164, 2017, doi: 10.1016/j.cmpb.2016.12.006.

[53] M. Pal, “Random forest classifier for remote sensing classification.,” *Int. J. Remote Sens.*, vol. 26, pp. 217–222, 2005.

[54] A. A. Abdullah and N. S. F. Azz-Zahra Md Som, “Design of an intelligent diagnostic system for detection of knee injuries,” *Appl. Mech. Mater.*, vol. 339, pp. 219–224, 2013, doi: 10.4028/www.scientific.net/AMM.339.219.

[55] S. S. Mazlan, M. Z. Ayob, and Z. A. K. Bakti, “Anterior cruciate ligament (ACL) injury classification system using support vector machine (SVM),” *2017 Int. Conf. Eng. Technol. Technopreneurship, ICE2T 2017*, vol. 2017-Janua, no. September, pp. 1–5, 2017, doi: 10.1109/ICE2T.2017.8215960.

[56] 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,” pp. 149–155, 2020, doi: 10.1109/idsta50958.2020.9264030.

[57] F. Liu *et al.*, “Deep learning approach for evaluating knee MR images: Achieving high diagnostic performance for cartilage lesion detection,” *Radiology*, vol. 289, no. 1, pp. 160–169, 2018, doi: 10.1148/radiol.2018172986.

[58] N. K. Namiri *et al.*, “Deep Learning for Hierarchical Severity Staging of Anterior Cruciate Ligament Injuries from MRI,” *Radiol. Artif. Intell.*, vol. 2, no. 4, p. e190207, 2020, doi: 10.1148/ryai.2020190207.

[59] M. J. Awan, M. S. M. Rahim, N. Salim, M. A. Mohammed, B. Garcia-Zapirain, and K. H. Abdulkareem, “Efficient Detection of Knee Anterior Cruciate Ligament from Magnetic Resonance Imaging Using Deep Learning Approach,” *Diagnostics*, vol. 11, no. 1, 2021, doi: 10.3390/diagnostics11010105.

[60] Y. S. Jeon *et al.*, “Interpretable and Lightweight 3-D Deep Learning Model for Automated ACL Diagnosis,” *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 7, pp. 2388–2397, 2021, doi: 10.1109/JBHI.2021.3081355.

[61] Y. Dai, Y. Gao, and F. Liu, “Transmed: Transformers advance multi-modal medical image classification,” *Diagnostics*, vol. 11, no. 8, pp. 1–15, 2021, doi: 10.3390/diagnostics11081384.

[62] B. Astuto *et al.*, “Automatic deep learning–assisted detection and grading of abnormalities in knee MRI studies,” *Radiol. Artif. Intell.*, vol. 3, no. 3, 2021, doi: 10.1148/ryai.2021200165.

[63] J. C. Fu, C. C. Lin, C. N. Wang, and Y. K. Ou, “Computer-aided diagnosis for knee meniscus tears in magnetic resonance imaging,” *J. Ind. Prod. Eng.*, vol. 30, no. 2, pp. 67–77, 2013, doi: 10.1080/10170669.2012.761285.

[64] M. H. F. Zarandi, A. Khadangi, F. Karimi, and I. B. Turksen, “A Computer-Aided Type-II Fuzzy Image Processing for Diagnosis of Meniscus Tear,” *J. Digit. Imaging*, vol. 29, no. 6, pp. 677–695, 2016, doi: 10.1007/s10278-016-9884-y.

[65] V. Couteaux *et al.*, “Automatic knee meniscus tear detection and orientation classification with Mask-RCNN,” *Diagn. Interv. Imaging*, vol. 100, no. 4, pp. 235–242, 2019, doi: 10.1016/j.diii.2019.03.002.

[66] V. Roblot *et al.*, “Artificial intelligence to diagnose meniscus tears on MRI,” *Diagn. Interv. Imaging*, vol. 100, no. 4, pp. 243–249, 2019, doi: 10.1016/j.diii.2019.02.007.

[67] V. Pedoia, B. Norman, S. N. Mehany, M. D. Bucknor, T. M. Link, and S. Majumdar, “3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects,” *J. Magn. Reson. Imaging*, vol. 49, no. 2, pp. 400–410, 2019, doi: 10.1002/jmri.26246.

[68] T. Ronneberger, O.; Fischer, P.; Brox, “U-net: Convolutional networks for biomedical image segmentation.,” *Proc. Int. Conf. Med. Image Comput. Comput. Interv. Munich, Ger.*, pp. 234–241, 2015.

[69] B. Rizk *et al.*, “Meniscal lesion detection and characterization in adult knee MRI: A deep learning model approach with external validation,” *Phys. Medica*, vol. 83, no. January, pp. 64–71, 2021, doi: 10.1016/j.ejmp.2021.02.010.

[70] S. Targ, D. Almeida, and K. Lyman, “Resnet in Resnet: Generalizing Residual Architectures,” pp. 1–7, 2016, [Online]. Available: http://arxiv.org/abs/1603.08029.

[71] F. Liu *et al.*, “Fully Automated Diagnosis of Anterior Cruciate Ligament Tears on Knee MR Images by Using Deep Learning,” *Radiol. Artif. Intell.*, vol. 1, no. 3, p. 180091, 2019, doi: 10.1148/ryai.2019180091.

[72] P. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, “Gradient-based learning applied to document recognition.,” *Proc. IEEE*, vol. 86, pp. 2278–2324, 1998.

[73] W. Liu *et al.*, “SSD: Single shot multibox detector,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.

[74] P. H. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, “Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers.,” *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognition, Online*, pp. 6881–6890, 2021.

[75] I. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, “Attention is all you need.,” *Proc. Adv. Neural Inf. Process. Syst. Long Beach, CA, USA*, pp. 5998–6008, 2017.

[76] K. J. A. P. A. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding.,” *arXiv:1810.04805*, 2018.

[77] A. J. A. P. A. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, “Language models are few-shot learners.,” *Adv. Neural Inf. Process. Syst*, vol. 33, pp. 1877–1901, 2020.

[78] A. M. J. A. P. A. Kouw, W.M.; Loog, M.; Bartels, L.W.; Mendrik, “MR acquisition-invariant representation learning.,” *arXiv*, 2017.

[79] A. V. J. A. P. A. Billot, B.; Greve, D.; Van Leemput, K.; Fischl, B.; Iglesias, J.E.; Dalca, “A learning strategy for contrast-agnostic MRI segmentation.,” *arXiv*, 2020.

[80] V. Musahl and J. Karlsson, “Anterior Cruciate Ligament Tear,” *N. Engl. J. Med.*, vol. 380, no. 24, pp. 2341–2348, 2019, doi: 10.1056/nejmcp1805931.

[81] S. Warach, C. S. Kidwell, and A. E. Baird, “Magnetic Resonance Imaging,” *Stroke Pathophysiol. Diagnosis, Manag.*, pp. 461–479, 2004, doi: 10.1016/B0-44-306600-0/50024-9.

[82] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, no. December 2012, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[83] A. Siouras *et al.*, “Knee Injury Detection Using Deep Learning on MRI Studies: A Systematic Review,” *Diagnostics*, vol. 12, no. 2, pp. 1–21, 2022, doi: 10.3390/diagnostics12020537.

[84] L. Wang, J. Liao, and C. Xu, “Vehicle detection based on drone images with the improved faster R-Cnn,” *PervasiveHealth Pervasive Comput. Technol. Healthc.*, vol. Part F1481, pp. 466–471, 2019, doi: 10.1145/3318299.3318383.

[85] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization,” *Rev. do Hosp. das Cl??nicas*, vol. 17, pp. 331–336, 2016, [Online]. Available: http://arxiv.org/abs/1610.02391.

[86] “Roboflow.” https://roboflow.com/.

[87] G. Jocher, “Yolov5.” https://github.com/ultralytics/yolov5.

[88] YOLOX, “No Title,” 2021. https://github.com/Megvii-BaseDetection/YOLOX.

[89] D. D. Anderson *et al.*, “Post-traumatic osteoarthritis: Improved understanding and opportunities for early intervention,” *J. Orthop. Res.*, vol. 29, no. 6, pp. 802–809, 2011, doi: 10.1002/jor.21359.

[90] I. Papathanasiou *et al.*, “Molecular changes indicative of cartilage degeneration and osteoarthritis development in patients with anterior cruciate ligament injury Pathophysiology of musculoskeletal disorders,” *BMC Musculoskelet. Disord.*, vol. 17, no. 1, pp. 1–10, 2016, doi: 10.1186/s12891-016-0871-8.

[91] L. S. Lohmander, P. M. Englund, L. L. Dahl, and E. M. Roos, “The long-term consequence of anterior cruciate ligament and meniscus injuries: Osteoarthritis,” *Am. J. Sports Med.*, vol. 35, no. 10, pp. 1756–1769, 2007, doi: 10.1177/0363546507307396.

[92] N. Phelan, P. Rowland, R. Galvin, and J. M. O’Byrne, “A systematic review and meta-analysis of the diagnostic accuracy of MRI for suspected ACL and meniscal tears of the knee,” *Knee Surgery, Sport. Traumatol. Arthrosc.*, vol. 24, no. 5, pp. 1525–1539, 2016, doi: 10.1007/s00167-015-3861-8.

[93] R. Crawford, G. Walley, S. Bridgman, and N. Maffulli, “Magnetic resonance imaging versus arthroscopy in the diagnosis of knee pathology, concentrating on meniscal lesions and ACL tears: A systematic review,” *Br. Med. Bull.*, vol. 84, no. 1, pp. 5–23, 2007, doi: 10.1093/bmb/ldm022.

[94] C. Kokkotis *et al.*, “Identifying Gait-Related Functional Outcomes in Post-Knee Surgery Patients Using Machine Learning: A Systematic Review,” *Int. J. Environ. Res. Public Health*, vol. 20, no. 1, 2023, doi: 10.3390/ijerph20010448.

[95] A. Giannakidis *et al.*, “Fast Fully Automatic Segmentation of the Severely Abnormal Human Right Ventricle from Cardiovascular Magnetic Resonance Images Using a Multi-Scale 3D Convolutional Neural Network,” *Proc. - 12th Int. Conf. Signal Image Technol.*

Internet-Based Syst. SITIS 2016, pp. 42–46, 2017, doi: 10.1109/SITIS.2016.16.

[96] T. A. Bohoran *et al.*, "Resource efficient aortic distensibility calculation by end to end spatiotemporal learning of aortic lumen from multicentre multivendor multidisease CMR images," *Sci. Rep.*, vol. 13, no. 1, pp. 1–15, 2023, doi: 10.1038/s41598-023-48986-6.

[97] M. Dunnhofer, N. Martinel, and C. Micheloni, "Deep convolutional feature details for better knee disorder diagnoses in magnetic resonance images," *Comput. Med. Imaging Graph.*, vol. 102, no. November, p. 102142, 2022, doi: 10.1016/j.compmedimag.2022.102142.

[98] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," *Commun. ACM*, vol. 63, no. 12, pp. 54–63, 2020, doi: 10.1145/3381831.

[99] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, 2014, doi: 10.1016/j.sigpro.2013.12.026.

[100] A. Dempster, D. F. Schmidt, and G. I. Webb, "MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 248–257, 2021, doi: 10.1145/3447548.3467231.

[101] J. K. Lee, L. Yao, C. T. Phelps, C. R. Wirth, J. Czajka, and J. Lozman, "Anterior cruciate ligament tears: MR imaging compared with arthroscopy and clinical tests.," *Radiology*, vol. 166, no. 3, pp. 861–4, Mar. 1988, doi: 10.1148/radiology.166.3.3340785.

[102] M. Leppänen *et al.*, "Sagittal Plane Hip, Knee, and Ankle Biomechanics and the Risk of Anterior Cruciate Ligament Injury: A Prospective Study," *Orthop. J. Sport. Med.*, vol. 5, no. 12, p. 232596711774548, Dec. 2017, doi: 10.1177/2325967117745487.

[103] B. J. Claes S, Verdonk P, Forsyth R, "The 'ligamentization' process in anterior cruciate ligament reconstruction: what happens to the human graft? A systematic review of the literature.," *Am J Sport. Med.*, vol. 39(11), pp. 2476–2483, 2011.

[104] W. E. Kaplan Y, "When Is It Safe to Return to Sport After ACL Reconstruction? Reviewing the Criteria.," *Sport. Heal.*, vol. 11(4), pp. 301–305, 2019.

[105] K. E. Webster and J. A. Feller, "Who Passes Return-to-Sport Tests, and Which Tests Are Most Strongly Associated With Return to Play After Anterior Cruciate Ligament Reconstruction?," *Orthop. J. Sport. Med.*, vol. 8, no. 12, Dec. 2020, doi: 10.1177/2325967120969425.

[106] K. E. Webster and T. E. Hewett, "What is the Evidence for and Validity of Return-to-Sport Testing after Anterior Cruciate Ligament Reconstruction Surgery? A Systematic Review and Meta-Analysis," *Sport. Med.*, vol. 49, no. 6, pp. 917–929, Jun. 2019, doi: 10.1007/s40279-019-01093-x.

[107] H. Grindem, L. Snyder-Mackler, H. Moksnes, L. Engebretsen, and M. A. Risberg, "Simple decision rules can reduce reinjury risk by 84% after ACL reconstruction: the Delaware-Oslo ACL cohort study," *Br. J. Sports Med.*, vol. 50, no. 13, pp. 804–808, Jul. 2016, doi: 10.1136/bjsports-2016-096031.

[108] N. van Melick *et al.*, "Evidence-based clinical practice update: practice guidelines for anterior cruciate ligament rehabilitation based on a systematic review and multidisciplinary consensus," *Br. J. Sports Med.*, vol. 50, no. 24, pp. 1506–1515, Dec. 2016, doi: 10.1136/bjsports-2015-095898.

[109] S. J. Meredith *et al.*, “Return to sport after anterior cruciate ligament injury: Panther Symposium ACL Injury Return to Sport Consensus Group,” *Knee Surgery, Sport. Traumatol. Arthrosc.*, vol. 28, no. 8, pp. 2403–2414, Aug. 2020, doi: 10.1007/s00167-020-06009-1.

[110] H. E. Bourke, L. J. Salmon, A. Waller, V. Patterson, and L. A. Pinczewski, “Survival of the Anterior Cruciate Ligament Graft and the Contralateral ACL at a Minimum of 15 Years,” *Am. J. Sports Med.*, vol. 40, no. 9, pp. 1985–1992, Sep. 2012, doi: 10.1177/0363546512454414.

[111] J. D. Harris *et al.*, “Return to Sport After ACL Reconstruction,” *Orthopedics*, vol. 37, no. 2, Feb. 2014, doi: 10.3928/01477447-20140124-10.

[112] S. Korbin *et al.*, “PET/MRI reveals ongoing metabolic activity in ACL grafts one year post-ACL reconstruction,” *J. Exp. Orthop.*, vol. 7, no. 1, p. 40, Dec. 2020, doi: 10.1186/s40634-020-00258-2.

[113] S. E. Putnis *et al.*, “Magnetic Resonance Imaging Assessment of Hamstring Graft Healing and Integration 1 and Minimum 2 Years after ACL Reconstruction,” *Am. J. Sports Med.*, vol. 50, no. 8, pp. 2102–2110, Jul. 2022, doi: 10.1177/03635465221096672.

[114] S. J. Rabuck, M. G. Baraga, and F. H. Fu, “Anterior Cruciate Ligament Healing and Advances in Imaging,” *Clin. Sports Med.*, vol. 32, no. 1, pp. 13–20, Jan. 2013, doi: 10.1016/j.csm.2012.08.003.

[115] J. Y. Choi, J. K. Ha, Y. W. Kim, J. C. Shim, S. J. Yang, and J. G. Kim, “Relationships Among Tendon Regeneration on MRI, Flexor Strength, and Functional Performance After Anterior Cruciate Ligament Reconstruction With Hamstring Autograft,” *Am. J. Sports Med.*, vol. 40, no. 1, pp. 152–162, Jan. 2012, doi: 10.1177/0363546511424134.

[116] L. M. White, J. Kramer, and M. P. Recht, “MR imaging evaluation of the postoperative knee: ligaments, menisci, and articular cartilage,” *Skeletal Radiol.*, vol. 34, no. 8, pp. 431–452, Aug. 2005, doi: 10.1007/s00256-005-0914-y.

[117] A. Grassi *et al.*, “Magnetic resonance imaging after anterior cruciate ligament reconstruction: A practical guide,” *World J. Orthop.*, vol. 7, no. 10, p. 638, 2016, doi: 10.5312/wjo.v7.i10.638.

[118] P. Van Dyck, K. Zazulia, C. Smekens, C. H. W. Heusdens, T. Janssens, and J. Sijbers, “Assessment of Anterior Cruciate Ligament Graft Maturity With Conventional Magnetic Resonance Imaging: A Systematic Literature Review,” *Orthop. J. Sport. Med.*, vol. 7, no. 6, Jun. 2019, doi: 10.1177/2325967119849012.

[119] H. Li, J. Chen, H. Li, Z. Wu, and S. Chen, “MRI-based ACL graft maturity does not predict clinical and functional outcomes during the first year after ACL reconstruction,” *Knee Surgery, Sport. Traumatol. Arthrosc.*, vol. 25, no. 10, pp. 3171–3178, Oct. 2017, doi: 10.1007/s00167-016-4252-5.

[120] H.-Y. Li, H. Li, Z.-Y. Wu, J.-W. Chen, and S.-Y. Chen, “MRI-based tendon bone healing is related to the clinical functional scores at the first year after anterior cruciate ligament reconstruction with hamstring tendon autograft,” *Knee Surgery, Sport. Traumatol. Arthrosc.*, vol. 26, no. 2, pp. 615–621, Feb. 2018, doi: 10.1007/s00167-017-4568-9.

[121] S. F. DeFroda, R. M. ODonnell, P. D. Fadale, B. D. Owens, and B. C. Fleming, “The role of magnetic resonance imaging in evaluating postoperative ACL reconstruction healing

and graft mechanical properties: a new criterion for return to play?," *Phys. Sportsmed.*, vol. 49, no. 2, pp. 123–129, Apr. 2021, doi: 10.1080/00913847.2020.1820846.

[122] R. Lin *et al.*, "Randomized controlled trial of all-inside and standard single-bundle anterior cruciate ligament reconstruction with functional, MRI-based graft maturity and patient-reported outcome measures," *BMC Musculoskelet. Disord.*, vol. 23, no. 1, p. 289, Dec. 2022, doi: 10.1186/s12891-022-05231-x.

[123] P. M. Lutz, A. Achtnich, V. Schütte, K. Woertler, A. B. Imhoff, and L. Willinger, "Anterior cruciate ligament autograft maturation on sequential postoperative MRI is not correlated with clinical outcome and anterior knee stability," *Knee Surgery, Sport. Traumatol. Arthrosc.*, vol. 30, no. 10, pp. 3258–3267, Oct. 2022, doi: 10.1007/s00167-021-06777-4.

[124] Y. Zhang, S. Liu, Q. Chen, Y. Hu, Y. Sun, and J. Chen, "Maturity Progression of the Entire Anterior Cruciate Ligament Graft of Insertion-Preserved Hamstring Tendons by 5 Years: A Prospective Randomized Controlled Study Based on Magnetic Resonance Imaging Evaluation," *Am. J. Sports Med.*, vol. 48, no. 12, pp. 2970–2977, Oct. 2020, doi: 10.1177/0363546520951507.

[125] S. E. Putnis *et al.*, "Magnetic Resonance Imaging 1 Year After Hamstring Autograft Anterior Cruciate Ligament Reconstruction Can Identify Those at Higher Risk of Graft Failure: An Analysis of 250 Cases," *Am. J. Sports Med.*, vol. 49, no. 5, pp. 1270–1278, Apr. 2021, doi: 10.1177/0363546521995512.

[126] S. W. Flannery *et al.*, "Early MRI-based quantitative outcomes are associated with a positive functional performance trajectory from 6 to 24 months post-ACL surgery," *Knee Surgery, Sport. Traumatol. Arthrosc.*, vol. 31, no. 5, pp. 1690–1698, May 2023, doi: 10.1007/s00167-022-07000-8.

[127] S. D. Barber-Westin and F. R. Noyes, "Objective Criteria for Return to Athletics After Anterior Cruciate Ligament Reconstruction and Subsequent Reinjury Rates: A Systematic Review," *Phys. Sportsmed.*, vol. 39, no. 3, pp. 100–110, Sep. 2011, doi: 10.3810/psm.2011.09.1926.

[128] S. Della Villa *et al.*, "Clinical Outcomes and Return-to-Sports Participation of 50 Soccer Players After Anterior Cruciate Ligament Reconstruction Through a Sport-Specific Rehabilitation Protocol," *Sport. Heal. A Multidiscip. Approach*, vol. 4, no. 1, pp. 17–24, Jan. 2012, doi: 10.1177/1941738111417564.

[129] G. D. Myer *et al.*, "No Association of Time From Surgery With Functional Deficits in Athletes After Anterior Cruciate Ligament Reconstruction," *Am. J. Sports Med.*, vol. 40, no. 10, pp. 2256–2263, Oct. 2012, doi: 10.1177/0363546512454656.

[130] D. M. Narducci E, Waltz A, Gorski K, Leppla L, "The clinical utility of functional performance tests within one-year post-acl reconstruction: a systematic review.," *Int J Sport. Phys Ther.*, vol. 6(4), pp. 333–342, 2011.

[131] E. Svantesson *et al.*, "Clinical outcomes after anterior cruciate ligament injury: panther symposium ACL injury clinical outcomes consensus group," *Knee Surgery, Sport. Traumatol. Arthrosc.*, vol. 28, no. 8, pp. 2415–2434, Aug. 2020, doi: 10.1007/s00167-020-06061-x.

[132] C. Kokkotis *et al.*, "Identifying Gait-Related Functional Outcomes in Post-Knee Surgery Patients Using Machine Learning: A Systematic Review," *Int. J. Environ. Res. Public Health*, vol. 20, no. 1, p. 448, Dec. 2022, doi: 10.3390/ijerph20010448.

[133] G. A. Koumantakis, K. Tsoligkas, A. Papoutsidakis, A. Ververidis, and G. I. Drosos, "Cross-cultural adaptation and validation of the International Knee Documentation Committee Subjective Knee Form in Greek," *J. Orthop. Traumatol.*, vol. 17, no. 2, pp. 123–129, Jun. 2016, doi: 10.1007/s10195-015-0362-y.

[134] A. Panagopoulos *et al.*, "Cross-Cultural Adaptation of the Greek Versions of the Lysholm Knee Scoring Scale and Tegner Activity Scale," *Cureus*, Jul. 2020, doi: 10.7759/cureus.9372.

[135] J. N. Yfantopoulos and A. E. Chantzaras, "Validation and comparison of the psychometric properties of the EQ-5D-3L and EQ-5D-5L instruments in Greece," *Eur. J. Heal. Econ.*, vol. 18, no. 4, pp. 519–531, May 2017, doi: 10.1007/s10198-016-0807-0.

[136] M. Moutzouri, P. Tsoumpos, E. Billis, A. Papoutsidakis, and J. Gliatis, "Cross-cultural translation and validation of the Greek version of the Knee Injury and Osteoarthritis Outcome Score (KOOS) in patients with total knee replacement," *Disabil. Rehabil.*, vol. 37, no. 16, pp. 1477–1483, Jul. 2015, doi: 10.3109/09638288.2014.972583.

[137] L. H. Ingelsrud *et al.*, "Meaningful Change Scores in the Knee Injury and Osteoarthritis Outcome Score in Patients Undergoing Anterior Cruciate Ligament Reconstruction," *Am. J. Sports Med.*, vol. 46, no. 5, pp. 1120–1128, Apr. 2018, doi: 10.1177/0363546518759543.

[138] A. M. Biercevicz *et al.*, "MRI Volume and Signal Intensity of ACL Graft Predict Clinical, Functional, and Patient-Oriented Outcome Measures After ACL Reconstruction," *Am. J. Sports Med.*, vol. 43, no. 3, pp. 693–699, Mar. 2015, doi: 10.1177/0363546514561435.

[139] C.-Y. Wen, L. Qin, K.-M. Lee, M. W.-N. Wong, and K.-M. Chan, "Grafted Tendon Healing in Tibial Tunnel Is Inferior to Healing in Femoral Tunnel After Anterior Cruciate Ligament Reconstruction: A Histomorphometric Study in Rabbits," *Arthrosc. J. Arthrosc. Relat. Surg.*, vol. 26, no. 1, pp. 58–66, Jan. 2010, doi: 10.1016/j.arthro.2009.06.025.

[140] G. S. Berns, S. M. Howell, and T. E. Farley, "The accuracy of signal intensity measurements in magnetic resonance imaging as evaluated within the knee," *Magn. Reson. Imaging*, vol. 10, no. 4, pp. 573–578, Jan. 1992, doi: 10.1016/0730-725X(92)90008-N.

[141] A. Hakozaki, Y. Niki, H. Enomoto, Y. Toyama, and Y. Suda, "Clinical significance of T2*-weighted gradient-echo MRI to monitor graft maturation over one year after anatomic double-bundle anterior cruciate ligament reconstruction: A comparative study with proton density-weighted MRI," *Knee*, vol. 22, no. 1, pp. 4–10, Jan. 2015, doi: 10.1016/j.knee.2014.11.004.

[142] W. P. Yau and W. Lin, "Evaluation of Graft Maturation by MRI in Anterior Cruciate Ligament Reconstruction With and Without Concomitant Anterolateral Ligament Reconstruction," *Orthop. J. Sport. Med.*, vol. 12, no. 1, Jan. 2024, doi: 10.1177/23259671231223976.

Appendixes

Appendix A

Database	Terms
MEDLINE:	((((((((deep learning) OR (cnn)) OR (convolutional neural networks)) OR (machine learning))) OR (artificial intelligence)) OR (neural networks)) AND (((((knee injury) OR (ACL tear)) OR (Anterior crucial ligament)) OR (knee menisc*)) OR (knee cartilage))) AND (image)
CENTRAL:	((((((” deep learning”) OR (” machine learning”)) OR (” neural networks”)) OR (” CNN”)) AND (knee)) AND (image)) NOT (osteoarthritis)- in All Text— (Word variations have been searched)
EMBASE:	find= (deep learning OR neural networks OR machine learning) AND Title= (knee AND image NOT osteoarthritis)

Table 16: Terms for strategy search.

Appendix B

Introduction to YOLO

You Only Look Once (YOLO) is a groundbreaking family of deep learning models designed for real-time object detection in images and videos. Traditional object detection algorithms typically employ a two-stage process: they first generate region proposals where objects might be located and then classify these regions. In contrast, YOLO reframes object detection as a single regression problem, directly predicting class probabilities and bounding box coordinates from full images in one evaluation. This single-stage approach enables YOLO to achieve remarkable speed and efficiency, making it highly suitable for applications that require real-time processing.

The key innovation of YOLO lies in its ability to process an image in its entirety, capturing contextual information about classes and their appearance. By dividing the image into a grid and predicting bounding boxes and class probabilities for each cell, YOLO can detect multiple objects of different sizes within an image efficiently and accurately.

The YOLO Architecture

The core architecture of YOLO is based on a single convolutional neural network (CNN) that processes the entire input image. The network consists of several convolutional layers, which are responsible for feature extraction, followed by fully connected layers that predict output probabilities and bounding box coordinates.

Figure 3.1 illustrates the simplified architecture of the YOLO model.

1. **Input Image**: The model accepts a full-sized image, typically resized to a standard dimension (e.g., 448×448 pixels), to maintain consistency across different images.
2. **Convolutional Layers**: These layers apply filters to the input image to extract hierarchical features. Early layers might detect simple patterns like edges and textures, while deeper layers capture more complex structures such as shapes and object parts.
3. **Fully Connected Layers**: These layers interpret the extracted features to predict bounding boxes and class probabilities for each grid cell.
4. **Output Layer**: The final output is a tensor encoding the bounding boxes, objectness scores, and class probabilities for each grid cell.

How YOLO Works

YOLO operates by dividing the input image into an S×SS \times SS×S grid of cells. Each grid cell is responsible for detecting objects whose centers fall within it. For each grid cell, the model predicts:

- **Bounding Boxes**: Each cell predicts BBB bounding boxes, each defined by:
    - **Coordinates**: (x,y)(x, y)(x,y) coordinates relative to the grid cell, representing the center of the bounding box.
    - **Dimensions**: Width www and height hhh relative to the entire image.
    - **Confidence Score**: Indicates the likelihood that the bounding box contains an object and how accurate the bounding box is.
- **Class Probabilities**: Conditional probabilities for each of the CCC object classes, given that an object is present in the grid cell.

The model's final prediction combines the confidence scores and class probabilities to assign class-specific confidence scores to each bounding box.

Mathematical Formulation

The output tensor for the entire image has the shape $S \times S \times (B \times 5 + C)$, where:

- $S \times S$ represents the grid cells.
- $B \times 5$ accounts for the bounding box predictions (each with 5 values: x, y, w, h, and confidence).
- C represents the class probabilities.

Loss Function

YOLO uses a multi-part loss function that balances localization error, confidence error, and classification error. The loss function $\mathcal{L}$ is defined as:

$$
\begin{aligned}
\mathcal{L} =& \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + \left(h_i - \hat{h}_i\right)^2\right] \\
&+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left(C_i - \hat{C}_i\right)^2 \\
&+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i\right)^2 \\
&+ \sum_{i=0}^{S^2} 1_{i}^{\text{obj}} \sum_{c=1}^{C} (p_i(c) - \hat{p}_i(c))^2
\end{aligned}
$$

Where:

- 1_{ij}^{obj} is 1 if the j-th bounding box in cell i is responsible for predicting an object.
- $1_{ij}^{\text{noobj}} = 1 - 1_{ij}^{\text{obj}}$.
- (x_i, y_i, w_i, h_i) are the ground truth bounding box parameters.
- $\left(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i\right)$ are the predicted bounding box parameters.
- C_i is the ground truth confidence score, and $\hat{C}_i$ is the predicted confidence score.
- $p_i(c)$ is the ground truth probability of class c in cell i, and $\hat{p}_i(c)$ is the predicted probability.
- λ_{coord} and λ_{noobj} are hyperparameters that balance the loss components.

YOLOv5: Enhancements and Features

YOLOv5 represents a significant evolution of the YOLO architecture, introducing several enhancements to improve performance, efficiency, and ease of deployment. Although the numbering suggests continuity from previous versions, YOLOv5 was developed independently by Ultralytics and includes the following key features:

Architecture Components

1. **Backbone**: YOLOv5 uses a Cross Stage Partial Networks (CSP) architecture, which improves learning capability and reduces model size. The backbone is responsible for extracting essential features from the input image.

2. **Neck**: Incorporates structures like the Path Aggregation Network (PANet) and Feature Pyramid Network (FPN) to fuse features at different scales, enhancing the detection of objects of varying sizes.
3. **Head**: Utilizes anchor-based detection, predicting bounding boxes, objectness scores, and class probabilities.

Key Enhancements

- **Focus Layer**: Introduced to efficiently downsample the input image by splitting it into slices, capturing rich spatial information with less computational overhead.
- **Mosaic Data Augmentation**: A data augmentation technique that combines four images into one during training, increasing the diversity of the dataset and improving the model's generalization.
- **Adaptive Anchor Calculation**: Automatically computes optimal anchor boxes based on the training data, enhancing localization accuracy.
- **Auto Learning Bounding Box Anchors**: Employs K-means clustering to generate anchor boxes that match the dimensions of objects in the training set.
- **Enhanced Activation Functions**: Uses advanced activation functions like SiLU (Sigmoid Linear Unit) or Mish for better gradient flow and improved learning.
- **Smaller Model Variants**: Offers various model sizes (e.g., YOLOv5s, YOLOv5m, YOLOv5.l, YOLOv5x, and YOLOv5n for nano) to balance between speed and accuracy, catering to different deployment scenarios.

Mathematical Foundations of YOLO

YOLOv5 introduces modifications to the loss functions and training strategies to enhance detection performance and training stability.

Localization Loss

YOLOv5 adopts advanced IoU-based loss functions like Complete Intersection over Union (CIoU) or Distance-IoU (DIoU) loss for bounding box regression. These loss functions consider not only the overlap area but also the distance between the center points and the aspect ratio of the bounding boxes, leading to more precise localization.

CIoU Loss

$$\mathrm{CIoULoss} = 1 - \mathrm{IoU} + \frac{\rho^2(\mathbf{b}_\mathbf{p}, \mathbf{b}_{\mathrm{gt}})}{c^2} + \alpha v$$

Where:

- $\rho^2(\mathbf{b}_\mathbf{p}, \mathbf{b}_{\mathrm{gt}})$ is the squared Euclidean distance between the center points of the predicted box $\mathbf{b}_\mathbf{p}$ and the ground truth box $\mathbf{b}_{\mathrm{gt}}$.

- c is the diagonal length of the smallest enclosing box covering both $\mathbf{b_p}$ and $\mathbf{b}_{\mathrm{gt}}$.
- v measures the consistency of aspect ratio.
- α is a positive trade-off parameter.

Confidence and Classification Loss

YOLOv5 employs Binary Cross-Entropy (BCE) loss with logits for both confidence and classification, providing better performance for multi-class object detection tasks.

Total Loss

The total loss function combines these components with carefully tuned hyperparameters to balance the contributions from localization, confidence, and classification losses, ensuring robust training.

Applications and Advantages of YOLOv5

Advantages

- **Real-Time Performance**: Capable of processing images at high frames per second (FPS), making it suitable for applications that require immediate responses.
- **High Accuracy**: Achieves competitive mean Average Precision (mAP) scores on benchmark datasets, demonstrating its effectiveness in detecting objects accurately.
- **Scalability**: Offers different model sizes to accommodate various hardware capabilities, from powerful servers to edge devices with limited resources.
- **Ease of Use**: Provides out-of-the-box support and is implemented in PyTorch, making it accessible for researchers and developers.

Applications

- **Medical Imaging**: Assists in the detection of anomalies in medical scans, such as tumors, fractures, or ligament injuries, enhancing diagnostic workflows.
- **Autonomous Systems**: Used in self-driving cars and drones for detecting pedestrians, vehicles, and obstacles, contributing to safer navigation.
- **Surveillance**: Employed in security systems for monitoring and detecting suspicious activities or unauthorized access.
- **Industrial Automation**: Utilized in manufacturing for quality control, detecting defects in products, and automating inspection processes.

Relevance to ACL Injury Detection

In the context of this dissertation, YOLOv5's capabilities are harnessed to develop a lightweight and efficient deep learning pipeline for detecting healthy ACLs in MRI scans. By utilizing YOLOv5-Nano, the smallest variant of YOLOv5, the model maintains high detection performance while operating with minimal computational resources.

The pipeline focuses on:

- **Identifying Sagittal Slices with ACL Presence**: By detecting the ACL in specific slices, clinicians can focus their evaluation on relevant images, streamlining the diagnostic process.
- **Efficiency and Portability**: The lightweight nature of YOLOv5-Nano allows deployment on devices with limited processing power, facilitating widespread adoption in various clinical settings.
- **Integration with Hybrid Frameworks**: The model serves as a crucial component in a larger framework that includes novelty detection and time-series analysis, enhancing the overall diagnostic accuracy for ACL tears.

Conclusion

YOLO and its advanced version YOLOv5 have significantly impacted the field of object detection by providing fast and accurate models that can operate in real-time. Their unique approach of framing object detection as a single regression problem simplifies the pipeline and reduces computational overhead.

In this dissertation, the application of YOLOv5, particularly the YOLOv5-Nano variant, is instrumental in advancing the detection and diagnosis of ACL injuries from MRI images. By leveraging YOLOv5's strengths, we develop a system that not only achieves high accuracy but also addresses practical considerations such as computational efficiency and ease of deployment.

The subsequent chapters will delve into the technical implementation of this pipeline, discussing how YOLOv5 is integrated into our novel hybrid framework and how it contributes to improving patient care through enhanced imaging analysis.

Appendix C

Introduction to MiniRocket

Time series classification is a critical task in various domains, including finance, speech recognition, and medical diagnostics. In medical imaging, particularly with MRI scans, sequences of images capture temporal or spatial changes essential for accurate diagnosis. Traditional methods for time series classification often face challenges related to computational efficiency and scalability, especially when handling large datasets or requiring real-time analysis.

To address these challenges, Dempster, Schmidt, and Webb (2020) introduced ROCKET (RandOm Convolutional KErnel Transform), a method that achieved state-of-the-art accuracy with significantly reduced computational cost. Building upon this foundation, they developed **MiniRocket**, an enhanced algorithm offering even greater efficiency while maintaining high accuracy. MiniRocket is designed to provide near-deterministic results and is up to 75 times faster than ROCKET on large datasets. Its innovative approach involves using a small, fixed set of convolutional kernels and an efficient feature extraction method, making it highly suitable for applications like ACL injury diagnosis from MRI sequences.

The MiniRocket Architecture

MiniRocket operates by transforming input time series data into a feature space using convolutional operations with a predetermined set of kernels. These features are then used to train a simple linear classifier. Unlike ROCKET, which uses a large number of random convolutional kernels, MiniRocket employs a small, fixed set of 84 kernels with weights constrained to two values, typically −1-1−1 and 222. This design choice simplifies computations and enhances efficiency.

The architecture of MiniRocket involves several key components. First, convolutional kernels are applied to the input time series data. The use of fixed kernels eliminates the need for generating and storing random kernels, reducing computational overhead. Dilation and padding techniques are employed to capture patterns at different scales within the time series, allowing the kernels to sample points at intervals determined by the dilation factor. This enables the convolution operation to cover the entire time series without reducing its length.

After the convolution operation, MiniRocket utilizes the proportion of positive values (PPV) pooling method to extract features from the convolution outputs. PPV calculates the proportion of positive values in the convolution output, providing a robust statistical representation of the data. These PPV values form a feature vector representing the time series. Finally, a simple linear classifier, such as logistic regression or ridge regression, is trained on these feature vectors to perform the classification task.

How MiniRocket Works

MiniRocket's approach revolves around efficiently transforming the input time series into a discriminative feature space. The process begins with the convolutional transformation, where each input time series $X = [x_0, x_1, \dots, x_{n-1}]$ of length n is convolved with the fixed set of kernels $W = [w_0, w_1, \dots, w_{m-1}]$ of length m, with weights $w_j \in \{-1,2\}$. The convolution operation with dilation d is defined as:

$$C_i = \sum_{j=0}^{m-1} w_j \cdot x_{i+j-d}$$

for all valid positions iii where the kernel fits within the time series.

The use of dilations allows the kernels to capture various patterns and features across different scales within the time series. By spacing out the kernel elements, dilations enable the model to detect both short-term and long-term dependencies in the data.

After the convolution, MiniRocket computes the proportion of positive values (PPV) for each convolution output:

$$C_i = \sum_{j=0}^{m-1} w_j \cdot x_{i+j-d}$$

where L is the length of the convolution output, I is the indicator function (1 if the condition is true, 0 otherwise), and b is a bias term, typically set based on the convolution output distribution. By varying the bias b and using multiple kernels and dilations, MiniRocket generates a rich set of features that capture different characteristics of the time series.

To achieve exceptional speed, MiniRocket employs several efficient computation strategies. Using weights of −1-1−1 and 222 allows convolution operations to be performed using only additions and subtractions, minimizing computationally expensive multiplications. The fixed kernels eliminate the need for generating and storing random kernels, reducing overhead. Additionally, convolutions and PPV calculations for different kernels and dilations can be executed in parallel, further enhancing computational efficiency.

The PPV values from all kernels and dilations form a feature vector representing the time series. A linear classifier is then trained on these feature vectors to perform classification. Due to the discriminative power of the features extracted by MiniRocket, a simple linear classifier suffices to achieve high accuracy, simplifying the overall model and reducing computational requirements.

Advantages of MiniRocket

MiniRocket offers several significant advantages. First, it provides rapid computation, significantly reducing processing time and making it suitable for large datasets and real-time applications. The efficient use of arithmetic operations minimizes resource usage, resulting in low computational cost. Second, MiniRocket produces deterministic results due to the use of fixed kernels, ensuring consistency and reliability, an essential factor in medical applications where reproducibility is crucial. Third, despite its simplicity, MiniRocket achieves accuracy comparable to more complex, resource-intensive models, demonstrating its competitiveness. Additionally, MiniRocket's linear computational complexity allows it to scale effectively with data size, handling large datasets efficiently. Its straightforward design facilitates easy implementation, maintenance, and integration into existing systems.

Application in ACL Injury Diagnosis

In the context of ACL injury diagnosis from MRI sequences, MiniRocket offers compelling benefits. MRI sequences consist of multiple sagittal slices capturing the knee's internal structures. Analyzing these sequences as time series data allows the detection of patterns indicative of ACL injuries. MiniRocket processes these sequences to extract features that reflect the presence or absence of injuries.

When combined with object detection models like YOLOv5-Nano, which identifies individual slices where the ACL is present, MiniRocket analyzes the sequence of detections to determine the overall condition of the ligament. This holistic analysis enhances diagnostic accuracy by considering the ligament's appearance across multiple images rather than in isolation. Moreover, medical datasets often have an imbalance between healthy and injured cases. MiniRocket's feature extraction helps mitigate this issue by focusing on significant patterns, improving the model's ability to handle class imbalance.

Mathematical Foundations

The mathematical foundations of MiniRocket involve the convolution operation with dilation and the calculation of PPV. The convolution with dilation d is given by:

$$C_i = \sum_{j=0}^{m-1} w_j \cdot x_{i+j \cdot d}$$

This operation allows the kernel to sample points at intervals determined by the dilation, capturing patterns over different scales within the time series. The weights wjw_jwj are constrained to −1-1−1 and 2, simplifying the computation to basic arithmetic operations.

The PPV is calculated as:

$$\mathrm{PPV} = \frac{1}{L} \sum_{i=0}^{L-1} \mathbb{I}(C_i > b)$$

This statistical measure reflects the proportion of the convolution output exceeding the bias bbb, effectively capturing the distributional characteristics of the transformed time series. By varying bbb and the dilation ddd, MiniRocket captures a diverse set of features.

For each kernel k and dilation d, a PPV feature is computed, and the collection of these features forms the feature vector:

$$\mathbf{F} = [\text{ Feature }_{1,1}, \text{ Feature }_{1,2}, \ldots, \text{ Feature }_{K,D}]$$

where K is the number of kernels and D is the number of dilations. A linear classifier predicts the class label y based on the feature vector $\mathbf{F}$:

$$y = \operatorname{sign}(\mathbf{w}^{\top}\mathbf{F} + b)$$

where $\mathbf{w}$ is the weight vector learned during training, and b is the bias term. The simplicity of the linear classifier contributes to the overall efficiency of the model.

Considerations

While MiniRocket offers significant advantages, certain considerations are essential. The use of fixed kernels may not capture all possible patterns in highly diverse datasets. This limitation can be mitigated by the nature of the medical imaging data, where the patterns related to ACL injuries are consistent and well-defined. Additionally, the model relies on the linear separability of features, which may not hold in all cases. However, the integration with other models, such as the object detection capabilities of YOLOv5-Nano, enhances the overall robustness and accuracy of the diagnostic framework.

Relevance to the Dissertation

The inclusion of MiniRocket in this dissertation is driven by its ability to efficiently process MRI sequences for ACL injury diagnosis. Its speed and low computational cost make it suitable for real-time applications in clinical settings, where timely diagnosis is critical. The deterministic nature of MiniRocket ensures reliable and reproducible results, aligning with the stringent requirements of medical diagnostics. By integrating MiniRocket into a hybrid deep learning framework, this research leverages its strengths to advance the automated detection and classification of ACL injuries, ultimately contributing to improved patient care.

Annexes

Annex A: Study approval

MINUTES EXTRACT
of the 18th/08-11-2023 Ordinary Assembly of the Scientific Council
of the UNIVERSITY GENERAL HOSPITAL OF LARISSA

In Larissa today the 8 November 2023, day of the week Wednesday and hour 13:00, at the Office of the Scientific Council of the UNIVERSITY GENERAL HOSPITAL OF LARISSA, after the invitation of the President of the Scientific Council of the UNIVERSITY GENERAL HOSPITAL OF LARISSA with prot. no. 46907/06-11-2023, the Scientific Council convened in an ordinary assembly, which came to order with the decision of the Director of the UNIVERSITY GENERAL HOSPITAL OF LARISSA with prot. no. 15471/29-03-2023, to discuss matters of the agenda.

Present during the meeting were the following members of the Scientific Council:

1.	Vasilopoulos Georgios	Prof. Director of the Univ. Haematology Clinic	President
2.	Nepka Charitini	NHS Cytology Director	Ordinary Member
3.	Oikonomou Athanasios	Urology Lect. A'	Ordinary Member
4.	Komisopoulos Georgios	UE Radiophysicist	Ordinary Member
5.	Soldatou Evgenia	TE Medical Laboratories	Ordinary Member

Mr. A. Koutalos, Mr. G. Perifanos, Mrs. A. Malita, Mr. V. Lachanas were absent because of an impediment.

Arseniou Ioanna, TE Library Science, was present at the meeting as Council Secretary. After the quorum was ascertained, the President declares the initiation of the meeting and sets the following Agenda Subjects for discussion:

7th SUBJECT:
Submission for approval of the document with prot. no. 47761/25-10-23 with the subject: "Approval of the study with the title: 'Evaluation of the integrity of the anterior cruciate ligament graft with an MRI through an artificial intelligence model'" by Mr. G-P Chlatsis Res. Physician of the Orthopedics Clinic of the UGH of Larissa.

The subject in question, with the following content, is taken into consideration of the Scientific Council for approval:

Please approve the conduct of the scientific research: 'Evaluation of the integrity of the anterior cruciate ligament graft with an MRI through an artificial intelligence model' with scientific responsible Prof. Mr. Chantes.

The members of the S.C. having taken into consideration what is mentioned on the above document, as well as the attached to it submitted date, after a discussion:

Unanimously give

1. A positive opinion for the approval of the subject: "Approval study with the title: 'Evaluation of the integrity of the anterior cruciate ligament graft with an MRI through an artificial intelligence model', as submitted with the document with prot. no. 47761/25-10-23 by Mr. G-P Chalatsis Res. Physician of the Orthopedic Clinic of the UGH of Larissa.
2. A positive opinion of all the submitted data relative to the subject in question, namely
   - Research protocol
   - Statutory declaration
   - Consent form
   - Questionnaire
   - Approval from the Director of the Orthopedic Clinic

Validation of the present on the same day.

THE PRESIDENT OF THE S.C.

Vasilopoulos Georgios
(signature)

THE MEMBERS
Nepka Charitini
Oikonomou Athanasios
Komisopoulos Georgios
Soldatou Evgenia

Exact Minutes Extract
of the 18th/08-11-2023
S.C. Ordinary Assembly
Larissa, 08-11-2023
The S.C. Secretary
Arseniou Ioanna

Exact Translation from the attached Greek Original

Larissa 21/11/2023

The certified translator

Sokratis Kyriazis

NETWORK OF TRANSLATION CENTERS
ERMINIA
SOKRATIS KYRIAZIS
TEL. & FAX VAT
2410 626055 045564280

ΑΠΟΣΠΑΣΜΑ ΠΡΑΚΤΙΚΟΥ
της 18ης /08-11-2023 Τακτικής Συνεδρίασης του Επιστημονικού Συμβουλίου
του ΠΑΝΕΠΙΣΤΗΜΙΑΚΟΥ ΓΕΝΙΚΟΥ ΝΟΣΟΚΟΜΕΙΟΥ ΛΑΡΙΣΑΣ

Στη Λάρισα σήμερα **8 Νοεμβρίου 2023**, ημέρα της εβδομάδας **Τετάρτη** και ώρα **13.00 μ.μ.**, στο Γραφείο του Επιστημονικού Συμβουλίου του ΠΑΝΕΠΙΣΤΗΜΙΑΚΟΥ ΓΕΝΙΚΟΥ ΝΟΣΟΚΟΜΕΙΟΥ ΛΑΡΙΣΑΣ, μετά την αριθ. πρωτ. 49607/06-11-2023 πρόσκληση του Προέδρου του Επιστημονικού Συμβουλίου του ΠΑΝΕΠΙΣΤΗΜΙΑΚΟΥ ΓΕΝΙΚΟΥ ΝΟΣΟΚΟΜΕΙΟΥ ΛΑΡΙΣΑΣ, συνήλθε σε **τακτική συνεδρίαση** το Επιστημονικό Συμβούλιο, το οποίο συγκροτήθηκε βάσει της αριθ. πρωτ. 15471/29-03-2023 απόφασης του Διοικητή του ΠΑΝΕΠΙΣΤΗΜΙΑΚΟΥ ΓΕΝΙΚΟΥ ΝΟΣΟΚΟΜΕΙΟΥ ΛΑΡΙΣΑΣ, για να συζητήσει επί θεμάτων της Ημερήσιας Διάταξης.

Παρόντα κατά τη συνεδρίαση είναι τα ακόλουθα μέλη του Επιστημονικού Συμβουλίου:

1.	Βασιλόπουλος Γεώργιος	Καθ. Διευθυντής Παν. Αιματολογικής Κλινικής	Πρόεδρος
2.	Νέπκα Χαριτίνη	Διευθύντρια ΕΣΥ Κυτταρολογίας	Τακτικό μέλος
4.	Οικονόμου Αθανάσιος	Επιμ. Α΄Ουρολογίας	Τακτικό μέλος
6.	Κομισόπουλος Γεώργιος	ΠΕ Ακτινοφυσικών	Τακτικό μέλος
7.	Σολδάτου Ευγενία	ΤΕ Ιατρικών Εργαστηρίων	Τακτικό μέλος

Οι κ. Α. Κουτάλος, κ. Γ. Περήφανος, κ. Α. Μαλίτα, κ. Β. Λαχανάς απουσίαζαν λόγω κωλύματος.

Στη συνεδρίαση παρέστη ως Γραμματέας του Συμβουλίου η Αρσενίου Ιωάννα, ΤΕ Βιβλιοθηκονομίας. Αφού διαπιστώθηκε απαρτία ο Πρόεδρος κηρύσσει την έναρξη της συνεδρίασης και θέτει προς συζήτηση τα παρακάτω θέματα της Ημερήσιας Διάταξης:

....

ΘΕΜΑ 8°:

Κατάθεση προς έγκριση του με αριθ. πρωτ. 47761/25-10-23 εγγράφου με θέμα: «Έγκριση μελέτης με τίτλο: 'Εκτίμηση της ακεραιότητας του μοσχεύματος πρόσθιου χιαστού συνδέσμου με μαγνητική τομογραφία μέσω ενός μοντέλου τεχνητής νοημοσύνης'» από τον κ. Γ-Π Χαλάτση Ειδ. Ιατρό της Ορθοπαιδικής Κλινικής του ΠΓΝ Λάρισας.

Τίθεται υπόψη του Επιστημονικού Συμβουλίου προς έγκριση το εν λόγω θέμα, με το ακόλουθο περιεχόμενο:

Παρακαλώ όπως μου εγκρίνεται την διενέργεια της επιστημονικής έρευνας: 'Εκτίμηση της ακεραιότητας του μοσχεύματος πρόσθιου χιαστού συνδέσμου με μαγνητική τομογραφία μέσω ενός μοντέλου τεχνητής νοημοσύνης' με επιστημονικό υπεύθυνο τον Καθ. κ Χαντέ.

Τα μέλη του Ε.Σ., λαμβάνοντας υπόψη τα όσα αναφέρονται στο ανωτέρω έγγραφο, καθώς και τα συνημμένα σε αυτό κατατεθέντα στοιχεία, μετά από διαλογική συζήτηση:

Ομόφωνα, γ ν ω μ ο δ ο τ ο ύ ν

1. θετικά για την έγκριση του θέματος: «Έγκριση μελέτης με τίτλο: 'Εκτίμηση της ακεραιότητας του μοσχεύματος πρόσθιου χιαστού συνδέσμου με μαγνητική τομογραφία μέσω ενός μοντέλου τεχνητής νοημοσύνης'», όπως κατατέθηκε με το αριθ. πρωτ. 47761/25-10-23 έγγραφο από τον κ. Γ-Π Χαλάτση Ειδ. Ιατρό της Ορθοπαιδικής Κλινικής του ΠΓΝ Λάρισας.
2. θετικά για την έγκριση όλων των σχετικών με το εν λόγω θέμα κατατεθέντων στοιχείων, ήτοι:
   - ερευνητικό πρωτόκολλο,
   - υπεύθυνη δήλωση,
   - έντυπο συγκατάθεσης,
   - ερωτηματολόγιο,
   - έγκριση από τον Διευθυντή της Ορθοπαιδικής Κλινικής

<u>Αυθημερόν επικύρωση της παρούσης.</u>

....

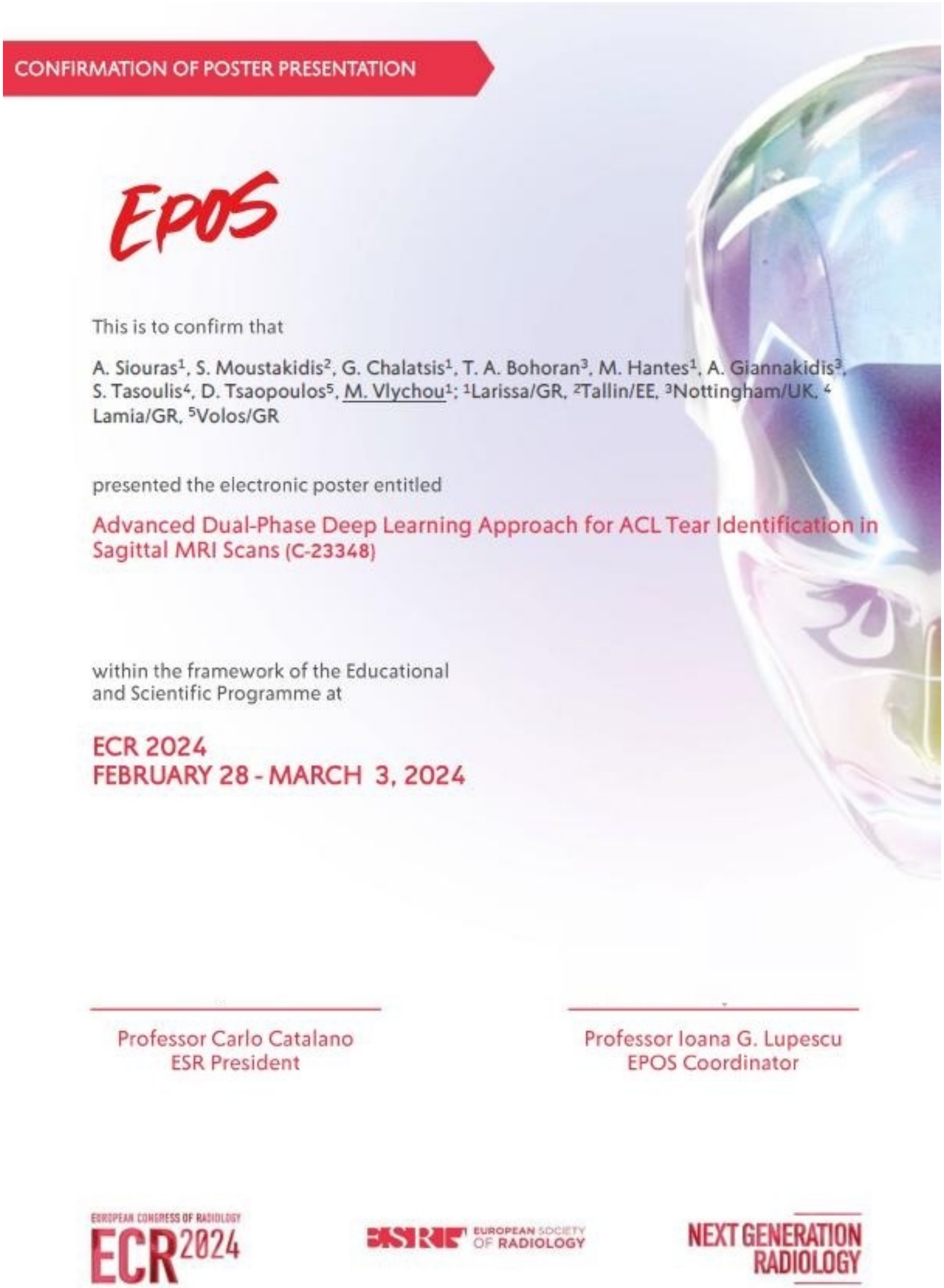
Ο ΠΡΟΕΔΡΟΣ ΤΟΥ Ε.Σ.

Βασιλόπουλος Γεώργιος

ΤΑ ΜΕΛΗ
Νέπκα Χαριτίνη
Οικονόμου Αθανάσιος
Κομισόπουλος Γεώργιος
Σολδάτου Ευγενία

Ακριβές Απόσπασμα Πρακτικού
της 18^{ης} /08-11-2023
Τακτικής Συνεδρίασης Ε.Σ.
Λάρισα, 08-11-2023
Η Γραμματέας του Ε.Σ.
Αρσενίου Ιωάννα

Annex B: E-poster


CONFIRMATION OF POSTER PRESENTATION

EPOS

This is to confirm that

A. Siouras[1], S. Moustakidis[2], G. Chalatsis[1], T. A. Bohoran[3], M. Hantes[1], A. Giannakidis[3], S. Tasoulis[4], D. Tsaopoulos[5], M. Vlychou[1]; [1]Larissa/GR, [2]Tallin/EE, [3]Nottingham/UK, [4]Lamia/GR, [5]Volos/GR

presented the electronic poster entitled

Advanced Dual-Phase Deep Learning Approach for ACL Tear Identification in Sagittal MRI Scans (C-23348)

within the framework of the Educational and Scientific Programme at

ECR 2024
FEBRUARY 28 - MARCH 3, 2024

Professor Carlo Catalano
ESR President

Professor Ioana G. Lupescu
EPOS Coordinator

EUROPEAN CONGRESS OF RADIOLOGY ECR 2024

ESR EUROPEAN SOCIETY OF RADIOLOGY

NEXT GENERATION RADIOLOGY