



POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

HTA for the Adoption of AI-Based MRI Reconstruction Software: Enhancement and Image Quality Quantification

TESI DI LAUREA MAGISTRALE IN
BIOMEDICAL ENGINEERING
INGEGNERIA BIOMEDICA

Author: **Simone Musumeci**

Student ID: 10714877
Advisor:
Prof. Giuseppe Baselli
Ing. Andrea Parisi
Academic Year: 2024-25

Abstract

In recent years, deep learning-based MRI reconstruction technologies, typically integrated with conventional techniques such as parallel imaging and compressed sensing, have demonstrated the potential to improve diagnostic performance while reducing acquisition times.

This thesis presents a local Health Technology Assessment (HTA) evaluating a vendor-agnostic AI-based MRI reconstruction software from clinical, safety, economic, and regulatory perspectives. The assessment included analysis of clinical effectiveness in terms of diagnostic confidence, based on available clinical evidence on Swift-MR, as well as evaluation of safety considerations related to AI-based reconstruction systems. An economic analysis, including a volume sensitivity analysis, was performed to assess financial sustainability under different productivity scenarios.

The regulatory framework was examined within the European context, assessing the positioning of the technology under the AI Act within MDR, and GDPR. Overall, clinical effectiveness and safety were demonstrated, and the economic evaluation suggests potential sustainability in medium-to-high volume settings.

Key-words: Health technologies assessment, Deep Learning reconstruction, MRI, Artificial Intelligence Act.

Abstract in italiano

Negli ultimi anni, le tecnologie di ricostruzione RM basate su deep learning, tipicamente integrate con tecniche convenzionali come il parallel imaging e compressed sensing, hanno dimostrato il potenziale di migliorare le performance diagnostiche riducendo i tempi di acquisizione.

Questa tesi presenta un Health Technology Assessment (HTA) locale volto a valutare un software vendor-agnostic di ricostruzione RM basato su intelligenza artificiale dal punto di vista clinico, di sicurezza, economico e regolatorio. L'efficacia clinica è stata analizzata in termini di confidenza diagnostica, sulla base delle evidenze cliniche disponibili relative a SWIFT-MR, mentre sono stati esaminati i principali profili di rischio associati ai sistemi di ricostruzione basati su AI. È stata inoltre condotta un'analisi economica comprensiva di volume sensitivity analysis per valutare la sostenibilità finanziaria in diversi scenari produttivi.

L'inquadramento regolatorio è stato analizzato nel contesto europeo, considerando l'AI Act all'interno del MDR e il GDPR. Nel complesso, l'efficacia clinica e la sicurezza risultano dimostrate, con una sostenibilità economica potenziale nei contesti a medio-alto volume di attività.

Parole chiave: Valutazione delle tecnologie sanitarie, Ricostruzione con Intelligenza Artificiale, Risonanza Magnetica, Regolamento sull'Intelligenza Artificiale.

Executive summary

Introduction and Objective

Deep learning-based MRI reconstruction software has shown significant potential in enhancing image quality while reducing acquisition times. By training neural networks on large high-quality datasets, these systems enable improved signal-to-noise ratio and diagnostic performance when integrated with established reconstruction techniques such as parallel imaging and compressed sensing. Conventional MRI requires extensive k-space sampling to ensure adequate image quality, resulting in prolonged acquisition times. Deep learning-based reconstruction aims to mitigate this limitation by enabling accelerated acquisitions without compromising diagnostic reliability.

This thesis presents a multidomain local Health Technology Assessment (HTA) conducted to support the decision-making process regarding the adoption of the technology at Ospedale di Erba. The assessment was developed according to an integrated framework evaluating clinical effectiveness, safety, economic impact, organizational implications, and regulatory positioning.

The analysis was based on the real annual MRI volumes of the healthcare facility and on data derived from a one-month pivotal trial conducted within the hospital. The objective was to address the policy question of whether the adoption of the technology provides a tangible clinical advantage in terms of image quality and diagnostic performance, as well as a sustainable economic benefit for the institution.

Technologies and current clinical gold standard

Currently, accelerated MRI relies primarily on two established strategies: parallel imaging, which leverages the spatial sensitivity profiles of multiple receiver coils, and compressed sensing, which exploits the inherent sparsity of MR images to enable reconstruction from undersampled k-space data. While effective, both approaches face limitations related to signal-to-noise ratio degradation, reconstruction complexity, and the need to balance acceleration with diagnostic reliability.

Deep learning–based reconstruction has emerged as an advanced solution to overcome these constraints. SwiftMR is based on a U-Net–derived architecture trained on pairs of fully sampled k-space data subjected to multiple degradation strategies. These include noise addition, uniform and random undersampling, k-max reduction, elliptical sampling, and partial Fourier acquisition. By learning to recover high-quality images from degraded inputs, the model aims to enable accelerated acquisitions while preserving structural details and diagnostic information.

The architecture is designed to be vendor-agnostic and operates as a standardized reconstruction module. Contextual training strategies allow denoising and sharpness optimization tailored to clinical imaging requirements. The undersampling patterns used during training are aligned with those adopted in conventional parallel imaging and compressed sensing pipelines, enabling recovery of spatial resolution along encoding directions.

From an integration perspective, the system is fully interoperable with existing MRI hardware and hospital IT infrastructure, including DICOM-compliant PACS systems. Reconstruction is performed on a secure cloud platform, ensuring scalability while maintaining compatibility with established clinical workflows.

Validation testing demonstrated effective denoising performance and recovery capability consistent with the training objectives. In addition to technical validation, this work includes a clinical assessment of diagnostic confidence to evaluate the real-world impact of the technology in the hospital setting.

Clinical effectiveness and Safety

Clinical validation of the software was analyzed through peer-reviewed evidence comparing deep learning–based reconstruction with standard-of-care MRI techniques. Across studies, deep learning reconstruction demonstrated improved image quality and increased diagnostic confidence when compared with conventional reconstruction methods, either at equal acquisition time or under accelerated protocols.

In neuroimaging, (Oh et al. 2025) reported increased sensitivity in the detection of small brain metastases (<5 mm) on T1-weighted black-blood imaging using deep learning reconstruction. However, potential false positives were associated with unsuppressed vascular signals mimicking metastatic lesions. In vascular imaging, (Kim et al. 2024) showed that deep learning reconstruction significantly enhanced 4D TWIST MR angiography, narrowing the quality gap with time-of-flight imaging while reducing

acquisition time and increasing field of view, particularly relevant in emergency settings such as acute ischemic stroke. Residual venous contamination related to dynamic acquisition was occasionally amplified by the reconstruction process.

In epilepsy imaging, (Suh et al. 2024) demonstrated increased diagnostic confidence in detecting temporal and hippocampal lesions. Nonetheless, isolated cases of incoherent signal intensity between adjacent slices were reported, potentially reducing specificity and increasing false positive interpretation.

Regarding accelerated acquisitions, (Lee et al. 2023) for knee MRI and (Yoo et al. 2023) for spine MRI reported scan time reductions of 41% and 32.3%, respectively, with non-inferior diagnostic confidence compared to standard protocols. In both musculoskeletal applications, image sharpness and denoising improvements were observed. However, enhanced sharpness occasionally increased false positive findings, and in some cases amplified residual parallel imaging artifacts.

Importantly, most reported safety-related effects did not reach statistical significance in clinical validation studies. Nevertheless, the correlation between observed imaging artifacts and known deep learning failure modes highlights the importance of integrating technological understanding with clinical evidence when assessing safety.

Based on literature review and technological analysis of the SwiftMR architecture, five domains of AI-related risk were identified and evaluated within the HTA framework (Table 1), with risk levels classified as low, medium, or high depending on potential clinical impact and detectability.

Table 1: Assessment of the safety domains for SwiftMR.

Domains	Definition	Assessment
Domain Shifts	Performance degradation in different acquisition setting	Low-Moderate
Hallucinations	Plausible but not-existing image features	Moderate (task-dependent)
Texture Alteration and oversmoothing	Suppression of fine structural details	Low-Moderate

Silent Failure	Plausible but inaccurate reconstruction	Moderate
Adversarial Perturbation Instability	Sensitivity of small perturbation in undersampled k-space	Low

Overall, the HTA indicates that the integration of deep learning–based reconstruction is supported by consistent evidence of improved image quality and non-inferior diagnostic performance in accelerated acquisition scenarios. When interpreted within a structured safety framework that accounts for known algorithmic limitations and ensures appropriate clinical oversight, the identified risks remain controllable and proportionate to the expected benefits.

From an HTA perspective, the combined assessment of clinical effectiveness and safety domains supports the adoption of the technology in medium-to-high volume clinical settings, provided that governance mechanisms and continuous performance monitoring are implemented.

Economic and Organizational Impact

The economic evaluation was conducted using real-world MRI activity data from Ospedale di Erba for the year 2025. A baseline acceleration scenario of 30% was defined as a conservative estimate, reflecting routine clinical and organizational constraints. This assumption was supported by results from a one-month pivotal trial, which demonstrated an average acceleration potential of approximately 40%.

To assess the robustness of the investment under varying productivity conditions, a volume sensitivity analysis was performed across acceleration scenarios ranging from 20% to 40%. The model incorporated the acquisition cost of the software (€71,260), an assumption of 240 working days per year, and the average weighted reimbursement per MRI examination derived from national tariff nomenclature (Sistema Sanitario Nazionale), adjusted according to the hospital’s actual examination mix.

Across all simulated scenarios, including the most conservative 20% acceleration case, the investment demonstrated economic sustainability. In the baseline scenario, projected additional annual revenue reached approximately €285,840, with a payback period of three months and a positive return on investment (ROI).

Table 2: Volume sensitivity analysis of SwiftMR.

Scenario	ROI	Additional revenue	Payback period
Base case (30%)	401.12%	€285840	3 months
20%	267.42%	€190560	5 months
40%	534.83%	€381120	2 months

The consistency of positive ROI and short payback time across all sensitivity scenarios supports the robustness of the economic case and indicates that the investment remains financially advantageous even under conservative operational assumptions.

Regulatory and Governance Considerations

From a regulatory perspective, the AI Act introduces specific requirements for artificial intelligence systems classified as high-risk, including those embedded in medical devices. AI-based MRI reconstruction software therefore falls within the intersection of the AI Act and the Medical Device Regulation (MDR), requiring full compliance with both regulatory frameworks. In addition, the use of a cloud-based reconstruction platform necessitates adherence to the General Data Protection Regulation (GDPR) to ensure secure data transmission, storage, and processing.

CE marking remains mandatory under MDR, while conformity with the AI Act introduces additional obligations related to risk management, transparency, human oversight, and post-market monitoring. The AI Act entered into force in August 2024, with specific provisions for high-risk AI systems becoming applicable from 2 August 2026. For AI systems integrated into medical devices, an extended transitional period applies until 2 August 2027. Evidence collected during the pivotal trial at Ospedale di Erba, together with findings from the literature, confirms that deep learning reconstruction may amplify patient-related motion artifacts or residual artifacts associated with parallel imaging and compressed sensing when acceleration is pushed to extreme levels. These findings reinforce the regulatory requirement for

structured human oversight and continuous post-market surveillance, as explicitly mandated by the AI Act.

Conclusions

Overall, the multidomain HTA—integrating technological analysis, clinical validation, safety assessment, economic modeling, and regulatory evaluation—addresses the research question by demonstrating a net benefit associated with the adoption of deep learning–based MRI reconstruction in medium-to-high volume clinical settings.

Clinical evidence supports improved image quality and maintained or enhanced diagnostic confidence, while identified safety risks remain proportionate and manageable under appropriate clinical supervision and governance frameworks.

Economic modeling, based on real-world hospital data, demonstrates financial sustainability and operational efficiency gains, particularly in routine MRI environments with substantial examination volumes.

Within a compliant regulatory framework and supported by structured human oversight, the adoption of the technology is therefore justified from a service-quality, patient-care, and institutional sustainability perspective.

Contents

Abstract.....	i
Abstract in italiano	iii
Executive summary	v
Introduction and Objective	v
Technologie and current clinical gold standard	v
Clinical effectiveness and Safety	vi
Economic and Organizational Impact.....	viii
Regulatory and Governance Considerations	ix
Conclusions	x
Contents.....	xi
Introduction	14
1 Regulatory Framework and HTA.....	17
1.1. Regulatory reference of EU MDR 2017/745.....	18
1.1.1. Quality Management System	22
1.1.2. Software Lifecycle	23
1.1.3. Risk Management.....	25
1.1.4. Post-Market Surveillance	26
1.1.5. Usability.....	27
1.1.6. Data Integrity and Interoperability	29
1.1.7. Information to be supplied by manufacturer.....	30
1.1.8. Unique Device Identification (UDI)	31
1.2. GDPR within MDR Framework.....	32
1.2.1. Cross reference between MDR and GDPR.....	34
1.2.2. IT networks and workflow surrounding data security	37
1.3. Regulatory reference of AI and AI Act	40

1.3.1.	Definition and scope	41
1.3.2.	Classification of AI product based on risk.....	42
1.3.3.	Documentation and Conformity Assessment for High-risk AI Systems	44
1.3.4.	Application on SwiftMR, limitations and final considerations	53
1.4.	HTA in the clinical context.....	56
2	Health problem and current use of the technology	63
2.1.	Clinical context: MRI in current practice	63
2.2.	Current reconstruction practice: gold standard and established acceleration.....	64
2.2.1.	Gold standard of conventional recontruction.....	64
2.2.2.	Established acceleration techniques	64
2.3.	Real and potential Use in Clinical Setting	70
3	Description and Technical Features.....	77
3.1.	System overview and Architecture.....	79
3.2.	Deep Learning algorithm and learning process	81
3.3.	Technical implementation.....	90
4	Safety	95
4.1.	Scientific and Clinical Evidence on Safety and Tolerability.....	95
4.1.1.	Technology-Related Risk Domains in Deep Learning MRI Image Reconstruction	96
4.1.2.	Clinical Safety Evidence of SwiftMR.....	102
4.2.	Organizational Perception of the Safety Dimension	104
5	Clinical effectiveness.....	109
5.1.	Clinical validation of the SwiftMR software based on scientific evidence	110
5.2.	Clinical validations of other AI-based MRI reconstruction software (FastScan and Enhanced Image Quality).....	141
5.3.	Clinical Effectiveness of AI-Based MRI Acceleration Techniques: Image-Based and Physics-Based Approaches.....	145
6	Economic and Financial Impact of the Adoption of AI-based MRI Reconstruction Software.....	153

6.1. Investment and Financial Considerations for the Healthcare Facilities	153
6.2. Potential Increase in MRI productivity and Cost-Effectiveness implications.....	155
6.3. Economic sensitivity analysis based on three Utilization scenarios (low, medium, high)	159
7 Conclusion and future developments.....	163
Bibliography.....	167
List of Figures	175
List of Tables	177
Acknowledgments	179

Introduction

Magnetic Resonance Imaging (MRI) provides superior soft-tissue contrast compared to other diagnostic imaging modalities. This capability derives from its underlying physical principle: the interaction between a strong static magnetic field and hydrogen protons within biological tissues, which induces precession around the magnetic field direction and enables signal acquisition primarily from hydrogen nuclei contained in water and lipid molecules (Brown et al. 2014).

MRI allows the acquisition of multiparametric information of anatomical, functional, diffusion-weighted, perfusional, metabolic, and vascular nature using a single scanning apparatus. By varying acquisition parameters and pulse sequences, it is possible to selectively exploit the different relaxation properties of protons in tissues, generating image contrasts sensitive to specific microstructural and physiological characteristics (van Beek et al. 2018).

The principal imaging modalities include T1-weighted, T2-weighted, and proton density-weighted sequences for structural and anatomical imaging; functional MRI (fMRI), based on the Blood Oxygen Level-Dependent (BOLD) signal, for functional brain imaging (Ogawa et al. 1990); diffusion-weighted imaging (DWI) and diffusion tensor imaging (DTI) for the assessment of water diffusion and anisotropic fiber tract analysis (Le Bihan et al. 2001); perfusion techniques such as perfusion-weighted imaging (PWI) and arterial spin labeling (ASL) for the evaluation of cerebral blood flow (Detre et al. 1992) magnetic resonance angiography (MRA) for vascular assessment; and magnetic resonance spectroscopy (MRS) for metabolic characterization of tissues (Brown et al. 2014).

MRI is widely used in radiological practice not only for its diagnostic versatility but also for its favorable safety profile. Unlike other imaging modalities, MRI does not use ionizing radiation, thereby reducing potential biological risks for patients (Kanal et al. 2013). This characteristic, together with its superior soft-tissue contrast, has significantly contributed to its widespread clinical adoption and continuous technological advancement.

Despite its diagnostic potential, MRI has historically been limited by relatively long acquisition times. Scan duration represents a critical constraint in routine clinical practice, affecting both patient throughput and institutional capacity to meet increasing demand (Streit et al. 2021) . Extended examination times may also

negatively impact patient comfort, potentially inducing anxiety or claustrophobia and increasing the likelihood of motion-related artifacts, which degrade image quality. Physiological motion, including respiration and cardiac activity, further complicates image acquisition and often requires synchronization or motion-compensation strategies.

In clinical practice, radiologists and MRI technologists continuously balance scan duration against the diagnostic accuracy required by the clinical question. Prolonged acquisitions may enhance image quality but increase patient discomfort, whereas shorter acquisitions reduce motion artifacts and improve tolerability but may compromise signal-to-noise ratio (SNR).

Technological developments in both hardware components and reconstruction software have significantly improved MRI efficiency and image quality. Techniques such as parallel imaging exploit spatial encoding provided by radiofrequency coils to reduce the number of k-space lines acquired (Pruessmann et al. 1999), while compressed sensing relies on mathematical sparsity constraints to reconstruct images from undersampled data (Lustig, Donoho, and Pauly 2007). These methodologies have enabled substantial reductions in acquisition time and are now implemented in major commercial MRI systems.

More recently, artificial intelligence (AI)–based reconstruction methods, particularly deep learning networks trained on large datasets of medical images, have emerged as promising tools to further accelerate MRI acquisition while enhancing image quality. Deep learning reconstruction has been shown to improve spatial resolution, SNR, and perceived sharpness, potentially surpassing traditional standard-of-care reconstruction techniques in certain contexts (Hammernik et al. 2017); (Knoll et al. 2019).

The training of AI models on large and representative datasets, including multiple anatomical regions routinely examined in clinical practice, enables broad applicability in radiological workflows. In this context, Ospedale di Erba, acquired in 2023 by the private healthcare group Lifenet Healthcare, identified SwiftMR (AIRS Medical) as a potential candidate for optimizing MRI scan times and improving image quality through AI-based reconstruction.

The objective of this project is to conduct a Health Technology Assessment (HTA) following the methodology of the European Network for Health Technology Assessment (EUnetHTA), applying the HTA Core Model to provide structured and multidisciplinary decision support for institutional adoption of the technology (EUnetHTA 2008). The policy question underlying this study is the following:

Is there a real advantage in acquiring and implementing the AI-based software SwiftMR (AIRS Medical) at Ospedale di Erba, in terms of improving service quality for patients and generating economic benefit for the institution?

The decision to evaluate this technology arises from the interest in comparing the current (as-is) organizational configuration with a potential future (to-be) scenario, specifically within the Radiology Department and the interconnected clinical units involved in the diagnostic pathway.

Subsequently, the methodological framework adopted for the Health Technology Assessment is presented, including the prioritization process of the HTA domains according to the local decision-making needs. This structured approach ensures that the evaluation focuses on the most relevant dimensions for the institutional context, in alignment with the principles of the EUnetHTA Core Model.

Before proceeding to the assessment of the specific HTA domains, it is essential to frame SwiftMR within the European regulatory landscape governing medical devices. As a deep learning–based MRI reconstruction software bearing CE marking, SwiftMR must comply with the relevant European regulations applicable to medical devices, particularly in terms of safety, performance, and post-market surveillance obligations. For this reason, a dedicated regulatory analysis was conducted to contextualize the technology within the current legal framework.

1 Regulatory Framework and HTA

The rapid diffusion of artificial intelligence (AI) software in the medical field over the last decade has significantly transformed the regulatory landscape for medical devices. In response to this evolution, European authorities have recognised the necessity of establishing a comprehensive regulatory framework to ensure the safety, quality, and effectiveness of AI-based medical technologies. Within the European Union, medical devices incorporating AI technologies are currently governed by a combined regulatory framework composed of the Medical Device Regulation (EU) 2017/745 (MDR) (European Union 2017), the General Data Protection Regulation (GDPR), and the recently introduced Artificial Intelligence Act (AI Act).

Together, these regulations define the legal, technical, and organisational requirements applicable to AI-based medical software throughout its entire life cycle. This chapter aims to analyse the regulatory requirements applicable to AI-based medical software in order to assess full compliance with the current European regulatory framework. In particular, the chapter focuses on the legal and technical obligations relevant for the Health Technology Assessment (HTA) of the SwiftMR software.

The chapter is organised as follows. First, the primary regulatory framework governing the medical device under investigation is analysed, namely the Medical Device Regulation (EU) 2017/745, which constitutes the core legal reference for software qualification and classification. Particular attention is paid to the classification rules applicable to medical software, with specific reference to Rule 11 of Annex VIII. Subsequently, the international and European technical standards adopted by the manufacturer and declared within the Declaration of Conformity are examined, distinguishing between harmonised standards and state-of-the-art technical standards.

The second part of the chapter addresses the General Data Protection Regulation (GDPR), focusing on data governance aspects related to the implementation of IT networks and cloud-based infrastructures within the hospital environment. Specific attention is given to data management processes for training, validation, and testing of AI models. Finally, the chapter introduces the Artificial Intelligence Act, analysing its interaction with the MDR framework and its implications for conformity assessment, CE marking, and post-market obligations. The integration of these three

regulatory pillars provides the legal and regulatory foundation for the subsequent Health Technology Assessment of the SwiftMR software.

1.1. Regulatory reference of EU MDR 2017/745

SwiftMR is an AI-based standalone software developed by AIRS Medical, intended for the acceptance, enhancement, and transfer of magnetic resonance (MR) images in DICOM format. The software is designed to perform noise reduction and image sharpness enhancement for MR images of the brain, spine, knee, ankle, shoulder, and hip.

According to the intended purpose declared by the manufacturer, SwiftMR does not provide autonomous diagnostic decisions but operates as a post-processing image enhancement tool aimed at improving image quality and reconstruction from under sampled MRI acquisitions. Under the Medical Device Regulation (EU) 2017/745, the classification of medical software is primarily determined by its intended purpose and its impact on diagnostic or therapeutic decision-making. In this context, Rule 11 of Annex VIII, Chapter III applies.

Rule 11 states:

“Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as Class IIa, expect if such decision have an impact that may cause:

- Death or an irreversible deterioration of health (class III)”;
- Serious deterioration of health or surgical intervention (class IIb)”.

Furthermore, software intended to monitor physiological processes is classified as Class IIa, unless it is intended for monitoring vital physiological parameters where variations could result in immediate danger to the patient, in which case it is classified as Class IIb.

Based on this regulatory framework, SwiftMR is classified as a Class IIa medical device, as it supports the diagnostic process indirectly by enhancing image quality, without autonomously generating diagnostic outputs or directly guiding therapeutic decisions.

Table 1.1: Intended pupose of SwiftMR

Description	
Purpose	Fast scans acquisition, enhancing image quality and reconstruction from low sample MRI.

Input/output	MR DICOM images / Enhanced MR DICOM images
User	Clinical radiologist
Effect on diagnosis	It does not direct influence diagnostic decision but enhance image quality or reconstruction that could have an important impact in the final diagnosis.

To obtain CE marking under MDR 2017/745, AIRS Medical has adopted a conformity assessment pathway supported by a set of international and European technical standards. These standards ensure compliance with essential safety and performance requirements and represent both harmonised standards under the MDR and recognised state-of-the-art practices.

The Table 1.2 summarises the standards adopted by the manufacturer and declared within the Declaration of Conformity. These normative references will be used in the subsequent HTA analysis to assess regulatory robustness, risk management, software lifecycle governance, and clinical integration of the SwiftMR solution.

Table 1.2: Declaration of conformity for SwiftMR

Reference	Standard Title
EU MDR 2017/745	Regulation (EU) 2017/745 of the European parliament and of the council
(EN) ISO 13485:2016/A11:2021	International Standard ISO 13485 Medical devices – Quality management systems – Requirements for regulatory purposes
IEC 62304:2006/ Amd 1:2015	Medical device software – Software life-cycle processes
ISO 14971:2019	Medical devices — Application of risk management to medical devices

ISO/TR 14971:2020	Medical devices — Guidance on the application of ISO 14971
ISO/TR 20416:2020	Medical devices — Post-market surveillance for manufacturers
ISO 62366-1:2015/AMD 1:2020	Medical Devices — Part 1: Application of usability engineering to medical devices
IEC/TR 62366-2:2016	Medical devices — Part 2: Guidance on the application of usability engineering to medical devices
IEC 80001-1:2021	Application of risk management for IT-networks incorporating medical devices Part 1: Safety, effectiveness and security in the implementation and use of connected medical devices or connected health software
IEC/TR 80001-2-1:20012	Application of risk management for IT-networks incorporating medical devices — Part 2-1: Step by Step Risk Management of Medical IT-Networks; Practical Applications and Examples
IEC/TR 80001-2-2:2012	Application of risk management for IT-networks incorporating medical devices —

	Part 2-2: Guidance for the communication of medical device security needs, risks and controls
IEC/TR 80002-1:2009	Medical device software — Part 1: Guidance on the application of ISO 14971 to medical device software
ISO/TR 80002-2:2017	Medical device software — Part 2: Validation of software for medical device quality Systems.
NEMA PS 3.1-3.20 (2022d)	Digital Imaging and Communications in Medicine (DICOM) Set
EN ISO 20417:2021	Medical devices - Information to be supplied by the manufacturer (ISO 20417:2021, Corrected version 2021-12)
EN ISO 15223-1:2021	Medical devices - Symbols to be used with information to be supplied by the manufacturer - Part 1: General requirements (ISO 15223-1:2021)
Commission Implementing Regulation (EU) 2021/2226	Commission Implementing Regulation (EU) 2021/2226 of 14 December 2021 laying down rules for the application of Regulation (EU) 2017/745 of the European

Parliament and of the Council as regards electronic instructions for use of medical devices.
--

1.1.1. Quality Management System

The international standard ISO 13485:2016 (ISO 2016), entitled “Medical devices – Quality management systems – Requirements for regulatory purposes”, defines the requirements for a quality management system (QMS) applicable to organisations involved in the lifecycle of medical devices. Within the European regulatory framework, the harmonised version EN ISO 13485:2016/A11:2021 is aligned with both the Medical Device Regulation (EU) 2017/745 (MDR) and the In Vitro Diagnostic Regulation (EU) 2017/746 (IVDR). Compliance with a harmonised standard confers a presumption of conformity with the relevant General Safety and Performance Requirements (GSPR) set out in European legislation.

Although the adoption of ISO 13485 is not mandatory, it is widely recognised as the reference standard for establishing a robust quality management system in the medical device sector. In the Italian context, the standard has been transposed as UNI CEI EN ISO 13485:2021, which implements the European harmonised version. ISO 13485 specifies the requirements for a quality management system that enables an organisation to consistently provide medical devices and related services that meet both customer expectations and applicable regulatory requirements.

The scope of the standard encompasses all phases of the medical device lifecycle, including design and development, production, storage, distribution, installation, servicing, and, where applicable, final disposal and recycling. The adoption of ISO 13485 supports regulatory compliance by ensuring that quality is systematically embedded throughout the product lifecycle. From a regulatory and HTA perspective, this standard represents a fundamental pillar for evaluating the organisational maturity and governance capabilities of manufacturers of AI-based medical software.

Several benefits arise from the implementation of ISO 13485 for the different stakeholders involved:

- Manufacturers benefit from an independent and structured assessment of their quality management processes by a Notified Body, which supports conformity assessment activities;
- Distributors and commercial entities obtain assurance that their operational processes comply with regulatory quality requirements;
- Healthcare professionals and patients benefit indirectly from increased confidence that medical devices are developed, manufactured, and maintained in accordance with established quality and safety principles.

In the specific context of medical device software, quality management activities are closely linked to risk management and software lifecycle processes. For this reason, ISO 13485 is commonly applied in conjunction with ISO 14971 and IEC 62304, supported by dedicated technical reports addressing software-specific aspects. The technical report IEC/TR 80002-1:2009 (IEC 2009), entitled “Medical device software – Part 1: Guidance on the application of ISO 14971 to medical device software”, provides practical guidance on how risk management principles should be adapted to software-based medical devices. The document follows the structure of ISO 14971 and tailors risk management activities to software-specific hazards, including:

- identification of software hazards and hazardous situations;
- classification of risks based on clinical consequences;
- implementation of risk control measures;
- traceability between software functions and risk management activities;
- definition of appropriate risk management documentation.

Complementarily, ISO/TR 80002-2:2017 (ISO 2017), “Medical device software – Part 2: Validation of software for medical device quality systems”, provides guidance on the validation of software used within the quality management system itself. This includes software supporting development, production, complaint handling, configuration management, and quality system automation, ensuring compliance with the requirements of ISO 13485 and ISO 14971. Together, these standards and technical reports establish a comprehensive quality and governance framework for medical device software, forming a critical component of the regulatory and HTA evaluation of AI-based solutions such as SwiftMR.

In the case of SwiftMR, the adoption of an ISO 13485-compliant quality management system, as declared by the manufacturer, provides assurance that the development, deployment, and maintenance of the AI-based reconstruction software are conducted within a controlled and traceable organisational framework. This aspect is particularly relevant in the HTA context, as it supports the evaluation of organisational robustness, reproducibility of performance, and long-term sustainability of the technology within clinical environments.

1.1.2. Software Lifecycle

The international standard IEC 62304:2006 (IEC 2006), entitled “Medical device software – Software life cycle processes”, defines the regulatory requirements applicable to each phase of the lifecycle of medical device software. The standard is internationally recognised and widely adopted by regulatory authorities, including alignment with expectations of the U.S. Food and Drug Administration (FDA), as well as by regulatory frameworks in Europe and other jurisdictions.

IEC 62304 applies to two principal categories of medical software: Software as a Medical Device (SaMD), which operates independently as a medical device, and Software in a Medical Device (SiMD), which is embedded within a hardware medical device. In the context of this project, SwiftMR clearly falls within the SaMD category, as it is a standalone AI-based software used for post-processing and enhancement of MR images. The structure of IEC 62304 is organised into nine chapters.

The initial chapters define the scope, terminology, and general requirements of the standard, while the remaining chapters focus on the regulation of the software lifecycle. The core lifecycle domains addressed by the standard include software development, software maintenance, software risk management, configuration management, and problem resolution. A fundamental aspect of IEC 62304 is its risk-based classification system, which categorises medical device software into three safety classes according to the potential severity of harm resulting from software failure:

- Class A – No injury or damage to health is possible;
- Class B – Injury is possible, but not serious;
- Class C- Death or serious injury is possible.

In the chapter 5 of IEC 62304 are underlying in detail all the different steps of the development of the project:

- 1 Software development planning;
- 2 Software requirements analysis;
- 3 Software architectural design;
- 4 Software detailed design;
- 5 Software unit implementation and verification;
- 6 Software integration and integration testing;
- 7 Software system testing;
- 8 Software release.

Through these processes, IEC 62304 establishes a structured and controlled framework for the development of medical device software, ensuring that safety and performance requirements are systematically addressed throughout the entire lifecycle. Risk management activities are integrated into each phase, with a strong emphasis on risk control and mitigation. Although IEC 62304 is not formally harmonised under the MDR, it is universally recognised as the state of the art for medical device software development and is routinely applied in conformity assessment procedures in Europe. The adoption of this standard contributes to the prevention of historical safety failures related to insufficient software governance. In the case of SwiftMR, compliance with IEC 62304, as declared by the manufacturer, provides assurance that the AI-based

reconstruction software has been developed, validated, and maintained within a controlled lifecycle framework. From an HTA perspective, this is particularly relevant for evaluating the reliability, reproducibility, and long-term maintainability of the software when deployed in clinical MRI workflows.

1.1.3. Risk Management

The manufacturer is required to demonstrate conformity with ISO 14971:2019 (ISO 2019), “Medical devices – Application of risk management to medical devices”, which establishes the principles and processes for systematic risk management throughout the entire lifecycle of a medical device. Within the European regulatory framework, risk management activities must be aligned with the Medical Device Regulation (EU) 2017/745, particularly with the General Safety and Performance Requirements outlined in Annex I. Risk management activities shall be continuously tracked and maintained throughout the product lifecycle and integrated with the software lifecycle processes defined in IEC 62304.

The manufacturer is required to provide evidence of a complete Risk Management File, comprising the risk management plan, hazard identification and analysis, risk estimation and evaluation, implementation of risk control measures, residual risk assessment, periodic reviews, and post-production data, including benefit–risk evaluation. ISO 14971 is intrinsically linked to IEC 62304, which governs the development and maintenance of medical device software. The risk classification of the software (Class A, B, or C), as defined by IEC 62304, determines the level of rigor required for both risk management activities and verification processes.

Consequently, software-related hazards must be systematically identified, evaluated, and mitigated in accordance with the assigned safety class. In the clinical context of AI-based MRI reconstruction software, risk management activities focus on a range of potential hazards that may impact diagnostic workflows, data integrity, and system availability. Table 1.3 summarises the most critical risk domains requiring mitigation, with particular attention to their relevance in relation to the software risk classification.

Table 1.3: Key risk domains according to ISO 14971 applied to AI-based MRI reconstruction software

Aspect	Description
Data/images	Reconstruction errors, misregistration, loss of DICOM data, incorrect patient could occur in misdiagnosis (high severity)

Algorithm parameters	Unpredictable behaviour of the software based on training, undetected signal saturation or aliasing artifacts
Interoperability	Incompatibility with the PACS, RIS, HIS system shall result in corrupt images
Usability	The setting of the parameters especially in emergent situation must follow an intuitive interface.
Cybersecurity	Unauthorized access or ransomwares attack may cause service interference and unavailable functioning of the software
Software updates	Risk re-evaluation and regression testing in accordance with IEC 62304.

Within the scope of this thesis, the application of ISO 14971 supports the assessment of risk governance and hazard control strategies across all stages of the software lifecycle. When applied in conjunction with IEC 62304, this framework contributes to the evaluation of the manufacturer's capability to manage foreseeable risks associated with AI-based medical software such as SwiftMR, without directly influencing clinical decision-making. The technical report ISO/TR 24971:2020 (ISO 2020a), "Medical devices – Guidance on the application of ISO 14971", provides practical guidance on the implementation and maintenance of a comprehensive risk management system. The document elaborates on the structure and content of the Risk Management File, detailing its key components, including planning, hazard analysis, risk evaluation, risk control, residual risk assessment, reviews, and post-production activities. The guidance provided by ISO/TR 24971 complements ISO 14971 and supports its effective application within the medical device software lifecycle.

1.1.4. Post-Market Surveillance

The technical report ISO/TR 20416:2020 (ISO 2020b), entitled "Medical devices – Post-market surveillance for manufacturers", provides a structured framework for the establishment and maintenance of a systematic post-market surveillance (PMS) system throughout the lifecycle of medical devices. The guidance outlined in the document is consistent with the principles defined in ISO 13485 and ISO 14971, supporting the integration of quality management and risk management activities within post-market processes.

The primary objective of ISO/TR 20416 is to support manufacturers in the systematic collection, analysis, and utilisation of post-market data, including user feedback, performance information, and incident reports. These activities enable continuous process improvement, ongoing risk evaluation, and verification of compliance with product safety and performance requirements, in alignment with the post-market surveillance obligations defined in Annex III of the MDR (EU) 2017/745. For AI-based medical software, post-market surveillance plays a particularly critical role due to the potential for performance variability across different clinical settings and user contexts. In the case of SwiftMR, compliance with ISO/TR 20416, as declared by the manufacturer, supports the structured monitoring of software performance, identification of potential anomalies, and evaluation of emerging risks following clinical deployment.

Post-market data collection from multiple users and clinical environments contributes to the continuous updating and versioning of the software, ensuring that the benefit–risk balance is periodically reassessed and maintained over time. From an HTA perspective, the presence of a structured PMS framework represents a key element for evaluating the long-term safety, sustainability, and organisational impact of AI-based MRI reconstruction software within real-world clinical practice.

1.1.5. Usability

The international standard IEC 62366-1:2015/Amd1:2020 (IEC 2015), entitled “Medical devices – Part 1: Application of usability engineering to medical devices”, defines a structured usability engineering process aimed at ensuring the safe and effective use of medical devices. The standard focuses on the interaction between users and devices, addressing use-related hazards, potential use errors, and the mitigation of risks associated with normal use conditions. Amendment 1, published in 2020, introduced clarifications and updates to key definitions, including the intended use environment, critical use scenarios, and the explicit integration of usability engineering activities with ISO 14971 risk management processes. The scope of IEC 62366-1 is limited to risks arising from normal use and foreseeable use errors, explicitly excluding risks related to abnormal use, defined as intentional misuse. Usability engineering according to IEC 62366-1 requires the systematic integration of human factors considerations into the device development process. The standard defines a sequence of interrelated activities that must be documented and aligned with the overall risk management framework.

Table 1.4: Key steps of the usability engineering process according to IEC 62366-1

Use specifications	The following definition is provided for the context of use, encompassing the user's activity, limitations, physical environment, and soft factors intended, such as stress, cognitive load, and timetable.
Use-related hazards	In this step, the potential danger use scenario that could result in a critical situation is examined, with particular attention paid to the critical task that could lead to significant damage. The 2020 amendment served to reinforce the definition of critical tasks and the importance of selecting the most appropriate use scenario in the summative test.
Users interface requirements/ usability specifications	The translation process involves the transformation of the use scenario into a set of specific requirements, including feedback, warning, notification, and interpretation modalities.
Formative evaluations	In the course of the implementation process, it is imperative to conduct user testing in order to develop a suitable user interface and implement iterative corrective measures.
Summative evaluations/ usability validation	In the concluding phase, a series of trials are to be conducted on clinicians in authentic settings. This will involve the administration of tests to a select number of critical cases, with the objective of simulating realistic conditions and evaluating the acceptability of the risk.

Usability engineering file	The documentation of the usability process encompasses a comprehensive range of materials, including use specifications, an analysis of risks associated with usage, requirements, the outcomes of test training, as well as summative, revision and iteration.
User's feedback and postproduction revision	The process is not terminated upon the realisation of the product; rather, it is imperative to monitor errors that occur on the field and subsequently update the usability analysis and user interfaces.

The technical report IEC/TR 62366-2:2016 (IEC 2016), entitled “Medical devices – Part 2: Guidance on the application of usability engineering to medical devices”, provides practical guidance for the implementation of the principles defined in IEC 62366-1. Although IEC/TR 62366-2 does not introduce mandatory requirements, it supports manufacturers in translating usability engineering concepts into practical and implementable solutions. In the context of SwiftMR, usability engineering is particularly relevant due to the interaction between radiologists, technologists, and AI-based image reconstruction parameters within time-sensitive clinical workflows. The application of IEC 62366 principles supports the evaluation of use-related risks, interface clarity, and workflow integration. From an HTA perspective, usability directly influences organizational efficiency, error prevention, and user acceptance, thereby representing a key non-clinical determinant of successful technology adoption.

1.1.6. Data Integrity and Interoperability

SwiftMR adopts the Digital Imaging and Communications in Medicine (DICOM) standard for medical image exchange, as defined by the NEMA PS 3.1–3.20 (2022) standard series (NEMA 2022). DICOM represents the internationally recognised reference standard for interoperability in medical imaging environments, enabling communication between imaging modalities, Picture Archiving and Communication Systems (PACS), Radiology Information Systems (RIS), and other healthcare IT infrastructures. The use of the DICOM standard supports several critical aspects of interoperability and data governance within clinical imaging workflows. In particular, DICOM enables:

- Interoperability: software must be integrated with PACS, RIS and MRI systems, as well other systems in the healthcare structures;
- Data format and transfer: the use of the DICOM standard ensures that other systems can correctly codify the information;
- Regulatory compliance and safety: the standard ensures both regulatory documentation and an integrated strategy for protecting data during exchange within an IT network.

Although the DICOM standard is not formally harmonised under the Medical Device Regulation (EU) 2017/745, its adoption is widely regarded as representing the state of the art for medical imaging interoperability. Compliance with DICOM supports the implementation of technical and organisational measures required to ensure data integrity and continuity of clinical workflows, particularly when medical devices are integrated within complex hospital IT networks. In the context of SwiftMR, adherence to the DICOM standard is essential to ensure seamless integration of AI-based image reconstruction outputs into existing radiological workflows without altering the original acquisition data or downstream diagnostic systems. From an HTA perspective, interoperability and data integrity directly impact organisational efficiency, workflow continuity, and the scalability of AI-based solutions within real-world clinical environments.

1.1.7. Information to be supplied by manufacturer

The international standard EN ISO 20417:2021 (ISO 2021a), entitled “Medical devices – Information to be supplied by the manufacturer”, specifies the requirements for information that manufacturers must provide with medical devices and accessories. The scope of the standard includes labels, instructions for use, symbols, warnings, precautions, and all associated documentation necessary to ensure the safe and effective use of the device. These requirements apply to all medical devices, including software as a medical device (SaMD). ISO 20417 defines the essential information elements required to support correct device identification, safe operation, and regulatory transparency. Key information requirements include:

- Identification of the device and of the manufacturer:
 - name, address, CE mark;
- Information for safety and correct use:
 - intended purpose, method of use, limitations and maintenance
 - warning, precaution and contraindications;
- Linguistic requirements and readability;
- Requirements for software as medical devices and digital devices;
- Standardization of symbols.

The standard EN ISO 15223-1:2021 (ISO 2021b), entitled “Medical devices – Symbols to be used with information to be supplied by the manufacturer – Part 1: General requirements”, provides a harmonized set of graphical symbols for use on device labels, packaging, instructions for use, and user interfaces. ISO 15223-1 is complementary to ISO 20417, as it specifies how required information should be represented through standardized symbols rather than textual descriptions. The adoption of internationally recognized symbols is particularly relevant for ensuring compliance with Annex I, Section 23 of the MDR, as it facilitates consistent interpretation of safety-related information while minimizing the need for extensive translations.

This approach contributes to improved usability, reduced risk of misinterpretation, and more efficient deployment of medical devices across different linguistic and organizational contexts. In the case of SwiftMR, compliance with ISO 20417 and ISO 15223-1, as declared by the manufacturer, supports the clear communication of intended use, operating conditions, and limitations of the AI-based software. From an HTA perspective, the adequacy and clarity of manufacturer-supplied information directly influence user training requirements, organizational readiness, and the safe integration of the software into routine clinical practice.

1.1.8. Unique Device Identification (UDI)

The Commission Implementing Regulation (EU) 2021/2226 of 14 December 2021 lays down detailed rules for the application of the Medical Device Regulation (EU) 2017/745 with regard to the Unique Device Identification (UDI) system for medical devices (European Commission 2021). The regulation specifies the technical and operational requirements for the assignment, registration, and management of UDI throughout the lifecycle of a medical device. This implementing regulation operationalize the UDI framework defined in Annex VI, Part C (Section 27) of the MDR and establishes harmonized procedures to ensure device traceability, post-market surveillance effectiveness, and regulatory transparency across the European Union. The UDI system is composed of two distinct elements:

- The UDI-DI is a device identifier that serves to identify the model of the device;
- The UDI-PI production identifier is a unique code that identifies the information of production.

Specific provisions apply to software as a medical device (SaMD). In particular, each significant software version that may impact safety or performance requires the assignment of a new UDI-DI, while minor updates or changes that do not affect regulatory characteristics may be managed through an updated UDI-PI. This distinction ensures traceability of clinically relevant software changes without imposing unnecessary administrative burden. The regulation also defines the entities authorized to issue UDI codes, known as issuing agencies, which include:

- GS1;
- HIBCC;
- ICCBBA;
- IFA GmbH.

Manufacturers are responsible for generating and maintaining the UDI, ensuring its correct assignment, and registering the required UDI data within the European Database on Medical Devices (EUDAMED). The regulation specifies the electronic format, mandatory data fields, validation mechanisms, and update procedures necessary to maintain accurate and up-to-date device information within the database. In the context of SwiftMR, implementation of the UDI system supports software version traceability, post-market surveillance activities, and the management of software updates within clinical environments. From an HTA perspective, UDI contributes to organizational transparency, facilitates incident reporting and recall procedures, and supports long-term monitoring of AI-based medical software deployed across multiple healthcare settings.

1.2. GDPR within MDR Framework

SwiftMR is an AI-based software as a medical device designed for accelerated MRI image reconstruction and enhancement. Its deployment requires the continuous exchange and processing of personal data within hospital IT networks and, where applicable, cloud-based platforms supporting image reconstruction workflows. As a result, the protection of personal data and the implementation of appropriate security measures represent critical prerequisites for the lawful and safe use of the technology in clinical environments. In addition to compliance with the Medical Device Regulation (EU) 2017/745, AI-based medical software operating within digital and cloud infrastructures must adhere to the requirements of the General Data Protection Regulation (GDPR) (European Union 2016). While the MDR focuses on safety, performance, and regulatory conformity of medical devices, the GDPR governs the lawful processing of personal data, establishing a complementary regulatory pillar for the deployment of AI-based healthcare technologies.

The provision of harmonised and state-of-the-art ISO/IEC standards, together with comprehensive technical documentation, supports the practical implementation of these regulatory requirements, as discussed in Section 1.1. The General Data Protection Regulation, formally identified as Regulation (EU) 2016/679, was adopted on 27 April 2016 and became fully applicable on 25 May 2018, replacing Directive 95/46/EC. As a regulation, the GDPR is directly applicable in all Member States and establishes a uniform legal framework governing the processing and free movement of personal data within the European Union. The GDPR formally recognises the protection of

personal data as a fundamental right, as enshrined in Article 8 of the Charter of Fundamental Rights of the European Union.

Table 1.5: Fundamental principle of the EU regulation of GDPR

Principles	Technical and juridical description
Lawfulness, Fairness and Transparency	Personal data must be processed lawfully, fairly, and in a transparent manner with respect to the data subject.
Purpose limitation	Data must be collected for specified, explicit, and legitimate purposes and not further processed in a manner incompatible with those purposes.
Data minimization	Data processing must be limited to what is adequate, relevant, and necessary for the intended purpose.
Accuracy	Personal data must be accurate and kept up to date; inaccurate data must be rectified or erased without delay
Storage Limitation	Data must be retained only for the period necessary to fulfil the purposes of processing.
Integrity and confidentiality	Personal data must be processed in a manner that ensures appropriate security, including protection against unauthorised or unlawful processing, accidental loss, destruction, or damage, through appropriate technical and organisational measures.
Accountability	The data controller is responsible for demonstrating compliance with all GDPR principles.

From an HTA perspective, adherence to these principles is essential for evaluating the organisational, legal, and ethical sustainability of AI-based medical technologies such

as SwiftMR, particularly when deployed within complex clinical and cloud-based infrastructures.

1.2.1. Cross reference between MDR and GDPR

This section focuses on the technical and organisational standards that support the practical implementation of GDPR requirements within the MDR regulatory framework. Rather than reiterating the legal principles of data protection, the analysis addresses how GDPR compliance is operationalised for software as a medical device (SaMD) deployed within hospital IT infrastructures and cloud-based environments. For AI-based medical software such as SwiftMR, these standards represent the state of the art for data protection, cybersecurity, and IT governance, and play a key role in aligning GDPR obligations with MDR requirements related to software lifecycle management, risk control, and post-market activities. The following paragraphs describe the relevant normative references and summarise their contribution to compliance with both regulatory frameworks within the context of the Health Technology Assessment.

The increasing digitalisation of healthcare systems has led to the widespread integration of medical devices and software into complex hospital IT networks. AI-based software solutions, such as SwiftMR, rely on continuous data exchange between imaging modalities, hospital information systems, and, where applicable, cloud-based infrastructures. While this integration enables advanced clinical workflows, it also introduces new categories of risk related to network reliability, cybersecurity, data integrity, and system availability, which extend beyond traditional device-level safety considerations. In this context, risk management can no longer be limited to the medical device itself but must also address the IT environment in which the device operates. The interaction between medical software, hospital networks, and external services creates shared responsibilities between manufacturers, healthcare organisations, and IT service providers. This shift has highlighted the need for a structured framework to manage risks arising from the use of IT networks incorporating medical devices. To address this challenge, the international standard IEC 80001-1:2021 (IEC 2021), entitled “Application of risk management for IT networks incorporating medical devices – Part 1: Safety, effectiveness, and security in the implementation and use of connected medical devices or connected health software”, defines a comprehensive approach to IT risk management within healthcare organisations. The standard builds upon earlier editions and reflects the evolution of connected medical software and cybersecurity threats. IEC 80001-1 establishes high-level requirements for organisations involved in the deployment and operation of medical devices within IT networks, covering the phases before, during, and after system integration. The framework addresses three interrelated objectives: safety, effectiveness, and security, recognising that failures in IT infrastructure may indirectly impact clinical performance and patient care. Within a healthcare facility

implementing SwiftMR as a software-based medical device, the application of IEC 80001-1 supports a structured approach to IT risk management, which may include:

definition of the healthcare IT network architecture, including servers, workstations, interfaces, PACS, RIS, HIS, firewalls, and network segmentation;

identification of IT-related risks associated with system integration, such as loss of connectivity, latency, data corruption, unauthorised access, malware or ransomware attacks, and software malfunctions;

- evaluation and prioritisation of identified risks based on their likelihood and potential impact;
- implementation of appropriate risk control measures, including authentication mechanisms, encryption, network segmentation, logging and monitoring, intrusion detection systems, backup strategies, and redundancy;
- continuous monitoring and periodic review of IT risks in response to software updates, infrastructure changes, and emerging threats;
- cooperation and clear allocation of responsibilities among stakeholders, including the hospital IT department, clinical personnel, and the software manufacturer.

IEC 80001-1 does not prescribe a predefined level of acceptable risk; instead, it requires organisations to define risk acceptance criteria in accordance with their internal policies and regulatory obligations. The effective implementation of the standard typically requires multidisciplinary collaboration between clinical staff, IT specialists, and manufacturers.

The IEC 80001-1:2021 standard provides a structured reference framework that supports the alignment between the Medical Device Regulation (EU) 2017/745 and the General Data Protection Regulation (GDPR) in the context of IT networks incorporating medical devices. Although the standard is not formally harmonised under either regulatory framework, it represents the current state of the art for managing risks related to safety, effectiveness, and security within healthcare IT infrastructures. In this assessment, the standard is used to frame the evaluation of IT-related risk management processes associated with connected medical software. Its application supports the implementation of cybersecurity and network risk controls consistent with Annex I, Section 17 of the MDR, while also contributing to the adoption of appropriate technical and organisational measures in line with Articles 25 and 32 of the GDPR, concerning data protection by design and security of processing.

While IEC 80001-1 provides the structural and organisational foundation for IT risk management, the practical implementation of security and data protection controls required by the General Data Protection Regulation (GDPR) relies on additional international standards. In this regard, the ISO/IEC 27000 family represents the state

of the art for translating high-level risk governance principles into concrete technical and organisational measures.

In particular, ISO/IEC 27001:2022 defines the requirements for an Information Security Management System (ISMS), supporting the systematic identification, treatment, and monitoring of information security risks. Building upon this foundation, ISO/IEC 27701:2019 extends the ISMS framework to include privacy-specific controls through the establishment of a Privacy Information Management System (PIMS), enabling the operational implementation of GDPR principles such as accountability, data minimisation, and data protection impact assessment. Given the cloud-based deployment of SwiftMR, standards addressing cloud-specific risks play a complementary role. ISO/IEC 27017:2015 provides guidance on information security controls for cloud services, clarifying shared responsibilities between cloud service providers and data controllers, while ISO/IEC 27018:2019 focuses on the protection of personally identifiable information in public cloud environments, supporting compliance with GDPR requirements related to security of processing, breach notification, and international data transfers. Finally, ISO/IEC 27799:2016 adapts information security controls to healthcare environments, ensuring that the confidentiality, integrity, and availability of clinical data are preserved across healthcare information systems. This standard reinforces the link between information security governance and patient safety, complementing the risk-based approach established by IEC 80001-1 within the MDR framework. Together, IEC 80001-1 and the ISO/IEC 27000 family form a coherent and layered compliance architecture, in which IT risk governance, cybersecurity, privacy protection, and cloud security are addressed in an integrated manner. From an HTA perspective, this layered approach supports the evaluation of organisational readiness, data governance maturity, and long-term sustainability of AI-based medical software deployed in real-world clinical settings.

In summary, compliance of the SwiftMR AI-based software with the two European regulatory frameworks is achieved through a cross-compliance approach integrating the requirements of the Medical Device Regulation (EU) 2017/745 and the General Data Protection Regulation (GDPR). Rather than relying on a single regulatory reference, this approach is supported by a coherent set of ISO and IEC international standards, which collectively address IT risk management, information security, privacy protection, cloud governance, and healthcare-specific cybersecurity requirements. Table 1.6 summarises how the standards adopted by AIRS Medical contribute to compliance with both regulatory pillars, highlighting their respective coverage with respect to MDR safety and performance requirements and GDPR data protection obligations.

Table 1.6: Cross compliance of GDPR and MDR of the ISO/IEC standard adopted by AIRS Medical for development of SwiftMR.

Standard	Purpose	Coverage MDR	Coverage GDPR
IEC 80001-1:2021	IT system management risk with connected medical devices	Annex I section 17.2-17.4	Article 25-32
ISO/IEC 27001:2022	ISMS- Information security managment	Section 17 (measure of IT safety in the QMS)	Article 32
ISO/IEC 27701:2019	Privacy estension (PIMS)	Section 23	Article 5-30
ISO/IEC 27017:2015	Safety of cloud service	Section 17	Article 28-32
ISO/IEC 27018:2019	personal data protection in the public cloud	Section 17	Article 6-28-33-34
ISO/IEC 27799:2016	Cybersecurity in health care	Section 17	Article 9-32

1.2.2. IT networks and workflow surrounding data security

The normative references described in Section 1.2.1 are fundamental for the implementation of SwiftMR, an AI-based software solution designed for healthcare IT environments. This section provides a detailed overview of the data protection workflow, describing how data are transmitted, processed, and managed within the network infrastructure. It also outlines the security measures implemented to ensure compliance with GDPR requirements and cybersecurity standards (AIRS Medical 2025).

Once the connection between the MRI system and the PACS infrastructure is established, SwiftMR performs image reconstruction on a cloud-based platform and subsequently transmits the enhanced images back to the local system. Throughout this process, regulatory requirements are applied to ensure compliance with the core GDPR principles, particularly regarding the confidentiality, integrity, and protection of sensitive clinical data.

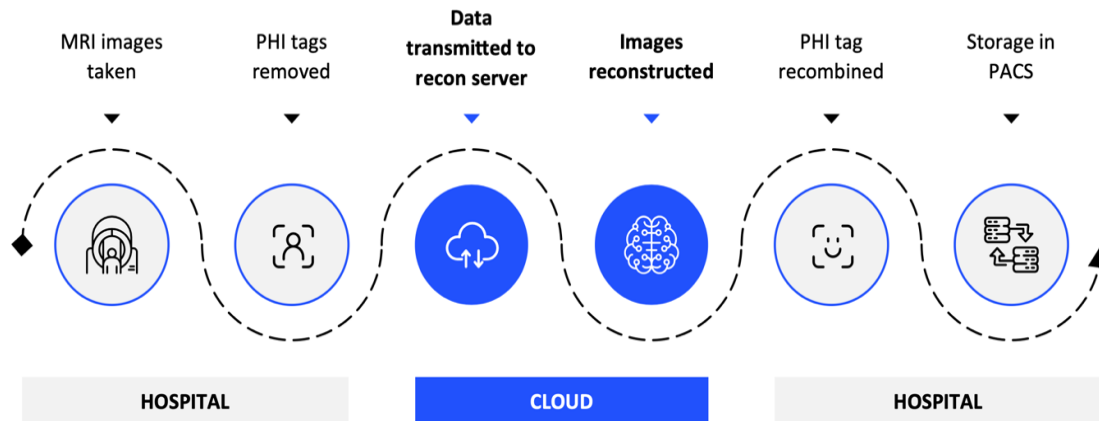


Figure 1.1: Cloud-Based product Workflow. Reproduced from the technical documentation of (AIRS Medical 2025)

Image acquisition can occur in two different ways: either through low-sampled k-space data acquisition or via standard-of-care (SOC) acquisition modalities. The latter approach aims to improve image quality in terms of signal-to-noise ratio, spatial resolution, and other parameters that enhance diagnostic performance. Regardless of the acquisition method, images are generated in DICOM format and exported to a predefined PACS receiver, ensuring secure data handling during transmission and cloud-based AI reconstruction. Data anonymisation is performed by removing DICOM tags associated with protected health information (PHI).

Following image reconstruction by the deep learning algorithm, the processed images are transmitted back to the local PACS environment, where the original DICOM tags can be securely reattached for authorised re-identification. All data transmitted within the cloud-based IT infrastructure are encrypted to ensure confidentiality and integrity. This section describes the security controls adopted to comply with GDPR requirements and to protect sensitive clinical information throughout the data lifecycle. These security measures include transport encryption, network access control lists (ACLs), application-level security controls, and identity and access management mechanisms. Together, these elements safeguard communication integrity and protect sensitive data from unauthorised access.

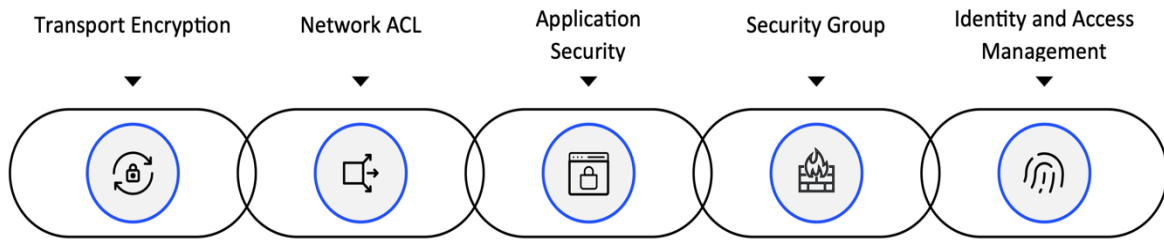


Figure 1.2: Cloud based security measures reproduced from technical implementation of (AIRS Medical 2025)

Transport encryption is applied to all communications to ensure secure transmission across potentially untrusted networks, preventing eavesdropping, tampering, and data interception. The adopted protocol is TLS version 1.2 or higher. Network ACLs regulate both internal and external traffic at the subnet level, allowing or denying access based on predefined security rules. This additional layer of control ensures that legitimate traffic is permitted while potential threats are effectively blocked. Application security measures include strict access control policies and real-time monitoring systems designed to detect suspicious activities promptly and trigger immediate alerts. This contributes to maintaining system integrity and enables rapid responses to potential security incidents. Security groups provide granular control over workloads and database access. Acting as virtual firewalls, they regulate authorised access to system resources while preventing unauthorised or potentially harmful traffic, thereby reinforcing overall network security.

Identity and access management (IAM) mechanisms ensure that only authorised users can access sensitive data. Access privileges are managed through a hierarchical role-based permission structure designed to minimise risk exposure. To ensure data confidentiality and integrity during both transmission and storage, industry-standard encryption techniques are adopted. Login credentials are protected using a one-way PBKDF2 hashing protocol, while stored data are encrypted using the symmetric AES-256 encryption standard, widely recognised as a global benchmark for secure data protection. In accordance with GDPR Article 32 on data processing security, the cloud-based IT platform incorporates robust backup and recovery strategies. These measures ensure data availability, system resilience, rapid recovery capabilities, and data integrity in the event of system disruptions or data loss.

The backup strategy follows industry best practices and consists of two main components:

- **Automated Backups:** Regular automated backups are scheduled to capture the state of the data and ensure updated recovery points;

- **Immutable Backups:** This is a more robust form of data backup in non-change format, ensuring an additional level of security, protection and integrity of the data, both for external parties and the administrative service.

SwiftMR also implements a structured disaster recovery plan aimed at minimising service disruptions. This plan includes:

- **Recovery Time Objective and Recovery Point Objective:** clearly defining metrics ensures that services are restored in a timely manner and that data loss is minimised during anomalies;
- **Failover Mechanism:** an automated system has been implemented to redirect traffic, workloads and databases to backup environments during usage.

To further enhance reliability, regular integrity checks and validation procedures are performed on backup data. Multiple backup versions are maintained at different time intervals to ensure data recoverability. Backup security is reinforced through several additional strategies. Role-based access controls and IAM systems restrict backup access to authorised personnel only. Continuous monitoring and alert systems allow rapid detection of unauthorised access attempts or operational anomalies. Furthermore, all backup data, both in transit and at rest, are encrypted to ensure maximum protection. Security monitoring represents another critical component of the infrastructure. This includes detecting insider threats, identifying unauthorised access to backup networks or services, and responding according to established security policies.

Additionally, security incident monitoring ensures overall service reliability through the timely identification, containment, and mitigation of both internal and external cybersecurity threats.

1.3. Regulatory reference of AI and AI Act

Regulation (EU) 2024/1689, known as the Artificial Intelligence Act (AI Act), represents the first comprehensive and harmonised regulatory framework worldwide governing the design, development, placing on the market, use, and market surveillance of artificial intelligence systems within the European Union.

The primary objective of the AI Act is to ensure that AI systems are developed and deployed in a manner that is safe, trustworthy, and respectful of fundamental rights, health, and safety, while simultaneously fostering innovation by providing a clear, predictable, and risk-proportionate regulatory environment, particularly for small and medium-sized enterprises and start-ups. The regulation is grounded in an anthropocentric approach to artificial intelligence, reflecting core European values such as human dignity, ethical responsibility, transparency, and accountability, as enshrined in Article 2 of the Treaty on European Union and the Charter of Fundamental Rights of the European Union. In this context, principles such as

transparency, explainability, and meaningful human oversight constitute essential requirements for the lawful use of AI systems.

From a regulatory perspective, the AI Act is conceived as a horizontal instrument that complements existing sector-specific legislation, notably the Medical Device Regulation (EU) 2017/745 (MDR) and the General Data Protection Regulation (GDPR). While the GDPR primarily addresses the protection of personal and sensitive data, particularly within cloud-based and distributed IT infrastructures, the AI Act introduces additional requirements related to algorithmic risk, governance, and oversight.

For AI-based medical devices requiring CE marking, compliance with the AI Act must therefore be achieved in conjunction with MDR and GDPR requirements, resulting in a unified but more complex regulatory pathway for a single product. This section is based on the provisions of Regulation (EU) 2024/1689 of the European Parliament and of the Council on artificial intelligence (Artificial Intelligence Act), which constitutes the primary regulatory reference for the governance of artificial intelligence systems within the European Union (Artificial Intelligence Act 2024).

1.3.1. Definition and scope

According to Article 1(1) of Regulation (EU) 2024/1689 (Artificial Intelligence Act), the purpose of the Regulation is to improve the functioning of the internal market and to promote the development and uptake of human-centric and trustworthy artificial intelligence, while ensuring a high level of protection of health, safety, and fundamental rights, including democracy, the rule of law, and environmental protection, against the potential harmful effects of AI systems within the European Union. Within the context of AI-based medical software such as SwiftMR, this definition has several relevant implications. First, the human-centric approach mandated by the AI Act requires that AI systems operate under appropriate human oversight, ensuring that clinical supervision remains central in the decision-making process. Second, the protection of health and safety is closely linked to the adoption of established regulatory and quality assurance frameworks, including quality management systems, risk management processes, and cybersecurity and data protection measures, as defined by ISO 13485, ISO 14971, and the Medical Device Regulation (EU) 2017/745 (MDR).

Finally, the protection of fundamental rights, particularly with respect to personal and sensitive health data, is intrinsically connected to compliance with the General Data Protection Regulation (GDPR). The AI Act establishes a comprehensive regulatory framework that defines the main elements of the AI market and the corresponding compliance requirements. These include the identification of the actors involved, the conditions under which AI systems must be developed, deployed, and evaluated, the classification of AI systems according to a risk-based approach, and the definition of

prohibited practices. Full compliance with the Regulation is required for AI systems classified as high-risk. With regard to the scope of application, the Regulation applies to providers placing AI systems on the European market, irrespective of whether they are established within the Union, as well as to deployers using AI systems within the EU. According to Article 3(4) of the AI Act, a deployer is defined as “a natural or legal person, public authority, agency, or other body that uses an AI system under its authority, except where the AI system is used in the course of a personal, non-professional activity”.

The AI Act follows a risk-based regulatory approach, identifying four main risk categories, alongside a separate classification for general-purpose AI systems subject to specific requirements. AI practices deemed unacceptable are explicitly prohibited, while high-risk AI systems are subject to the most stringent compliance obligations. This aspect is particularly relevant for SwiftMR, which, as AI-based software integrated into a medical device, falls within the scope of high-risk AI systems and is therefore subject to full regulatory compliance. The Artificial Intelligence Act entered into force on 1 August 2024, with a phased implementation period allowing stakeholders time to adapt to the new requirements. Prohibited AI practices under Article 5 will become enforceable from 2 February 2025, while full compliance obligations for high-risk AI systems will apply from 2 August 2026. AI systems integrated into medical devices benefit from an extended transitional period until 2 August 2027 to ensure full alignment with the new regulatory and quality requirements. Although transitional periods are provided for, early understanding and integration of the AI Act requirements are essential. In particular, compliance assessment processes and technical documentation, including the risk management file, must be progressively updated to ensure alignment with the AI Act in conjunction with the MDR and IVDR frameworks.

1.3.2. Classification of AI product based on risk

The Artificial Intelligence Act adopts a risk-based regulatory approach for the classification of AI systems, whereby regulatory obligations are proportionate to the potential impact of the system on health, safety, and fundamental rights. Certain AI practices are explicitly prohibited under Article 5 of the AI Act, including forms of subliminal manipulation, exploitation of individual or group vulnerabilities, social scoring, real-time remote biometric identification in public spaces, and other practices deemed incompatible with the protection of fundamental rights. AI systems falling within these prohibited categories are excluded from the European market. Beyond prohibited practices, Article 6 of the AI Act defines the conditions under which AI systems are classified as high-risk. In particular, AI systems are considered high-risk when they are intended to be used as a safety component of a product, or when the AI system itself constitutes a product subject to harmonised Union legislation listed in Annex I of the Regulation.

Medical devices regulated under Regulation (EU) 2017/745 (MDR) and in vitro diagnostic medical devices regulated under Regulation (EU) 2017/746 (IVDR) fall within this category and are therefore automatically classified as high-risk AI systems when they incorporate artificial intelligence. Consequently, AI-based medical software, including software integrated into or constituting a medical device, is subject to the full set of requirements established by the AI Act for high-risk systems. In this context, no exemptions from the high-risk classification are foreseen for medical devices, given their potential impact on patient safety and clinical decision-making. In addition to high-risk systems identified through harmonised legislation (Annex I), the AI Act also defines a list of high-risk use cases in Annex III, covering specific application domains such as biometric identification, critical infrastructure, education and vocational training, employment, access to essential services, law enforcement, and migration and border control. For AI systems not subject to harmonised Union legislation, classification under Annex III depends on whether the system directly affects health, safety, or fundamental rights in decision-making processes. In such cases, manufacturers are required to conduct a prior conformity assessment and to register the system in accordance with Article 49(2) of the AI Act. Article 7 of the AI Act empowers the European Commission to amend Annex III in order to include additional AI systems where their risk profile is considered equivalent to or higher than that of existing high-risk systems, thereby ensuring the dynamic adaptation of the regulatory framework to technological evolution.

The Regulation also introduces a specific category for general-purpose AI models, defined in Article 3(63) as AI models capable of performing a wide range of tasks and being integrated into multiple downstream applications. General-purpose AI systems are not classified as high-risk per se but are subject to specific obligations related to transparency, technical documentation, and dataset summaries, as set out in Articles 52 to 56. However, where a general-purpose AI system is integrated into a high-risk product or used in a prohibited practice, it must comply with the full requirements applicable to high-risk AI systems. Finally, the AI Act distinguishes between limited-risk and minimal-risk AI systems. Limited-risk systems are subject primarily to transparency and information obligations, while minimal-risk systems fall outside the scope of the AI Act and remain regulated under general EU legislation. This graduated classification reflects the legislator's intention to concentrate regulatory efforts on high-risk applications, which are considered most critical for the protection of fundamental rights and the anthropocentric approach underpinning the Regulation.

Table 1.7: Classification of the AI systems according to the AI Act.

Category	Article	Description
Unacceptable risk	Art. 5	Social scoring, subliminal manipulation

High-risk AI	Art 6, annex I and III	harmonisation with European regulations listed in Annex 1, Section A, and scope of application in Annex III.
Limited risk	Art. 50-52	The impact on health, safety and fundamental rights is not direct.
Minimal risk	Exclusion from the other classification	Not subject to the current regulation
General purpose AI	Art. 52-56, Art. 3 (63)	multimodal generators, large language models

1.3.3. Documentation and Conformity Assessment for High-risk AI Systems

According to Regulation (EU) 2024/1689, conformity of high-risk artificial intelligence systems requires the preparation and maintenance of specific technical and regulatory documentation, as well as the completion of an appropriate conformity assessment procedure.

This section outlines the technical, procedural, and administrative requirements necessary to demonstrate compliance of high-risk AI systems with the provisions of the AI Act. In particular, the Regulation mandates the availability of a comprehensive set of compliance documents, including technical documentation, a risk management framework, data governance and data quality documentation, a post-market monitoring plan, and an EU declaration of conformity.

Depending on the applicable conformity assessment route, notified bodies may be involved in the evaluation of the documentation and of the overall compliance of the AI system. Upon successful completion of the conformity assessment procedure, the CE marking may be affixed to the AI system, serving as evidence of conformity with Union harmonisation legislation, where applicable, including Regulation (EU) 2017/745 on medical devices.

Conformity requirements

According to Article 8 of Regulation (EU) 2024/1689, entitled “Requirements of conformity”, high-risk AI systems shall comply with the requirements set out in the AI Act and, in addition, be developed and implemented in accordance with the state of the art in the field of artificial intelligence. Compliance must be supported by an

appropriate risk management framework, as further specified in Article 9 of the Regulation.

For AI systems that are subject to harmonised Union legislation listed in Annex I, Section A, such as Regulation (EU) 2017/745 (MDR) and Regulation (EU) 2017/746 (IVDR), conformity documentation may be integrated in order to demonstrate compliance with both regulatory frameworks. In such cases, the conformity assessment procedure, including the involvement of a notified body where applicable, may be streamlined, provided that all applicable requirements under the AI Act and the sector-specific legislation are adequately addressed.

Risk management assessment

In accordance with Article 9 of Regulation (EU) 2024/1689, manufacturers of high-risk AI systems are required to establish, document, and continuously update a risk management system throughout the entire lifecycle of the AI system. Risk management is conceived as an iterative and continuous process, aimed at identifying, evaluating, and mitigating risks associated with the intended use of the AI system, as well as reasonably foreseeable misuse.

The risk management process comprises several key steps, including: (i) the identification and analysis of known and reasonably foreseeable risks that the high-risk AI system may pose to health, safety, and fundamental rights when used for its intended purpose; (ii) the estimation and evaluation of such risks, including scenarios of reasonably foreseeable misuse; (iii) the evaluation of residual risks based on the analysis of data collected after the system has been placed on the market, in line with the post-market monitoring requirements set out in Article 72 of the AI Act; and (iv) the adoption and implementation of appropriate risk control and mitigation measures aimed at reducing risks to an acceptable level.

Pursuant to Article 9(6) of the AI Act, the risk management system shall include testing procedures designed to evaluate risks at different stages of development prior to placing the AI system on the market or putting it into service. These tests are intended to support the identification of the most effective risk mitigation measures and to ensure that any residual risks are assessed and deemed acceptable in light of the intended purpose of the system.

In the case of AI-based medical software such as SwiftMR, the requirements of the AI Act are closely aligned with the existing risk management framework established under ISO 14971:2019 for medical devices and its technical guidance provided by ISO/TR 24971:2020. This alignment enables the integration of AI-specific risk considerations within an already established medical device risk management process, ensuring consistency with both the AI Act and the Medical Device Regulation (EU) 2017/745 (MDR).

Data governance

Article 10 of Regulation (EU) 2024/1689 establishes specific requirements related to data governance and data quality for high-risk AI systems, recognising these elements as critical determinants of AI model reliability and performance. The objective of these provisions is to ensure that training, validation, and testing datasets are appropriate for the intended purpose of the AI system and are managed in a manner that minimises risks to health, safety, and fundamental rights.

In particular, datasets used for training, validation, and testing shall be relevant, representative of the target population, sufficiently complete, and free, as far as possible, from errors, distortions, and biases that could lead to discriminatory outcomes, compromise safety, or negatively affect health. These requirements are intended to reduce the risk of systematic bias and to support the robustness and generalisability of high-risk AI systems in their intended clinical context.

Where the correction or mitigation of biases requires the processing of sensitive personal data, such processing must be carried out in full compliance with the General Data Protection Regulation (GDPR). In this context, Article 10(5) of the AI Act explicitly refers to the application of GDPR provisions, including safeguards related to the lawful processing of special categories of personal data. Appropriate technical and organisational measures, such as pseudonymisation, access control, and logging mechanisms, must be implemented to ensure that access to sensitive data is strictly limited to authorised personnel and purposes.

Furthermore, the use of sensitive data for bias detection and correction shall be limited to what is strictly necessary and supported by appropriate documentation. Once the intended corrective or validation purpose has been achieved, such data should no longer be retained, in accordance with the principles of data minimisation and storage limitation under the GDPR.

The verification of data governance measures, data protection safeguards, and dataset integrity must be completed prior to placing the AI system on the market or putting it into service. This ensures alignment between the requirements of the AI Act and the GDPR and supports the overall conformity assessment process for high-risk AI systems.

Technical documentation

The requirements for technical documentation of high-risk AI systems are set out in Article 11 of Regulation (EU) 2024/1689. For such systems, technical documentation shall be established prior to placing the AI system on the market or putting it into service and shall be kept up to date throughout the system's lifecycle. The documentation must be sufficiently detailed and clear to enable notified bodies and competent national authorities to assess compliance with the requirements of the AI Act.

Where a high-risk AI system constitutes, or is integrated into, a product subject to harmonised Union legislation listed in Annex I, the technical documentation serves as a comprehensive repository of all relevant information required to demonstrate conformity with both the AI Act and the applicable sector-specific regulation. The minimum content of the technical documentation is specified in Annex IV of the AI Act, which outlines the information necessary for the conformity assessment process.

Annex IV is structured into nine sections covering, among other aspects, general system information, technical design and development, data management and validation, risk management, performance evaluation, and cybersecurity measures. In particular, the documentation shall include general information such as the intended purpose of the AI system, the identity of the provider, system versioning and traceability, interfaces with other hardware, software or AI systems, distribution and deployment modalities, hardware requirements, user interfaces for deployers, and instructions for use.

The technical documentation must also provide a detailed description of the design and development of the AI system, including the underlying algorithms, model architecture, key design choices, and assumptions relevant to the interaction with natural persons or specific groups. This includes information on training and pre-training methodologies, datasets used for training, validation and testing, data labelling and cleaning procedures, outlier detection methods, and the computational resources employed.

Furthermore, validation and testing procedures shall be documented using appropriate performance and safety metrics aligned with the intended purpose of the system. This includes evidence of robustness, accuracy, and the assessment of potential discriminatory impacts. The documentation shall also address cybersecurity measures implemented to protect the AI system throughout its lifecycle.

In accordance with Article 9, the technical documentation must be supported by a comprehensive risk management system covering all applicable harmonised standards and technical specifications. In addition, a copy of the EU declaration of conformity, as required under Article 47, shall be included. These elements collectively support post-market monitoring and surveillance activities, as set out in Article 72 of the AI Act.

Record keeping and logging

High-risk AI systems are required to support comprehensive documentation of the entire product lifecycle through appropriate record-keeping and logging mechanisms. These mechanisms enable traceability of system behaviour and facilitate the monitoring of performance in accordance with the system's intended purpose.

In line with Articles 72 and 74 of Regulation (EU) 2024/1689, providers of high-risk AI systems shall establish a post-market monitoring system supported by logging

capabilities that allow for the continuous collection, documentation, and analysis of relevant performance data following the placement of the system on the market or its putting into service.

Such logging and monitoring activities are essential to detect potential malfunctions, performance degradation, or unintended effects that may pose risks to health, safety, or fundamental rights. Where serious incidents or malfunctioning occur, providers are required to report them in accordance with Article 73 of the AI Act. These obligations ensure that high-risk AI systems remain under continuous regulatory oversight throughout their operational lifecycle and support timely corrective actions where necessary.

Transparency and information to deployers

Article 13 of Regulation (EU) 2024/1689 establishes specific requirements concerning transparency and the provision of information to deployers of high-risk AI systems. In accordance with these provisions, the output of the AI system shall be interpretable and usable by deployers on the basis of clear, accurate, and comprehensive information supplied by the provider.

High-risk AI systems must therefore be designed in such a way that deployers are able to understand and appropriately interpret the system's outputs within the context of its intended purpose. To this end, the provider is required to supply detailed instructions for use, including information on the system's capabilities and limitations, performance characteristics, validation procedures, and the metrics used to assess performance. Where relevant, information on the quality, modality, and expected characteristics of input data shall also be provided.

The level of transparency and information provided shall be proportionate to the role and responsibilities of both the provider and the deployer, as defined within the regulatory framework of the AI Act. This ensures that deployers are adequately informed to use the system safely and effectively, without exceeding its intended scope of application.

In addition, Article 14 of the AI Act requires that high-risk AI systems be subject to appropriate human oversight measures. These measures shall ensure that the system's operation can be effectively supervised by natural persons, including the ability to interpret outputs, intervene where necessary, and override or suspend the system in the event of anomalies or unexpected behaviour.

Transparency and human oversight obligations are further supported by record-keeping and logging mechanisms, as required under Article 12 of the AI Act, which enable traceability of system behaviour and support both accountability and post-market monitoring activities.

Humans' surveillance

In accordance with Article 14 of Regulation (EU) 2024/1689, appropriate human oversight is a fundamental requirement for high-risk AI systems and represents a key safeguard against risks arising from the system's operation. Human oversight is intended to ensure that the outputs of the AI system do not lead to automated or uncritical decision-making and that residual risks identified through the risk management process are effectively mitigated.

High-risk AI systems shall therefore be designed and developed with suitable human-machine interfaces that enable effective supervision of the system throughout its use. The objective of human oversight is to allow natural persons to understand, interpret, and, where necessary, intervene in the system's operation when it is used for its intended purpose or in cases of reasonably foreseeable misuse. This requirement is particularly relevant in the context of residual risk, as identified through the risk management assessment.

Human oversight measures can be broadly distinguished into two complementary categories. The first category concerns oversight mechanisms implemented by the provider during the design and development phase of the AI system, prior to its placement on the market. These measures are embedded in the system's architecture and user interface and reflect the provider's strategy to enable safe and controlled use.

The second category relates to oversight exercised by the deployer during the operational use of the AI system. In this context, deployers must be adequately informed of the system's capabilities, limitations, and intended use, and must be able to recognise anomalies, malfunctions, or unexpected behaviour during deployment. Particular attention must be paid to the risks of automation bias and overreliance on AI-generated outputs.

Ultimately, the responsibility for decisions informed by the AI system remains with the human operator. Deployers shall retain the ability to critically assess and, where appropriate, disregard the system's output, as well as the capability to intervene or suspend the system's operation at any time. These safeguards ensure that high-risk AI systems support, rather than replace, human judgment in safety- and health-critical contexts.

Accuracy, Robustness and Cybersecurity

According to Article 15 of Regulation (EU) 2024/1689, high-risk AI systems shall be designed and developed in such a way as to achieve an appropriate level of accuracy, robustness, and cybersecurity throughout their entire lifecycle. These requirements aim to ensure that the system performs reliably under normal conditions of use and remains resilient to foreseeable disturbances or misuse.

The assessment of accuracy, robustness, and cybersecurity shall be based on metrics that are appropriate to the intended purpose of the AI system and proportionate to the risks associated with its use. In particular, the AI system shall demonstrate resilience

to errors, faults, or inconsistencies that may arise within the system itself, from its operating environment, or from interactions with natural persons.

Robustness of high-risk AI systems may be supported through the implementation of technical measures such as redundancy mechanisms, backup solutions, and fail-safe strategies, aimed at ensuring continuity of operation and controlled system behaviour in the presence of failures or anomalies.

Cybersecurity measures are implemented to protect the AI system against unauthorised access, manipulation, or interference by third parties, thereby safeguarding the integrity, availability, and reliability of the system's outputs. To this end, appropriate technical and organisational measures shall be adopted to prevent, detect, and mitigate vulnerabilities, including risks related to data poisoning, model poisoning, and model evasion attacks.

Quality Management System

According to Article 17 of Regulation (EU) 2024/1689, providers of high-risk AI systems are required to establish, implement, and maintain a quality management system (QMS) with the aim of ensuring continuous compliance with the requirements set out in the AI Act. The quality management system shall be structured in a systematic and organised manner and shall include documented policies, procedures, and written instructions covering all stages of the AI system's lifecycle. In particular, the QMS shall address the following elements:

- Design and development controls;
- Data management;
- Risk management;
- Testing and validation;
- Post-market monitoring;
- Document control and record-keeping.

As reflected by the scope of the quality management system, these processes are closely linked to the requirements set out in Articles 9 to 15 of the AI Act and are consistent with the risk management framework discussed in Section 2. The QMS therefore represents the organisational backbone through which the technical and regulatory requirements applicable to high-risk AI systems are systematically implemented and maintained.

CE conformity assessment

The Artificial Intelligence Act requires that, prior to placing a high-risk AI system on the market or putting it into service, the provider shall carry out an appropriate conformity assessment procedure. The conformity assessment framework for high-risk AI systems is set out in Articles 43 to 48 of Regulation (EU) 2024/1689 and includes

specific requirements concerning assessment procedures, documentation, the EU declaration of conformity, and CE marking.

For high-risk AI systems subject to harmonised standards pursuant to Article 40 or common specifications pursuant to Article 41, providers may follow different conformity assessment pathways. Where applicable, conformity may be demonstrated through internal control procedures as described in Annex VI of the AI Act, supported by the implementation of a quality management system and the availability of complete technical documentation. The involvement of a notified body, in accordance with Annex VII, is mandatory in specific circumstances. These include cases where harmonised standards are not available or not fully applied, where common specifications are not applied by the provider, or where harmonised standards have been published with restrictions. In such situations, the notified body performs an independent assessment of the conformity of the high-risk AI system with the requirements of the AI Act.

In the case of AI-based software regulated as a medical device, such as SwiftMR, the involvement of a notified body is typically required as part of the conformity assessment process, in alignment with the Medical Device Regulation (EU) 2017/745. This integrated approach ensures consistency between the conformity assessment procedures under the AI Act and sector-specific medical device legislation.

Upon successful completion of the conformity assessment procedure, the provider shall draw up a written EU declaration of conformity, which shall be kept for a period of ten years after the AI system has been placed on the market. Where the AI system is subject to multiple Union harmonisation acts, a single EU declaration of conformity may be issued covering all applicable requirements, provided that it remains accurate and up to date.

The CE marking shall be affixed to high-risk AI systems as evidence of conformity with the applicable Union harmonisation legislation. For digital AI systems, the CE marking shall be displayed on the system interface in a visible, legible, and accessible manner. Where a notified body is involved in the conformity assessment, its identification number shall be indicated alongside the CE marking, where applicable. Where the AI system is also subject to other Union harmonisation legislation, the CE marking shall indicate conformity with all applicable regulatory frameworks.

Post-market monitoring

In accordance with Article 72 of Regulation (EU) 2024/1689, providers of high-risk AI systems are required to establish, document, and implement a post-market monitoring system that is proportionate to the nature of the technology and to the level of risk associated with the AI system. The purpose of the post-market monitoring system is to ensure the continuous assessment of the AI system's performance and its ongoing compliance with the requirements set out in Chapter III, Section 2 of the AI Act.

The post-market monitoring system shall be capable of systematically collecting, documenting, and analysing relevant data provided by deployers or obtained from other appropriate sources throughout the operational lifetime of the AI system. These activities support the identification of performance degradation, malfunctions, or unintended effects that may emerge after the system has been placed on the market or put into service.

In accordance with data protection principles, the post-market monitoring system does not rely on the unrestricted collection of sensitive personal data. Where personal data are processed, such processing must be strictly limited to what is necessary and carried out in compliance with the applicable provisions of the General Data Protection Regulation (GDPR), as described in the relevant sections of the technical documentation.

Post-market monitoring constitutes an integral component of the overall regulatory framework applicable to high-risk AI systems and is closely linked to the technical documentation required under Article 11. The content and structure of the post-market monitoring plan are further detailed in Annex IV of the AI Act, which forms part of the documentation supporting the conformity assessment process. The monitoring system shall be implemented in accordance with the transitional provisions of the AI Act and maintained throughout the lifecycle of the AI system.

Table 1.8: summary of the conformity requirements and conformance assessment of the AI Act regulation

Requirements/Document	AI Act article	Purpose	Responsible Entity
Risk management file	Art. 9	Identify, Analyse, mitigate risk	Provider
Data governance	Art. 10	Quality of dataset	Provider
Technical documentation	Art. 11	System architecture and parameters	Provider
Logging and traceability	Art. 12	Data of performance of the system	Provider

Transparency and information to deployers	Art. 13	Correct usage of the model's outputs	Provider
Humans' surveillance	Art.14	Residual risk mitigation	Provider and deployers
Accuracy, robustness and cybersecurity	Art. 15	Integrity and protection of the data	Provider
Quality management system	Art. 17	Ensure process quality and conformity requirements	Provider
CE marking and Declaration of Conformity	Art. 43-48	Certify compliance	Provider
Post-market surveillance plan	Art. 72	Continuously ensure conformity requirements after placing on the market	Provider

1.3.4. Application on SwiftMR, limitations and final considerations

Regulation (EU) 2024/1689, commonly referred to as the Artificial Intelligence Act, establishes a comprehensive regulatory framework defining the principles, requirements, and governance mechanisms applicable to high-risk artificial intelligence systems. The Regulation combines horizontal obligations related to safety, transparency, and fundamental rights protection with detailed technical, procedural, and organisational requirements aimed at ensuring regulatory compliance throughout the lifecycle of AI systems. Within this regulatory context, SwiftMR, developed by AIRS Medical, is classified as a high-risk AI system, as it constitutes software as a medical device integrated into the diagnostic workflow. In accordance with Article 6(1) of the AI Act, AI systems that are incorporated into medical devices regulated under Regulation (EU) 2017/745 fall within the high-risk category and are therefore subject to the full set of obligations established by the Regulation.

Implementation of AI Act on SwiftMR

From a regulatory perspective, AI-based medical software such as SwiftMR is required to demonstrate compliance with both Regulation (EU) 2017/745 on medical devices (MDR) and Regulation (EU) 2024/1689 on artificial intelligence (AI Act). This dual compliance requirement arises from the fact that SwiftMR constitutes software as a medical device incorporating artificial intelligence within the diagnostic workflow. In accordance with Annex I, Section A and Article 6(1) of the AI Act, medical devices and in vitro diagnostic devices integrating AI are classified as high-risk AI systems and are therefore subject to the full set of obligations established by the Regulation.

The application of the AI Act requirements to SwiftMR can be structured across the following key regulatory domains:

1. Technical documentation and risk management (Articles 9–11).

The provider is required to establish comprehensive technical documentation describing the system architecture, intended purpose, hardware and software components, datasets, and training, validation, and testing procedures. In addition, a risk management file aligned with ISO 14971:2019 and ISO/TR 24971:2020 shall demonstrate how risks to patient safety, clinical performance, and data integrity are identified, evaluated, and mitigated throughout the system lifecycle.

2. Data governance and GDPR compliance (Article 10 and Regulation (EU) 2016/679).

Data used for training, validation, and clinical evaluation must be processed in compliance with the General Data Protection Regulation (GDPR). This includes the application of appropriate safeguards such as pseudonymisation, access control, and data quality management, ensuring that datasets are relevant, reliable, and, as far as possible, free from bias or distortion.

3. Transparency and human oversight (Articles 13–14).

The AI system is required to provide outputs that are transparent and interpretable by clinicians and radiologists, enabling informed clinical decision-making. Appropriate human oversight measures must ensure that AI-generated outputs are critically assessed and that the system is not relied upon in situations that fall outside its intended diagnostic scope.

4. Accuracy, robustness, and cybersecurity (Article 15).

SwiftMR is required to achieve an appropriate level of accuracy, robustness, and cybersecurity throughout its lifecycle. Compliance with relevant technical standards, such as ISO/IEC 27001:2022 and IEC 81001-5-1:2021, supports the implementation of measures aimed at protecting data integrity, system availability, and resilience against cyber threats in cloud-based environments.

5. Quality management system and conformity assessment (Articles 17 and 43–48).

The provider shall implement a quality management system incorporating AI-specific processes in accordance with ISO 13485:2016/A11:2021, ensuring version control, traceability, and lifecycle monitoring. Depending on the applicable conformity assessment route, the involvement of a notified body may be required to verify

compliance with the obligations set out in Annex I of the MDR and Chapter III of the AI Act, supporting the affixing of the CE marking.

6. Post-market monitoring and lifecycle management (Article 72).

A post-market monitoring system shall be established following the placement of the system on the market, enabling the collection and analysis of performance data, incidents, and malfunctions reported by deployers. These activities support the continuous evaluation of regulatory compliance and, where necessary, inform corrective actions or renewed conformity assessment procedures.

Limitations and practical challenges

The harmonised regulatory framework established by the Medical Device Regulation (MDR) and the Artificial Intelligence Act is characterised by a high degree of complexity. While this framework aims to ensure safety, transparency, and protection of fundamental rights, it also introduces significant practical challenges for providers, both in the development of new AI-based medical technologies and in the adaptation of existing products such as SwiftMR to the newly applicable regulatory requirements.

One of the primary challenges relates to the current lack of fully harmonised technical standards under the AI Act. At the time of implementation, several ISO/IEC standards intended to support compliance with the Regulation are still under development. As a result, providers are required to rely on state-of-the-art practices and existing technical standards to demonstrate conformity, which may lead to interpretative uncertainty and variability in regulatory expectations.

A further challenge arises from the need to ensure dual compliance with both the AI Act and the MDR. This requires the integration of additional AI-specific obligations into existing regulatory frameworks, including technical documentation, quality management systems, risk management files, post-market monitoring plans, and conformity assessment procedures. The alignment of these requirements increases the organisational and administrative burden on providers.

The management of adaptive AI systems represents an additional regulatory challenge. AI models that require retraining or continuous learning based on new data may be difficult to reconcile with traditional conformity assessment approaches, which are inherently static. Consequently, providers may be required to implement robust change management procedures and, where applicable, update CE marking documentation to reflect significant modifications to the AI system.

Data protection requirements under the General Data Protection Regulation (GDPR) further constrain the development and optimisation of AI models. In particular, limitations on the processing of sensitive health data, while essential for safeguarding fundamental rights, may restrict the availability of data for training and performance improvement, thereby affecting the evolution of AI-based medical software.

Finally, the cloud-based architecture of SwiftMR introduces additional challenges related to the allocation of responsibilities among multiple stakeholders involved in data processing, system maintenance, and deployment. Clear definition of roles and responsibilities among providers, deployers, and third-party service operators is essential to ensure effective governance and sustained compliance with European regulatory requirements.

Final considerations

The analysis conducted in this section demonstrates that the application of the European regulatory framework to SwiftMR requires alignment with the three main regulatory pillars previously identified as normative references, namely the Medical Device Regulation (MDR), the Artificial Intelligence Act (AI Act), and the General Data Protection Regulation (GDPR). In this context, the future actions of the provider, AIRS Medical, will play a crucial role in ensuring continued compliance with the obligations introduced by the AI Act as they become fully applicable.

From a health technology assessment perspective, any decision regarding the adoption of SwiftMR should take into account not only current regulatory compliance but also the provider's ability to maintain conformity with regulatory requirements that will be fully enforced by 2 August 2027. Particular attention should be paid to data governance practices, as the quality, management, and use of datasets for training, validation, and testing represent a central element of both conformity assessment and notified body evaluation for CE marking.

Overall, SwiftMR can be considered a technology with the potential to sustain regulatory compliance over time, provided that ongoing alignment with both the MDR and the AI Act is ensured. Such alignment supports the maintenance of safety, traceability, and trustworthiness principles throughout the product lifecycle, which are essential prerequisites for the long-term sustainability of AI-based medical technologies within the European regulatory landscape.

1.4. HTA in the clinical context

Health Technology Assessment (HTA) is a structured framework that has been continuously developed and put into practice to facilitate the thorough evaluation of medical technologies such as medical devices, in vitro diagnostic devices, and pharmaceutical products at different points in their life cycle. (Kristensen et al. 2017).

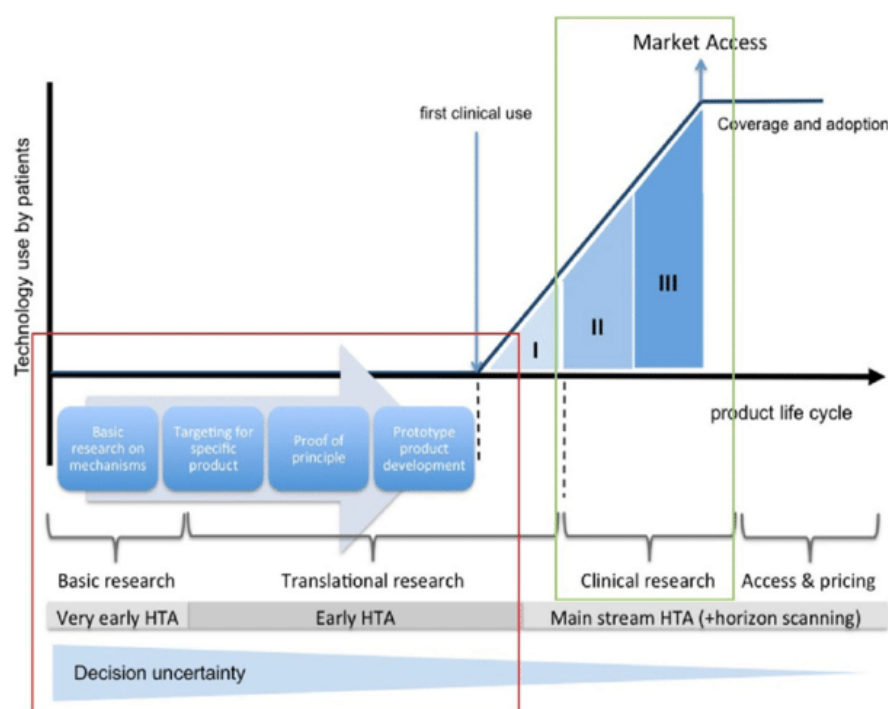


Figure 1.3: Definition of the different stage of product life cycle and HTA assessment (reproduced from: (Smith et al. 2019))

Health Technology Assessment (HTA) can be conducted at different stages of the product life cycle. As illustrated in Figure 1.3, the process may begin with a Very Early HTA, performed before the actual technological development takes place. This stage focuses on assessing the potential value, feasibility, and anticipated impact of the innovation, supporting decision-making in the earliest conceptual phase. Once a prototype becomes available, the process typically advances to the Early Assessment (Smith et al., 2019). This phase is conducted when the technology is already in a pre-market stage and aims to evaluate its preliminary clinical performance, usability, and potential cost-effectiveness. The purpose of this step is to refine the technology and ensure alignment with both clinical and non-clinical requirements—such as patient safety, workflow integration, and organizational feasibility—before market introduction. Finally, the Mainstream HTA represents the comprehensive assessment of a fully developed technology, usually performed at or near market entry. This evaluation integrates clinical, economic, ethical, and organizational perspectives to determine the overall value and readiness of the product for real-world adoption. This is the approach adopted in the present project for the evaluation of the SwiftMR software solution.

HTA aims to ensure a transparent, systematic, and standardized evaluation covering various aspects, both clinical and non-clinical, corresponding to clinical and non-clinical assessments, respectively. Each technology is subject to specific models tailored to address distinct policy questions based on its classification.

Upon closer examination of the various HTA domains, the initial four are generally considered clinical domains, often forming part of a Relative Effectiveness Assessment (REA), which could sometimes take the form of a rapid HTA, depending on the analysis's breadth and depth. These four domains are then followed by an evaluation of multidisciplinary domains that, while linked to the clinical aspects, are more focused on the administrative, economic, and organizational aspects of the healthcare environment.

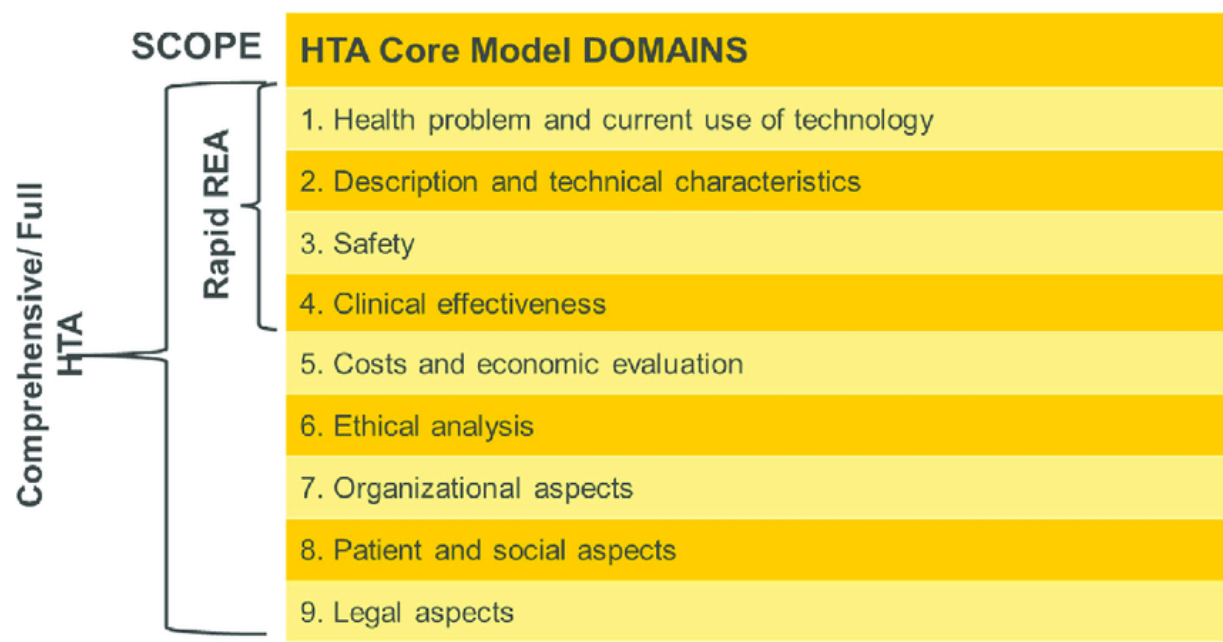


Figure 1.4: HTA domains of EUnetHTA (reproduced from: (Kristensen et al. 2017)).

Following the canonical workflow in the evaluation HTA, first the prioritization of the 9 dimensions, related to the Core model of EUnetHTA (Radaelli et al. 2014), interviews a panel of health care experts in the fields, in the way to define the relevant domain of most interest. For each of the domain is assigned a score from 1 (less importance and priority) to 9 (the most relevant and priority).

A panel of 12 experts was selected within the Ospedale di Erba. The group is composed of: the Chief of Radiology, the Deputy Chief of Radiology, two MRI technicians with over 10 years of experience in the hospital, a Clinical Engineer, the Data Protection Officer (DPO), the Head of the Prevention and Protection Service (RSPP), an Emergency Department Nurse, the IT Officer, the Operations Director, and the Medical Director. Experts were chosen from both clinical and administrative areas to cover all relevant information and ensure a comprehensive and reliable evaluation. The panel is multidisciplinary and from a single centre, reflecting the need to address all criteria for the assessment of the AI software within the hospital’s clinical environment and to consider both the advantages and limitations of the technologies under investigation. Clinical needs are addressed through the daily experience of the

radiology staff and technicians, while the administrative offices contribute insights regarding economic and management aspects of the technologies.

According to Lampe et al. (2009), there are two possible ways to apply the EUnetHTA Core Model as evident in the figure below. The first is to perform a comprehensive and multidisciplinary assessment following the standardized European framework for the evaluation of health technologies. The second is a more flexible approach, adapted to local contexts and aimed at supporting specific decision-making processes.

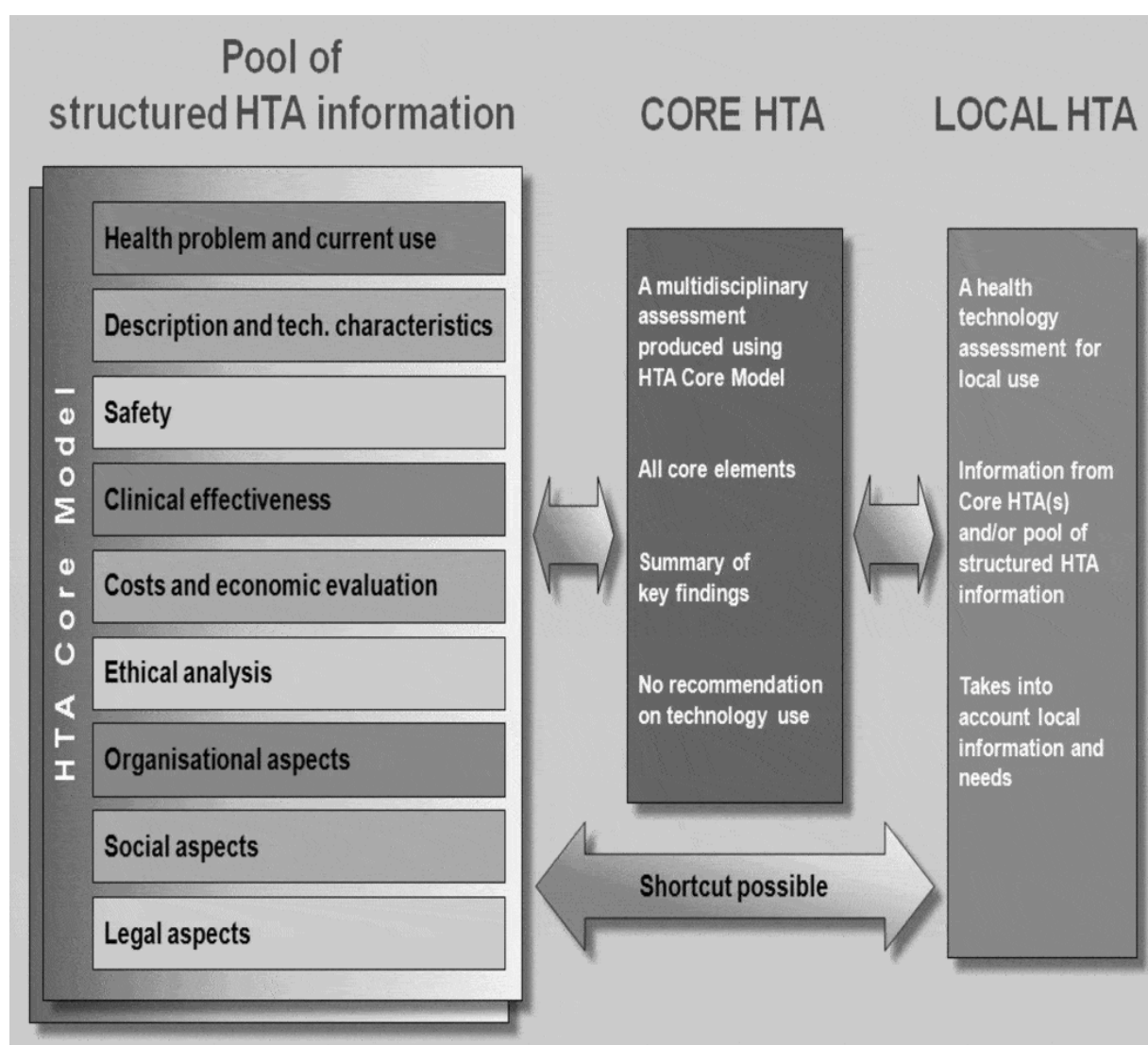


Figure 1.5: The core HTA structure (reproduced from: (Lampe et al. 2009))

In this project, the second approach has been adopted: in accordance with the Core Model, only the most relevant domains have been selected and analyzed within the HTA.

A local HTA has been carried out, meaning that only the domains considered relevant and significant to the specific context of acquisition and implementation within the hospital setting were included in the analysis. The selection of these domains was based on the prioritization table presented in Table 1.9, where each domain was rated on a scale from 1 to 9. The mean score for each domain was then calculated, and a threshold value of 3 was applied. Domains with a mean score below this threshold were deemed not relevant and therefore excluded from the local HTA, as they were considered non-transferable to the context of this study.

Table 1.9: Prioritization of the nine domains of the Health Technologies Assessment

Dimension	Priority
Clinical Effectiveness	8.17
Safety	7.92
Health problem and current use of the technologies	7.50
Description and technical characteristic of technology	6.83
Costs and economical evaluation	6.50
Organizational aspects	2.83
Ethical analysis	2.50
Patient and social aspects	2.17
Legal aspects	2.17

According to the panel of experts involved in the evaluation, the analysis includes five domains of the EUnetHTA Core Model: Health problem and current use of the technology, Description and technical characteristics of the technology, Clinical effectiveness, Safety, and Economic and financial aspects.

On this basis, the patient and social, ethical, legal, and organisational domains were not developed as independent sections of the local HTA, as they were considered to fall below the predefined relevance threshold and were therefore judged by the expert panel to have limited impact on local decision-making. Nevertheless, these domains

were not excluded from the assessment but were addressed transversally across other chapters of the analysis. In particular, organisational aspects were integrated into the economic evaluation in order to quantify workflow reorganisation in MRI services and the organisational impact of diagnostic performance on economic outcomes. Legal considerations were addressed within the regulatory framework discussed in Chapter 1, while ethical aspects were incorporated through the analysis of the Artificial Intelligence Act, with specific reference to human oversight, transparency, and risk mitigation requirements. Finally, patient and social aspects were implicitly considered within the clinical and diagnostic effectiveness evaluation, reflecting their indirect influence on patient pathways and service delivery.

In the context of the Ospedale di Erba, the analysis was conducted to provide the general management with all the relevant tools to support the decision-making process. The adoption of the SwiftMR software is consistent with the hospital's strategic policy aimed at enhancing the quality of patient care, increasing operational efficiency and revenue, and aligning with the institution's long-term strategic investment plan. For these reasons, the application of an Health Technology Assessment (HTA) represents the most appropriate methodological approach to address the formulated policy questions and to guide evidence-based decision making.

2 Health problem and current use of the technology

2.1. Clinical context: MRI in current practice

Magnetic Resonance Imaging (MRI) has been one of the most significant diagnostic technologies since the 1980s, when the first 1.5-tesla scanners were introduced to clinical practice (Kabasawa 2021). Its primary strengths lie in the exceptional soft-tissue contrast, spatial resolution, and signal-to-noise ratio (SNR) that remain unattainable with other imaging modalities (Jäger and Reiser 2001).

A major advantage of MRI lies in its non-ionising physical principles, which considerably enhance patient safety compared with X-ray-based techniques. Moreover, MRI supports multiparametric analysis, enabling both morphological and functional assessment of organs. This allows clinicians to evaluate vascularisation, contrast-agent pharmacokinetics, and water-proton density within healthy and pathological tissues (Chikui et al. 2012), (Hagiwara et al. 2023).

The multiparametric capability is achieved through different pulse sequences that exploit distinct relaxation mechanisms of hydrogen nuclei. T1-weighted sequences rely on longitudinal relaxation, whereas T2-weighted sequences are based on transversal relaxation; proton-density-weighted images complement these by adjusting repetition and echo times (Panda et al. 2017).

The integration of these sequences provides a comprehensive representation of anatomical and physiological features across various regions of the body, including the brain, abdomen, cardiovascular system, and musculoskeletal structures, making MRI a cornerstone of modern diagnostic imaging (Hagiwara et al. 2023).

However, despite its diagnostic versatility, MRI remains a time-intensive and resource-demanding procedure, where acquisition duration, motion artefacts, and noise levels can limit efficiency and image quality. These challenges form the basis for technological innovation in image reconstruction methods, which aim to accelerate scans and improve signal quality without compromising diagnostic reliability (Zaitsev, Maclaren, and Herbst 2015).

2.2. Current reconstruction practice: gold standard and established acceleration

2.2.1. Gold standard of conventional reconstruction

Magnetic Resonance Imaging reconstruction conventionally relies on the application of the Inverse Fourier Transform to data acquired in the k-space domain, which is widely regarded as the gold standard in current clinical practice. The process involves three key encoding mechanisms: slice selection, frequency encoding, and phase encoding, which together enable the systematic excitation and spatial localisation of the body volume. These mechanisms are achieved through the use of the gradient coils integrated within the MRI scanner hardware (Bernstein, King, and Zhou 2004).

Data acquisition may be performed using various strategies, most commonly through two-dimensional sampling in k-space. In this configuration, a single slice-selection gradient, a frequency-encoding gradient, and a phase-encoding gradient are applied. The total number of phase-encoding steps determines the spatial resolution in that direction. Depending on the pulse sequence design, the spatial domain may be sampled along Cartesian or radial/polar trajectories (Lin et al. 2003).

Three-dimensional acquisitions are also feasible, in which the volume is excited as a whole rather than slice by slice. However, this approach requires multiple phase-encoding steps and longer recovery times between excitations, leading to a substantial increase in total scan duration. Consequently, despite its robustness and diagnostic reliability, this conventional reconstruction framework is constrained by extended acquisition times and susceptibility to motion artefacts, motivating the search for more efficient reconstruction strategies (Zaitsev, Maclaren, and Herbst 2015).

2.2.2. Established acceleration techniques

In contemporary clinical practice, MRI acceleration techniques are essential to reduce scan times, improve patient comfort, and maintain throughput compatible with clinical workflow requirements. The two primary strategies currently implemented to achieve these objectives are Parallel Imaging and Compressed Sensing. Parallel Imaging leverages the spatial sensitivity information of multiple receiver coils to reconstruct undersampled data, while Compressed Sensing exploits the mathematical sparsity inherent in MR images to recover high-quality reconstructions from incoherently undersampled k-space data. These methods, often integrated into vendor-specific reconstruction pipelines, collectively represent the current gold standard for accelerated MRI acquisition in clinical environments (Hamilton, Franson, and Seiberlich 2017; Lustig, Donoho, and Pauly 2007).

Parallel imaging

Parallel imaging reconstruction exploits the use of special hardware called phase array coils, which contain multiple receiver channels. Each coil in the phase array is more sensitive to the volume close to it and less sensitive to magnetisation further away. This spatial sensitivity is reflected in the sensitivity profile of each coil in the phase array. In this configuration, the signal-to-noise ratio (SNR) is higher close to the coil but decreases consistently away from the receiver. To cover the entire field of view (FOV) of the scanned volume, the strategies of the phased array coil are exploited.

Each coil acquires a signal from the entire excited volume, but this signal is weighted according to the receiver's sensitivity profile. In the image domain, this concept is realised by multiplying the signal by the coil's local sensitivity map. In k-space, it is realised by convolving the Fourier transform of the image with the spectrum of the local sensitivity maps. This is similar to taking different views of the same object and then using the different information from each viewpoint to reconstruct the situation. Information on position from the phased array coil could reduce the need for gradient encoding for localisation in the volume and reduce the scan time by a factor of R by downsampling the k-space. Downsampling could only be performed in the direction of coil sensitivity variation, with the maximum amount of acceleration being the number of receivers in the phased array coils.

There are two types of parallel imaging reconstruction. The first type is based on reconstructing the image in its own image space. The second is based on reconstruction in the k-space domain. Both methods accelerate the scanning of the patient's body due to the limitations of the MRI discussed above. In both methods, the acquisition involves downsampling the k-space and reconstructing the image to recover the spatial information thanks to the logic of the phased array coil.

Sensitivity encoding (SENSE)

Further investigation of the reconstruction method most commonly used in clinical settings reveals that sensitivity encoding (SENSE), a parallel imaging technique, exploits local coil information to derive pixels superimposed in the downsampled reconstructed image. The number of pixels that can be unfolded from the aliasing image is theoretically equal to the number of phased array coils, but in reality, it is less than this value. A matrix of local sensitivity is estimated in a pre-scan phase and then used to derive the unfolded pixels by inverting the sensitivity matrix. The acceleration factor R directly influences the number of unfolded pixels that can be reconstructed in the final image (Pruessmann et al. 1999).

The classic SENSE method is based on local sensitivity stability, which is acquired during the pre-scan to estimate the sensitivity matrix, and then the reconstruction is performed from the aliasing pixels. The sensitivity map estimation is performed dynamically by exploiting phase accumulation in k-space during fast acquisition, so

the scan is fast, and spatial information is dynamically acquired. This method is called TSENSE.

Other methods, such as mSENSE, exploit auto-calibration for coil sensitivity to acquire lines of k-space near the centre, providing a low-resolution image reference. A common feature of parallel imaging is that the SNR decreases in the reconstructed image compared to the fully sampled one. In particular, the degree of noise is determined by a g-factor that depends on many variables, including the number of coils, the plan orientation, and the array configuration. SNR is inversely proportional to g-factor and to the square root of the acceleration factor R, a concept known as Fourier averaging in signal processing. This is an inevitable feature of all parallel imaging.

Parallel imaging is one of the most widely used methods for reconstructing fast scan images. The other method is GRAPPA. MRI vendors usually customise this solution directly in the image post-processing and the acceleration factor is 2 or 3 (Hamilton et al. 2017).

Generalized Partially Parallel Acquisitions (GRAPPA)

GRAPPA parallel imaging methods have demonstrated the ability to substantially reduce scan times without compromising diagnostic image quality. In contrast to fully sampled k-space acquisitions, which yield high-quality images at the cost of long acquisition times, GRAPPA offers a practical balance between efficiency and fidelity, making it a preferred technique in routine clinical practice (Griswold et al. 2002).

The reconstruction of missing k-space lines is performed directly in the Fourier domain. Specifically, each unacquired line is estimated as a linear combination of acquired data from neighbouring coils in the phased-array system. The weights of these linear combinations are defined by the GRAPPA kernel, which functions as a calibration filter applied across all positions in the Cartesian acquisition grid.

Typically, the kernel consists of a finite set of source points distributed along the readout and phase-encoding directions. As image quality requirements increase, a larger number of source points are used to estimate the target points, thereby increasing the number of coefficients that must be calibrated. The calibration process relies on the Auto-Calibration Signal (ACS), where additional k-space lines are acquired near the centre of the k-space to estimate the kernel weights. Once calibrated, the kernel is propagated throughout the entire dataset to reconstruct the full k-space.

GRAPPA has been shown to be more robust than SENSE, as it does not depend directly on pre-estimated coil sensitivity maps but rather on the correlations between coil signals. Nevertheless, both techniques share the same theoretical framework of parallel imaging, and in both cases the Signal-to-Noise Ratio (SNR) remains inversely related to the geometry factor (g-factor), which in GRAPPA can be analytically derive from the kernel configuration.

Self-Consistent Parallel Imaging (SPIRiT)

The latest evolution of parallel imaging is represented by SPIRiT (Iterative Self-Consistent Parallel Imaging Reconstruction), a hybrid technique that combines elements of both SENSE and GRAPPA. In this approach, the missing k-space lines resulting from accelerated acquisitions are estimated through a convolutional kernel applied across neighbouring coils, as in GRAPPA, while the reconstruction is formulated as an iterative inverse problem akin to SENSE (Lustig and Pauly 2010).

The algorithm iteratively minimises a loss function composed of two regularisation terms: one enforcing calibration consistency, ensuring coherence with the estimated k-space convolution kernel, and the other enforcing data consistency, which guarantees agreement between the measured and reconstructed k-space samples during each iteration.

A notable strength of SPIRiT lies in its flexibility: it performs the reconstruction on a Cartesian grid while remaining compatible with arbitrary acquisition trajectories. This characteristic marks a pivotal transition toward the discussion of non-Cartesian parallel imaging, where the extension of such reconstruction methods poses significant computational and methodological challenges.

Non-Cartesian parallel imaging

Non-Cartesian data acquisition offers several advantages over conventional Cartesian sampling, although it is important to note that most parallel imaging frameworks are originally formulated for Cartesian k-space reconstruction. The benefits of non-Cartesian trajectories—such as spiral, radial, or PROPELLER (blade) sampling—are twofold. First, they reduce coherent aliasing artefacts by distributing undersampling errors more uniformly across the image. Second, they allow for higher acceleration factors (R) compared with Cartesian acquisitions.

In these trajectories, skipped k-space samples are distributed in both the k_x and k_y directions, rather than being limited to the phase-encoding axis as in conventional downsampling. The reduced acquisition time results from the more efficient coverage of the k-space plane, while oversampling of the central region further improves SNR and motion robustness.

The main drawback of non-Cartesian acquisition lies in its computational complexity. Image reconstruction requires specialised algorithms such as the Non-Uniform Fast Fourier Transform (NUFFT) and gridding-based interpolation, which are subsequently followed by the Discrete Fourier Transform (DFT).

In summary, while non-Cartesian parallel imaging offers significant potential for faster and more robust MRI acquisition, its implementation introduces additional computational demands. The following section revisits the three established methods—SENSE, GRAPPA, and SPIRiT—in light of these considerations,

highlighting their respective strengths and limitations within non-Cartesian frameworks.(Hamilton et al. 2017)

The conjugate gradient SENSE (CG SENSE) is the counterpart of the same methods applied to non-Cartesian data. The computational burden is substantial due to the necessity of performing aliasing unfolding of pixels to differentiate between numerous alternatives, not solely in a single direction. The primary disadvantages are analogous to those of the conventional approach, in that the quality of the image is contingent on the precision of the sensitivity local mapping estimation.

The implementation of GRAPPA in non-Cartesian acquisitions is based on the observation that the interpolation kernel used to estimate missing k-space points may vary across different segments of the acquisition, due to geometric discrepancies and the varying relationships between local regions of k-space. Prior to the scan, a full-sampled calibration dataset is therefore required to determine the appropriate kernels for subsequent reconstruction.

A trade-off between time efficiency and reconstruction accuracy arises when considering whether a single kernel can be applied across broader segments of k-space that share similar geometric characteristics.

GRAPPA has been successfully extended to a variety of non-Cartesian trajectories, including radial, variable-density and uniform-density spiral, rosette, and PROPELLER (blade) configurations. However, these implementations are inherently more computationally demanding than the conventional Cartesian GRAPPA, as they require additional processing steps to account for non-uniform sampling and trajectory-dependent geometry.

SPIRiT methods applied to parallel imaging with non-Cartesian data are characterised by a relatively straightforward reconstruction strategy, in which the undersampled non-Cartesian k-space is first transformed into a fully sampled Cartesian grid through convolution-based interpolation. The reconstruction is then performed within this Cartesian domain, following the same iterative framework described previously.

Maintaining data consistency throughout the iterative process ensures that the reconstructed k-space lines remain aligned with the originally acquired samples, thereby preserving the fidelity of the measured data and minimising reconstruction artefacts.

Parallel imaging techniques are now routinely employed in clinical MRI practice and are integrated by scanner vendors directly within the image post-processing pipeline. As a result, reconstructed images are typically available immediately after acquisition, facilitating an efficient clinical workflow.

Despite their widespread use, several limitations persist. The degree of acceleration is fixed by the vendor and cannot be freely adjusted by the user, and the implementation

of non-Cartesian trajectories remains computationally demanding. Moreover, in routine MRI protocols, Cartesian sampling represents only a portion of the acquisition schemes, as many clinical sequences employ alternative trajectory types such as spiral, radial, or PROPELLER sampling.

Compressed sensing

Compressed Sensing (CS) represents a major advance in MRI reconstruction, introduced by the work of (Lustig et al. 2007). The approach exploits the intrinsic sparsity of MR images in specific transform domains, such as wavelet or total variation, allowing accurate image reconstruction from highly undersampled k-space data. In contrast to parallel imaging, which uses coil sensitivity information to interpolate missing samples, CS relies on the mathematical property that most image coefficients are negligible once transformed into an appropriate sparse representation.

The reconstruction is formulated as a nonlinear optimisation problem, in which the objective is to minimise a sparsity-promoting term while ensuring consistency with the acquired k-space data. The problem can be expressed as:

$$\min_x \|\Psi x\|_1 \text{ subject to } \|F_u x - y\|_2 < \varepsilon \quad (2.1)$$

where x is the image to be reconstructed, y represents the undersampled k-space data, F_u is the undersampled Fourier operator, Ψ is a sparsifying transform, and ε controls data fidelity. The use of the L1 norm encourages sparsity in the transformed domain, while iterative reconstruction algorithms—such as Iterative Shrinkage Thresholding (IST), FISTA, or TwIST—are used to achieve convergence.

This methodology allows the reconstruction of diagnostic-quality MR images with acceleration factors typically ranging from four- to eight-fold compared to conventional acquisitions. Random undersampling introduces incoherent aliasing, which is effectively removed during the iterative reconstruction, leading to a marked reduction in structured artefacts. Furthermore, CS can be combined with parallel imaging methods, resulting in hybrid approaches such as CS-SENSE or CS-GRAPPA, which further enhance image quality and reconstruction speed.

However, the benefits of CS come at the cost of increased computational complexity and the need for parameter tuning, as the balance between sparsity enforcement and data fidelity must be empirically optimised. In addition, the choice of the sparsifying transform and regularisation weights significantly affects the final image quality. Despite these limitations, Compressed Sensing has become a standard component of modern MRI reconstruction pipelines and serves as a conceptual bridge toward the development of AI-based reconstruction techniques such as SwiftMR, which aim to

achieve similar acceleration and denoising capabilities through data-driven learning rather than explicit mathematical sparsity constraints.

2.3. Real and potential Use in Clinical Setting

The present study aims to characterise the real baseline utilisation of MRI examinations within the context of current clinical practice at the Ospedale di Erba. To this end, institutional activity data collected between January and October were analysed.

During this period, a total of 4,085 MRI examinations were performed using a single 1.5 T MRI scanner operating approximately 240 days per year for 12 hours per day. Assuming a standard working schedule, this corresponds to an average throughput of approximately 20 MRI examinations per working day. Breast MRI examinations are performed in a dedicated weekly session lasting approximately half a morning considering approximately 10 examinations per week. While this dedicated session introduces a limited recurrent constraint on scheduling, it does not substantially alter the overall daily throughput when averaged across the full working calendar.

As a result, days with higher productivity (e.g. 24–25 examinations) compensate for days with reduced availability due to the breast MRI session, yielding an average of about 20 examinations per day. However, breast MRI examinations are not considered eligible for AI-based reconstruction using SwiftMR, as specified in the regulatory framework and GDPR-related conformity documentation, where the software is not declared for use in breast imaging protocols.

Consequently, although breast MRI contributes to the operational workload of the scanner, the baseline analysis for the evaluation of SwiftMR focuses exclusively on the subset of examinations that are technically and regulatorily impactable by the device. Table 2.1 reports the baseline distribution of MRI examinations eligible for SwiftMR implementation. The volumes are derived from institutional activity data collected between January and October 2025 and include musculoskeletal, abdominal–pelvic, spine and neuro MRI examinations. Average examination durations were estimated based on routine clinical MRI protocols currently adopted at the institution and are consistent with values reported in the literature for standard 1.5 T MRI examinations .

Table 2.1: Baseline volumes of MRI for the current use of technologies

Examination categories	Average duration	Examination per years/estimation
Musculoskeletal	25 min	1090

(MSK)		
Abdomen/Pelvis	35 min	660
Brain	30 min	1620
Spine	35 min	713
Thorax	35 min	2
Total		4085

The distribution of examinations across anatomical regions is as follows: approximately 26.7 % musculoskeletal (MSK), 16.2 % abdominal/pelvic, 39.7 % brain, 17.5 % Spine. Breast MRI examinations are excluded from this percentage distribution for two reasons. Firstly, a dedicated session is scheduled once per week specifically for breast imaging. Secondly, the SwiftMR software is not designed for breast MRI enhancement, and no substantial improvements in reconstruction time or image quality are currently observed in this application.

Musculoskeletal and abdominal examinations are expected to benefit substantially from reduced scan time and improved diagnostic image quality, particularly in terms of signal-to-noise ratio (SNR). In addition, neuroimaging examinations—including both brain and spine MRI—are also likely to benefit from AI-based reconstruction, especially in sequences requiring high spatial resolution or extended acquisition times, where noise reduction and motion robustness are critical.

To refine the analysis and provide a more detailed understanding of clinical activity, the broad anatomical categories were further subdivided into clinically relevant subtypes. For each category, subgroups were defined according to standard anatomical districts routinely adopted in MRI clinical practice.

Table 2.2: Volumes of MRI examination for the subtype MSK

Subtype of MSK	Volume (%)	Examination per years/estimation
Knee	58.2 %	634
Shoulder	18.9 %	206
Ankle/foot	13.6 %	148
Wrist/hand	3.2 %	35

Hip (coxofemoral/femur)	1.6 %	17
Arm/ upper lower limb	4.6 %	50
Total MSK		1090

Table 2.3: Volumes of MRI examination for the subtype abdominal/pelvis

Subtype abdominal/ pelvis	Volume (%)	Examination per years/estimation
Upper abdomen	47.4%	313
Lower abdomen	34.4%	227
Pelvis	18.2%	120
Total abdominal/pelvis		660

Table 2.4: Volumes of MRI examination for the subtype brain MRI

Subtype brain MRI	Volume (%)	Examination per years/estimation
Brain (encephalon)	77 %	1248
Brain + Angio MRI	13.2 %	214
Pituitary, Sellar region	5.3 %	86
Other brain (trunk, facial mass)	4.5%	72
Total brain MRI		1620

Table 2.5: Volumes of MRI examination for the subtype spine MRI

Subtype spine MRI	Volume (%)	Examination per years/estimation
-------------------	------------	-------------------------------------

Lumbar spine	55%	392
Cervical spine	26.8 %	191
Dorsal spine	13.5 %	96
Complete spine/ other	4.7 %	34
Total spine MRI		713

This section outlines the baseline utilisation of the MRI scanner installed within the hospital, accompanied by a detailed overview of the corresponding anatomical districts. Institutional activity data collected between January and October were used to estimate the average daily throughput, which was approximately 20 MRI examinations per working day. Assuming a standard operational schedule of 240 working days per year, this corresponds to an annualised baseline capacity of approximately 4,800 MRI examinations. Three prospective scenarios were defined to evaluate the potential increase in MRI throughput following the implementation of accelerated reconstruction technologies such as SwiftMR. Starting from the annualised baseline, a moderate improvement (+20%) would raise the capacity to approximately 5,760 examinations per year (24 examinations/day). A more substantial enhancement (+30%) would yield around 6,240 examinations per year (26 examinations/day), while a high-performance scenario (+40%) could achieve approximately 6,720 examinations per year (28 examinations/day). These projections illustrate the potential operational gains achievable through workflow optimisation and AI-assisted reconstruction.

Table 2.6: Potential use in clinical setting for the adoption of SwiftMR

Scenario	% Increase	Exams/year	Exams/day (240 days)	Description
Baseline (current use)	—	4800	20	Current clinical throughput, based on routine protocols and average scan durations.
Moderate efficiency	+20 %	5760	24	Achievable through moderate

				acceleration (e.g., reduced scan time by 10–15 %) and fewer repeat scans.
Enhanced efficiency	+30 %	6240	26	Reflects more significant workflow optimisation and wider adoption of AI-based reconstruction across sequences.
High-performance scenario	+40 %	6720	28	Represents an ambitious yet plausible case combining advanced reconstruction (SwiftMR) and optimal scheduling efficiency.

In conclusion, the evaluation of the three utilisation scenarios is fully aligned with the Health Technology Assessment (HTA) framework, as it represents a structured sensitivity analysis aimed at exploring the range of potential operational outcomes associated with the introduction of accelerated reconstruction technologies. The actual impact on hospital performance will depend on multiple contextual factors, including resource allocation, scheduling strategies, and the degree of adaptation of clinical workflows.

In the subsequent sections, these scenarios will be assessed from both an economic and a clinical effectiveness perspective, in order to quantify their potential benefits in terms of cost optimisation, workflow efficiency, and patient-related outcomes.

While established reconstruction techniques, namely Parallel Imaging and Compressed Sensing, remain integral components of vendor-specific MRI systems, further reductions in acquisition time are expected through the adoption of AI-based

image enhancement solutions such as SwiftMR. By enabling the reconstruction of high-quality images from reduced k-space sampling, these technologies represent a complementary advancement to current standards of care rather than a replacement.

3 Description and Technical Features

The following domain of the HTA explores the technological advances related to the adoption of new medical devices—in this case, the DL-based software SwiftMR. Through a technological features analysis, this section examines the advantages and disadvantages of the system compared with current solutions used in clinical routine, following the framework described in (Lampe et al. 2009) where the domains of the EUnetHTA core model are defined. The section is structured around the general characteristics and outputs of the AI software, followed by the description of the overall architecture and its integration within the hospital IT network and the MRI scanner. The functioning of the DL model, the data collection strategy, and the evaluation of data quality are also addressed within this domain.

The technical prerequisites for implementing the hospital IT network and transmitting images to the cloud platform administered by AIRS Medical are fundamental for the clinical perspective. These aspects directly influence the diagnostic quality delivered to the radiology department, as well as the overall capability to accelerate scanning and enhance image reconstruction across multiple anatomical regions and MRI pulse sequences. The architecture of the DL model and the characteristics of the training dataset influence the model's performance in terms of image-quality enhancement and the multidimensional reconstruction approach developed by AIRS Medical. These elements represent the technical foundation of the SwiftMR commercial product.

The technical description presented in this chapter is based on manufacturer-provided technical documentation and white paper material made available upon request. These sources are used exclusively to describe the system architecture, workflow integration, and operational modalities of the SwiftMR software, without implying independent clinical validation.

The DICOM-based approach adopted by the DL algorithm is a strategic asset for the company, as it guarantees vendor agnosticism, independence from anatomical region, and independence from contrast type and pulse sequence. This technical feature distinguishes SwiftMR from other solutions that use k-space-based DL reconstruction approaches. The rationale for this choice will be discussed in detail after the fundamentals of the DL architecture have been presented.

The software can be used in two fundamental modalities thanks to its DICOM-based approach:

- Fast-scan image quality enhancement
- Image quality enhancement in standard-of-care acquisitions.

In the first scenario, clinicians perform a fast scan of the anatomical region of interest. This time reduction is achieved by sacrificing lines of k-space, for example by increasing Δk , reducing the kmax high-frequency content in the encoding directions, applying elliptical sampling, or using partial Fourier sampling. The standard reconstruction pipeline of the MRI vendor is then applied, together with the conventional acceleration techniques such as parallel imaging or compressed sensing. However, when the degree of acceleration exceeds the capabilities of these standard methods, the resulting images may be degraded. At this point, the DICOM-based DL algorithm intervenes with its multidimensional enhancement capabilities learned during training. The software denoises the image, enhances super-resolution, restores edge sharpness, and compensates for blurring caused by low-resolution acquisition. In other words, the diagnostic quality is brought to a level that is equal to or superior to the standard-of-care acquisition. An example of a brain T1-weighted image acquired on a 3 T MRI scanner is shown in the figure below. It is evident that an acceleration reduction of approximately 54% was achieved, while image quality improved across multiple dimensions, including noise level, super-resolution in all encoding directions, and edge sharpness.

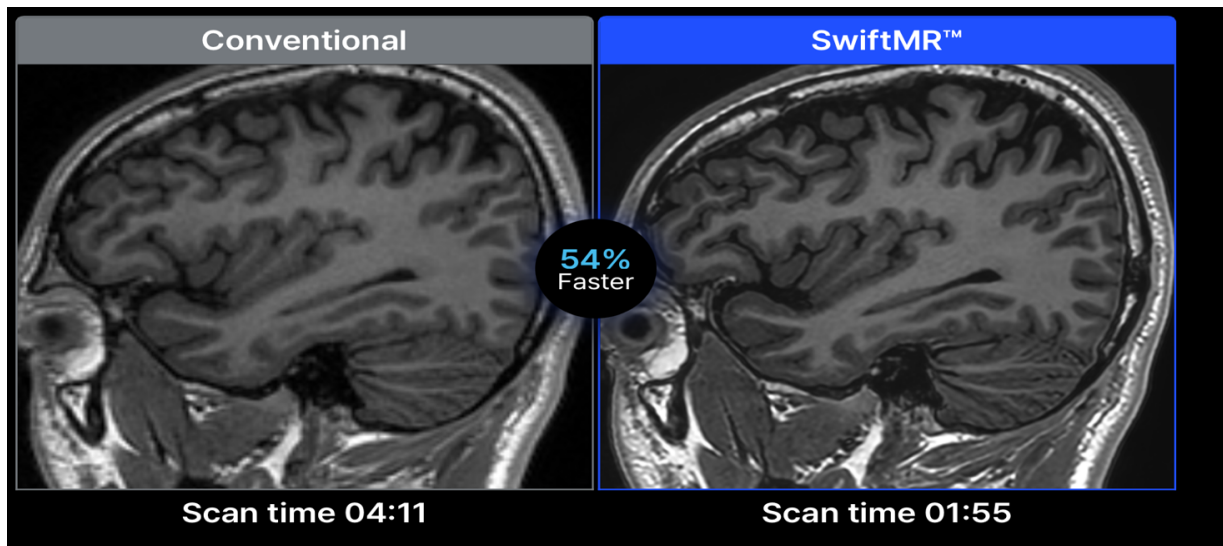


Figure 3.1: Brain T1WI on a 3T scanner and reduction in 54 % scanning time. Image reproduced (AIRS Medical 2024b).

The second possibility is to maintain the standard acquisition protocol and still enhance image quality in the specific domains required by the clinician. An example is shown for a liver T2-weighted image, where the acquisition time remained fast while the resulting image quality was clearly superior.

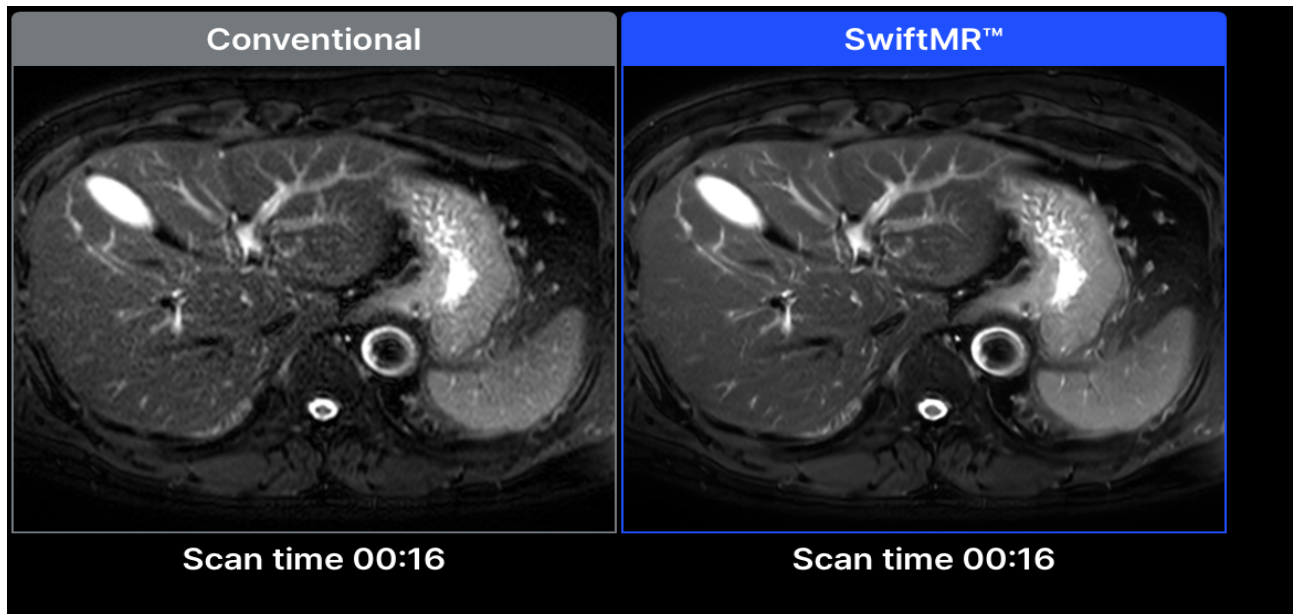


Figure 3.2: Liver T2WI pulse sequence from a 3 T MR scanner. Image reproduced from: (AIRS Medical 2024b).

3.1. System overview and Architecture

According to the technical white paper and peer reviewed publications (Seungwook and Yang 2024) released by AIRS Medical, SwiftMR is a deep learning (DL)-based software developed for MRI image reconstruction. The solution is designed to substantially reduce scan time while maintaining the diagnostic quality required for the clinical standard of care. Unlike many other AI-based reconstruction systems, According to the vendor documentation SwiftMR operates directly on reconstructed DICOM images in the image domain rather than on raw data or k-space.

The reduction of scan time in clinical MRI is a critical objective, as it enables the acquisition of multiparametric information and enhances the efficiency of diagnostic workflows. However, the temporal optimisation of scanning must always balance signal-to-noise ratio (SNR), spatial resolution, and artefact management with diagnostic accuracy and patient comfort. Striking this balance remains essential for radiology departments seeking to align technical performance with clinical needs (Saloner, Liu, and Haraldsson 2015).

As a vendor-provided document, the white paper (Seungwook and Yang 2024) is used in this section exclusively to describe the system architecture, software workflow, and integration within clinical MRI environments, without implying independent clinical validation.

The reconstruction process in SwiftMR is fully automated and occurs in the background. Once the network connection between the MRI scanner, the Picture

Archiving and Communication System (PACS), and the AIRS Medical cloud platform is established, images follow a defined transfer pathway and are returned to the hospital system for clinical use.

DICOM compliance is a key design feature of SwiftMR. The system processes standard DICOM image series within hospital IT networks and ensures patient anonymisation during transmission to and from the cloud. This design guarantees seamless integration with existing radiological workflows and preserves data security in compliance with international standards.

The SwiftMR solution is a vendor-agnostic software tool, meaning that it operates independently of the specific hardware manufacturer of the MRI scanner, the anatomical region under examination, or the pulse sequence employed. This flexibility allows seamless integration within heterogeneous imaging environments and across multiple MRI platforms.

The principal MRI sequences referenced in this study include T1, T2, T2*, PD, FLAIR, DWI, and MRA. These represent the standard acquisition protocols utilised for image reconstruction and quality enhancement.

The system's implementation follows a protocol that encompasses all major anatomical districts. The neuro category includes both brain and spine imaging; the musculoskeletal (MSK) category covers the shoulder, wrist, hip, knee, and ankle; and the body category comprises the breast, cardiac, abdominal, and genitourinary regions. This structure reflects the eight standard clinical protocols supported by SwiftMR for deep learning-based reconstruction and denoising.

The eight anatomical protocols implemented by SwiftMR are summarised in the Table 3.1.

Table 3.1: Eight anatomical protocols for the SwiftMR AI based software

N°	Anatomical protocols	Anatomical district included
1	Brain	Standard neuro, FLAIR, DWI
2	Spine	Cervical, thoracic, lumbar
3	MSK Upper Limb	Shoulder, wrist
4	MSK Lower Limb	Hip, ankle, knee

5	Abdomen	Liver, kidney, pancreas
6	Pelvis- Genitourinary	Prostate, uterus, bladder
7	Cardiac	Cine, perfusion
8	Breast	DCE breast MRI

3.2. Deep Learning algorithm and learning process

The subsequent portion outlines the technical details of the deep learning (DL) framework utilized to improve image quality in fast-scan MRI scans compared to the current standard of care. The deep learning framework described in this section illustrates the architectural and methodological principles underlying the main features of the SwiftMR software developed by AIRS Medical. The description is based on a vendor-authored research publication (Jeong et al. 2024) which is used to explain the rationale of the approach, while acknowledging that the commercial implementation may incorporate additional engineering optimizations and constraints. In the realm of DL-based MRI reconstruction as expressed in (Lebel and Healthcare 2020), two primary methods are prevalent in both commercial and research environments:

- K-space-based reconstruction: This approach involves feeding raw k-space data into the model, which then predicts missing or corrupted k-space samples before the traditional vendor reconstruction process;
- Image-based reconstruction: Here, both the input and output are DICOM images generated by vendor-reconstructed images. This method is vendor-agnostic and directly applicable in clinical workflows.

The focus in this study is on the latter approach (image-based reconstruction during inference), although the training datasets are derived from raw k-space. To create the training pairs, fully sampled raw k-space undergoes degradation through various methods, including noise addition, undersampling (both uniform and random), kmax reduction, elliptical sampling, and partial Fourier undersampling. Both the degraded and original k-space data are then processed through the conventional vendor reconstruction pipeline to produce pairs of degraded and fully sampled images for supervised learning.

The core of the network is a modified U-Net, with two significant enhancements:

- **Standardization Module:** This module preprocesses input DICOM images to normalize and standardize them across different scanner vendors. Operations such as axis transposition, removal of zero-padded regions, Lanczos interpolation for consistent in-plane dimensions (1024 for 2D and 768 for 3D sequences), and normalization to mitigate vendor-specific scaling are carried out;
- **Contextual Data Pathway:** An additional input pathway injects contextual data derived from scan parameters into the network. These data encompass information such as noise reduction level, kmax scaling for frequency and phase, partial Fourier factors, and other acquisition-specific parameters. This feature enables the model to dynamically adjust its enhancement based on the acquisition conditions.

Two distinct models are trained: one for 2D acquisitions, processing a single input slice, and another for 3D acquisitions, handling a stack of seven adjacent slices to achieve super-resolution along the slice-encoding direction. In both scenarios, the network produces a single enhanced slice. This architecture enables SwiftMR to work directly on DICOM images during inference while maintaining compatibility with various scan parameters and vendors, owing to the comprehensive k-space-based degradation simulations conducted during training.

Training phase and collection of pairs samples

The dataset used for the training phase of the deep learning (DL) model was collected from MRI scanners produced by three different vendors. A total of 1.5 million fully sampled raw MRI datasets were acquired specifically for this purpose from volunteers. The dataset consisted of approximately 1 million 2D slices and 2 million 3D slices, spanning a broad range of pulse sequences, contrasts, acquisition parameters, and anatomical regions. This diversity ensures that the training process is as comprehensive and generalizable as possible, covering the full spectrum of clinical scenarios. Fully sampled raw k-space data served as the ground truth. To generate the paired inputs for supervised learning, each fully sampled dataset was synthetically degraded in the k-space domain across multiple independent dimensions. After degradation, both the degraded and the original k-spaces were reconstructed using the standard vendor-specific reconstruction pipeline, producing the degraded/target image pairs used to train the model. The degradations covered five key k-space domains, enabling the model to learn multidimensional enhancement tasks:

- kmax scaling in all encoding directions (frequency, phase, and slice), modulating the amount of high-frequency content and enabling super-resolution;

- Δk (delta-k) modifications, simulating changes in sampling density to prevent aliasing and enforce the correct field of view (FOV);
- Partial Fourier factors, introducing asymmetric k-space acquisition and the associated directional blurring;
- Elliptical sampling factors, altering peripheral k-space coverage to mimic elliptical scanning strategies;
- Noise reduction factors, simulating different SNR levels and enabling the model to learn tunable denoising.

In the Figure 3.3, a representative training pair is shown. As SwiftMR is an image-based reconstruction model at inference, both the input and target for training must be reconstructed into DICOM images using the vendor's standard reconstruction pipeline. On the left, the fully sampled k-space is displayed, representing the ground-truth acquisition. On the right, the corresponding degraded k-space is shown, after the application of several synthetic degradation processes, including noise addition, k-max undersampling, elliptical undersampling, and partial Fourier undersampling. These k-space manipulations are subsequently passed through the conventional vendor reconstruction workflow to produce the degraded image used for training.

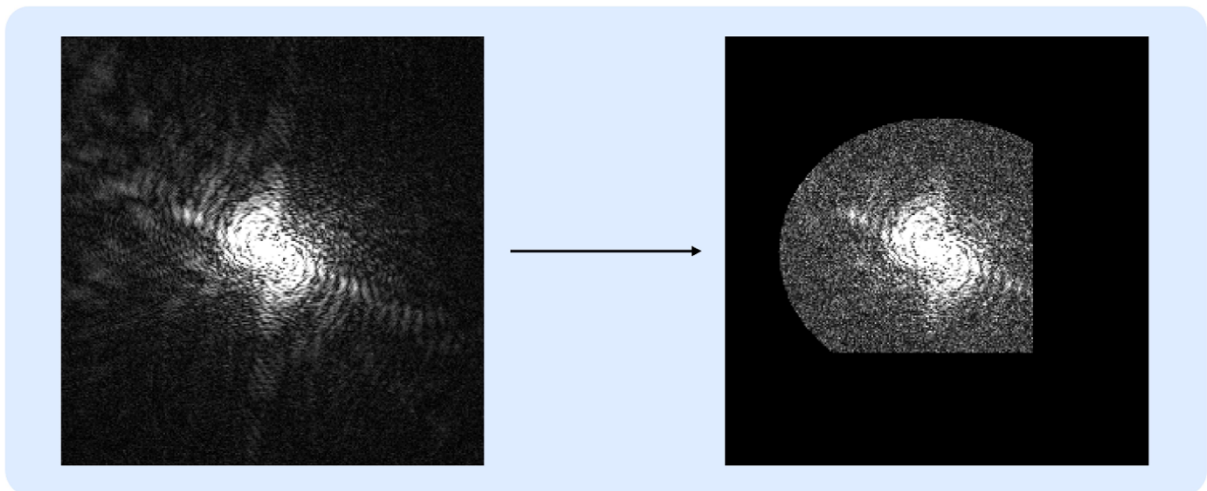


Figure 3.3: Pair of training samples in the k-space domain. Image reproduces from: (Jeong et al. 2024a).

These dimensions extend far beyond traditional DL-based MRI reconstruction methods, which typically simulate only uniform or random undersampling of k-space (Heckel et al. 2024). In contrast, the SwiftMR training pipeline systematically explores all combinations of the above degradations, ensuring that the model is exposed to the full range of multidimensional acquisition scenarios encountered in clinical practice. As mentioned earlier, the degradation is performed directly in the k-space domain. The degraded k-space and the original fully sampled k-space are both passed through the vendor's full reconstruction pipeline, typically including anti-ringing filters,

parallel imaging (PI) reconstruction, compressed sensing (CS), and coil/channel combination.

This ensures that the final DICOM images used as inputs and targets during training accurately reflect the variability introduced by real-world reconstruction workflows across different vendors. Through this strategy, the model learns to enhance image quality in a multidimensional manner, addressing super-resolution, blurring from partial Fourier, aliasing from Δk changes, and noise suppression. Particularly, the noise level factor becomes a crucial parameter in clinical use, as it is supplied through the contextual data array to finely control the level of denoising performed by the DL software.

Standardized module and contextual data

To cater to the unique characteristics and objectives of the deep learning model, the conventional U-Net architecture is enhanced with two key elements. The initial element, situated before the input is fed into the network, consists of a normalization module. This module ensures that the model can effectively extract the pertinent features of each image slice to enhance image quality.

The normalization module comprises four essential steps:

- Image transposition;
- Elimination of unnecessary zero padding by cropping;
- Lanczos in-plane interpolation;
- Pixel intensity normalization.

Image transposition is essential for aligning the phase-encoding direction uniformly from left to right. This alignment is critical for the model to enhance image quality and conduct super-resolution accurately across all encoding directions. The cropping of unnecessary zero padding addresses potential asymmetric fields of view resulting from some vendors applying zero padding as part of their reconstruction process. Since these areas do not contain real data, they are removed.

Lanczos in-plane interpolation standardizes the number of columns in the frequency-encoding direction, ensuring consistent pixel spacing. The 2D model utilizes 1024 columns, while the 3D model employs 768 pixels. Lastly, intensity normalization is crucial to provide the model with a consistent dynamic range, as different scanner vendors may apply varying linear intensity mappings in the DICOM format.

The second enhancement involves contextual data. This data is introduced at an intermediate stage through a fully connected layer with a specific activation function, resulting in the context-enhanced U-Net. The content of the contextual data directly influences the model's utilization in clinical settings, especially in modulating the noise-level factor. During training, this factor is determined mathematically, under the assumption that the model cannot fully learn independently added noise.

Additional contextual parameters encompass the scan parameters of the fully sampled data and the degradation parameters applied during training. These inputs serve as guidelines for the model, specifying the desired output relative to the standard of care that must be achieved regardless of acceleration factors or the initial image quality processed by the software.

In essence, the contextual data comprises:

- The noise-level factor;
- The scan parameters of the input image;
- Descriptors of the target fully sampled image, such as kmax, the super-resolution direction, and the partial Fourier factor.

The two components contributing to the CE U-Net architecture are depicted in the accompanying Figure 3.4.

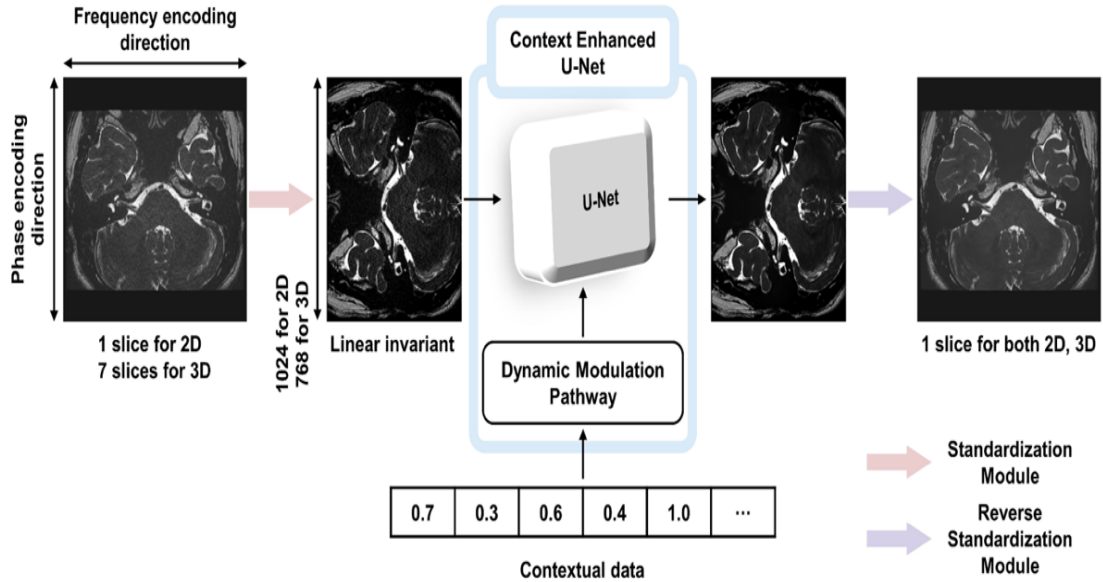


Figure 3.4: Scheme of the U-net model and the standardized module and the contextual data. Image reproduces from: (Jeong et al. 2024).

Multi-dimensional image enhancement and noise level factor

The training of the model, using pairs composed of fully sampled data and their corresponding low-sample versions generated through specific degradation strategies, enables the network to learn through a supervised approach how to enhance image quality within a multidimensional framework. The exploitation of these potential sources of improvement is managed by the clinical user, who can adjust the contextual data according to the diagnostic goal. To verify that the model has correctly learned the different tasks and sources of multidimensional enhancement, several tests were performed after the training phase. The first acquisition used to

validate the multidimensional enhancement was a brain T2-weighted image obtained with a 3D constructive interference in steady state (CISS) sequence on a 1.5 T scanner from a specific vendor. The original image was then reconstructed by the DL model five times, each time with different contextual data values, producing five images with sequential increases in image quality, each corresponding to a different dimension of improvement:

1. Noise reduction only;
2. Adding frequency kmax dimension;
3. Adding phase partial Fourier dimension;
4. Adding phase kmax dimension;
5. Adding slice kmax dimension.

The same original image was passed through the model multiple times, each time with different contextual data values, in order to target specific sequential dimensions of image-quality enhancement.

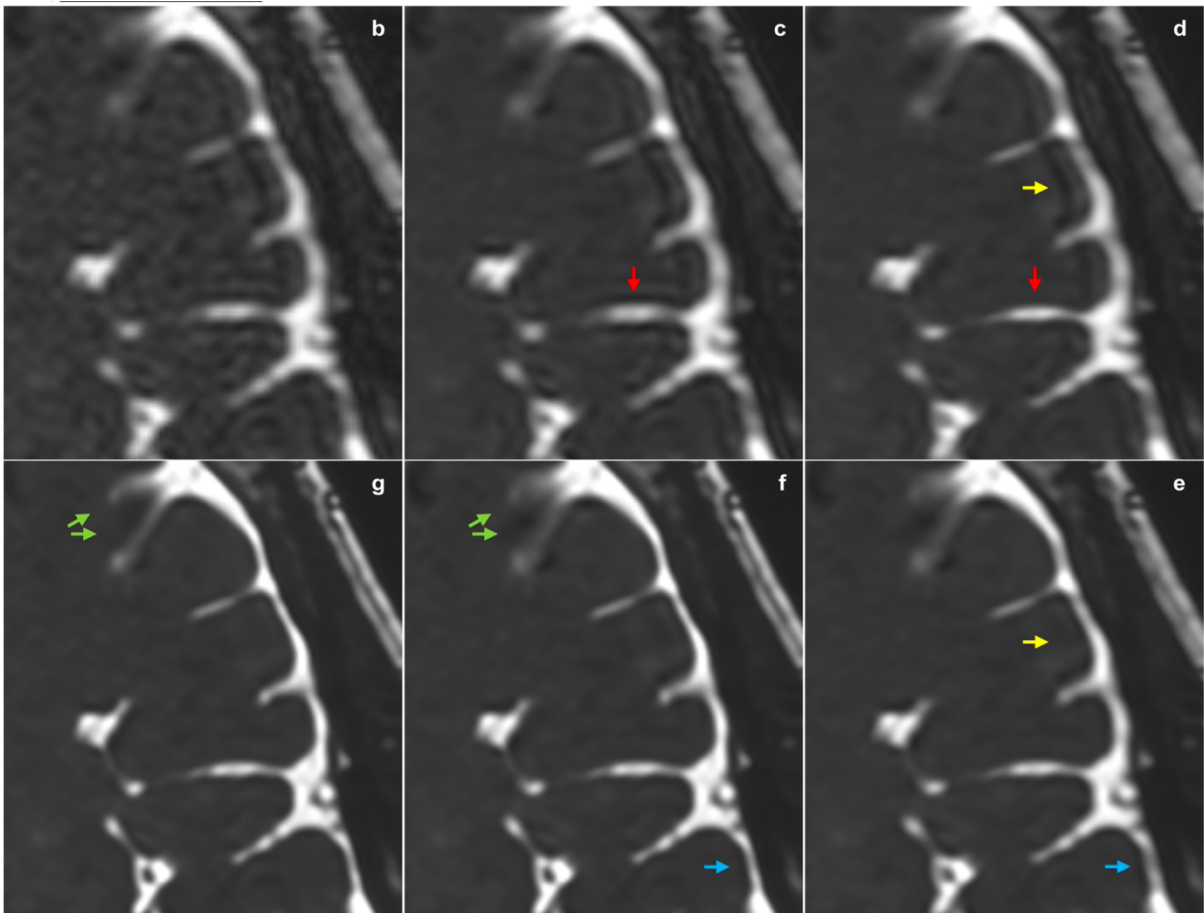


Figure 3.5: the original image and the five DL reconstructed with enhancement quality.
Image reproduces from: (Jeong et al. 2024)

The Figure 3.5 above shows, in panel b), the original image zoomed into a specific ROI, where it is possible to appreciate the improvement along the targeted enhancement dimension selected through the contextual data. Panel c) shows the same image reconstructed with a noise-level factor equal to 3, and it is evident from the ROI that the standard noise level is reduced proportionally to the applied factor.

The following images underwent the same noise-level correction together with the additional enhancement dimensions introduced at each reconstruction step.

In panel d), the frequency $k_{\max}$ dimension is added, and the perceived resolution along the direction indicated by the red arrows increases, as demonstrated by sharper edges and an increased amount of high-frequency information.

Panel e) includes the phase partial Fourier dimension, and the blurring induced by asymmetric k -space sampling is corrected.

Panel f) incorporates the phase $k_{\max}$ dimension, resulting in super-resolution in the horizontal direction, as indicated by the blue arrows.

Finally, panel g) includes the slice $k_{\max}$ dimension, where the information embedded in multiple adjacent slices is utilized, consistent with the 3D model processing seven slices centered on the target one.

These images are analyzed qualitatively but clearly demonstrate that the DL model behaves coherently with the training phase, and that the use of contextual data allows clinicians to obtain the desired diagnostic quality under different scanning conditions, saving time and improving patient comfort without sacrificing image quality. The next section is dedicated to the quantitative analysis of noise reduction in the DL model. Within the contextual data array, the noise-level factor is crucial for controlling the denoising strength. Increasing the noise-level factor results in a proportional reduction of noise between the original and the DL-reconstructed image. To demonstrate that the model has learned this task effectively, two quantities are considered:

- Mean signal intensity
- Noise standard deviation

First of all, a brain T1-weighted turbo field echo (TFE) pulse sequence was acquired on a 1.5 T MR scanner, and five images were reconstructed using different values of the noise-level factor, starting from 1 and increasing by 0.5 up to 3. The region of the pons was selected as the ROI for computing the quantitative metrics for each reconstructed image. This region was chosen because, as demonstrated by (Uetani et al. 2021) it exhibits homogeneous intensity and therefore provides a reliable area for noise estimation. The noise-level factor was computed as the ratio between the noise level of the standard (original) image and that of the DL-reconstructed image. The quantitative results are reported in the Figure 3.6.

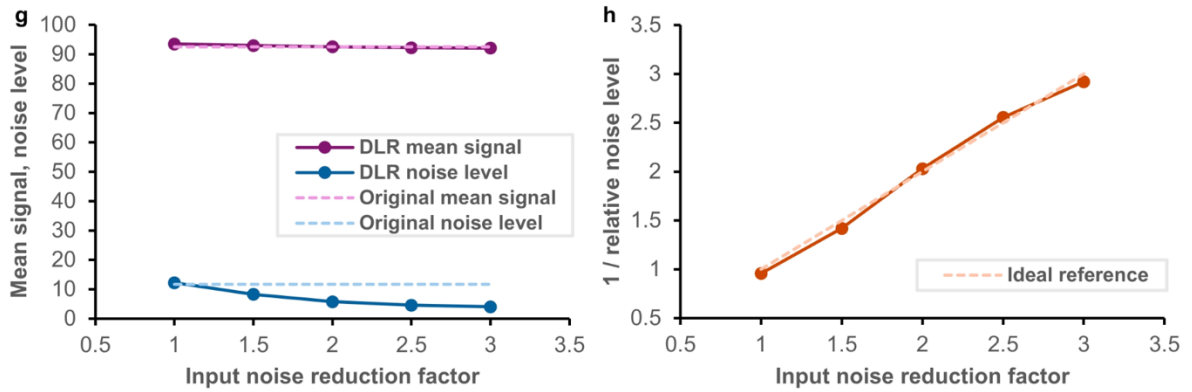


Figure 3.6: Means signal and noise standard deviation in the ROI at different noise level factor. Image reproduces from: (Jeong et al. 2024).

On the left, the mean signal remains constant as the noise-level factor increases, indicating that the model behaves correctly with respect to the denoising task. The noise level, shown by the blue line, decreases as expected.

On the right graph, the empirical computation of the noise-level factor—defined as the ratio between the standard deviation of noise in the original and in the reconstructed images—closely follows the theoretical values specified in the contextual data.

This is an important result, demonstrating that the model has effectively learned the denoising task and performs reliably in the clinical environment. Super-resolution enhancement in the encoding directions. The enhancement of spatial resolution in the three encoding directions—phase, frequency, and slice—is made possible by the task learned by the model when the kmax dimension is increased in the corresponding direction. This result was qualitatively assessed in the multidimensional enhancement section, where a visible comparison between the original and the DL-reconstructed images clearly showed the differences.

However, to confirm the model's super-resolution capability, a quantitative evaluation of edge sharpness is required. For this purpose, a brain MR angiography time-of-flight (TOF) pulse sequence acquired on a 1.5 T scanner was used, and a maximum intensity projection (MIP) was generated from both the original and the DL-reconstructed images for edge-sharpness computation. The DL reconstruction was performed using different tunable noise-level factors: 1, 1.5, 2, 2.5, and 3. The super-resolution enhancement in the frequency-encoding direction was assessed in the A2 segment of the anterior cerebral artery on a sagittal MIP image. The P1 segment of the posterior cerebral artery and the S1 segment of the superior cerebellar artery were chosen to evaluate the super-resolution in the phase and slice encoding directions, respectively, using coronal MIP images. The images acquired with a noise-level factor of 2.5 are shown in the Figure 3.7.

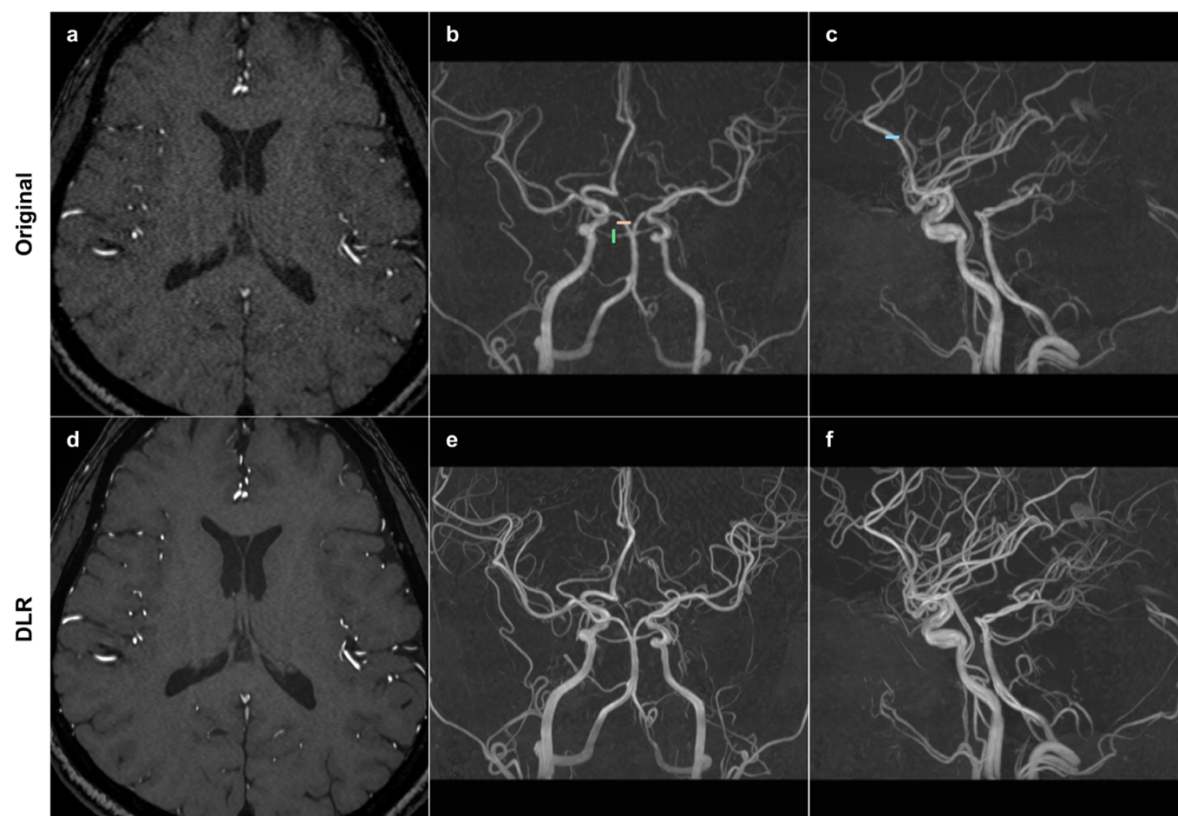


Figure 3.7: Brain MR angiography TOF, MIP images with coronal and sagittal acquisition. Image reproduced from (Jeong et al. 2024).

The first row of the figure above shows, respectively, the original image, the coronal MIP, and the sagittal MIP. The second row presents the same views, but reconstructed using the DL model; the MIP images were generated after the model processed the original acquired data. The method implemented for computing edge sharpness is the one described by (Lebel and Healthcare 2020) and is based on calculating the maximum derivative of pixel intensity along a line perpendicular to the arterial edge. This procedure was applied in the three directions of interest in the arteries mentioned above, for both the original and the DL-reconstructed images, and the corresponding ratio between the two sharpness values was computed. The results are shown in the Figure 3.8, and it is evident that, regardless of the noise-reduction factor used, the super-resolution capability of the model is enhanced, with an average improvement of approximately 2.5 across all encoding directions in favor of the DL-reconstructed images.

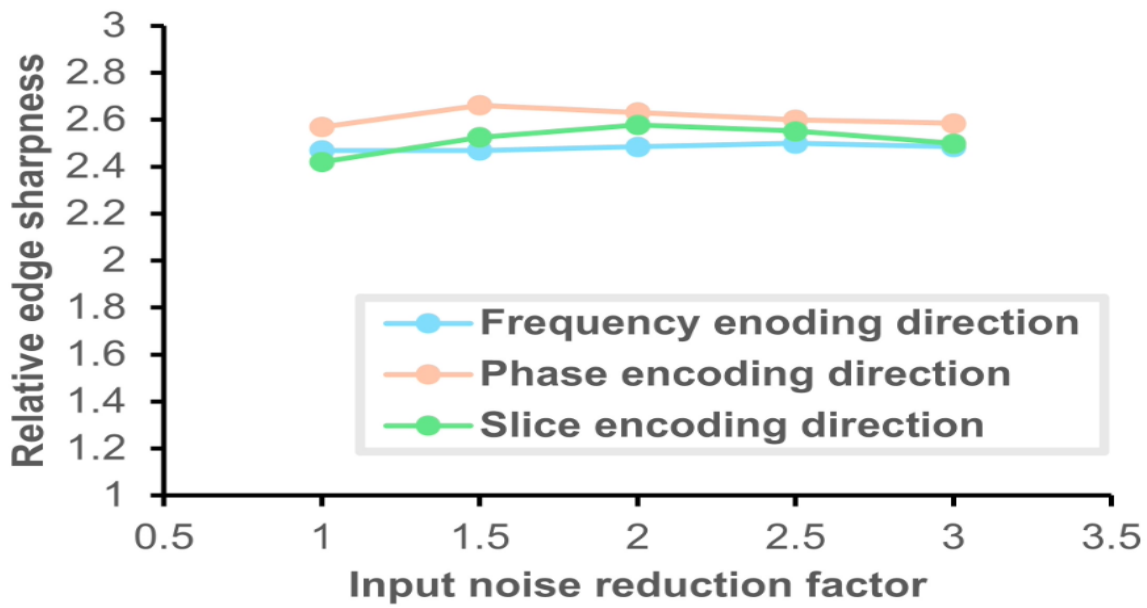


Figure 3.8: Relative edges sharpness between DL reconstructed image and the original one.
Image reproduced from (Jeong et al. 2024).

In conclusion, the training phase and the full range of simulated k-space degradations used for generating the training pairs provide the multidimensional image-quality enhancement that was defined during the learning process. This has been demonstrated both qualitatively and quantitatively in the previous sections. In the clinical environment, scan times can be reduced, and the degree of acceleration depends on the diagnostic accuracy required by the clinician. The performance of the model will require further evaluation in real clinical workflows. The DICOM-based DL reconstruction implemented by the model ensures vendor independence and guarantees consistent performance across a wide range of scanning parameters, including voxel size, field of view, use of compressed sensing, slice thickness, elliptical sampling factors, partial Fourier factors, and acquisition bandwidth.

3.3. Technical implementation

This section delineates the technical implementation of the SwiftMR product within the IT networks of the hospital, with particular reference to the MRI hardware scanner, the PACS system of the radiology department, and the cloud platform service provided and maintained by AIRS Medical. The description is based on the manufacturer-provided technical implementation documentations and white papers (AIRS Medical 2024a), (AIRS Medical 2025) made available upon request, used exclusively to describe system deployment, data flow, cybersecurity measures, and GDPR compliance.

The subsequent sections are the primary focus of this study:

1. the integration between the software and the IT networks of the hospital;

2. the elaboration phases;
3. transmission of the images and protection of data.

SwiftMR has been designed to integrate seamlessly with existing MRI systems, regardless of vendor or configuration. It takes DICOM images acquired under various scanning conditions as input and produces output images in the same format, with substantially reduced noise and enhanced sharpness.

The software can be applied in different acquisition contexts. In some cases, it enables scanning with a limited k-space sampling compared with the standard of care, thereby reducing scan time. In other scenarios, it enhances images of the same anatomical regions to improve diagnostic quality when acquisition time is constrained by clinical workflow requirements.

The first stage of the installation process consists of configuring a secure DICOM communication channel between the MRI console, the Picture Archiving and Communication System (PACS), and the SwiftMR cloud platform. Once this connection is established, the reconstruction and image enhancement are executed automatically in the background within the SwiftMR cloud environment.

Users can monitor the processing status through a web-based dashboard that displays the workflow state, queue progression, and processing logs in real time. The system is entirely software-based and requires no modification to existing MRI hardware or PACS infrastructure, ensuring seamless integration within the hospital network.

The SwiftMR workflow consists of four main phases. Phases 1, 2, and 4 are executed within the hospital IT network, while Phase 3 is performed on the SwiftMR cloud server. This configuration ensures a secure and efficient integration between local infrastructure and the cloud-based processing environment.

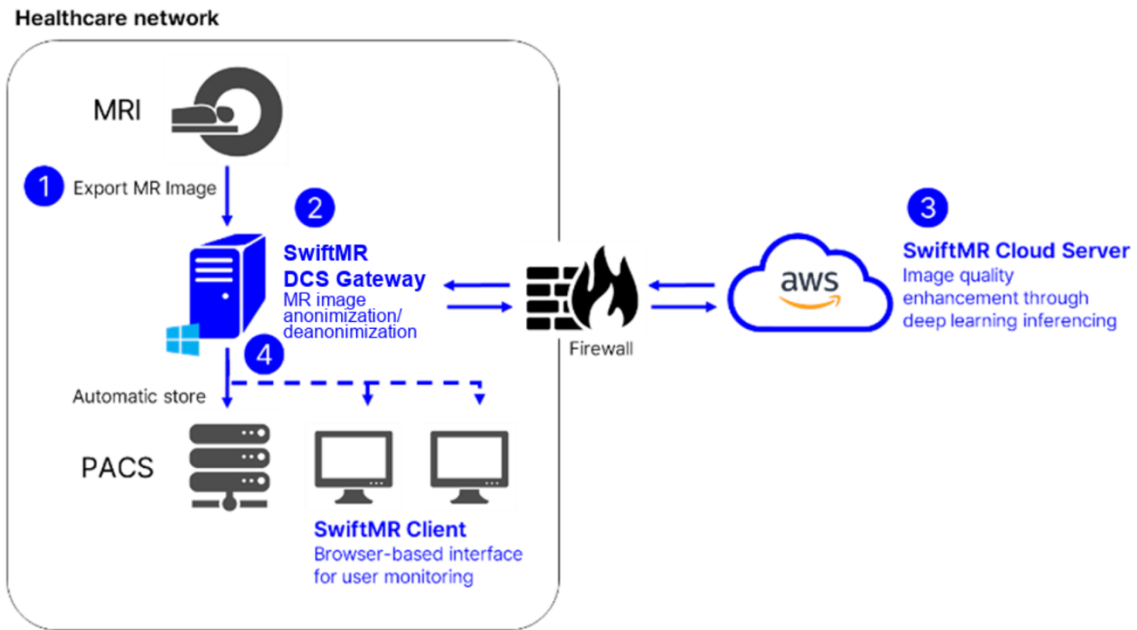


Figure 3.9: SwiftMR Cloud-Hosted Workflow (reproduced from (AIRS Medical 2024a))

First, it is essential to describe the main phases of the SwiftMR workflow in order to define the role of the DCS Gateway, which serves as an intermediary node between the hospital IT network and the AIRS Medical cloud services, as illustrated in Figure 3.9.

Phase 1 – Image export

After MRI acquisition, the DICOM images are exported to predefined destinations. In addition to the hospital PACS, the SwiftMR DCS Gateway is configured as a network node to receive the DICOM series for cloud-based reconstruction.

Phase 2 – Data anonymisation and transmission

The DCS Gateway includes integrated software that automatically removes all DICOM tags containing Protected Health Information (PHI), retaining only pixel data and minimal metadata required for reconstruction. Each anonymised study is assigned a unique Process ID. All data are transmitted to the SwiftMR cloud via end-to-end HTTPS/TLS encryption.

Phase 3 – AI-based reconstruction

Image enhancement and reconstruction are performed on the AWS cloud platform, which hosts the SwiftMR deep learning architecture. These servers are located in GDPR-compliant data centres in Singapore, US-East, Dubai, and Frankfurt. The reconstruction process typically takes less than two minutes for a standard study (fewer than 200 images, 2D acquisition), although processing time may increase for 3D

volumetric datasets. All temporary data are automatically and irreversibly deleted within 24 hours of processing completion.

Phase 4 – Reassociation and PACS archiving

The enhanced DICOM images are returned to the DCS Gateway, where the PHI tags are re-associated with the corresponding study. The completed images are then transferred to the hospital PACS for archiving, ensuring seamless reintegration into the existing radiological workflow.

In summary, SwiftMR ensures full compliance with the General Data Protection Regulation (GDPR), particularly regarding data integrity, encryption and decryption algorithms, and overall protection against unauthorised access or modification. The security framework and privacy protocols, as detailed in Section 1.2.2, also apply to the present technical implementation within the hospital IT network.

The system architecture enables efficient integration between the hospital's IT infrastructure and the AIRS Medical cloud environment. Interoperability is achieved through standard DICOM communication between the MRI scanner, the PACS, and the SwiftMR platform, ensuring adherence to international data protection and cybersecurity standards.

The automated workflow, covering export, anonymisation, AI-based reconstruction, and reintegration, operates in a fully autonomous manner while maintaining high performance and data integrity. As a software-only solution, SwiftMR functions independently of the MRI hardware, providing an innovative technological response to the challenges addressed in this study.

The technical features described herein constitute the foundation for the subsequent evaluation of clinical effectiveness and validation of the system in real-world clinical settings.

4 Safety

According to the EUnetHTA HTA Core Model for Diagnostic Technologies (EUnetHTA 2008), the safety domain for diagnostic technologies aims to assess both qualitative and quantitative aspects of safety, considering the potential severity of harm associated with the use of the technology and its impact on the clinical and diagnostic workflow. In the context of AI-based diagnostic tools, such as SwiftMR, safety assessment extends beyond traditional device-related risks and includes potential effects on image reliability, diagnostic confidence, and workflow integration.

In order to address the requirements of the EUnetHTA Core Model for this specific domain, the present section is structured into two complementary parts. The first part is based on a secondary analysis of the scientific literature, with the objective of identifying potential safety-related issues associated with deep learning-based MRI reconstruction techniques. This includes evidence derived from clinical validation studies of SwiftMR as well as from studies addressing other deep learning reconstruction approaches, with specific attention to reported risks, limitations, and failure modes.

The second part is based on primary qualitative evidence collected within the Ospedale di Erba, aimed at exploring the perceived safety of the technology in a real-world clinical setting. This assessment is conducted through expert-based questionnaires administered locally and is intended to complement the available scientific evidence by providing contextual insights into operational and organisational aspects of safety, rather than to generate quantitative safety metrics.

4.1. Scientific and Clinical Evidence on Safety and Tolerability

The safety assessment of deep learning-based MRI reconstruction within the Health Technology Assessment (HTA) framework requires a structured evaluation of both algorithm-related risks and available clinical evidence. This subsection is therefore organised into two complementary components.

The first part focuses on technology-related safety considerations inherent to deep learning MRI reconstruction algorithms. Based on current scientific literature, the main safety domains associated with these approaches are defined and discussed from a theoretical perspective. These domains reflect intrinsic characteristics of data-driven

reconstruction methods and are not product-specific, but rather representative of the broader class of deep learning-based reconstruction systems. Following this general overview, a qualitative HTA-oriented risk assessment is performed for SwiftMR, assigning low, moderate and high risk levels to each domain based on the technical implementation, training strategy, and mitigation measures described in Chapter 3.

The second part shifts the perspective toward clinical safety evidence. In this context, five peer-reviewed studies evaluating SwiftMR in different anatomical and acquisition settings are examined specifically through a safety lens. Rather than reassessing overall diagnostic effectiveness, the analysis focuses on reported safety-related findings, including variations in specificity, false-positive and false-negative interpretations, artefact persistence, and changes in lesion conspicuity. This dual approach allows for the integration of theoretical algorithmic risk domains with real-world clinical observations.

4.1.1. Technology-Related Risk Domains in Deep Learning MRI Image Reconstruction

The first part focuses on Technology-Related Safety Considerations in Deep Learning Reconstruction, analysing the main risk domains recurrently described in the scientific literature. In particular, the following five domains are examined:

- Domain Shifts;
- Hallucinations;
- Adversarial Perturbation and Instability;
- Texture Alteration, Oversmoothing and Detail Loss;
- Silent Failure and Uncertainty Estimation.

Each domain is discussed in relation to the intrinsic characteristics of deep learning reconstruction algorithms and their potential consequences on reconstruction fidelity, robustness, and diagnostic reliability.

Following the literature-based analysis, these safety domains are contextualised specifically to SwiftMR. Based on the detailed description of the technical implementation and training strategy provided in Chapter 3, a qualitative HTA-oriented risk characterisation is performed. For each identified domain, a risk level (low, moderate, or high) is assigned, together with the mitigation strategies implemented by the manufacturer, including training data diversity, raw k-space learning, protocol validation, and clinical performance assessment.

Domain shifts

Domain shift, or distribution shift, refers to the limited ability of a trained deep learning model to accurately reconstruct images that differ from those encountered during training. In the context of MRI reconstruction, generalization describes the

capability of the model to produce reliable outputs when applied to previously unseen acquisition settings. When new input data deviate from the statistical distribution of the training dataset, reconstruction performance may degrade. This variability may arise from differences in MRI hardware, vendor-specific acquisition protocols, coil configurations, or anatomical regions, leading to models that perform well on source data but less reliably in other clinical contexts (Heckel et al. 2024).

The problem of robustness under distribution shifts has been explored in several studies. (Johnson et al. 2021) investigated the impact of small structural changes, the addition of noise to k-space measurements, and variations in the number of receiver coils, demonstrating that many deep learning reconstruction models are sensitive to such perturbations and may exhibit performance degradation under altered acquisition conditions. Similarly, (Darestani, Chaudhari, and Heckel 2021) analysed robustness under out-of-distribution data and showed that both trained and untrained networks were affected by distribution shifts, highlighting intrinsic generalisation limits of deep learning reconstruction systems rather than deficiencies in specific implementations.

Distribution shifts may also arise from variations in sampling strategies. Models trained on specific undersampling schemes may not generalize reliably to alternative sampling patterns or acceleration factors. In MRI, this includes differences in k-space sampling masks, vendor-specific reconstruction pipelines, and the integration of parallel imaging (PI) or compressed sensing (CS) strategies. Since PI and CS introduce characteristic artefact patterns and noise distributions, a deep learning model trained under specific sampling configurations may encounter altered error structures when applied to different acceleration schemes or vendor-dependent pipelines (Avidan and Freiman 2023).

Overall, domain shift represents a technology-related safety consideration inherent to deep learning-based MRI reconstruction. While extensive and diverse training datasets can mitigate this risk, the possibility of performance degradation under unvalidated acquisition conditions cannot be fully excluded, particularly in heterogeneous real-world clinical environments.

Hallucinations

The term hallucination in deep learning reconstruction refers to the generation of false but realistic-looking image features that are not supported by the acquired data (Heckel et al. 2024). In medical imaging, hallucinations may arise when the reconstruction model relies excessively on learned priors, particularly in the presence of distribution shifts between training and test data (Bhadra et al. 2021). In such cases, the network may produce visually plausible structures that reflect patterns learned during training rather than true anatomical information.

Importantly, hallucinations may not be detectable through conventional quantitative image quality metrics. Models can demonstrate improvements in signal-to-noise ratio (SNR) or structural similarity index (SSIM) while simultaneously introducing hallucinated artefacts that are not present in the original measurements (Muckley et al. 2021). This highlights a fundamental limitation of pixel-based or perceptual similarity metrics, which may fail to capture clinically relevant distortions.

(Cohen, Luck, and Honari 2018) further demonstrated that distribution-matching loss functions used in medical image translation frameworks can promote the generation of artificial features, as the model is incentivised to reproduce global statistical properties of the target distribution rather than ensuring strict data fidelity. Similarly, Badra et al. discussed hallucination in the context of tomographic image reconstruction and introduced the concept of hallucination maps to visualise and understand the influence of prior information in regularised reconstruction models.

From a theoretical perspective, (Gottschling et al. 2020) analysed the phenomenon of in-distribution hallucination, showing that even when test data originate from the same distribution as the training dataset, reconstruction models may still generate artificial structures due to the intrinsic ambiguity of inverse problems. This reinforces the idea that hallucinations are not solely a consequence of distribution shift, but may also emerge from the underlying ill-posed nature of reconstruction tasks.

The issue of hallucination is therefore particularly relevant from a safety perspective, as the introduction of realistic but non-existent structures may mislead clinical interpretation and potentially contribute to false positive or false negative diagnoses.

Adversarial attack and instabilities

Adversarial attacks refer to small, often imperceptible perturbations intentionally applied to input data in order to induce significant degradation in model output. In the context of deep learning-based MRI reconstruction, adversarial perturbations are not primarily considered a realistic clinical threat, but rather a methodological tool used to evaluate algorithmic robustness (Heckel et al. 2024). These perturbations simulate worst-case input variations and allow the stability limits of reconstruction networks to be investigated.

(Antun et al. 2020) demonstrated through simulations that small, adversarially selected perturbations applied to undersampled k-space measurements can result in severe reconstruction artefacts. Their findings also highlighted a trade-off between robustness and reconstruction performance, suggesting that models optimized for high perceptual quality may exhibit increased sensitivity to worst-case perturbations. Importantly, the study showed that classical sparsity-based reconstruction methods are also sensitive to adversarial perturbations, indicating that instability is not exclusive to deep learning approaches.

More broadly, (Antun et al. 2020) analysed instabilities of deep learning methods in inverse problems and demonstrated that small, structured perturbations in measurement data can lead to disproportionately large changes in reconstructed outputs. These findings underscore that stability guarantees in deep learning reconstruction are not inherently ensured by training alone and remain an area of ongoing research.

From a safety perspective, adversarial perturbation studies should therefore be interpreted as stress tests of algorithmic robustness rather than direct clinical risk scenarios. In real-world MRI practice, similar effects may arise from acquisition variability, noise fluctuations, or protocol deviations rather than malicious manipulation, reinforcing the importance of validation under diverse acquisition conditions and human oversight in routine use.

Silent Failure and Uncertainties Estimation

Deep learning-based MRI reconstruction operates within the framework of an ill-posed inverse problem, where multiple high-quality image solutions may correspond to the same undersampled k-space measurements (Heckel et al. 2024). In such settings, the reconstruction model relies on learned priors derived from the training data to select a plausible solution among many mathematically valid possibilities. While this strategy enables substantial improvements in acceleration and image quality, it also introduces the risk of silent failure, whereby the model produces a visually convincing reconstruction that may not fully reflect the underlying anatomical reality.

Unlike conventional algorithmic failures that result in obvious artefacts or reconstruction breakdown, silent failures may manifest as subtle deviations in texture, structure, or lesion conspicuity without triggering explicit warning signals. This concern has been discussed in the broader literature on deep learning for inverse problems, where stability and robustness limitations have been demonstrated even in well-trained networks (Antun et al. 2020); (Darestani et al. 2021). Because MRI reconstruction can admit multiple plausible solutions, quantifying reconstruction uncertainty has been proposed as a strategy to enhance reliability and user trust, particularly in clinical settings (Heckel et al. 2024). Although uncertainty estimation techniques are still under active development, their conceptual role highlights the importance of human oversight and external validation in mitigating silent failure risks.

Texture alteration, Oversmoothing and Detail Loss

Deep learning-based MRI reconstruction methods rely on learned priors derived from training data to suppress noise and recover undersampled information. While this approach substantially improves visual image quality and signal-to-noise ratio, it may

also lead to texture alteration and oversmoothing, particularly when strong denoising behaviour dominates the reconstruction process (Aghabiglou and Eksioglu 2022). Oversmoothing refers to the suppression not only of stochastic noise but also of subtle anatomical variations and fine structural details, resulting in images that appear cleaner but may lack high-frequency information relevant for diagnosis. This phenomenon has been discussed in the context of deep learning-based MRI reconstruction and denoising as a potential limitation of models trained with pixel-wise loss functions or perceptual similarity objectives ((Hammernik et al. 2017); Knoll et al. 2019). When the learned prior is overly dominant, the network may favour smooth, visually plausible outputs over strict fidelity to the acquired k-space data, thereby attenuating small lesions, reducing edge sharpness in diagnostically relevant regions, or modifying tissue texture patterns.

Several studies have highlighted that improvements in conventional image quality metrics (e.g., PSNR or SSIM) do not necessarily correlate with preservation of subtle diagnostic features, particularly in low-contrast or low-SNR scenarios ((Antun et al. 2020); (Hossain et al. 2024)). In post-processing deep learning denoising frameworks, similar effects have been observed, where aggressive noise suppression resulted in reduced conspicuity of small structures or altered appearance of tissue interfaces. These findings underscore an intrinsic trade-off in deep learning reconstruction between noise suppression and detail preservation. Importantly, this limitation does not imply systematic diagnostic failure but represents a technology-related safety consideration, especially in tasks requiring high sensitivity to small or low-contrast abnormalities.

Following the theoretical analysis of the five technology-related safety domains associated with deep learning-based MRI reconstruction, a qualitative risk characterisation was performed specifically for SwiftMR. Based on the technical implementation, training strategy, and algorithmic architecture described in Chapter 3, each safety domain was evaluated in terms of its potential relevance to the product. A qualitative risk level (low, moderate, or high) was assigned within the HTA framework, together with a justification of the corresponding mitigation strategies implemented by the manufacturer. No domain was classified as high risk, as identified concerns are considered inherent to the class of deep learning reconstruction systems and are mitigated through design, validation, and controlled clinical integration.

Safety Domain	Theoretical description	Qualitative Risk for SwiftMR	Mitigation Strategy (Based on Chapter 3)
---------------	-------------------------	------------------------------	--

Domain Shifts	Performance degradation when applied to acquisition settings differing from training data (hardware, vendor protocols, anatomical regions, sampling schemes).	Low–Moderate	Large multivendor dataset; heterogeneous anatomical and protocol coverage; multidimensional k-space degradation simulation; vendor-specific reconstruction pipeline integration.
Hallucinations	Generation of plausible but non-existent image features due to strong learned priors or data mismatch.	Moderate (task-dependent)	Supervised training with fully sampled k-space references; controlled degradation modelling; validation against standard-of-care reconstruction; radiologist comparison.
Adversarial Perturbation Instability	Sensitivity to small structured perturbations in undersampled k-space, potentially leading to severe reconstruction artefacts.	Low	Exposure to diverse k-space perturbations during training; robustness under controlled acquisition conditions; standard clinical workflow supervision.
Texture Alteration, Oversmoothing and Detail Loss	Suppression of fine structural detail due to strong denoising	Low–Moderate	Quantitative evaluation of edge sharpness and

	behaviour or learned priors.		signal preservation; raw k-space-based reconstruction; contextual noise- level control.
Silent Failure and Uncertainty Estimation	Plausible but inaccurate reconstructions due to the ill-posed nature of inverse problems without explicit failure signalling.	Moderate	Extensive supervised training; comparison with fully sampled references; validation against standard reconstruction; mandatory human interpretation.

Table 4.1: Risk assessment of the technology safety related domains.

4.1.2. Clinical Safety Evidence of SwiftMR

While the previous subsection addressed safety from a technology-oriented perspective, focusing on algorithmic and model-related risks, the following analysis considers safety within a clinical context. A number of peer-reviewed studies evaluating SwiftMR were identified in the literature, and their findings are examined specifically to extract clinical safety-related considerations, including potential impacts on diagnostic reliability, lesion interpretation, and reader confidence.

In this chapter, these studies are examined specifically through a safety-oriented lens, with attention to reported diagnostic limitations, potential false positive or false negative findings, and any discrepancies compared to standard-of-care reconstructions. The comprehensive evaluation of diagnostic performance and clinical effectiveness of these technologies will be presented in detail in Chapter 5. Here, the focus remains strictly on safety-related aspects emerging from the available clinical evidence.

(Oh et al. 2025) evaluated the use of SwiftMR for thin-slice T1-weighted black-blood imaging in the detection of brain metastases. From a safety perspective, the enhanced image sharpness associated with deep learning reconstruction was reported to increase false-positive findings, particularly due to unsuppressed vascular signals. This effect may complicate the differentiation between small vessels and metastatic lesions, requiring cautious radiological interpretation.

(Kim et al. 2024) evaluated on deep learning reconstruction applied to 4D TWIST MR angiography assessed the integration of SwiftMR within a dynamic angiographic

acquisition framework. From a safety perspective, residual venous contamination was observed, reflecting intrinsic characteristics of the dynamic sequence that were not fully eliminated by the deep learning reconstruction. While this represents a potential interpretative limitation in vascular assessment, it also suggests that the algorithm does not artificially suppress physiologically acquired signal components, maintaining consistency with the original acquisition dynamics.

(Suh et al. 2024) evaluated coronal T2-weighted turbo spin-echo imaging with 1 mm slice thickness for the detection of temporal lobe epilepsy-related hippocampal and temporal lobe lesions assessed the use of SwiftMR in high-resolution thin-slice acquisitions. From a safety perspective, an increased false-positive rate and reduced specificity were reported, mainly due to focal signal abnormalities visible on a single thin slice but not confirmed across adjacent contiguous slices. This phenomenon may lead to overinterpretation of hippocampal lesions, particularly when relying on isolated signal changes without multiplanar or multi-slice confirmation.

(Lee et al. 2023) evaluated knee MRI using a fast-imaging protocol with approximately 41% scan time reduction assessed the performance of SwiftMR in accelerated musculoskeletal imaging. From a safety perspective, the authors reported occasional underestimation of cartilage abnormalities and overestimation of signal intensity, likely related to the level of denoising applied by the deep learning reconstruction. Additionally, artefacts associated with aggressive parallel imaging acceleration were not fully corrected by the algorithm and could persist in the reconstructed images, potentially affecting image interpretation in selected cases.

(Yoo et al. 2023) evaluated lumbar spine MRI using T1-weighted and T2-weighted turbo spin-echo sequences reconstructed with SwiftMR assessed the impact of deep learning reconstruction combined with parallel imaging and compressed sensing, achieving an approximate 32.3% scan time reduction. From a safety perspective, a reduction in specificity with an increased false-positive rate was reported, particularly in the detection of annular fissures. These findings were attributed to noise modulation and enhanced image sharpness associated with the denoising effect, which may lead to overinterpretation of subtle signal alterations.

Overall, the safety-related findings reported across the SwiftMR clinical validation studies can be interpreted within the framework of the previously defined technology-related safety domains. In particular, increased false-positive rates and reduced specificity observed in selected scenarios may reflect effects associated with lesion conspicuity alteration and denoising-related image modulation, while persistence of acquisition-dependent artefacts highlights the intrinsic limitations of sequence design rather than algorithm-induced distortions. These observations reinforce the importance of contextualising clinical findings within the broader technological risk domains previously outlined.

4.2. Organizational Perception of the Safety Dimension

In order to complement the literature-based safety analysis with primary contextual evidence, an organisational-level assessment was conducted within the hospital setting directly involved in the HTA decision-making process. This qualitative component aims to explore how the safety of the technology is perceived in routine clinical practice, particularly with respect to operational reliability, workflow integration, and diagnostic confidence.

Two key professionals from the Radiology Unit of Ospedale di Erba were interviewed: the Head of Radiology and a senior MRI technologist with more than ten years of experience. Their roles were selected to capture both the diagnostic and operational perspectives of MRI practice. A structured set of open-ended questions was administered to investigate perceived safety aspects of SwiftMR, including potential limitations, workflow implications, interpretative challenges, and confidence in the reconstructed images.

Prior to the interviews, the software had been implemented within the unit for a one-month pilot phase, allowing its use in routine clinical activity under real-world conditions. The interviews were conducted following this trial period, enabling participants to reflect on direct practical experience rather than hypothetical expectations. This approach ensured that the qualitative assessment was grounded in actual operational exposure to the technology within the local hospital context.

The responses were analysed qualitatively to identify recurring themes and relevant safety-related considerations. This section therefore presents the main insights emerging from these expert interviews, providing complementary evidence to the scientific literature and contributing to a context-specific understanding of safety within the local clinical environment.

Building upon the methodological framework described above, the following section presents a structured synthesis of the evidence emerging from the interviews. Rather than reporting responses individually, the findings have been organized according to recurring thematic patterns identified across participants. This approach allows for a clearer interpretation of converging perceptions and shared safety-related considerations within the local clinical context. The extracted themes were grouped into six analytical domains, reflecting the main areas in which safety implications were discussed: general reliability perception, diagnostic confidence, artifacts and technical stability, workflow and operational safety, human supervision and responsibility, and parameters of use and technical perception. For each domain, the analysis highlights the common elements emerging from both professionals, with the aim of identifying areas of alignment between diagnostic and operational perspectives.

It is important to note that while most domains reflect convergent insights between the Head of Radiology and the Senior MRI Technologist, the section on parameters of

use and technical perception derives exclusively from the technologist's perspective. This distinction is methodologically relevant, as it captures a dimension of safety specifically related to hands-on system interaction and technical implementation in daily MRI practice.

The following subsections therefore present a thematically organized synthesis of the interview findings, emphasizing shared perceptions where applicable and explicitly distinguishing role-specific insights when necessary.

General reliability perception

Across both professional perspectives, the AI-based reconstruction was perceived as comparable to the standard reconstruction in terms of overall reliability. The system was considered technically stable and consistent in its output, without deviations that would compromise routine clinical practice.

Importantly, no specific safety risks were identified in the implementation and use of the technology within the local clinical setting. The absence of observed reliability-related concerns suggests that the integration of SwiftMR did not introduce additional operational vulnerabilities compared to conventional reconstruction methods. Overall, the technology was perceived as maintaining baseline safety standards in everyday MRI practice.

Diagnostic confidence

From both the diagnostic and operational perspectives, the use of AI-based reconstruction did not negatively influence diagnostic confidence. The professionals reported that, in routine clinical use, the technology did not generate uncertainty in image interpretation nor raise concerns regarding the validity of the visualized findings.

On the contrary, the perceived subjective image quality was described as equal or, in some cases, superior to the standard reconstruction. Importantly, the integration of SwiftMR did not lead to situations in which identified findings were questioned due to reconstruction-related doubts. This suggests that the technology preserved interpretative reliability while potentially enhancing perceived image clarity.

Artifacts and technical stability

A relevant limitation emerged in relation to the management of image artefacts. While the system was generally perceived as technically stable, it was observed that, in the presence of artefacts, AI-based reconstruction tended to amplify their visual impact, regardless of their origin (e.g., patient-related motion or other acquisition-related factors). When compared to standard reconstruction under equivalent artefact conditions, the AI-reconstructed images were perceived as presenting a greater degradation in overall quality. This suggests that the technology may be more sensitive to pre-existing image disturbances, potentially reducing image clarity in suboptimal acquisition scenarios.

Although this limitation does not inherently compromise safety, it represents an important operational consideration. The amplification of artefacts may affect image interpretability in specific cases and should therefore be acknowledged as a constraint in the clinical deployment of the technology.

Workflow and operational safety

Both professionals reported a substantial reduction in examination time associated with the use of AI-based reconstruction, estimated at approximately 40% on average compared to standard reconstruction protocols. This reduction translated into a measurable acceleration of workflow processes within the MRI unit.

From an operational safety perspective, the shortening of scan time was perceived as beneficial, contributing to improved efficiency and potentially reducing patient discomfort and motion-related issues. The optimization of workflow did not introduce additional procedural complexity or safety concerns, suggesting that the integration of the technology was compatible with existing organizational processes.

Overall, the time efficiency gain was interpreted not only as an economic or productivity advantage but also as a factor indirectly supporting operational stability and patient management within routine clinical practice.

Human supervision and responsibility

Both professionals emphasized that human supervision remains fundamental when using AI-based reconstruction. A systematic review of reconstructed images, regardless of whether they are generated through standard or AI-based techniques, was considered good clinical practice prior to diagnostic reporting. The technology was therefore perceived as a supportive tool rather than an autonomous decision-making system, with full interpretative responsibility remaining under medical supervision.

However, particular attention was deemed necessary in more delicate clinical scenarios, such as dynamic abdominal protocols requiring patient cooperation. In these contexts, respiratory motion artefacts are frequently present, and given the observed tendency of AI-based reconstruction to amplify artefacts, the potential impact on image interpretability may be more pronounced. These situations were identified as requiring heightened vigilance to ensure that reconstruction-related amplification does not interfere with accurate clinical assessment.

Overall, the integration of SwiftMR was not perceived as diminishing professional responsibility; rather, it reinforced the importance of maintaining structured human oversight, especially in clinically sensitive acquisition settings.

Parameters of use and technical perception

From a technical-operational perspective, the implementation of AI-based reconstruction required a structured calibration phase. Each anatomical district was

optimized using a specific denoising level, determined through iterative testing of different parameter configurations. This process aimed to identify a balanced compromise between acquisition time reduction and preservation of diagnostic image quality.

Time reduction was primarily achieved by adjusting acquisition-related parameters such as the number of signal averages and compressed sensing factors. Importantly, variations in these parameters did not result in qualitative discrepancies in the AI-reconstructed images, suggesting a stable reconstruction performance under different acquisition settings.

Particular attention was also devoted to maintaining an adequate signal-to-noise ratio (SNR), in order to avoid the generation of artefactual formations that could subsequently be amplified by the AI reconstruction process. This technical precaution reflects a proactive approach to risk mitigation, ensuring that parameter optimization did not inadvertently introduce image distortions.

Overall, the technology was perceived as technically manageable, provided that parameter selection was carefully calibrated and monitored.

Safety domains discussion

Overall, the safety domain was assessed through a multi-layered approach integrating theoretical risk analysis, secondary clinical evidence, and primary contextual data collected within the hospital setting. From a technology-related risk perspective, the HTA framework allowed for the identification of potential model-associated vulnerabilities, particularly those linked to training data validity, undersampling strategies, and reconstruction processes. These risks were considered predominantly low to moderate and largely mitigated by the structured training phase of the model, the quality of the case-based undersampling masks, and the controlled post-processing reconstruction pipeline implemented by SwiftMR.

The review of peer-reviewed clinical studies and case-based literature highlighted isolated concerns, including occasional alterations in lesion conspicuity related to signal intensity modulation or subtle sub-tissue representation changes. However, these findings were not supported by statistically robust evidence and remain areas requiring further investigation rather than established safety threats.

The primary qualitative evidence collected following a one-month pilot implementation provided context-specific insight into real-world use. Both diagnostic and operational perspectives converged on the importance of maintaining structured human supervision. Diagnostic confidence and perceived image quality were not reported as compromised by AI-based reconstruction. On the contrary, workflow acceleration and scan-time reduction were identified as operational benefits, contributing to improved efficiency.

At the same time, a relevant technical limitation emerged in relation to artefact amplification, particularly when aggressive acceleration strategies (e.g., parallel imaging and compressed sensing adjustments) are applied. This finding underscores the need for careful parameter calibration to balance acquisition speed with artefact control, especially in motion-sensitive protocols.

From an HTA perspective, the safety profile of SwiftMR within the local context appears consistent with the standard of care in MRI practice. The technology does not eliminate the need for human oversight, nor does it fundamentally alter professional responsibility. Rather, it introduces an innovation whose safe use depends on appropriate technical calibration, structured supervision, and awareness of context-specific limitations.

5 Clinical effectiveness

In this chapter, the effectiveness of SwiftMR in clinical settings is discussed within the context of Health Technology Assessment (HTA) for AI-based diagnostic software. SwiftMR is considered a diagnostic medical tool as it directly impacts the enhancement and quality of magnetic resonance images. Its ability to modify image reconstruction significantly influences diagnostic accuracy, image clarity, and clinical decision-making. The assessment of its clinical effectiveness is a crucial aspect of the HTA process. According to the HTA Core Model for diagnostic technologies, evidence from randomized controlled trials, clinical research, and structured reader or performance evaluation studies is essential to validate its clinical efficacy, reliability, and practicality (EUnetHTA 2008). (Oh et al. 2025)

To ensure methodological consistency and transparency, the chapter is organized as follows. Firstly, the clinical validation of SwiftMR is evaluated by analyzing five peer-reviewed scientific articles found in reputable bibliographic databases, which serve as supporting evidence for the software. Secondly, clinical data concerning other image-based post-processing software solutions are examined to place the performance of SwiftMR in the current technological context. Thirdly, a few studies exploring alternative methods involving software integrated directly into MRI acquisition parameters are reviewed. These methods rely on physics-based or model-based predictive techniques and offer new options beyond traditional image-based post-processing.

Finally, a comparative discussion is presented, weighing the existing evidence supporting image-based post-processing solutions against emerging technologies that may soon be clinically accessible. This assessment aims to guide decision-making on whether to procure an image-based post-processing solution immediately or to wait for potentially validated alternatives in the future. Throughout each section, a standardized structure is used to analyze each study, including a PICO framework, details on study design and methods, a summary of key findings, and an evaluation of clinical significance. Whenever possible, relevant figures and qualitative findings from the original studies are included to reinforce the analysis.

5.1. Clinical validation of the SwiftMR software based on scientific evidence

This section addresses the clinical validation of SwiftMR within the framework of the Health Technology Assessment (HTA) of an AI-based diagnostic software. Clinical validation represents a critical component for demonstrating the clinical performance, safety, and intended use of AI-based medical devices, in accordance with current regulatory and methodological standards. For this purpose, five peer-reviewed scientific articles provided by the manufacturer were identified and independently retrieved from recognized scientific databases.

These studies were selected as relevant evidence supporting the clinical validity of SwiftMR in its intended diagnostic context. The following table summarizes the analyzed articles, including the article title, bibliographic reference, and the database from which each publication was retrieved.

Subsequently, each study is analyzed individually. For every article, a PICO framework is presented to provide a concise overview of the study design, population, and objectives. This is followed by a structured analysis of the study methodology, main outcomes, and clinical relevance. Where appropriate, figures extracted from the publications are included to illustrate key results and performance metrics of the evaluated software.

Table 5.1: Scientific paper for the clinical validation of SwiftMR

Article title	Reference	Database
Deep learning-based image enhancement for improved black blood imaging in brain metastasis	(Oh et al. 2025)	Pubmed
Highly accelerated knee magnetic resonance imaging using deep neural network (DNN)-based	(Lee et al. 2023)	Pubmed

reconstruction: prospective, multi-reader, multi-vendor study			
Deep Learning-Based High-Resolution Magnetic Resonance Angiography (MRA) Generation Model for 4D Time-Resolved Angiography with Interleaved Stochastic Trajectories (TWIST) MRA in Fast Stroke Imaging	(Kim et al. 2024)	Pubmed	
Improving Diagnostic Performance of MRI for Temporal Lobe Epilepsy With Deep Learning-Based Image Reconstruction in Patients With Suspected Focal Epilepsy	(Suh et al. 2024)	Pubmed	
Deep learning-based reconstruction for acceleration of lumbar spine MRI: a prospective comparison with standard MRI	(Yoo et al. 2023)	Pubmed	

Article 1: Deep learning-based image enhancement for improved black blood imaging in brain Metastasis

This study primarily focuses on the diagnosis and identification of brain metastases. Black-blood (BB) MR imaging is considered the reference standard for lesion detection in this clinical context. Common BB imaging techniques include variable refocusing flip-angle schemes that promote effective intravoxel dephasing, as well as preparation pulses such as double inversion recovery (Lindenholz et al. 2018). Compared with conventional contrast-enhanced T1-weighted imaging (T1WI), BB sequences suppress vascular signal, thereby increasing lesion-to-background contrast. In standard contrast-enhanced T1WI, both metastatic lesions and blood vessels appear hyperintense, which may reduce lesion conspicuity. BB imaging exploits blood flow properties to selectively attenuate vessel signal, improving the differentiation between vascular structures and metastatic lesions. According to current clinical practice, thin-slice acquisitions are recommended to improve the detection of small metastatic lesions (Yamatogi et al. 2011). However, increasing spatial resolution is associated with a trade-off, as thinner slices may lead to increased image noise and susceptibility to artifacts. The objective of this study is to assess the performance of an AI-based image enhancement solution, such as SwiftMR, in terms of image quality and diagnostic performance, particularly in the context of high-resolution BB MR imaging.

PICO

Table 5.2: PICO for the paper (Oh et al. 2025)

P	I	C	O
126 patients with brain metastasis and 121 without Sequence: 3D contrast enhanced T1-weighted black blood on a 3T MR scanner	Deep-learning based image enhancement (SwiftMR)	MRI images without post-processing with DL software	Image quality: • Quantitative: SNR, CNR on lesions. • Diagnostic performance: two radiologist graded the overall image quality and the artifacts • sensitivity and false positive rate

of lesions detection

Study design

Two experienced neuroradiologists, each with more than 15 years of clinical experience, retrospectively reviewed all MR images and annotated all metastatic lesions. The radiologists independently reviewed both the original MR images and the post-processed images, including pre- and post-contrast 3D T1-weighted gradient-echo sequences and 3D T1-weighted black-blood imaging. These sequences constituted the acquisition protocol used in the study. The ground truth, used as the standard reference for accuracy-related analyses, was defined based on the assessment of an expert neuroradiologist. Quantitative image quality was assessed using signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR). Due to data availability, only 126 patients with metastasis were included in the quantitative analysis. For CNR evaluation, all metastatic lesions identified in each image were considered. CNR was calculated using the signal intensity of the lesion region of interest (ROI), the signal intensity of normal white matter, and the standard deviation of the background noise, according to the Equations (5.1), (5.2).

$$SNR = \frac{S_{lesion}}{\sigma_{background}} \quad (5.1)$$

$$CNR = \frac{S_{lesion} - S_{white\ matter}}{\sigma_{background}} \quad (5.2)$$

Images from the 126 patients and the corresponding metastatic lesions were analyzed on both the original MR images and the deep learning (DL)-reconstructed images and subsequently compared. Qualitative assessment of overall image quality and image artifacts was performed using a five-point Likert scale. Image evaluation was conducted in a blinded manner and in randomized order. Regarding diagnostic performance, the two evaluators assessed lesion conspicuity using a three-point scale on both DL-reconstructed and original images. Lesion detectability was then compared against the study ground truth. Sensitivity and specificity of lesion detection with respect to the reference standard were calculated accordingly. Statistical analyses were selected based on the type of outcome measure. Quantitative image quality metrics, including signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), and the five-point Likert scale scores, were analyzed using the Wilcoxon signed-rank test. For diagnostic performance, lesion-based analyses were conducted to assess sensitivity and lesion conspicuity scores. Each metastatic lesion was considered independent of the patient and was analyzed using logistic regression and ordinal logistic regression models with generalized estimating equations (GEE). These models accounted for within-subject correlation due to multiple lesions per patient and clustering related to

the two radiologist readers. For patient-based diagnostic performance analysis, specificity and false-positive lesion counts were evaluated using Poisson regression models. In the patient-based approach, each subject was classified according to the presence or absence of metastatic lesions. For summary of the statistical test the Table 5.2 is reported.

Table 5.3: statistical test used for (Oh et al. 2025; Park et al. 2012)

Statistical test	Analysis of interest
Wilcoxon	Image quality (SNR, CNR, 5-point Likert scale)
Logistic regression (binary analysis) and GE	Sensitivity of lesions
Ordinal regression and GEE	Conspicuity (1-3)
Logistic regression	Specificity
Poisson regression	False positive count

Main results

Quantitative analysis demonstrated significantly higher signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) in the deep learning (DL)–enhanced images compared with the original images. Qualitative image quality, assessed using a five-point Likert scale, was rated as significantly superior by both readers in terms of overall image quality and noise reduction. No statistically significant differences were observed for image artifacts. Diagnostic performance analysis showed a higher sensitivity for lesion detection in the DL post-processed images compared with the comparator. When stratified by lesion size, a statistically significant difference in sensitivity was observed only for lesions smaller than 5 mm. Lesion conspicuity scores were significantly higher in DL-enhanced images for both readers across all lesion size subgroups (<5 mm, 5–10 mm, and >10 mm). Both readers consistently agreed that DL-enhanced images improved the detectability of small lesions and increased the highest conspicuity scores for metastatic lesions. False-positive counts and specificity did not show statistically significant differences between the comparator and intervention groups.

Based on the reported results, the deep learning–based image enhancement software demonstrates a consistent improvement in image quality, as reflected by significantly

higher signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) compared with the comparator images reconstructed without deep learning enhancement. These quantitative improvements are supported by qualitative assessments, with both readers reporting superior overall image quality and reduced noise levels. From a diagnostic perspective, the improvement in image quality translates into an increased ability to detect smaller metastatic lesions. In particular, the use of deep learning-enhanced images is associated with a higher sensitivity for lesions smaller than 5 mm compared with conventional images. This effect is clinically relevant, as small lesions represent a known diagnostic challenge in brain metastasis imaging and are more likely to be missed on standard contrast-enhanced sequences. Overall, these findings indicate that deep learning-based image enhancement may improve the identification of small brain metastases by enhancing lesion conspicuity and image quality, without adversely affecting specificity or increasing false-positive findings. The observed impact primarily supports the clinical validation of the software in the context of lesion detection, particularly for small lesions, while broader implications for downstream clinical outcomes remain outside the scope of the present evidence.

Figures and qualitative results

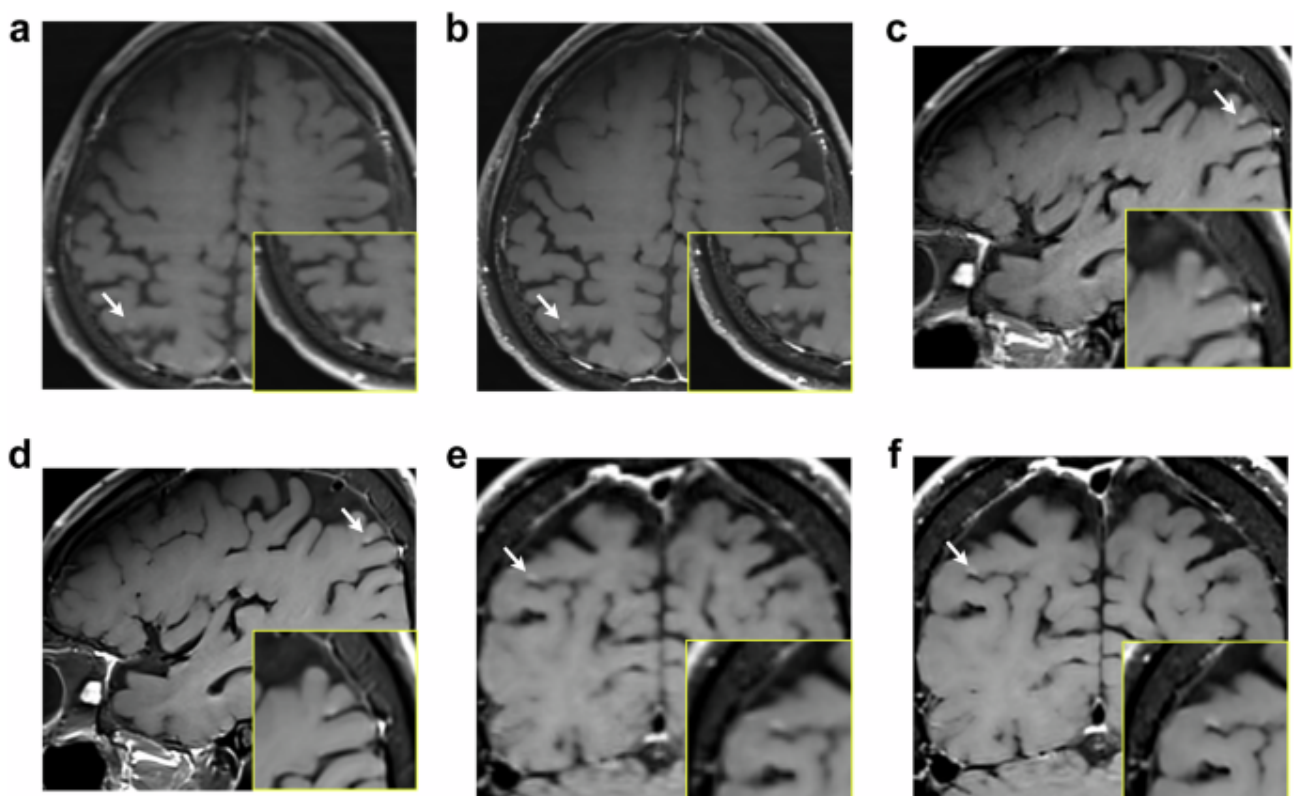


Figure 5.1: Case of 67 years old man with lung cancers. Image reproduced from: (Oh et al. 2025).

Figure 5.1 illustrates three brain metastatic lesions. Subfigures (a), (c), and (e) correspond to the comparator group, while subfigures (b), (d), and (f) show the deep learning (DL)-enhanced images from the intervention group. Compared with the

comparator images, DL-enhanced reconstructions demonstrate improved image quality and increased lesion conspicuity. This qualitative improvement supports the observed increase in diagnostic sensitivity for small lesions.

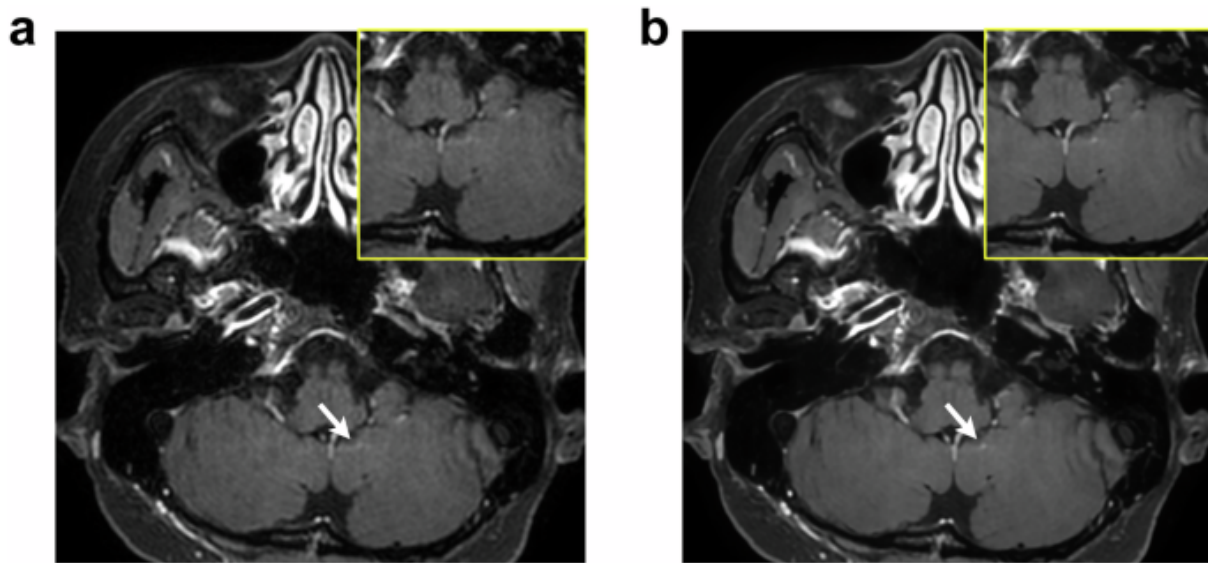


Figure 5.2: False positive case of brain metastasis in a 43 years old women with breast cancer. Image reproduced from: (Oh et al. 2025).

Figure 5.2 shows an example of a false-positive finding. On the left, the comparator image acquired without post-processing demonstrates a residual unsuppressed blood vessel in the black-blood sequence, which appears qualitatively less conspicuous and is less likely to be interpreted as a metastatic lesion. On the right, the deep learning-enhanced image shows increased vessel-to-background contrast, resulting in higher contrast-to-noise ratio of the unsuppressed vessel. This enhancement may increase the likelihood of misclassifying residual vascular signal as a lesion. This example illustrates a potential risk associated with image quality enhancement in black-blood imaging, namely the amplification of incompletely suppressed vascular structures. However, despite this qualitative observation, the overall study results did not show a statistically significant increase in false-positive counts or a reduction in specificity in the intervention group.

Clinical relevance and limitations

This study primarily aimed to improve image quality in black-blood MR imaging by increasing lesion-to-background contrast, as reflected by higher signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) at the lesion level. From a clinical standpoint, a potential concern associated with deep learning-based image enhancement is that improved background and vessel signal, particularly in incompletely suppressed blood vessels, could increase false-positive findings and reduce specificity, given the known risk of misinterpreting residual vascular signal as metastatic lesions in black-blood imaging (Park et al. 2021). However, this concern was not supported by the

study results, as no statistically significant differences were observed in false-positive counts or specificity between the intervention and comparator groups.

These findings indicate that the observed improvement in image quality translated into enhanced diagnostic sensitivity for small lesions without compromising diagnostic reliability, thereby supporting the clinical validation of the software in this specific diagnostic setting. Nevertheless, several limitations should be acknowledged.

The study was conducted on a relatively small patient cohort, which may have limited statistical power for some analyses, and it was performed at a single center, potentially affecting generalizability. Additionally, patients with a high metastatic burden (≥ 10 metastases) were excluded, which may limit applicability to more advanced disease stages. The ground truth was defined by expert radiologist assessment rather than histopathological confirmation, which was not feasible in most cases due to the lack of surgical resection. Finally, the evaluation was limited to a single image-based post-processing software, without comparison to other similar AI-based solutions.

Article 2: Highly accelerated knee magnetic resonance imaging using deep neural network (DNN)-based reconstruction: prospective, multi-reader, multi-vendor study

This study provides clinical evidence supporting the use of deep neural network-based image reconstruction (SwiftMR) in accelerated 2D Fast Spin Echo knee MRI. The primary objective is to demonstrate that a substantial reduction in acquisition time—approximately 41%—can be achieved while maintaining quantitative and subjective image quality, as well as diagnostic performance, comparable to conventional standard-of-care acquisitions. The study therefore aims to assess whether acceleration enabled by deep learning reconstruction preserves clinical effectiveness in routine knee MRI.

PICO

Table 5.4: PICO of the study (Lee et al. 2023)

P	I	C	O
45 patients evaluated for an arthroscopic knee Orthopaedic surgeon	Accelerated 2D fast spin echoes knee MR enhanced with DNN image-based AI software	Conventional 2D fast spin echoes knee MR images	Image quality: • Quantitative (SNR, CNR) • Qualitative (5-point Likert scale) Diagnostic performance:

meniscal, cruciate and cartilage lesions

Study design

The research conducted was a prospective study involving multiple readers and vendors to validate its clinical findings. A total of 45 patients, who required knee MRI scans, were enrolled in the study in advance. The patients were evenly distributed among three distinct 3.0 T MRI machines from different manufacturers, with 15 patients assigned to each machine. The evaluation of the images was carried out by three musculoskeletal radiologists, each with varying levels of experience in the field: 6 years, 1 year, and 1 year, respectively. These radiologists independently reviewed the imaging data sets. Each patient underwent knee MRI scans that included traditional and accelerated acquisitions reconstructed using software based on deep neural networks. The imaging protocol comprised four sets of 2D Fast Spin Echo (FSE) images, covering various planes and contrasts:

- 1. Axial T2-weighted fat-saturated images;
- 2. Coronal T2-weighted fat-saturated images;
- 3. Sagittal T2-weighted images;
- 4. Axial T1-weighted images.

Each set of images consisted of a series of 2D slices with a specific slice thickness, collectively encompassing the knee's anatomical volume. The assessment of image quality involved the use of Signal-to-Noise Ratio (SNR) and Contrast-to-Noise Ratio (CNR). SNR calculations were based on circular regions of interest (ROI) placed within the femoral bone marrow on the central slice, with signal intensity derived from the mean ROI value and image noise estimated from the background noise's standard deviation, serving as the denominator in the SNR computation. CNR calculations were conducted on the same central slice for each sequence, utilizing tissue-specific ROIs based on the imaging plane and contrast.

Subjective image quality evaluations were performed by the three radiologists using a five-point Likert scale. Each MRI scan was evaluated as a whole, rather than on a sequence-by-sequence basis. The assessment criteria included:

- 1 Overall image quality of anatomical structures;
- 2 Perceived image noise;
- 3 Presence of imaging artifacts.

Scores ranged from 1 (poor) to 5 (excellent). Diagnostic performance was evaluated concerning the identification and grading of musculoskeletal issues, such as meniscal lesions (medial and lateral meniscus), ligamentous lesions (including ACL, PCL,

MCL), and cartilaginous lesions. A reference standard (ground truth) was established by two additional expert radiologists, who reviewed traditional FSE images alongside the patients' medical records.

Diagnostic performance for meniscal and ligamentous lesions was assessed in terms of sensitivity, specificity, accuracy, and the Area under the ROC curve (AUC). For cartilage lesions, which were graded on an ordinal scale, the diagnostic agreement between traditional and DNN-reconstructed images was evaluated using the Pearson correlation coefficient, rather than classification-based measures. Paired statistical comparisons were conducted between conventional and DNN-reconstructed images for quantitative image quality metrics (SNR and CNR) and subjective Likert scores using paired t-tests. Inter-reader agreement was assessed utilizing the Pearson correlation coefficient.

Main results

An analysis of image quality using quantitative measures revealed that images reconstructed with deep neural network technology exhibited either similar or notably improved Signal-to-Noise Ratio (SNR) and Contrast-to-Noise Ratio (CNR) compared to standard 2D Fast Spin Echo scans. These enhancements were consistently evident across various sequences and imaging orientations. Subjective evaluations of image quality further corroborated these results. When rated on a five-point Likert scale, images reconstructed with deep neural networks received higher or equivalent scores for the overall quality of anatomical details, perceived noise levels, and the presence of imaging artifacts, indicating an enhancement or preservation of perceptual image quality.

Regarding diagnostic accuracy, images reconstructed with deep neural networks demonstrated performance on par with traditional MRI in assessing critical musculoskeletal structures such as meniscal tears, ligament damage, and cartilage irregularities. Various metrics including sensitivity, specificity, accuracy, and AUC values remained consistently high among different readers and anatomical regions.

Nevertheless, a couple of instances were identified in which discrepancies in cartilage assessment were noted in the images reconstructed with deep neural networks, with one case showing an overestimation and another an underestimation of a cartilage lesion. Despite these isolated occurrences, the overall diagnostic capability for cartilage evaluation aligned well with standard imaging practices.

In summary, the findings suggest that leveraging deep neural network reconstruction techniques allows for significant reduction in scanning time while upholding both objective and subjective image quality metrics and maintaining diagnostic accuracy comparable to traditional knee MRI, with minor issues observed in cartilage evaluation.

Figures and qualitative results

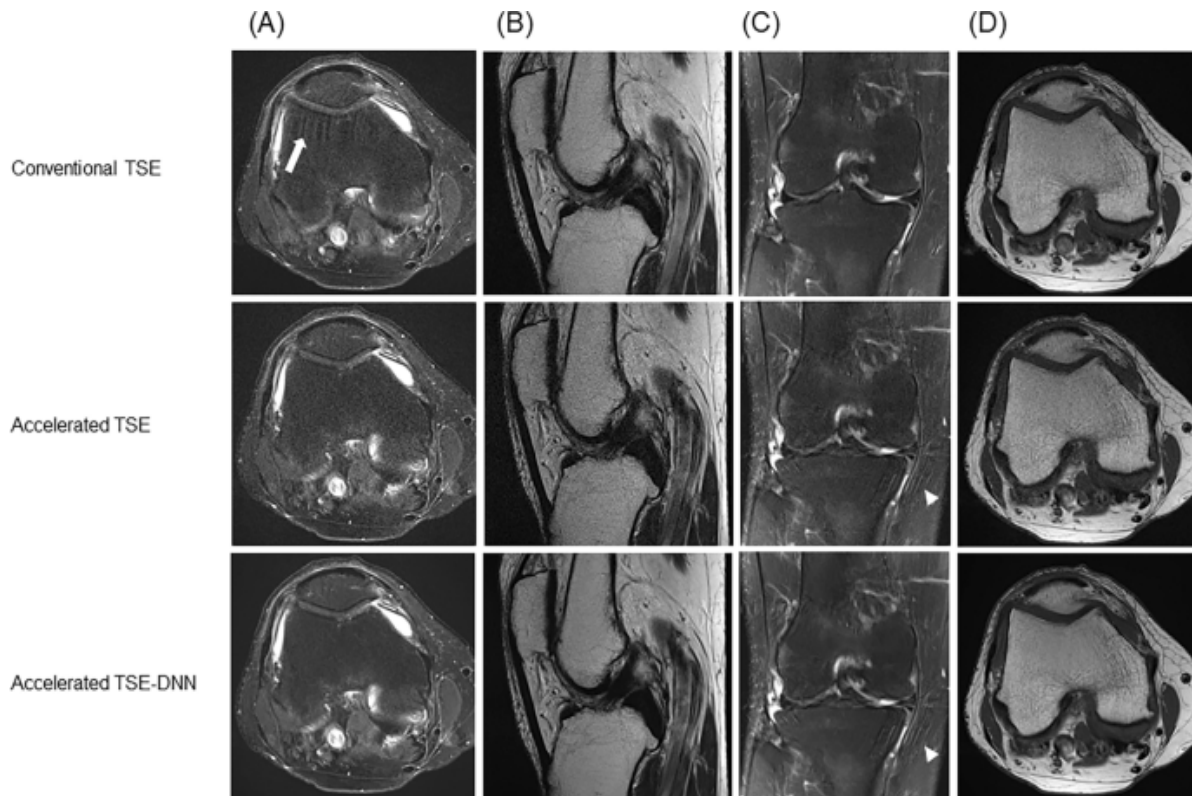


Figure 5.3: Conventional and reconstructed images of accelerated sequences: a 64 years old woman with knee pain. Images reproduced from: (Lee et al. 2023)

In Figure 5.3, there is a visual comparison presented for the four 2D Fast Spin Echo sequences outlined in the research design, depicted in columns A to D (axial T2-weighted fat-saturated, sagittal T2-weighted, coronal T2-weighted fat-saturated, and axial T1-weighted images). The initial row displays traditional scans with arrows identifying artifacts caused by motion. In the subsequent rows, representing faster scans and accelerated scans reconstructed using advanced deep learning post-processing, motion artifacts are notably diminished. Nevertheless, arrows point out the presence of artifacts linked to parallel imaging in these accelerated scans. This illustration underscores the balance between reducing motion artifacts by speeding up scan times and the remaining artifacts associated with parallel imaging methods.

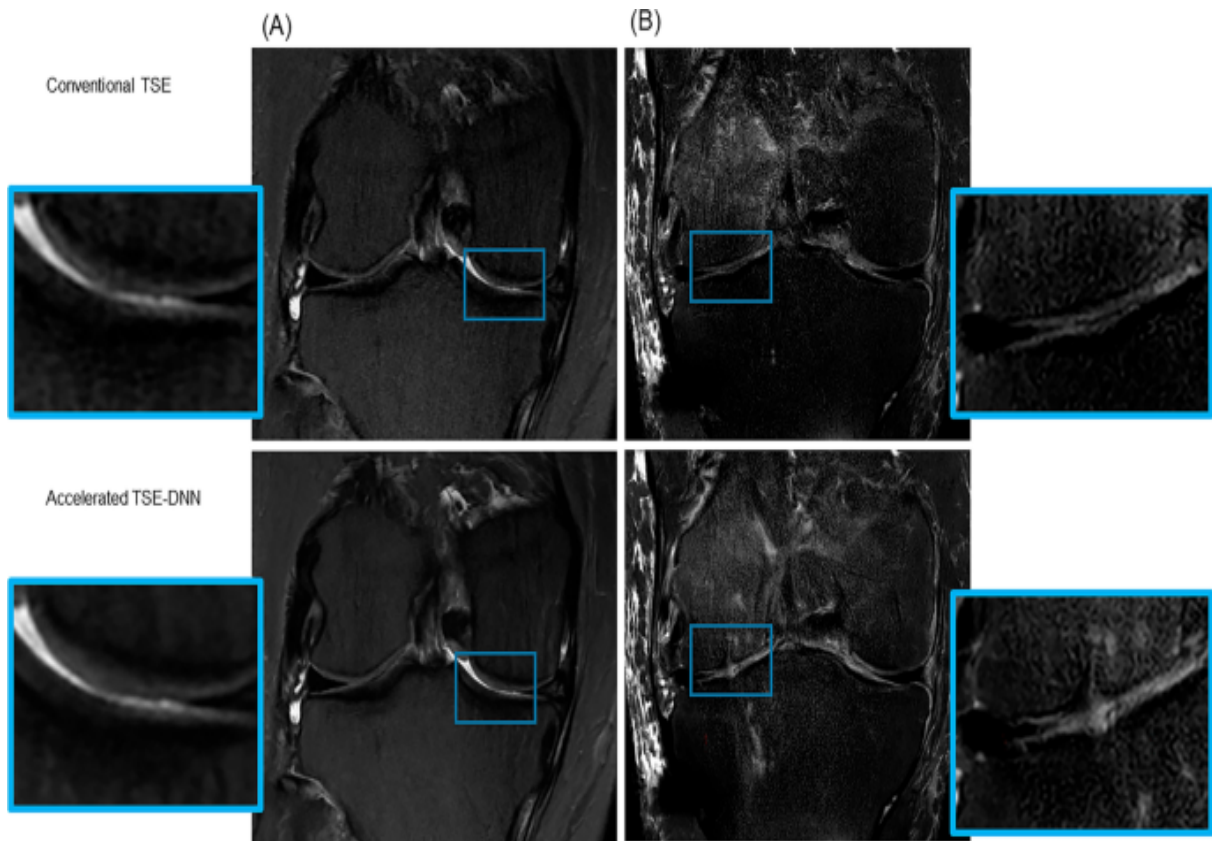


Figure 5.4: Cartilage grade on conventional and reconstructed images of accelerated sequence. A 54 years-old female with knee pain (A) and 41 years-old man (B). Image reproduced from: (Lee et al. 2023)

The illustration provided showcases two instances of discrepancies related to cartilage following reconstruction using a deep neural network.

In the initial scenario (A), a situation of underestimation is depicted. A 54-year-old woman with knee discomfort exhibits cartilage fissures on traditional imaging at the inner femoral condyle. In the corresponding 2D FSE-DNN image, the fissures appear somewhat smoothed out, resulting in an underestimation of cartilage damage. This outcome is attributed to the noise reduction and smoothing effects of the deep learning reconstruction, potentially reducing the visibility of minor irregularities on the cartilage surface.

In the subsequent scenario (B), an overestimation case is presented. A 41-year-old man with knee pain displays cartilage signal alterations without a clear structural flaw at the outer tibial plateau on standard imaging. Post deep neural network reconstruction, the cartilage exhibits heightened T2-weighted signal intensity, misinterpreted as a more severe cartilage defect, leading to an overestimation. In this instance, the deep learning reconstruction seems to amplify signal intensity in the cartilage on fluid-sensitive sequences, possibly magnifying pathological observations.

Overall, these examples emphasize how deep neural network reconstruction can impact cartilage evaluation differently: by potentially masking subtle fissures through

smoothing effects and by amplifying signal intensity on T2-weighted sequences, potentially overestimating cartilage irregularities. While based on only two cases, these findings emphasize the necessity for tailored optimization of noise reduction levels, particularly for delicate and signal-responsive tissues like articular cartilage.

Clinical relevance and limitations

The study successfully increased the speed of 2D Fast Spin Echo knee MRI using parallel imaging methods on three 3.0 T MRI systems. Following this acceleration technique, a deep neural network was utilized for image enhancement without altering the initial data collection process. The reconstructed images still bear traces of artifacts associated with parallel imaging, as evident in Figure 5.3.

Despite this reliance on parallel imaging characteristics, the deep neural network reconstruction notably reduced scan times by an average of 41%, while upholding image quality metrics and diagnostic accuracy. This reduction in scan duration is advantageous in clinical settings, enhancing patient comfort, operational efficiency, and potentially economic gains. Moreover, shorter scan times minimize motion-related artifacts, particularly beneficial for patients experiencing knee pain who may find it challenging to remain motionless during lengthy scans.

Nevertheless, there is a trade-off to consider. While motion artifacts are mitigated by accelerated scan times, parallel imaging artifacts persist and may become more noticeable. This underscores the importance of optimizing acquisition parameters further, possibly by integrating compressed sensing with parallel imaging, a strategy not explored in this study.

In terms of diagnostic accuracy, the study maintained overall precision for identifying meniscal and ligamentous lesions; however, discrepancies were noted in cartilage-specific assessments post deep neural network reconstruction. These discrepancies suggest that noise reduction and smoothing effects could obscure subtle cartilage defects or amplify signal intensity in fluid-sensitive sequences. Hence, fine-tuning denoising parameters, especially for delicate tissues like articular cartilage, is crucial.

Several study limitations warrant acknowledgment. Firstly, this was a single-center study with a modest sample size, potentially limiting generalizability. Secondly, although the study involved multiple MRI systems from different vendors, the acquisition parameters were not entirely standardized, mirroring real-world variability but introducing inter-scanner differences. Thirdly, the reference standard relied on expert radiological consensus and clinical data rather than systematic arthroscopic validation, impacting diagnostic certainty. Lastly, a direct comparison between conventional and deep learning-enhanced images by paired readers was not conducted, which could have provided a more robust evaluation of performance equivalence.

In conclusion, this study offers valuable insights supporting the application of deep neural network reconstruction for accelerated knee MRI, showcasing sustained clinical efficacy with notable workflow enhancements while delineating the technological and methodological constraints guiding its application.

Article 3: Deep Learning-Based High-Resolution Magnetic Resonance Angiography (MRA) Generation Model for 4D Time-Resolved Angiography with Interleaved Stochastic Trajectories (TWIST) MRA in Fast Stroke Imaging

This research project compared time-of-flight MR angiography (TOF-MRA) with four-dimensional time-resolved contrast-enhanced MR angiography obtained through the 4D-TWIST-MRA technique to examine acute ischemic stroke and cervico-cranial arterial stenosis. The main aim of the investigation was to evaluate the effect of a reconstruction software based on deep neural networks (SwiftMR) on the quality of 4D-TWIST-MRA images. Specifically, the study contrasted images reconstructed using deep learning with traditional TOF-MRA, which is considered the standard in clinical practice because of its excellent spatial resolution and signal strength (Powers et al. 2019). The assessment consisted of both subjective and objective evaluations of image quality, along with an examination of the accuracy of the reconstructed images in urgent clinical situations requiring prompt and confident decision-making.

Understanding the practical significance of the imaging methods in the comparator and intervention groups requires clarification on the role of time-of-flight MR angiography (TOF-MRA) in current clinical practice. TOF-MRA is widely recognized as the standard for assessing large vessel blockages (LVO) and is pivotal in deciding whether endovascular thrombectomy is needed for patients undergoing cranio-cervical artery evaluations (Schellinger et al. 1999). Its reputation as the preferred method is mainly due to its excellent resolution and signal quality, enabling precise imaging of intracranial arteries. Nevertheless, TOF-MRA has drawbacks, such as lengthy scan times and limited imaging range, making it less suitable for urgent situations like acute ischemic stroke management.

To address these drawbacks, an alternative approach is the use of four-dimensional time-resolved contrast-enhanced MR angiography with the Time-resolved angiography with interleaved stochastic trajectories technique (4D-TWIST-MRA). By leveraging improvements in coil technology and k-space sampling, this method offers quicker scans and broader imaging coverage compared to traditional contrast-enhanced MRA. Moreover, its dynamic nature allows for capturing multiple stages of arterial flow. However, these advantages may result in a trade-off with lower resolution and signal quality compared to TOF-MRA, potentially impacting diagnostic reliability, especially for smaller blood vessels and subtle abnormalities.

In this clinical setting, employing deep neural network algorithms for enhancing image quality shows promise in overcoming the inherent limitations of 4D-TWIST-MRA. By enhancing images without changing scan protocols, AI-based reconstruction

has the potential to boost diagnostic accuracy and clinical utility in scenarios where swift and dependable decision-making is critical.

PICO

Table 5.5: PICO for the paper (Kim et al. 2024)

P	I	C	O
520 patients for ischemic stroke or intracranial vessel stenosis evaluation	4D-TWIST-MRA reconstructed with SwiftMR based AI-software	4D-TWIST-MRA conventional clinical acquisition and TOF-MRA	Image quality: - Quantitative (SNR) - Qualitative (5-point Likert scale) Diagnostic performance (Ischemic acute stroke) - Diagnostic confidence level - Time for diagnostic evaluation

Study design

The study was structured as a retrospective review involving a total of 520 patients. Among the cases, a combination of three imaging techniques was examined in 76.3% (n = 397): time-of-flight MR angiography (TOF-MRA), four-dimensional time-resolved contrast-enhanced MR angiography captured using the Time-resolved angiography with interleaved stochastic trajectories method (4D-TWIST-MRA), and deep learning-enhanced 4D-TWIST-MRA produced by an AI-based reconstruction tool (SwiftMR). The primary aim of this initial patient group was to assess the enhancement in image quality facilitated by AI-driven image reconstruction.

An evaluation of image quality was carried out through quantitative analysis by determining the signal-to-noise ratio (SNR) on Maximum Intensity Projection (MIP) images from the M1, M2, and M3 segments of the middle cerebral artery, as well as the basilar artery. These SNR values were then compared among the three imaging techniques. Moreover, a qualitative appraisal of image quality was conducted using a five-point Likert scale by two veteran neuroradiologists with 11 and 13 years of

experience, respectively. The assessment criteria covered overall image quality, noise reduction, sharpness, vascular clarity, and venous interference level.

Additionally, to gauge the potential enhancement in diagnostic accuracy with deep learning-driven reconstruction compared to the established clinical benchmark, aneurysm detection was examined. The rate of aneurysm identification was computed using TOF-MRA as the benchmark, and for aneurysms detected using all three imaging methods, the maximum aneurysm size was compared.

Concurrently, the second phase of the research concentrated on the remaining 23.7% of the patient pool ($n = 123$), which comprised individuals assessed retrospectively in an acute ischemic stroke scenario. In this subset, solely contrast-enhanced 4D-TWIST-MRA images were accessible and were juxtaposed with their deep learning-enhanced versions. The evaluation in this context centered on diagnostic efficacy parameters, encompassing confidence in diagnosing large vessel occlusion and the time required for decision-making, both evaluated using a five-point Likert scale.

This subsequent investigation was specifically tailored to explore how enhancements in image quality translated into clinically relevant advantages in time-sensitive conditions, thereby scrutinizing the practical clinical applicability of the AI-driven reconstruction tool. The statistical analysis encompassed various methods such as Student's t-test for normally distributed continuous variables, Mann-Whitney U test for non-normally distributed continuous variables, chi-square test or Fisher's exact test for categorical variables, and analysis of variance (ANOVA) for comparing image quality metrics and SNR among the three imaging techniques.

Main results

A study focusing on the comparison of image quality revealed a notable enhancement in signal-to-noise ratio (SNR) with the use of deep learning-enhanced reconstruction for 4D-TWIST-MRA in comparison to the traditional method. This initial quantitative discovery highlighted the effective enhancement of image signal characteristics through deep neural network reconstruction.

Furthermore, an evaluation of image quality using a five-point Likert scale showcased marked enhancements in various aspects when comparing deep learning-reconstructed 4D-TWIST-MRA with the conventional approach. Notably, attributes such as overall image quality, noise reduction, sharpness, and vascular visibility received higher ratings post deep learning reconstruction. Conversely, there was no significant difference in venous contamination, indicating that the enhancement was related to image quality rather than noise reduction during reconstruction.

Regarding aneurysm evaluation, while there wasn't a statistically significant improvement in aneurysm detection rate between the two methods, measurements of

aneurysm size from deep learning–reconstructed images closely matched those from TOF-MRA. Unlike the conventional method, which tended to underestimate aneurysm size, the deep learning reconstruction provided measurements more aligned with the clinical reference standard.

Direct comparison between deep learning–reconstructed 4D-TWIST-MRA and TOF-MRA revealed that image quality parameters such as SNR and qualitative Likert-scale ratings were still lower in the former across all aspects. Nevertheless, the utilization of deep learning narrowed the performance gap, bringing 4D-TWIST-MRA closer to TOF-MRA in terms of image quality.

Lastly, in the assessment of acute ischemic stroke cases, a notable enhancement in diagnostic confidence for large vessel occlusion was observed using deep learning–reconstructed images, as indicated by Likert scale assessments. Additionally, a decrease in diagnostic reading time was noted, underscoring the clinical benefits of improved image quality through deep learning reconstruction in critical decision-making scenarios.

Figures and qualitative results

The diagram in Figure 5.8 demonstrates a direct visual comparison of how aneurysms are shown in the three types of imaging methods studied in the initial part of the research, which focused on assessing image quality. Panel A displays traditional 4D-TWIST-MRA, panel B exhibits 4D-TWIST-MRA reconstructed using deep learning, and panel C presents time-of-flight MR angiography (TOF-MRA). The illustration reveals a gradual enhancement in the depiction of aneurysms among the three techniques, with more distinct outlining and an apparent increase in aneurysm size observed from traditional 4D-TWIST-MRA to deep learning–enhanced 4D-TWIST-MRA, and finally to TOF-MRA. While no notable variations in aneurysm detection rates were detected between the methods, there was a progressive growth in the measured aneurysm size, indicating enhanced precision in size estimation with deep learning reconstruction and closer agreement with the TOF-MRA benchmark.

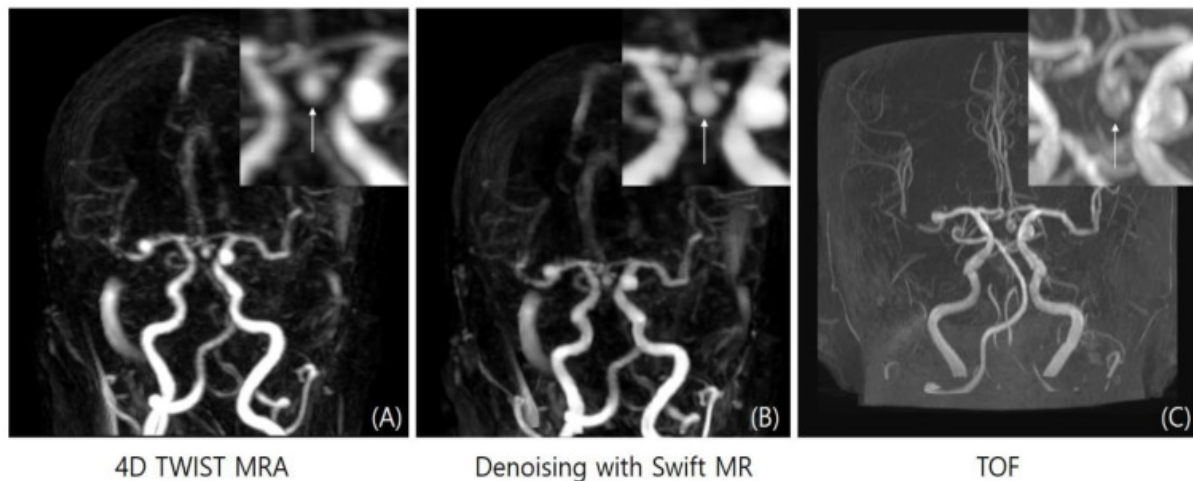


Figure 5.5: Aneurism visualization in the three image modalities. Image reproduced from: (Kim et al. 2024).

Comparison of the three imaging methods using Maximum Intensity Projection (MIP) images is illustrated in Figure 5.9. Panels A and B depict conventional 4D-TWIST-MRA and deep learning-enhanced 4D-TWIST-MRA respectively, showing prominent signal intensity in the bilateral sigmoid sinuses, highlighted by arrows. Conversely, panel C displaying the TOF-MRA image lacks this distinctive feature. The presence of elevated venous signal in panels A and B is attributed to venous contamination associated with contrast-enhanced time-resolved MRA, elucidating the absence of notable enhancement in the venous contamination parameter as per the five-point Likert scale image quality evaluation. However, qualitative assessments revealed substantial enhancements in background noise reduction, sharpness, and vascular conspicuity with deep learning reconstruction. This enhancement in overall image clarity underscores the effectiveness of deep neural network-based reconstruction in preserving authentic venous signal while enhancing image quality aspects.

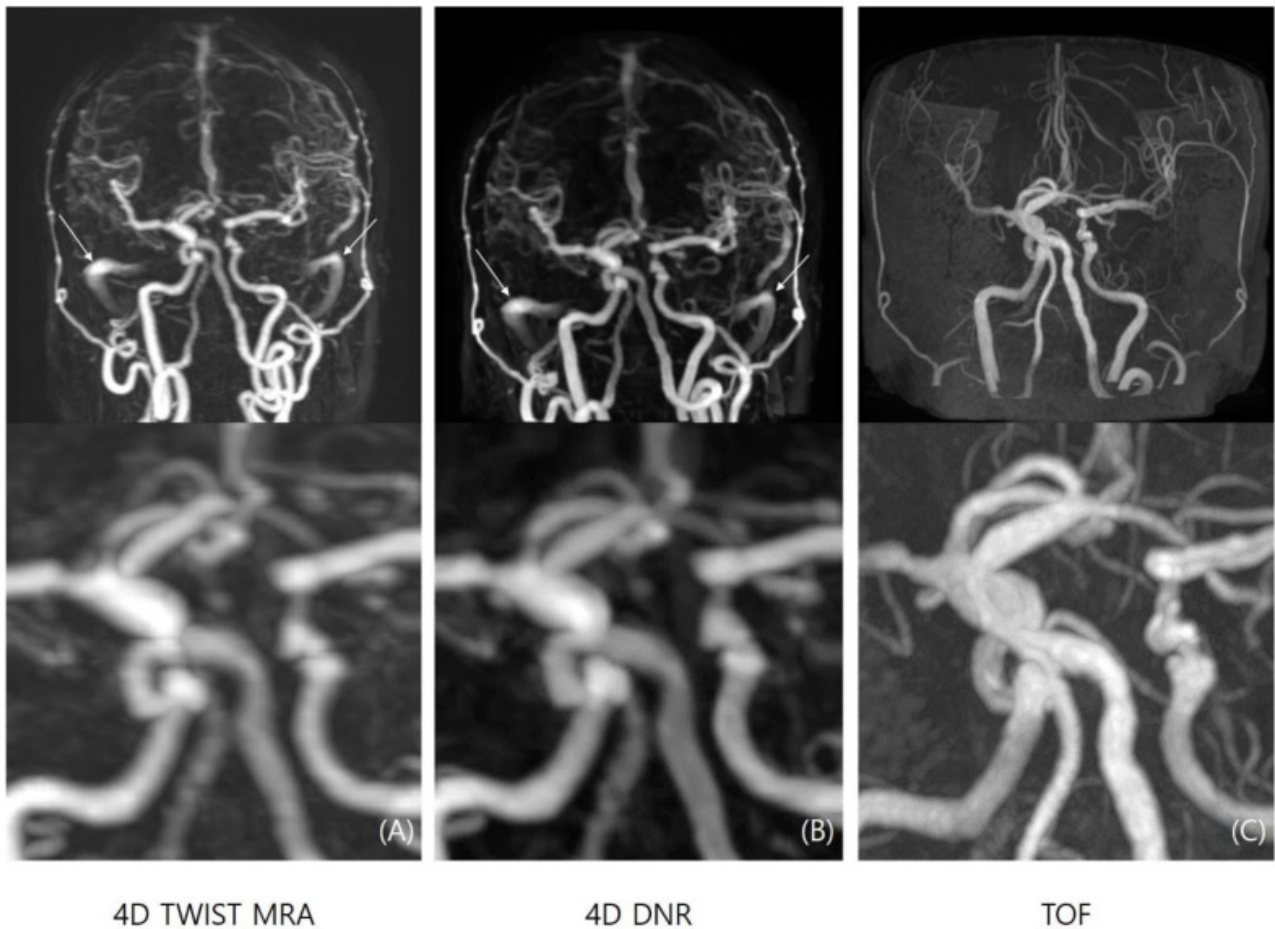


Figure 5.6: Maximum intensity projection of the three image modalities. Image reproduced from: (Kim et al. 2024).

Clinical relevance and limitations

In summary, SwiftMR consistently improved the overall image quality of 4D time-resolved contrast-enhanced MR angiography. The deep learning–reconstructed 4D-TWIST-MRA exhibited a higher signal-to-noise ratio and better qualitative image quality compared to the conventional 4D-TWIST-MRA. This approach also narrowed the performance difference with time-of-flight MR angiography, which is the clinical reference standard due to its superior spatial resolution and signal-to-noise ratio. Nevertheless, time-of-flight MR angiography has limitations such as longer acquisition times and a smaller field of view, making it less suitable for urgent clinical situations.

In this context, the deep learning–enhanced 4D-TWIST-MRA is valuable as it combines enhanced image quality with the inherent advantages of time-resolved contrast-enhanced MRA, including faster acquisition, wider field of view, and dynamic visualization of multiple arterial phases. These attributes are particularly crucial in

acute ischemic stroke imaging where quick and accurate vascular assessment is critical.

Despite its advantages, there are some limitations to consider. For instance, venous contamination persists after deep learning reconstruction, especially in later arterial phases and in MIP images. This issue is intrinsic to contrast-enhanced time-resolved MRA and cannot be solely corrected through post-processing, as venous signal reflects genuine physiological enhancement rather than noise from reconstruction. The deep neural network used was trained to enhance denoising and resolution while maintaining data fidelity, thus not altering the fundamental physical acquisition process. Though this limitation impacts venous contamination, it can be viewed as a strength in terms of safety and reliability by reducing the risk of artificial signal removal or distortion.

Another limitation relates to aneurysm detection. Despite improving the accuracy of aneurysm size measurement, there was no significant enhancement in aneurysm detection rate compared to conventional 4D-TWIST-MRA. Notably, missed aneurysms were mostly small lesions with a diameter ≤ 2 mm. This suggests that while deep learning-enhanced 4D-TWIST-MRA may be beneficial for evaluating and measuring aneurysms above a certain size threshold, its sensitivity to very small aneurysms is limited.

Further limitations of the study include its retrospective nature and single-center scope, potentially limiting the generalizability of the findings. Additionally, digital subtraction angiography was not used as the reference standard, and image quality evaluation primarily relied on MIP images rather than cross-sectional assessment.

In conclusion, this study underscores that deep learning-based reconstruction significantly boosts the image quality and clinical applicability of 4D-TWIST-MRA, especially in urgent scenarios, while upholding the inherent advantages and constraints of the original acquisition method.

Article 4: Improving Diagnostic Performance of MRI for Temporal Lobe Epilepsy With Deep Learning-Based Image Reconstruction in Patients With Suspected Focal Epilepsy

The primary focus of this research is to tackle a significant obstacle in diagnosing temporal lobe epilepsy (TLE) by employing an advanced image reconstruction tool called SwiftMR, based on deep learning technology. Traditional diagnosis of TLE relies on spotting irregularities in the hippocampus and/or temporal lobe on magnetic resonance imaging scans. Yet, these irregularities are often subtle, intricate, and small, causing a considerable portion of patients, around 20–40%, to be labeled as having negative MRI results during routine clinical assessments (Cascino et al. 1991).

As outlined in the Harmonized Neuroimaging of Epilepsy Structural Sequences (HARNESS-MRI) guidelines, it is recommended to use high-resolution imaging with

turbo spin echo sequences featuring slice thicknesses of 2 mm or less to enhance the visualization of the internal structure of the hippocampus (Bernasconi et al. 2019).

However, thinner slices typically lead to a decline in signal clarity and overall image sharpness, potentially limiting the effectiveness of these precise acquisitions. This study seeks to explore whether leveraging SwiftMR, a deep learning-based image reconstruction tool, can address these challenges by elevating image quality and subsequently enhancing the diagnostic accuracy of MRI scans for individuals with temporal lobe epilepsy.

PICO

Table 5.6: PICO of the study: (Suh et al. 2024)

P	I	C	O
117 patients with suspect of TLE (34 with pathology and 83 without)	1.5 mm slice thickness DLR oblique coronal turbo spin echoes sequence (TSE)	• 1.5 mm slice thickness oblique coronal TSE • 3 mm slice thickness oblique coronal TSE (standard protocol)	Image quality: • Quantitative (SNR, CNR) • Qualitative (3-point scale of lesion findings) Diagnostic performance • Accuracy, sensitivity and specificity (on diagnosis of TLE) • AUC and ROC curves (on lesion of hippocampus and temporal lobe)

Study design

This research was conducted retrospectively at a single center and involved applying specific criteria for including and excluding patients. Individuals with a history of previous surgeries or lacking high-resolution 1.5-mm slice-thickness images were not considered. Following these criteria, a total of 117 patients were selected for the study.

All participants underwent oblique coronal two-dimensional turbo spin echo T2-weighted imaging using three distinct protocols: (1) 1.5-mm slice thickness reconstructed through a deep learning-based method (referred to as SwiftMR), considered as the intervention; (2) 1.5-mm slice thickness without deep learning reconstruction; and (3) standard MRI with 3-mm slice thickness using identical sequence parameters, serving as the comparator protocol.

The study outcomes were categorized into two sections: diagnostic performance and image quality assessments. Diagnostic performance encompassed metrics such as accuracy, sensitivity, specificity, and detection sensitivity specifically for patients with temporal lobe epilepsy. The agreement between different readers was evaluated using Cohen’s kappa statistics. Moreover, receiver operating characteristic (ROC) curves were utilized, and the corresponding areas under the curve (AUC) were calculated based on lesion detection scores to assess the discriminatory capability of each MRI protocol for identifying hippocampal and temporal lobe lesions separately.

Image quality evaluation involved both quantitative and qualitative analyses. Quantitative assessment included determining the signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR). Qualitative image quality was judged using a three-point scale to assess overall subjective image quality, presence of flow artifacts, motion artifacts, image sharpness, and structural clarity.

For diagnostic performance evaluations, the reference standard was set by two seasoned neurologists with 20 and 35 years of clinical practice, respectively. The final diagnosis and classification of epileptogenic conditions were established through a thorough review of clinical records, laboratory results, electroencephalography findings, surgical reports when accessible, and MRI results. The interpretation of images and qualitative analysis was carried out independently by two neuroradiologists, each with a two-year background, who were blinded to any clinical details.

The statistical analysis is summarized in the tables 5.3 and 5.4. Two separate tables are presented, one for diagnostic performance and one for image quality. Each table also outlines the rationale for the statistical tests employed, which is essential for the correct interpretation of the results presented in the subsequent section.

Table 5.7: Statistical analysis for diagnostic performance

Statistical Analysis	Rationale
Logistic regression with generalized estimation equation	To compare diagnostic performance across the two readers among MRI protocols considering the correlation of the observation of the patients

McNemar’s test	Compare two protocols in each reader
ROC and AUC	For lesions finding scores for Hippocampus and temporal lobe
Cohen’s Kappa statistic	Inter reader agreement

Table 5.8: Statistical analysis for image quality assessment

Statistical Analysis	Rationale
Friedman test	For lesion findings and quality score images
Paired sampled t-tests	SNR and CNR

Main results

The research findings reveal that utilizing deep learning–based reconstruction for 1.5-mm slice-thickness fast spin echo imaging leads to a notable enhancement in sensitivity when contrasted with the standard 3-mm slice-thickness MRI approach. Despite this heightened sensitivity, the overall accuracy of diagnosis remains similar across the two methodologies, while specificity does not exhibit a noteworthy enhancement in comparison to the standard procedure.

Upon comparing the 1.5-mm MRI reconstructed using deep learning to the non-deep learning reconstructed 1.5-mm MRI, a consistent occurrence of false-positive results emerged. Specifically, both readers frequently detected 12 false-positive instances in the deep learning–reconstructed images, in contrast to 19 instances in the non-deep learning reconstructed images. The primary reasons behind these false-positive readings encompassed: (1) perceived absence of internal hippocampal structure on a lone slice despite appearing normal on adjacent slices; (2) anatomical diversities in the hippocampus; (3) restricted assessment of the hippocampus due to artifacts; and (4) radiological irregularities, such as enlarged perivascular spaces, which were not clinically diagnosed as temporal lobe epilepsy.

Regarding the quality of images, the adoption of deep learning–based reconstruction yielded significant enhancements in both signal-to-noise ratio and contrast-to-noise ratio when juxtaposed with the respective comparator methodologies.

Figures and qualitative results

In the following section, two illustrative cases are presented to exemplify the main findings described above. The first case highlights the improvement in diagnostic sensitivity achieved with deep learning-based reconstruction, with particular emphasis on the increased detection sensitivity observed in the subgroup of patients with temporal lobe epilepsy. The second case illustrates the occurrence of false-positive findings associated with deep learning reconstruction, providing insight into the potential sources of reduced specificity.

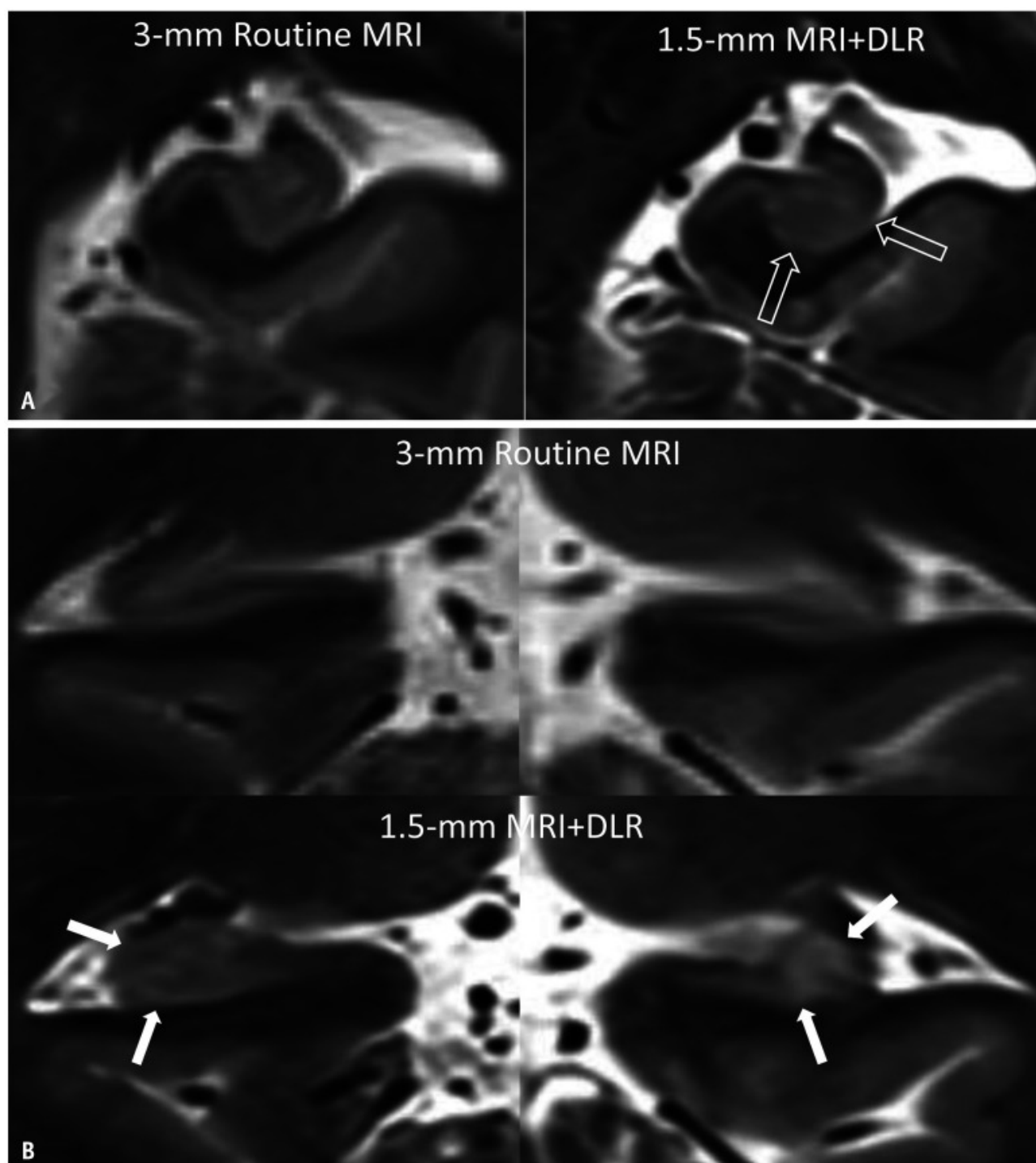


Figure 5.7: increasing in sensitivity detection in 1.5 mm slice thickness sequence

In Figure 5.7 (A), there is a scenario where the blurring of the hippocampal striation in the left hippocampus is faintly noticeable on a regular MRI scan, making it somewhat suspicious but not definitively diagnostic. Conversely, the same issue is distinctly highlighted on the reconstructed 1.5-mm MRI through deep learning, showcasing a clearer depiction of the loss of internal hippocampal striation.

Moving on to Figure 5.7 (B), the standard MRI scan displays a normal hippocampal volume, signal intensity, and intact internal structure of both hippocampal tails. However, upon examining the 1.5-mm MRI with deep learning reconstruction, there is a noticeable increase in signal intensity and a blurred hippocampal striation seen on both sides, indicating abnormalities that were not initially visible on the standard imaging procedure.

These instances exemplify how the application of deep learning reconstruction improves the identification of subtle hippocampal irregularities by enhancing image quality and structural visibility. Consequently, lesions that might be ambiguous or overlooked on regular MRI scans are more easily identifiable, thereby boosting the sensitivity in detecting hippocampal sclerosis in individuals with temporal lobe epilepsy.

Illustrated in the Figure 5.7 below are four instances illustrating false-positive results linked to the application of deep learning in image reconstruction. As discussed earlier, these scenarios showcase the key issues contributing to the decreased accuracy noted in the 1.5-mm MRI reconstructed using deep learning.

In the first scenario, denoted as Case A, a slightly blurred depiction of the internal structure of a normal-sized and shaped left hippocampus is observed in the 1.5-mm MRI reconstructed using deep learning. This subtle blurring, confined to one slice only, might give the impression of a loss of internal hippocampal structure, potentially leading to a misdiagnosis. Yet, the regular MRI shows the intact internal structure, indicating an incorrect positive identification.

Case B presents a similar occurrence involving a hippocampus with an incomplete inversion. The deep learning-reconstructed image displays a blurred internal structure, while the standard MRI exhibits the normal architecture of the hippocampus. The patient's electroencephalography results were normal, supporting the notion that the perceived anomaly is likely an anatomical variation rather than an actual pathology.

In Case C, the false-positive detection is primarily linked to image deterioration due to motion artifacts. Despite the enhancement in noise characteristics through deep learning reconstruction, the persistence of motion artifacts may create an illusion of internal hippocampal structure loss, leading to a less confident image interpretation. Conversely, the artifact-free standard MRI showcases the preserved hippocampal architecture.

Lastly, Case D showcases a radiological irregularity unrelated to temporal lobe epilepsy. A notable perivascular space was misinterpreted as encephalomalacic changes in both the deep learning-reconstructed and standard MRI images. The absence of abnormal electroencephalogram findings confirms that this radiological feature does not correspond to an epileptogenic lesion, underscoring the significance of clinical correlation in image analysis.

These examples underscore that false-positive results associated with deep learning-based reconstruction primarily stem from single-slice internal hippocampal structure loss, anatomical variances, imaging artifacts, or non-epileptogenic radiological anomalies. These factors elucidate the decline in specificity despite the general enhancement in detection sensitivity.

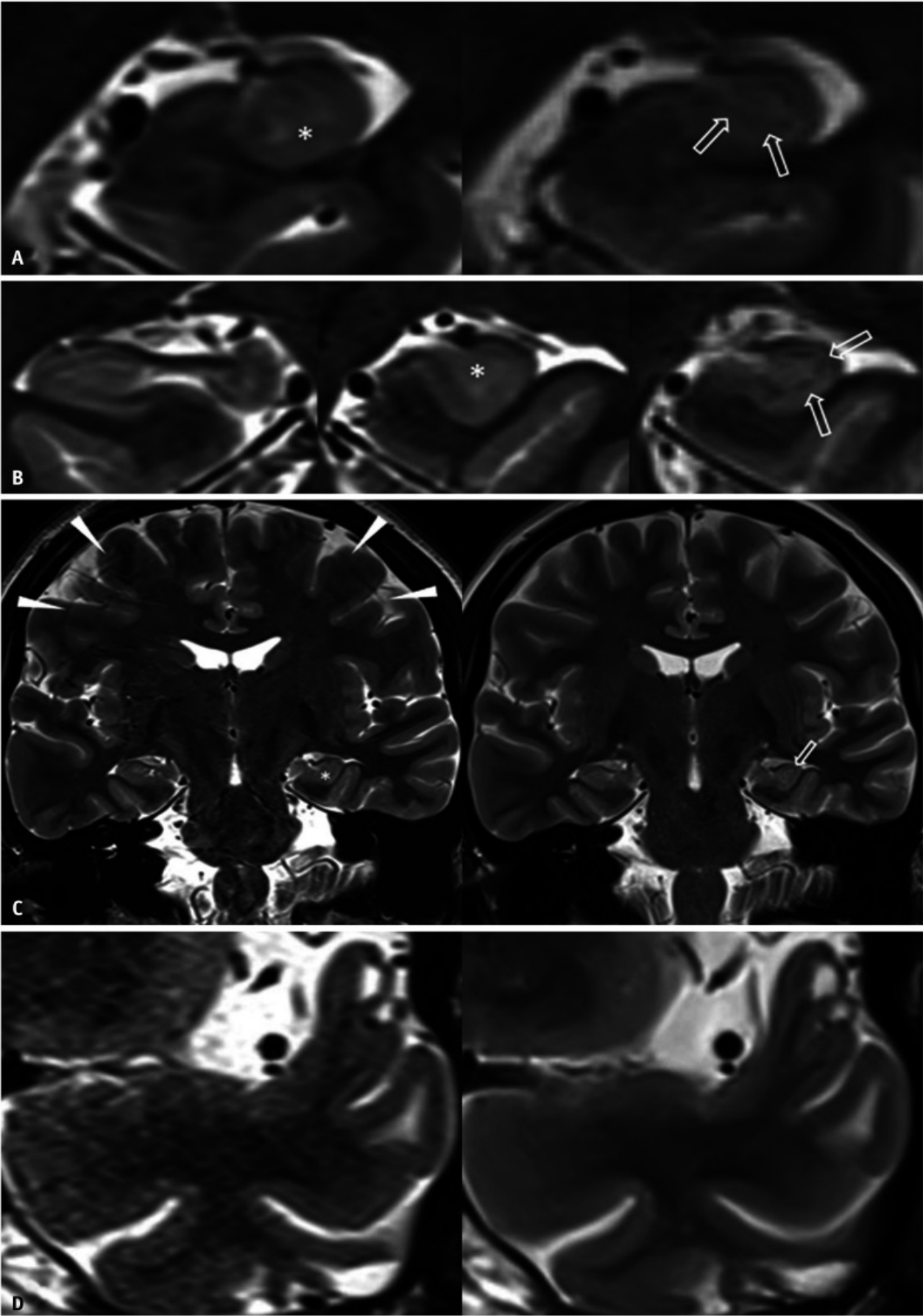


Figure 5.8: False positive example induced by DLR algorithm

Clinical relevance and limitations

Utilizing deep learning technology for image enhancement has shown a notable enhancement in image quality pertinent to clinical settings. This enhancement is evidenced through the enhancement of signal-to-noise ratio, contrast-to-noise ratio, and effective spatial resolution. By addressing the compromise between reduced slice thickness and image quality, deep learning reconstruction facilitates the utilization of thin-slice MRI scans without the customary decline in image clarity. This advancement in technology is particularly advantageous in identifying subtle abnormalities in the hippocampus and temporal lobe that are often intricate, diminutive, and challenging to detect through traditional MRI scans. Consequently, this enhancement leads to a more accurate detection of temporal lobe epilepsy, notably hippocampal sclerosis. The improved visualization of lesions could potentially expedite and strengthen the diagnosis of temporal lobe epilepsy, thereby potentially decreasing the number of cases where MRI scans fail to detect abnormalities. Notably, the enhanced non-invasive lesion identification could reduce the necessity for invasive diagnostic procedures like intracranial electrode placement or surgical exploration by offering clearer localization of the epileptogenic focus during the preoperative assessment.

Despite the benefits mentioned, the improved sensitivity also brings about notable drawbacks in a clinical context. When utilizing thin-slice imaging in conjunction with deep learning reconstruction, there is a possibility of identifying anomalies that are visible on one slice but not consistently present on neighboring slices. This occurrence, in combination with variations in hippocampal anatomy and lingering imaging imperfections, can lead to inaccurate positive diagnoses. Noteworthy recurring false-positive results include the loss of internal hippocampal structure on a single slice, anatomical variations, restricted assessment due to motion or flow artifacts, and radiological irregularities unrelated to epilepsy. These observations underscore the importance of exercising caution when interpreting images reconstructed using deep learning and underscore that such reconstruction methods should complement rather than substitute thorough clinical and electroencephalographic evaluations.

It is important to recognize several constraints in this research. Firstly, the study's reliance on a retrospective approach and data from a single institution may restrict the applicability of the findings. Additionally, a small subset of patients underwent surgery and had histopathological confirmation, leading to the reference standard being established through clinic radiological consensus rather than consistent pathological verification. Furthermore, while thin-slice imaging enhances lesion identification, the influence of deep learning reconstruction on diagnostic accuracy necessitates further exploration through broader, planned, multi-site investigations. Subsequent studies should emphasize prospective validation, standardized interpretation guidelines, and evaluation of clinical outcomes to more clearly delineate

the utility of deep learning–based reconstruction in standard epilepsy imaging protocols.

Article 5: Deep learning–based reconstruction for acceleration of lumbar spine MRI: a prospective comparison with standard MRI

In the study, a deep learning–based reconstruction (DLR) software was applied to enhance the image quality of turbo spin-echo (TSE) T1-weighted and T2-weighted sequences in a lumbar spine MRI protocol. The fast acquisition protocol, combining parallel imaging and compressed sensing, resulted in reduced image quality prior to DL reconstruction. The aim of the study was to demonstrate that a reduction in acquisition time of approximately 32.3% compared with the standard protocol could still provide non-inferior qualitative and quantitative image quality for anatomically relevant structures in degenerative lumbar spine disease, as well as non-inferior diagnostic performance.

PICO

Table 5.9: PICO of (Yoo et al. 2023)

P	I	C	O
50 patients with 6 degenerative lumbar spine disease: • Disc abnormalities • Annular fissure • Central canal stenosis • Spinal instability • Neural foramina stenosis • Schmorl’s node	TSE T1WI and T2WI fast acquisition DLR with two values of denoising level	TSE T1WI and T2WI lumbar spine standard protocol	Image quality: • Quantitative (SNR, CNR) • Qualitative (5-point Likert scale of eight anatomical structure of lumbar spine MRI, overall image impression, presence of artifacts) Diagnostic performance • Sensitivity and specificity of degenerative

	lumbar spine diseases
--	--------------------------

Study design

The study was prospective and single-center and involved the use of 1.5 T MRI scanners from two different vendors. Images acquired with a fast acquisition protocol were reconstructed using a deep learning-based algorithm with two different levels of denoising and were compared with images obtained using the standard lumbar spine MRI protocol.

Degenerative lumbar spine disease served as the reference standard and was established based on consensus labelling by two expert radiologists, who evaluated all available clinical and radiological information. Five intervertebral disc levels and six vertebral levels were considered per patient, resulting in a total of 250 intervertebral disc levels and 300 vertebral levels across the study population, depending on the specific pathology analysed. Image analysis was performed by four blinded readers in two separate parts.

The first part consisted of a qualitative assessment of image quality for eight anatomical structures using a five-point Likert scale, as well as evaluation of overall image quality and artifact presence. The second part assessed diagnostic performance, with sensitivity and specificity calculated with respect to the reference standard. Quantitative image analysis included evaluation of signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR), with additional subgroup analyses according to MRI vendor. Statistical analysis involved paired t-tests for quantitative measurements, Wilcoxon signed-rank tests for overall image quality, artifact presence, and anatomical structure assessment, and generalized estimating equations for sensitivity and specificity.

Main results

Regarding quantitative image quality, both deep learning-reconstructed image sets demonstrated higher signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) compared with the standard reconstruction, on both T1-weighted and T2-weighted sequences. This improvement was observed consistently across the two denoising levels applied in the DL reconstruction.

Qualitative image assessment showed similar findings. Image quality scores for the evaluated anatomical structures, assessed using a five-point Likert scale, were overall comparable or slightly higher for DL-reconstructed images than for standard images, with no relevant increase in image artifacts. With respect to diagnostic performance, no statistically significant differences were observed in sensitivity and specificity between DL-reconstructed images and standard images across the evaluated lumbar degenerative pathologies.

However, a consistent trend toward increased sensitivity and decreased specificity was noted for several conditions. This pattern was particularly evident for annular fissure detection, where DL reconstruction led to higher sensitivity at the expense of reduced specificity.

Figures and qualitative results

Figure 5.9 illustrates that image quality obtained with the accelerated protocol and DL-based reconstruction is increased or at least non-inferior compared with the standard protocol. Moreover, despite the reduction in acquisition time, diagnostic performance in terms of sensitivity and specificity is not significantly different. In particular, mild central canal stenosis at the L3–L4 level and moderate central canal stenosis at the L4–L5 level are clearly visible across all three image modalities.

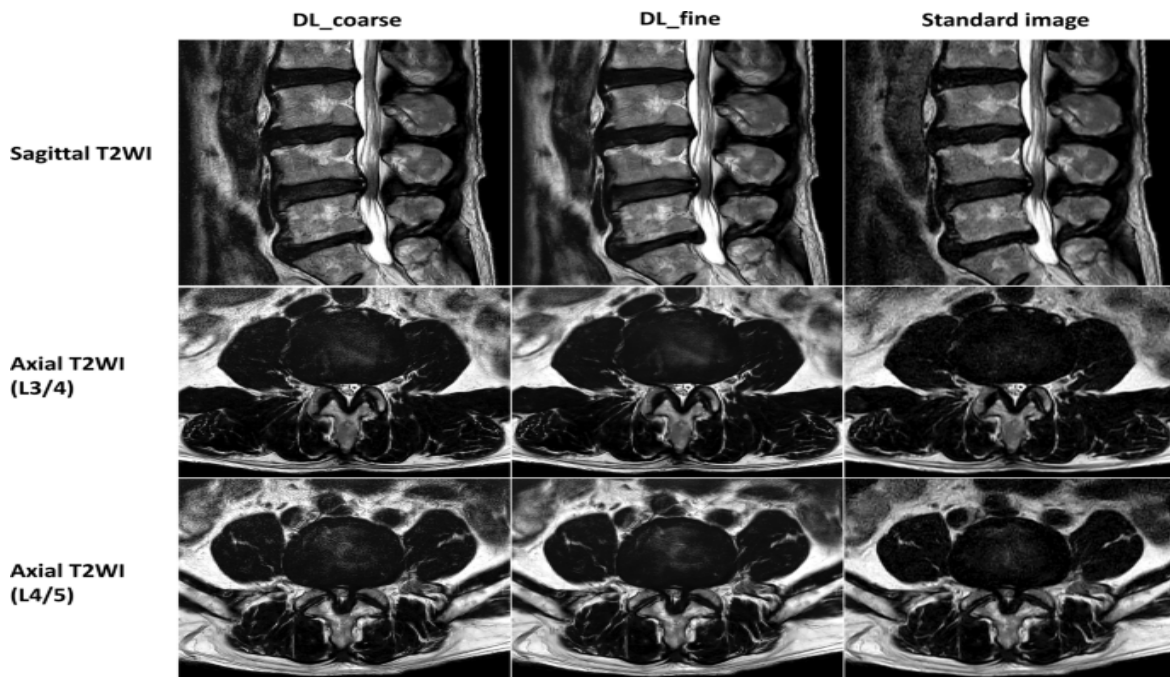


Figure 5.9: Comparison of three image modalities for assessment of mild centre canal stenosis of L3/4 and moderate at L4/5

Clinical relevance and limitations

From a clinical perspective, this study demonstrates that the use of a deep learning, based reconstruction software enables a substantial reduction in MRI acquisition time (approximately 32.3%) while maintaining non-inferior image quality and diagnostic performance compared with the standard lumbar spine MRI protocol. This effect was observed across both denoising levels provided by the software, allowing flexibility in image appearance according to radiologist preference without compromising diagnostic reliability.

The clinical implications of these findings are twofold. First, shorter acquisition times may reduce motion-related artifacts, which are particularly relevant in patients with lumbar spine degeneration who often experience pain and difficulty maintaining a stable position during scanning. Second, the reduction in scan time may improve scanner throughput, thereby increasing cost-effectiveness and operational efficiency in clinical practice. Although overall diagnostic performance was not statistically inferior to the standard protocol, a consistent trend toward increased sensitivity and reduced specificity was observed across several degenerative lumbar spine conditions.

According to the authors, this effect is likely attributable to improved image sharpness and denoising, which may enhance lesion conspicuity. However, this increased detectability may also lead to overinterpretation, resulting in a higher number of false-positive findings. This trade-off was particularly evident in the evaluation of annular fissures, highlighting the need for careful clinical interpretation and appropriate selection of denoising strength.

The strengths of this study include its prospective design and the use of MRI scanners from two different vendors, supporting the generalizability of the results and suggesting a degree of vendor independence of the DL-based reconstruction algorithm. Furthermore, the algorithm was trained on a large, multi-vendor dataset, which may contribute to its robust performance across different acquisition systems. Several limitations should be acknowledged. The reference standard was based on consensus reading by two expert neuroradiologists using all available clinical and radiological information, without systematic confirmation by an independent gold standard. Additionally, the relatively small sample size of 50 patients limits the statistical power of the study and precludes formal non-inferiority testing of diagnostic performance.

5.2. Clinical validations of other AI-based MRI reconstruction software (FastScan and Enhanced Image Quality)

Following the detailed analysis of peer-reviewed clinical validation studies specifically focused on SwiftMR, a broader literature search was conducted to identify comparable clinical studies evaluating other artificial intelligence-based MRI reconstruction software. These studies share a common clinical objective, namely the reduction of acquisition time while preserving image quality, or the improvement of image quality to enhance diagnostic performance in challenging clinical scenarios. This section is therefore based on a narrative review of the available clinical literature, presenting and comparing multiple commercial AI-based reconstruction solutions currently available on the market.

The clinical evidence supporting these technologies is analyzed in relation to the findings previously discussed for SwiftMR, with the aim of contextualizing its level of clinical maturity within the current technological landscape. Both vendor-independent, image-based post-processing solutions operating exclusively on DICOM images and vendor-dependent approaches integrated into the MRI acquisition protocol were included. The latter often rely on physics-based or k-space-based reconstruction strategies. Through this comparative analysis, the clinical advantages and limitations of these two approaches are discussed, highlighting their respective implications for diagnostic reliability, workflow integration, and clinical deployment.

The main competing deep learning-based MRI reconstruction solutions, including both image-based and k-space-based approaches, were identified based on a previously published review article (Yoo and Choi 2024), which provides an overview of currently available commercial and research-grade methods.

Table 5.10 summarizes the different AI-based MRI reconstruction products and their respective vendors, together with the corresponding references analyzed for the clinical validation of methods pursuing the same clinical goals as SwiftMR.

Table 5.10: Different DL-based MR image reconstruction

Vendor	Product	Type	Reference
Subtle Medical	SubtleMR	DICOM-based (vendor neutral)	(Rudie et al. 2022),(Suzie Bash et al. 2021).
GE Healthcare	AIR Recon DL	k-space based (vendor specific)	(Lee et al. 2021),(Suzie Bash et al. 2021),(S. Bash et al. 2021),(Kim et al. 2021),(Iwamura et al. 2024),(Kim et al. 2023)

Starting with Subtle Medical’s product, SubtleMR is a DICOM-based, vendor-neutral solution and therefore represents a direct competitor to SwiftMR. Similar to SwiftMR, it operates as a post-processing image-based software designed to perform denoising and image quality enhancement, with the dual aim of supporting accelerated MRI protocols and improving diagnostic confidence.

Its clinical performance was evaluated in a prospective brain MRI study focusing on T2-FLAIR sequences. The accelerated acquisition combined with SubtleMR

reconstruction achieved an acquisition time reduction of approximately 45% compared with the standard protocol. Quantitative analysis showed a reduction in the full width at half maximum (FWHM) of the septum pellucidum compared with the fast acquisition protocol, indicating improved spatial resolution and edge sharpness for small anatomical structures (Rudie et al. 2022). Although image quality did not fully reach that of the standard protocol, SubtleMR reconstruction substantially improved resolution relative to the fast scan alone.

(Suzie Bash et al. 2021) revealed that using a faster method led to a significant increase in the speed of collecting 3D T1-weighted brain MRI images by 60%. The measurements of volumetric biomarkers obtained through this accelerated process were found to be similar to those achieved using the conventional protocol. It was a prospective, multicenter, multi-reader study conducted across six different MRI scanners, further demonstrating the vendor-agnostic nature of the deep learning approach and its robust performance in a real-world clinical setting.

(S. Bash et al. 2021) prospective, multicenter, multi-reader clinical study demonstrates that a spine MRI protocol accelerated by approximately 40% of the acquisition time preserves diagnostic integrity and image quality compared with the standard of care. Image quality was rated as comparable or superior, with improvements in perceived signal-to-noise ratio and artifact reduction, supporting the feasibility of its implementation in routine clinical practice.

In summary, SubtleMR is an image-based, vendor-neutral MRI reconstruction software whose peer-reviewed clinical validation is primarily focused on brain imaging and spine applications. While the underlying technology has been investigated during research activities across different anatomical districts, robust clinical validation through peer-reviewed studies remains limited outside the neuroradiological domain. Therefore, the applicability of SubtleMR to other anatomical regions is currently supported mainly by technical feasibility rather than by published clinical evidence.

Moving to AIR Recon DL, a vendor-specific deep learning-based reconstruction software integrated into GE HealthCare MRI scanners, this approach represents a case-based reconstruction method. (Lee et al. 2021) compared a standard brain MRI protocol using 3-mm slice thickness contrast-enhanced T1-weighted images with an accelerated protocol employing 1-mm coronal thin-section contrast-enhanced T1-weighted imaging. The thinner slices were used to reduce voxel signal averaging while preserving diagnostic contrast. Results demonstrated that AIR Recon DL improved image quality for diagnostic purposes, enabling clearer delineation of the normal pituitary stalk and gland, with a higher perceived signal-to-noise ratio and reduced artifacts. Additionally, the improved spatial resolution led to increased detection of pituitary adenomas, resulting in higher diagnostic sensitivity compared with the standard protocol. Overall, the findings highlight the trade-off between voxel size and

spatial resolution, demonstrating that deep learning–based reconstruction can mitigate signal loss associated with thinner slices while enhancing diagnostic performance.

Moreover, another study of (Kim et al. 2021) conducted in a comparable setting using the same AIR Recon DL software evaluated 3-mm versus 1-mm contrast-enhanced T1-weighted imaging. The results demonstrated improved diagnostic performance of the 1-mm thin-section protocol for the identification of residual tumor and cavernous sinus invasion. In particular, detection of cavernous sinus invasion showed higher diagnostic accuracy, with an overall increase in sensitivity compared with the standard 3-mm protocol.

(Iwamura et al. 2024), based on the thin-slice brain imaging hypothesis, compared conventional 5-mm 2D T2-weighted images with 1-mm 2D T2-weighted images reconstructed using the same deep learning–based reconstruction software, AIR Recon DL. In patients with multiple sclerosis, the thin-slice protocol demonstrated a significantly higher lesion detection rate compared with the standard 5-mm images. Notably, differences in lesion detection were particularly evident in the brainstem, where the higher spatial resolution enabled improved visualization of small lesions.

Finally, the application of this commercial deep learning–based reconstruction software has also been investigated in pediatric brain imaging. In a cohort of 107 pediatric patients undergoing conventional T2-weighted brain MRI, the study of (Kim et al. 2023) demonstrated a significant improvement in overall image quality, with reduced noise and enhanced gray–white matter sharpness. In addition, truncation artifacts were reduced, and lesion conspicuity was improved, with clearer visualization reported in 36 lesions across the study population.

Overall, the analysis of the selected deep learning–based MRI reconstruction software provides a comprehensive overview of the current state of clinical validation over the last five years. Image-based, DICOM-compatible solutions such as SwiftMR and SubtleMR share key technological characteristics, including vendor independence and post-processing deployment, allowing flexible integration into existing clinical workflows. These solutions can be applied either to enhance image quality while preserving standard acquisition protocols or to support accelerated imaging strategies when combined with parallel imaging and compressed sensing, while maintaining diagnostic performance compared with the standard of care. This approach is particularly relevant in high-volume clinical settings, such as musculoskeletal imaging, where multiple acquisitions are often required and image quality directly impacts diagnostic reliability. In contrast, vendor-integrated deep learning solutions based on k-space reconstruction, such as AIR Recon DL, represent a case-based approach tightly coupled to the scanner hardware and acquisition pipeline. These systems intervene directly in the acceleration process and are inherently vendor-dependent, offering optimized performance within specific scanner ecosystems.

Together, these findings demonstrate that deep learning has become a clinically validated technological enabler for MRI acceleration and image quality improvement, with two distinct and viable technological pathways currently available to healthcare institutions.

5.3. Clinical Effectiveness of AI-Based MRI Acceleration Techniques: Image-Based and Physics-Based Approaches

Building on the clinical evidence presented in the previous sections, the following analysis aims to comparatively evaluate the clinical effectiveness of deep learning-based MRI reconstruction technologies within an HTA framework. This section synthesizes the scientific evidence collected across image-based, vendor-independent solutions and vendor-integrated, k-space-based approaches, with the objective of identifying the most clinically relevant outcomes and technological trade-offs. Particular attention is given to differences in integration strategies, generalizability across scanners, impact on image quality and diagnostic performance, and suitability for routine clinical implementation. By consolidating the most significant findings from the reviewed studies, this comparative assessment provides a structured basis for concluding the clinical effectiveness domain of the HTA and supports informed decision-making regarding the adoption of deep learning-based MRI reconstruction technologies in hospital settings.

To facilitate a concise and decision-oriented synthesis of the clinical evidence, the following tables summarize the main findings of the five clinical studies evaluating SwiftMR. Given the heterogeneity of clinical indications and study objectives, the evidence was grouped according to the primary intended benefit of the deep learning reconstruction, distinguishing between image quality-driven applications at equivalent scan time and acceleration-driven applications aimed at reducing acquisition time.

The first table focuses on studies in which SwiftMR was applied to improve image quality while maintaining standard acquisition times, highlighting the clinical context, workflow integration, and the main advantages and trade-offs observed. In particular, these studies demonstrate improved signal-to-noise ratio and lesion conspicuity, often associated with increased diagnostic sensitivity, while also emphasizing potential interpretative challenges such as the amplification of sequence-related artifacts and a higher risk of false-positive findings.

The second table summarizes studies in which SwiftMR was primarily used to support scan time reduction, typically in combination with parallel imaging and compressed sensing. In these settings, the deep learning reconstruction preserved diagnostic performance while enabling faster acquisitions, although residual artifacts related to

aggressive acceleration strategies remained evident. Together, these tables provide a structured overview of the clinical benefits and limitations of SwiftMR, enabling a clear comparison across different use cases and supporting the subsequent comparative assessment with alternative deep learning reconstruction approaches.

Table 5.11: Main benefit and limitation of the SwiftMR clinical study validation in the enhancement of image quality

Study	Clinical context	Workflow integration	Main benefit	Main limitation
(Oh et al. 2025)	Detection of brain metastases using contrast-enhanced T1-weighted black-blood imaging with thin-slice acquisition, reconstructed with deep learning-based image reconstruction to improve lesion conspicuity and sensitivity compared with standard brain MRI protocols.	The deep learning reconstruction is integrated into the standard brain MRI workflow as a post-processing step, without modifying acquisition protocols or clinical decision pathways.	Increased signal-to-noise and contrast-to-noise ratios, with improved subjective image quality, associated with significantly higher sensitivity for the detection of small brain metastases (<5 mm).	Increased false-positive findings due to enhanced sharpness of unsuppressed vascular signal following deep learning reconstruction, potentially complicating differentiation between small vessels and metastatic lesions.
(Kim et al. 2024)	Comparative evaluation of 4D TWIST MR angiography reconstructed with DLR and Time-of-Flight MRA for cerebrovascular	In the acute ischemic stroke workflow, deep learning-reconstructed 4D TWIST MR angiography is integrated as	Deep learning reconstruction significantly improves image quality compared with conventional 4D TWIST reconstruction,	Residual venous contamination related to the dynamic acquisition strategy persists despite deep learning reconstruction,

	assessment, focusing on arterial visualization for acute and pre-surgical evaluation.4D TWIST provides extended field-of-view and dynamic arterial phase information but at the expense of spatial resolution, motivating the application of deep learning-based reconstruction to improve image quality.	the primary vascular imaging sequence, providing image quality comparable to Time-of-Flight MRA while avoiding additional acquisitions in time-critical settings.	achieving image quality closer to, although not fully equivalent to, Time-of-Flight MRA. Despite this, diagnostic confidence and reading efficiency are improved, with reduced interpretation time compared with non-deep learning-reconstructed 4D TWIST imaging.	and sensitivity remains lower for the detection of very small aneurysms compared with Time-of-Flight MRA.
(Suh et al. 2024)	Detection of temporal lobe epilepsy-related hippocampal and temporal lobe lesions using oblique coronal T2-weighted turbo spin-echo imaging with 1-mm slice thickness reconstructed with deep learning,	In the standard epilepsy MRI workflow, subtle hippocampal and temporal lobe lesions may remain undetected on conventional 3-mm slice thickness imaging. The integration of thin-slice (1-mm) acquisitions	Improved image quality with increased sensitivity for the detection of hippocampal and temporal lobe lesions associated with temporal lobe epilepsy.	An increased false-positive rate, with reduced specificity, related to the detection of focal signal abnormalities visible on a single thin slice but not consistently confirmed across contiguous slices, potentially leading to overinterpretation of hippocampal lesions.

compared with standard 3-mm slice thickness protocols.	reconstructed with deep learning is incorporated into dedicated epilepsy protocols to enhance lesion visualization without altering the overall clinical workflow.
--	--

Table 5.12: Main benefit and limitation of the SwiftMR clinical study validation in fast scan protocol

Study	Clinical context	Workflow integration	Main benefit	Main limitation
(Lee et al. 2023)	Evaluation of knee MRI using a fast imaging protocol with approximately 41% scan time reduction, reconstructed with deep learning, compared with the standard protocol to assess image quality and diagnostic confidence for cartilage, anterior and posterior	The accelerated knee MRI protocol reconstructed with deep learning is integrated into routine clinical practice to reduce acquisition time, improving patient comfort and minimizing motion-related artifacts, while increasing scanner	Image quality is maintained or improved compared with the standard protocol, achieving comparable or superior subjective image quality and comparable diagnostic accuracy for knee structures.	Potential limitations include underestimation of cartilage abnormalities in some cases and overestimation of signal intensity related to the level of denoising applied by the deep learning reconstruction. In addition, artifacts associated with aggressive parallel imaging acceleration are not corrected by the deep learning

	cruciate ligaments, medial and lateral collateral ligaments, and meniscus.	efficiency and supporting greater cost-effectiveness at the system level.		algorithm and may persist in the reconstructed images.
(Yoo et al. 2023)	Evaluation of lumbar spine MRI using T1-weighted and T2-weighted turbo spin-echo sequences reconstructed with deep learning, combined with parallel imaging and compressed sensing, achieving an overall acquisition time reduction of approximately 32.3% compared with the standard protocol.	The deep learning–accelerated protocol is integrated into routine lumbar spine MRI as a direct replacement of the standard examination, reducing acquisition time while preserving the same imaging workflow and clinical interpretation process.	Comparable or improved subjective image quality compared with the standard protocol, with increased detection sensitivity across the evaluated lumbar degenerative conditions.	Reduced specificity with an increased false-positive rate, particularly for annular fissures, related to enhanced image sharpness and denoising effects that may lead to overinterpretation of subtle signal alterations.

Beyond the detailed evaluation of SwiftMR, additional evidence from other commercial deep learning reconstruction solutions was considered to contextualize the observed clinical findings and assess the generalizability of image-based and k-space based approaches.

The clinical evidence reported for SubtleMR in Section 5.2 supports the same image-based, post-processing deep learning approach adopted by SwiftMR. These solutions

operate as denoising and image enhancement tools that do not alter the physical acquisition process, but improve image sharpness and lesion conspicuity, resulting in increased sensitivity for lesion detection. Although some studies suggest a potential increasing in false-positive findings, no statistically significant differences were consistently demonstrated. Overall, the evidence from SubtleMR reinforces the validity of DICOM-based, vendor-independent deep learning post-processing solutions, highlighting their generalizability across different scanners and clinical environments. In addition, SubtleMR studies further support the feasibility of accelerated MRI acquisitions without sacrificing diagnostic performance, complementing the image quality improvements observed with SwiftMR.

The clinical evidence on AIR Recon DL supports the effectiveness of deep learning–based reconstruction in reducing image noise and enhancing image sharpness. This vendor-specific, case-based approach is directly integrated into the scanner reconstruction pipeline, enabling tight coupling with parallel imaging and compressed sensing acceleration strategies that are already part of the reference standard acquisition workflow. While this integration allows effective mitigation of acquisition-related noise and artifacts, the strong vendor dependency limits the generalizability and scalability of the software across heterogeneous imaging environments.

In light of the clinical evidence and the technical characteristics discussed above, a final HTA-oriented table is presented to summarize the main advantages and limitations of image-based (DICOM-based) and physics-based (k-space–integrated) deep learning MRI reconstruction approaches.

Table 5.13: HTA assessment of clinical validation of DL software

	Image-based (DICOM)	DL	Physic based (k-space)
Examples	SwiftMR, SubtleMR		AIR Recon DL
Integration	Post-processing		Scanner-integrated
Vendor neutrality	High		Low
Workflow impact	Minimal		Embedded
Artifact behaviour	Preserved		Partially mitigated
Risk of over smoothing	Possible		Lower
Clinical flexibility	High		Limited

Scalability (HTA)	High	Medium-Low
-------------------	------	------------

Overall, the available clinical evidence demonstrates that deep learning–based MRI reconstruction techniques provide clinically meaningful improvements in image quality and/or acquisition efficiency across multiple anatomical districts. Image-based, vendor-independent solutions show consistent benefits in terms of lesion conspicuity and diagnostic confidence, while preserving existing clinical workflows, whereas case-based, scanner-integrated approaches offer optimized reconstruction within specific vendor ecosystems at the expense of reduced generalizability. These findings confirm the clinical effectiveness of deep learning MRI reconstruction and provide a robust foundation for the subsequent HTA domain of economic and implementation considerations

6 Economic and Financial Impact of the Adoption of AI-based MRI Reconstruction Software

In accordance with the EUnetHTA HTA Core Model for Diagnostic Technologies (EUnetHTA 2008) the economic domain aims to assess the financial implications and resource consequences associated with the adoption of a health technology. Within this framework, the evaluation does not merely consider direct acquisition costs, but examines whether the technology generates measurable economic value in relation to its operational context.

In the case of SwiftMR, the economic assessment is intended to determine whether the adoption of deep learning-based MRI reconstruction may provide financial advantages for the healthcare facility, considering both the investment required and the potential productivity gains achievable in routine clinical practice.

This section is structured into three complementary parts. The first subsection (6.1) describes the nature of the investment and the underlying economic rationale, contextualising the relevance of operational efficiency in MRI services, particularly within the local setting of Ospedale di Erba. The second subsection (6.2) develops the economic modelling framework, defining the baseline operational assumptions, reimbursement structure, and financial indicators used to evaluate the investment within a local HTA perspective. Finally, the third subsection (6.3) presents a quantitative analysis based on three utilization scenarios (low, medium, and high), performing a sensitivity analysis of the projected economic outcomes through ROI estimation, monthly payback calculation, and cumulative revenue assessment. This structured approach allows for a transparent and locally contextualised evaluation of the economic sustainability of SwiftMR within the healthcare facility.

6.1. Investment and Financial Considerations for the Healthcare Facilities

Magnetic Resonance Imaging (MRI) represents a capital-intensive diagnostic service, characterised by high fixed costs related to equipment acquisition, maintenance contracts, technical infrastructure, and specialised personnel (Reyes-Santias et al.

2023); (Ladapo et al. 2018). Once installed, MRI systems generate substantial fixed expenditures irrespective of the number of examinations performed, making operational efficiency a central determinant of economic sustainability (Streit et al. 2021).

At the same time, the demand for MRI examinations has progressively increased, driven by multiple interrelated factors, including the growing prevalence of chronic diseases, broader access to diagnostic pathways, increased awareness of preventive healthcare, and continuous technological advancements expanding clinical indications (Kabasawa 2022). In this context, the optimisation of scanner time and workflow organisation becomes economically relevant, as the effective utilisation of available machine hours directly influences service capacity and financial performance.

For capital-intensive assets such as MRI systems, agenda saturation constitutes a critical economic variable (van Beek et al. 2018). The ability to maximise productive scanner time, without compromising diagnostic quality, determines whether the fixed costs associated with the technology are efficiently absorbed. Within this framework, software solutions such as SwiftMR may be interpreted not merely as technical upgrades, but as potential operational instruments aimed at enhancing throughput and improving resource allocation (Brix et al. 2024).

SwiftMR is a software-driven investment solution designed to work seamlessly with current MRI scanner setups without the need for hardware alterations. It utilizes a cloud-based system that integrates into the existing imaging process, enabling the reconstruction algorithm to process MRI data without requiring additional spending on imaging machinery. The yearly cost for utilizing this platform is €71,260 and is considered an operational expense (OPEX) due to its recurrent nature and service-oriented focus, contrasting with the typical capital expenses associated with purchasing new MRI systems that involve significant initial outlays and extended amortization periods.

In addition to the direct yearly fee, potential indirect expenses should be taken into account. These could involve the technical incorporation of the AI reconstruction software into the current digital framework, focusing on the compatibility between the MRI hardware, the PACS system, and the cloud-based platform offered by AIRS Medical. Moreover, costs related to training MRI technologists and, if needed, radiologists on operating the system should also be considered. While these indirect costs are usually less compared to acquiring software component, they play a crucial role in a thorough economic assessment.

From an economic standpoint, the value proposition of SwiftMR lies in its ability to shorten the average time taken for each MRI scan by expediting the reconstruction of images. By reducing the duration of scans, it can potentially open up more slots for

examinations to be conducted within the same timeframe, thus boosting the overall daily capacity without requiring additional hardware investments. Since MRI infrastructure involves substantial fixed costs, any rise in the number of completed scans contributes to a more efficient utilization of existing capital and operational outlays. In this scenario, improved time efficiency doesn't directly cut fixed expenses but can bolster revenue generation by accommodating more scans within the already designated scanner hours. Therefore, the anticipated economic impact of adopting SwiftMR centers on enhancing productivity rather than cutting costs. The financial viability of this initiative hinges on how effectively time savings can be translated into an uptick in examination numbers and, consequently, additional revenue for the healthcare facility.

The investment discussed earlier needs to be translated into specific operational and financial measures. To assess the practical benefits of the proposed productivity improvement, it is essential to convert the anticipated decrease in scan time into clear assumptions about the daily capacity for examinations. Section 6.2 establishes a framework for operational modeling by outlining the baseline utilization of scanners, working days annually, and potential enhancements in daily examination numbers due to integrating SwiftMR into the clinical process. These assumptions are vital for projecting additional monthly and yearly revenue based on different productivity scenarios.

Following this, Section 6.3 quantifies these estimates through a methodical analysis of return on investment (ROI) and calculation of the payback period. By translating time savings into financial outcomes, the analysis seeks to ascertain if the ongoing operational costs of SwiftMR are justifiable in the specific healthcare environment

6.2. Potential Increase in MRI productivity and Cost-Effectiveness implications

The model for the economic evaluation of the adoption of a SwiftMR-like deep learning reconstruction software is built on a deterministic framework based on clearly defined operational and financial assumptions. The annual investment cost for the software is estimated at € 71,260 per year.

The main operational assumptions are the following:

- 5 working days per week;
- 4 weeks per month;
- 12 months per year;
- Baseline productivity of 20 MRI examinations per day, derived from the real 2025 activity volumes of Ospedale di Erba (see Chapter 2).

Under these assumptions, total annual working days are 240 per year.

The base-case scenario assumes a 30% average productivity increase per examination, enabled by deep learning reconstruction through reduced acquisition times and workflow optimization. Starting from a baseline of 20 exams per day, the productivity gain translates into an increase in daily scan capacity proportional to this percentage improvement.

To assess the robustness of the investment under uncertainty in baseline operational conditions, a scan volume sensitivity analysis was performed. Two additional scenarios were considered:

- +20% variation in baseline scan volume;
- +40% variation in baseline scan volume.

This sensitivity analysis allows testing the resilience of the investment under lower and higher throughput conditions, reflecting possible variability in demand, scheduling efficiency, or referral patterns.

MRI Reimbursement Estimation

MRI reimbursement values were derived from the Italian National Health Service (Servizio Sanitario Nazionale, SSN) nomenclator, as Ospedale di Erba operates as an accredited hospital within the national reimbursement framework.

The average reimbursement per examination was estimated using a structured multi-step approach:

1. Identification of the most representative examinations by volume within each macro-category:
 - Brain (e.g., encephalon with and without contrast, brainstem);
 - Abdomen and pelvis (upper and lower abdomen, with and without contrast);
 - Musculoskeletal (shoulder and knee);
 - Spine (lumbar-sacral and cervical);
2. Computation of weighted average reimbursement within each macro-category, using the real percentage distribution of subcategories from Ospedale di Erba.
3. Computation of a final global weighted average reimbursement, weighted again by the relative volume share of each macro-category.

Based on this methodology, the final weighted average reimbursement per MRI examination for Ospedale di Erba is €198.50.

Table 6.1 summarizes the input parameters used for the sensitivity analysis, including baseline scans per day, productivity increase assumptions, reimbursement value, annual investment cost, and working days per year.

Table 6.1: Model input summary

Input	Value	Source
Number of working days per year	240	Real data
Number of MR scanner	1	Real data
Daily scan volume without SwiftMR	20	Estimated from the real data
Increase in scan volume with SwiftMR	30% 2 scenario: 20% increase vs 40% increase	Assumption
SwiftMR retail price	71,260 €	AIRS Medical
MRI reimbursement rate	198.50 €	Estimated from the real data

Model outputs

The selection of the economic output indicators was driven by the need to evaluate the investment from three complementary perspectives: revenue generation capacity, profitability, and capital recovery timing. The objective was not only to estimate the potential increase in turnover, but also to assess the financial sustainability and robustness of the investment under different operational conditions.

The economic outcomes evaluated in the models are:

- Incremental annual revenue, calculated as:

$$\begin{aligned}
 &\text{Incremental Annual Revenue} \\
 &\quad = \text{Additional Exams per Day} \\
 &\quad \times \text{Average Reimbursement} \\
 &\quad \times \text{Working Days per Year}
 \end{aligned}
 \tag{6.1}$$

- Return on investment (ROI), calculated as:

$$ROI = \frac{\text{Net Benefit}}{\text{Annual Investment Cost}}
 \tag{6.2}$$

- Payback period, calculated as:

$$\text{Payback Period} = \frac{\text{Investment Cost}}{\text{Annual Net Gain}}
 \tag{6.3}$$

The Incremental Annual Revenue was calculated in order to isolate the economic effect attributable exclusively to the productivity gain generated by the deep learning reconstruction software.

Importantly, only the additional examinations performed beyond the baseline daily capacity were considered in the computation. The baseline activity (20 examinations per day) represents the existing operational equilibrium of Ospedale di Erba and therefore does not generate incremental value attributable to the investment.

This approach ensures that the revenue estimation reflects only the marginal gain generated by the software adoption, avoiding any overestimation linked to pre-existing activity levels. In the subsequent section, monthly revenue projections are also presented in order to better illustrate the temporal distribution of the incremental financial benefits throughout the year.

The Return on Investment was selected as a key indicator to evaluate the profitability of the project relative to its cost.

ROI is calculated as the ratio between annual net benefit (incremental annual revenue minus annual investment cost) and the annual investment cost. This metric allows a direct assessment of how efficiently the invested capital generates financial returns.

The ROI is computed across the three productivity scenarios (+20%, base case, +40%) in order to test the robustness of the investment under different scan volume assumptions. This sensitivity analysis provides insight into how resilient the financial performance of the project remains under varying operational conditions.

The Payback Period was included to assess the time required to recover the initial annual investment through the incremental net gains generated by the software. This indicator provides an intuitive measure of financial risk, as it identifies the breakeven point beyond which the investment starts generating net positive returns. After the payback threshold is reached, subsequent months represent pure financial gain for the institution. The calculation is based on dividing the annual investment cost by the annual net gain derived from the additional examinations, and its interpretation is supported by the monthly revenue projections previously described.

The model described above provides a structured and institution-specific framework for evaluating the economic implications of adopting deep learning MRI reconstruction software at Ospedale di Erba. The following section presents the quantitative results of the three productivity scenarios according to the methodology and assumptions described.

6.3. Economic sensitivity analysis based on three Utilization scenarios (low, medium, high)

The results of the base-case analysis, assuming a 30% increase in average scanning productivity per examination, indicate a substantial economic benefit associated with the adoption of SwiftMR within a one-year time horizon.

Under these assumptions and within the operational context of Ospedale di Erba, the implementation of SwiftMR would generate a total incremental revenue of €285,840 over one year. After accounting for the annual investment cost, the project yields a Return on Investment (ROI) of 401.12%, demonstrating a strong profitability profile.

Furthermore, the investment is recovered within a very short timeframe, with a payback period of approximately three months, after which the remaining months of the year generate net positive financial returns for the institution.

The cumulative revenue trajectory and the evolution of the remaining balance over time are illustrated in Figure 6.1.

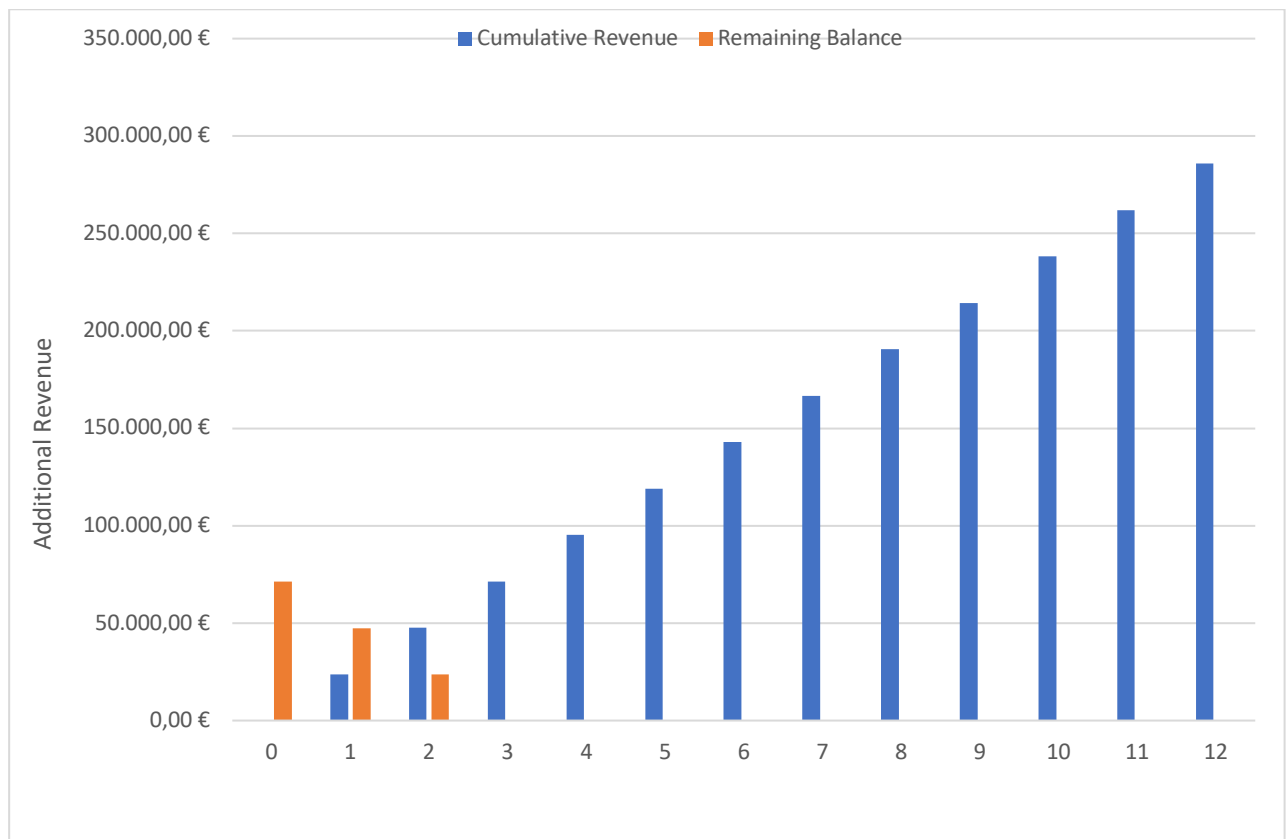


Figure 6.1: Monthly additional revenue from SwiftMR vs. cost

The baseline scan volume sensitivity analysis was conducted to evaluate the robustness of the investment under different productivity improvement assumptions. In addition to the base-case scenario (30% increase), two alternative scenarios were analyzed: a conservative 20% productivity increase and an optimistic 40% increase in scanning capacity per examination.

In the 20% productivity increase scenario, the model estimates an incremental annual revenue of €190,560. Under this assumption, the investment yields a ROI of 267.42%, with a payback period of approximately five months. Although lower than the base case, this scenario still demonstrates strong financial viability and relatively rapid capital recovery.

In contrast, the 40% productivity increase scenario results in a substantial economic impact, with an estimated incremental annual revenue of €381,120. In this case, the ROI increases to 534.83%, and the payback period is reduced to approximately two months, highlighting the strong financial leverage associated with higher productivity gains.

Overall, the sensitivity analysis confirms that the economic sustainability of the investment remains robust across all evaluated scenarios, with consistently high returns and short capital recovery times.

These findings suggest that even under conservative productivity assumptions, the adoption of deep learning reconstruction technology remains economically advantageous.

Scenario	ROI	Additional revenue	Payback period
Base case (30%)	401.12%	€285840	3 months
20%	267.42%	€190560	5 months
40%	534.83%	€381120	2 months

Table 6.2: Results by scenario

Limitations of the economic model

This economic evaluation is based on a deterministic productivity-driven framework and assumes that workflow optimization can be effectively translated into additional completed examinations. The model does not explicitly account for potential organizational constraints, such as scheduling rigidity, referral variability, or staff availability, which may influence the real-world conversion of time savings into increased throughput. Clinical effectiveness and diagnostic safety were evaluated separately within the HTA framework and are therefore assumed to be non-inferior to standard care for the purpose of this economic analysis. Additionally, the model adopts a one-year time horizon and does not incorporate indirect institutional benefits (e.g., waiting list reduction or competitive positioning), making the financial projections conservative and context-specific to Ospedale di Erba.

Final economic consideration for SwiftMR

The sensitivity analysis demonstrates that the most favorable scenario corresponds to the 40% productivity increase, which generates an incremental annual net revenue of €381,120 and achieves a payback period of approximately two months. Importantly, this scenario cannot be considered unrealistic. Clinical studies evaluated during the Health Technology Assessment process indicate that such reductions in acquisition time are achievable while maintaining image quality comparable to — or in some cases

superior to — standard clinical practice. Therefore, the 40% scenario represents a plausible upper-bound productivity improvement rather than an extreme assumption.

The base-case scenario (30% productivity increase) still produces a substantial incremental annual revenue of €285,840, with a payback period of approximately three months. This assumption may in fact be considered conservative. While technically feasible productivity gains may be higher, the 30% estimate reasonably accounts for potential workflow integration challenges, operational adaptation periods, and real-world clinical constraints that may limit the immediate realization of maximum efficiency gains.

Even under the most conservative scenario (20% productivity increase), the investment remains financially attractive. In this case, the model estimates an incremental annual revenue of €190,560, with a payback period of approximately five months. Although lower than the other scenarios, the economic return remains robust and the capital recovery time remains short within the one-year horizon.

Overall, the results consistently indicate that, under realistic and in some cases conservative assumptions, the adoption of SwiftMR at Ospedale di Erba is economically favorable. Notably, the model does not incorporate additional potential benefits such as reduced waiting lists, improved patient throughput, enhanced institutional competitiveness, or qualitative gains in image quality. Consequently, the presented estimates may be interpreted as financially prudent projections rather than optimistic overestimations.

7 Conclusion and future developments

The regulatory positioning of SwiftMR must be interpreted within three interrelated horizontal frameworks: the Medical Device Regulation (MDR), the General Data Protection Regulation (GDPR), and the Artificial Intelligence Act (AI Act). As a CE-marked medical device based on deep learning reconstruction and deployed through cloud integration with the hospital IT infrastructure, SwiftMR operates at the intersection of medical device compliance, data governance, and artificial intelligence regulation.

Under the MDR framework, the software is required to comply with the essential requirements applicable to medical devices, including the implementation of a Quality Management System (QMS), structured risk management procedures, and continuous post-market surveillance throughout the entire software lifecycle. These elements ensure that safety, performance, and clinical validity are maintained beyond initial certification (European Union 2017).

From a data governance perspective, the integration of SwiftMR's cloud-based platform with the hospital's IT network — including the PACS system and MRI hardware — entails the processing of sensitive patient data. Consequently, full compliance with the GDPR is required, ensuring lawful data processing, security safeguards, accountability, and adherence to the core principles of data protection established under European law (European Union 2016).

The introduction of the AI Act further strengthens the regulatory landscape. As a deep learning-based medical software, SwiftMR is expected to be classified as a high-risk AI system due to its medical purpose and clinical impact. Accordingly, it will be subject to specific requirements concerning transparency, risk management, technical documentation, human oversight, and robustness. With the AI Act entering into force progressively, full regulatory alignment will require harmonization between MDR and AI-specific compliance obligations (Artificial Intelligence Act 2024).

Beyond formal regulatory certification, the technical robustness of the software also plays a critical role in its adoption. The architecture has been trained and validated on large-scale, fully sampled raw MRI datasets (approximately three million acquisitions), aiming to ensure generalizability across anatomical districts and scanner vendors. To mitigate known deep learning reconstruction limitations, such as distribution shift, the undersampling strategies applied in k-space reconstruction are designed to remain consistent with current compressed sensing and parallel imaging

pipelines. This alignment supports technical stability across different acquisition environments (Jeong et al. 2024).

Within the scope of this project, both regulatory compliance and technical validation have been examined in detail. From a governance perspective, the adoption of SwiftMR appears aligned with current European regulatory requirements and with the forthcoming AI Act obligations, provided that institutional oversight and lifecycle monitoring remain actively implemented.

From a clinical perspective, the evidence emerging from the Peer reviewed studies used to validate SwiftMR consistently demonstrates image quality that is comparable — and in several contexts superior — to the standard-of-care reconstruction. Quantitative metrics such as Signal-to-Noise Ratio (SNR) and Contrast-to-Noise Ratio (CNR), together with subjective image quality assessments, showed favorable results across multiple anatomical districts and acquisition sequences. Most importantly, diagnostic confidence was never statistically inferior to the comparator and, in specific contexts, was reported as significantly improved.

Particularly relevant examples include contrast-enhanced MRI protocols for acute ischemic stroke, where deep learning reconstruction enhanced image clarity compared to conventional TWIST acquisitions (Kim et al. 2024), and hippocampal imaging in temporal lobe epilepsy, where lesion conspicuity improved (Suh et al. 2024). In black-blood imaging, diagnostic confidence for metastatic lesions smaller than 5 mm was statistically superior in comparison to the standard reconstruction pipeline (Oh et al. 2025).

Within the operational context of Ospedale di Erba, real-world data from 2025 indicate that the majority of MRI examinations are routine diagnostic studies, where the primary institutional need is not necessarily further image quality enhancement but rather a significant reduction in scan time. In this context, deep learning reconstruction provides measurable value through acceleration. For example, knee MRI studies reported an average scan time reduction of 41% without any reduction in diagnostic confidence (Lee et al. 2023), while spine examinations demonstrated a 32.3% reduction with comparable or improved confidence in specific pathologies such as annular fissures (Yoo et al. 2023).

These clinical findings justify the economic modelling approach adopted in this work. The sensitivity volume analysis, based on a 30% baseline productivity increase and tested under 20% (conservative) and 40% (optimistic) scenarios, is directly supported by published evidence and by the local pilot experience, which reached acceleration rates close to 40%. Even under the conservative 20% scenario, the investment remains economically favorable. In the base-case scenario (30%), the model projects an additional annual revenue of €285,840 and a payback period of approximately three months, confirming the financial sustainability of the adoption.

From a safety perspective, theoretical risks described in the deep learning literature — including domain shift and hallucination phenomena (Heckel et al. 2024), were not observed as clinically significant in the majority of evaluated contexts. Domain shift risk appears mitigated through architecture design and the use of large-scale, high-quality training datasets. However, isolated reports in the literature describe potential hallucination effects in cartilage assessment in the study of (Lee et al. 2023) or vessel-suppressed black-blood sequences (Oh et al. 2025), potentially increasing false positives. Furthermore, primary evidence collected at Ospedale di Erba during the pilot phase identified increased sensitivity to motion artifacts and reconstruction artifacts when undersampling strategies were excessively aggressive.

These findings confirm that human supervision remains fundamental. Structured clinical oversight and parameter calibration mitigate residual safety concerns while preserving the benefits of accelerated reconstruction.

This assessment presents several limitations inherent to its design as a Local Health Technology Assessment. First, the economic evaluation is based on a deterministic productivity-driven model with a one-year time horizon and assumes that time savings can be effectively translated into additional completed examinations without major organizational constraints. The analysis does not incorporate long-term clinical outcomes, cost-effectiveness in terms of QALYs, or potential indirect system-level effects such as reduced downstream diagnostic procedures.

Second, the clinical evidence supporting the adoption of SwiftMR is largely based on non-inferiority findings and image-quality metrics rather than hard patient-centered outcomes. Although diagnostic confidence was not statistically inferior and, in some contexts, superior to standard reconstruction, long-term prospective outcome data remain limited.

Finally, as a context-specific evaluation conducted within Ospedale di Erba, the results may not be directly generalizable to institutions with different case mixes, workflow structures, or reimbursement systems. Nevertheless, these limitations are consistent with the scope and objectives of a Local HTA and do not invalidate the overall conclusions, but rather define the boundaries within which they should be interpreted.

Looking forward, the integration of deep learning reconstruction systems such as SwiftMR should not be considered a static implementation but rather a dynamic process requiring continuous monitoring and governance. The forthcoming full application of the AI Act will further strengthen lifecycle obligations, emphasizing documentation, human oversight, and risk monitoring as integral components of institutional adoption.

From a clinical perspective, future evaluations could move beyond non-inferiority analyses toward outcome-based assessments, including potential reductions in repeat examinations, waiting times, and workflow inefficiencies. Longitudinal real-world data collection within Ospedale di Erba and across additional centers of the Lifenet

Healthcare group may provide further evidence regarding scalability and generalizability.

Technical optimization also remains an area for continued development. Adaptive calibration of undersampling strategies and structured audit of artifact-related findings could further enhance reconstruction robustness while maintaining accelerated acquisition protocols.

Ultimately, the long-term value of AI-based MRI reconstruction will depend not only on immediate economic returns but on the institution's ability to integrate technological innovation within a structured governance model that ensures safety, transparency, and sustained clinical performance.

Overall, this Health Technology Assessment demonstrates that the adoption of a deep learning-based MRI reconstruction system such as SwiftMR is clinically validated, economically advantageous, and compatible with safe use under appropriate governance and human oversight, as required by the evolving European regulatory framework.

In response to the policy question guiding this work, namely whether there is a real advantage in acquiring and implementing AI-based software systems in healthcare in terms of improving service quality for patients and generating economic benefit for institutions and the evidence collected within this assessment supports an affirmative conclusion. Within the evaluated local context, the implementation of SwiftMR appears capable of enhancing operational efficiency, maintaining or improving diagnostic quality, and generating sustainable economic returns, provided that regulatory compliance, technical calibration, and structured clinical supervision remain integral components of its deployment

Bibliography

A. Aghabiglou and E. M. Eksioglu, "Deep unfolding architecture for MRI reconstruction enhanced by adaptive noise maps," *Biomedical Signal Processing and Control*, vol. 78, p. 104016, 2022.

AIRS Medical, *DOC2308 SwiftMR A20-CL Implementation Guide (IT)*, Rev. 8, 2024.

AIRS Medical, "SwiftMR Image Gallery | Enhanced MRI Efficiency and Image Quality," [Online]. Available: <https://airsmed.com/en/image-gallery/>. [Accessed: 21-Feb-2026].

AIRS Medical, *Information Security Compliance White Paper (Cloud Version)*, 2025.

V. Antun, F. Renna, C. Poon, B. Adcock and A. C. Hansen, "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 48, pp. 30088–30095, 2020.

European Union, *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, Brussels, 2024.

N. Avidan and M. Freiman, "MA-RECON: Mask-aware deep neural network for robust fast MRI k-space interpolation," [Online]. Available: https://github.com/nitzanavidan/PD_Recon. [Accessed: 21-Feb-2026].

S. Bash et al., "Deep learning image processing enables 40% faster spinal MR scans which match or exceed quality of standard of care: A prospective multicenter multireader study," *Clinical Neuroradiology*, vol. 32, no. 1, pp. 197–206, 2021.

S. Bash et al., "Deep learning enables 60% accelerated volumetric brain MRI while preserving quantitative performance: A prospective, multicenter, multireader trial," *American Journal of Neuroradiology*, vol. 42, no. 12, pp. 2130–2137, 2021.

E. J. R. van Beek et al., "Value of MRI in medicine: More than just another test?," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 7, pp. e14–e25, 2018.

A. Bernasconi et al., "Recommendations for the use of structural magnetic resonance imaging in the care of patients with epilepsy: A consensus report from the International League Against Epilepsy Neuroimaging Task Force," *Epilepsia*, vol. 60, no. 6, pp. 1054–1068, 2019.

- M. A. Bernstein, K. F. King and X. J. Zhou, *Handbook of MRI Pulse Sequences*, Amsterdam: Elsevier, 2004.
- S. Bhadra, V. A. Kelkar, F. J. Brooks and M. A. Anastasio, "On hallucinations in tomographic image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3249–3260, 2021.
- D. Le Bihan et al., "Diffusion tensor imaging: Concepts and applications," *Journal of Magnetic Resonance Imaging*, vol. 13, no. 4, pp. 534–546, 2001.
- M. A. K. Brix et al., "Financial impact of incorporating deep learning reconstruction into magnetic resonance imaging routine," *European Journal of Radiology*, vol. 175, p. 111434, 2024.
- R. W. Brown et al., *Magnetic Resonance Imaging: Physical Principles and Sequence Design*, 2nd ed., Hoboken, NJ: Wiley, 2014.
- G. D. Cascino et al., "Magnetic resonance imaging-based volume studies in temporal lobe epilepsy: Pathological correlations," *Annals of Neurology*, vol. 30, no. 1, pp. 31–36, 1991.
- T. Chikui et al., "The principle of dynamic contrast enhanced MRI, the method of pharmacokinetic analysis, and its application in the head and neck region," *International Journal of Dentistry*, vol. 2012, p. 480659, 2012.
- J. P. Cohen, M. Luck and S. Honari, "Distribution matching losses can hallucinate features in medical image translation," in *Lecture Notes in Computer Science*, vol. 11070, pp. 529–536, 2018.
- D. Darestani, A. S. Chaudhari and R. Heckel, "Measuring robustness in deep learning based compressive sensing," in *Proceedings of Machine Learning Research*, vol. 139, pp. 2433–2444, 2021.
- J. A. Detre, J. S. Leigh, D. S. Williams and A. P. Koretsky, "Perfusion imaging," *Magnetic Resonance in Medicine*, vol. 23, no. 1, pp. 37–45, 1992.
- EUnetHTA, *HTA Core Model for Diagnostic Technologies, Version 1.0, Work Package 4*, 2008.
- European Commission, *Commission Implementing Regulation (EU) 2021/2226 laying down rules for the application of Regulation (EU) 2017/745 as regards the unique device identification system for medical devices*, Brussels, 2021.
- European Union, *Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation)*, Brussels, 2016.
- European Union, *Regulation (EU) 2017/745 on medical devices*, Brussels, 2017.
- N. M. Gottschling, V. Antun, A. C. Hansen and B. Adcock, "The troublesome kernel: On hallucinations, no free lunches and the accuracy-stability trade-off in

- inverse problems," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 8, no. 1, pp. 73–104, 2020.
- M. A. Griswold et al., "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002.
- A. Hagiwara et al., "Multiparametric MRI: From simultaneous rapid acquisition methods and analysis techniques using scoring, machine learning, radiomics, and deep learning to the generation of novel metrics," *Investigative Radiology*, vol. 58, no. 8, pp. 548–559, 2023.
- J. Hamilton, D. Franson and N. Seiberlich, "Recent advances in parallel imaging for MRI," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 101, pp. 71–95, 2017.
- K. Hammernik et al., "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, 2017.
- R. Heckel et al., "Deep learning for accelerated and robust MRI reconstruction," *MAGMA*, vol. 37, no. 3, pp. 335–356, 2024.
- M. B. Hossain et al., "A systematic review and identification of the challenges of deep learning techniques for undersampled magnetic resonance image reconstruction," *Sensors*, vol. 24, no. 3, p. 753, 2024.
- International Electrotechnical Commission, *IEC 62304: Medical device software – Software life cycle processes*, Geneva, 2006.
- International Electrotechnical Commission, *IEC/TR 80002-1: Medical device software – Guidance on the application of ISO 14971 to medical device software*, Geneva, 2009.
- International Electrotechnical Commission, *IEC 62366-1: Medical devices – Application of usability engineering to medical devices*, Geneva, 2015.
- International Electrotechnical Commission, *IEC/TR 62366-2: Medical devices – Guidance on the application of usability engineering to medical devices*, Geneva, 2016.
- International Electrotechnical Commission, *IEC 80001-1: Application of risk management for IT networks incorporating medical devices*, Geneva, 2021.
- International Organization for Standardization, *ISO 13485: Medical devices – Quality management systems – Requirements for regulatory purposes*, Geneva, 2016.
- International Organization for Standardization, *ISO 80002-2: Medical device software – Validation of software for medical device quality systems*, Geneva, 2017.
- International Organization for Standardization, *ISO 14971: Medical devices – Application of risk management to medical devices*, Geneva, 2019.

International Organization for Standardization, *ISO/TR 24971: Medical devices – Guidance on the application of ISO 14971*, Geneva, 2020.

International Organization for Standardization, *ISO 20416: Medical devices – Post-market surveillance for manufacturers*, Geneva, 2020.

International Organization for Standardization, *ISO 15223-1: Medical devices – Symbols to be used with information to be supplied by the manufacturer*, Geneva, 2021.

International Organization for Standardization, *ISO 20417: Medical devices – Information to be supplied by the manufacturer*, Geneva, 2021.

M. Iwamura et al., “Thin-slice two-dimensional T2-weighted imaging with deep learning-based reconstruction: Improved lesion detection in the brain of patients with multiple sclerosis,” *Magnetic Resonance in Medical Sciences*, vol. 23, no. 2, pp. 184–192, 2024.

L. Jäger and M. Reiser, “CT and MR imaging of the normal and pathologic conditions of the facial nerve,” *European Journal of Radiology*, vol. 40, no. 2, pp. 133–146, 2001.

G. Jeong et al., “All-in-one deep learning framework for MR image reconstruction,” [Online]. Available: <http://arxiv.org/abs/2405.03684>. [Accessed: 21-Feb-2026].

P. M. Johnson et al., “Evaluation of the robustness of learned MR image reconstruction to systematic deviations between training and test data for the models from the fastMRI challenge,” in *Lecture Notes in Computer Science*, vol. 12964, pp. 25–34, 2021.

H. Kabasawa, “MR imaging in the 21st century: Technical innovation over the first two decades,” *Magnetic Resonance in Medical Sciences*, vol. 21, no. 1, pp. 71–86, 2021.

E. Kanal et al., “ACR guidance document on MR safe practices: 2013,” *Journal of Magnetic Resonance Imaging*, vol. 37, no. 3, pp. 501–530, 2013.

B. K. Kim, S. H. You, B. Kim and J. H. Shin, “Deep learning-based high-resolution magnetic resonance angiography (MRA) generation model for 4D time-resolved angiography with interleaved stochastic trajectories (TWIST) MRA in fast stroke imaging,” *Diagnostics*, vol. 14, no. 11, 2024.

M. Kim et al., “Thin-slice pituitary MRI with deep learning-based reconstruction: Diagnostic performance in a postoperative setting,” *Radiology*, vol. 298, no. 1, pp. 114–122, 2021.

S. H. Kim et al., “Deep learning reconstruction in pediatric brain MRI: Comparison of image quality with conventional T2-weighted MRI,” *Neuroradiology*, vol. 65, no. 1, pp. 207–214, 2023.

- F. Knoll et al., "Deep learning methods for parallel magnetic resonance image reconstruction," [Online]. Available: <https://arxiv.org/abs/1904.01112>. [Accessed: 21-Feb-2026].
- F. B. Kristensen et al., "The HTA Core Model®—10 years of developing an international framework to share multidimensional value assessment," *Value in Health*, vol. 20, no. 2, pp. 244–250, 2017.
- J. A. Ladapo et al., "Economics of MRI operations after implementation of interpersonal skills training," *Journal of the American College of Radiology*, vol. 15, no. 12, pp. 1775–1782, 2018.
- K. Lampe et al., "The HTA Core Model: A novel method for producing and reporting health technology assessments," *International Journal of Technology Assessment in Health Care*, vol. 25, suppl. 2, pp. 9–20, 2009.
- R. M. Lebel and GE Healthcare, *Performance Characterization of a Novel Deep Learning-Based MR Image Reconstruction Pipeline*, 2020.
- D. H. Lee et al., "Deep learning-based thin-section MRI reconstruction improves tumour detection and delineation in pre- and post-treatment pituitary adenoma," *Scientific Reports*, vol. 11, 2021.
- J. Lee et al., "Highly accelerated knee magnetic resonance imaging using deep neural network (DNN)-based reconstruction: Prospective, multi-reader, multi-vendor study," *Scientific Reports*, vol. 13, 2023.
- F.-H. Lin et al., "Degenerate mode birdcage volume coil for sensitivity-encoded imaging," *Magnetic Resonance in Medicine*, vol. 50, no. 5, pp. 1107–1111, 2003.
- A. Lindenholz et al., "The use and pitfalls of intracranial vessel wall imaging: How we do it," *Radiology*, vol. 286, no. 1, pp. 12–28, 2018.
- M. Lustig, D. Donoho and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- M. Lustig and J. M. Pauly, "SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 457–471, 2010.
- M. J. Muckley et al., "Results of the 2020 fastMRI challenge for machine learning MR image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2306–2317, 2021.
- National Electrical Manufacturers Association, *Digital Imaging and Communications in Medicine (DICOM) Standard, PS 3.1–3.20*, Washington, DC, 2022.

- S. Ogawa et al., "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *Proceedings of the National Academy of Sciences*, vol. 87, no. 24, pp. 9868–9872, 1990.
- G. Oh et al., "Deep learning-based image enhancement for improved black blood imaging in brain metastasis," *European Radiology*, 2025.
- A. Panda et al., "Magnetic resonance fingerprinting—An overview," *Current Opinion in Biomedical Engineering*, vol. 3, pp. 56–66, 2017.
- J. Park et al., "Detection of small metastatic brain tumors: Comparison of 3D contrast-enhanced whole-brain black-blood imaging and MP-RAGE imaging," *Investigative Radiology*, vol. 47, no. 2, pp. 136–141, 2012.
- Y. W. Park et al., "Robust performance of deep learning for automatic detection and segmentation of brain metastases using three-dimensional black-blood and three-dimensional gradient echo imaging," *European Radiology*, vol. 31, no. 9, pp. 6686–6695, 2021.
- W. J. Powers et al., "Guidelines for the early management of patients with acute ischemic stroke: 2019 update," *Stroke*, vol. 50, no. 12, pp. e344–e418, 2019.
- K. P. Pruessmann, M. Weiger, M. B. Scheidegger and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999.
- G. Radaelli et al., "Implementation of EUnetHTA Core Model® in Lombardia: The VTS framework," *International Journal of Technology Assessment in Health Care*, vol. 30, no. 1, pp. 105–112, 2014.
- F. Reyes-Santias et al., "Cost analysis of magnetic resonance imaging and computed tomography in cardiology: A case study," *Healthcare*, vol. 11, no. 14, 2023.
- J. D. Rudie et al., "Clinical assessment of deep learning-based super-resolution for 3D volumetric brain MRI," *Radiology: Artificial Intelligence*, vol. 4, no. 2, 2022.
- D. Saloner, J. Liu and H. Haraldsson, "MR physics in practice: How to optimize acquisition quality and time for cardiac MR imaging," *Magnetic Resonance Imaging Clinics of North America*, vol. 23, no. 1, pp. 1–6, 2015.
- P. D. Schellinger et al., "A standardized MRI stroke protocol: Comparison with CT in hyperacute intracerebral hemorrhage," *Stroke*, vol. 30, no. 4, pp. 765–768, 1999.
- K. Seungwook and Y. Yang, *AI-Powered MR Image Enhancement Solution SwiftMR: Clinical Implications and Use Cases; Hurdles in Clinical MR Imaging*, 2024.

- V. Smith et al., "Defining the clinician's role in early health technology assessment during medical device innovation: A systematic review," *BMC Health Services Research*, vol. 19, 2019.
- U. Streit et al., "Analysis of core processes of the MRI workflow for improved capacity utilization," *European Journal of Radiology*, vol. 138, p. 109648, 2021.
- P. S. Suh et al., "Improving diagnostic performance of MRI for temporal lobe epilepsy with deep learning-based image reconstruction in patients with suspected focal epilepsy," *Korean Journal of Radiology*, vol. 25, no. 4, pp. 374–383, 2024.
- H. Uetani et al., "A preliminary study of deep learning-based reconstruction specialized for denoising in high-frequency domain," *Neuroradiology*, vol. 63, no. 1, pp. 63–71, 2021.
- S. Yamatogi et al., "Evaluation of small ischemic lesions after carotid artery stenting," *Neuroradiology*, vol. 53, no. 4, pp. 255–260, 2011.
- H. Yoo et al., "Deep learning-based reconstruction for acceleration of lumbar spine MRI: A prospective comparison with standard MRI," *European Radiology*, vol. 33, no. 12, pp. 8656–8668, 2023.
- R. E. Yoo and S. H. Choi, "Deep learning-based image enhancement techniques for fast MRI in neuroimaging," *Magnetic Resonance in Medical Sciences*, vol. 23, no. 3, pp. 341–351, 2024.
- M. Zaitsev, J. Maclaren and M. Herbst, "Motion artefacts in MRI: A complex problem with many partial solutions," *Journal of Magnetic Resonance Imaging*, vol. 42, no. 4, pp. 887–901, 2015.

List of Figures

Figure 1.1: Cloud-Based product Workflow. Reproduced from the technical documentation of (AIRS Medical 2025)	38
Figure 1.2: Cloud based security measures reproduced from technical implementation of (AIRS Medical 2025)	39
Figure 1.3: Definition of the different stage of product life cycle and HTA assessment (reproduced from: (Smith et al. 2019))	57
Figure 1.4: HTA domains of EUnetHTA (reproduced from: (Kristensen et al. 2017)).	58
Figure 1.5: The core HTA structure (reproduced from: (Lampe et al. 2009))	59
Figure 3.1: Brain T1WI on a 3T scanner and reduction in 54 % scanning time. Image reproduced (AIRS Medical 2024b).	78
Figure 3.2: Liver T2WI pulse sequence from a 3 T MR scanner. Image reproduced from: (AIRS Medical 2024b).....	79
Figure 3.3: Pair of training samples in the k-space domain. Image reproduces from: (Jeong et al. 2024a).....	83
Figure 3.4: Scheme of the U-net model and the standardized module and the contextual data. Image reproduces from: (Jeong et al. 2024).	85
Figure 3.5: the original image and the five DL reconstructed with enhancement quality. Image reproduces from: (Jeong et al. 2024)	86
Figure 3.6: Means signal and noise standard deviation in the ROI at different noise level factor. Image reproduces from: (Jeong et al. 2024).....	88
Figure 3.7: Brain MR angiography TOF, MIP images with coronal and sagittal acquisition. Image reproduced from (Jeong et al. 2024).	89
Figure 3.8: Relative edges sharpness between DL reconstructed image and the original one. Image reproduced from (Jeong et al. 2024).	90
Figure 3.9: SwiftMR Cloud-Hosted Workflow (reproduced from (AIRS Medical 2024a))	92
Figure 5.1: Case of 67 years old man with lung cancers. Image reproduced from: (Oh et al. 2025).	115

Figure 5.2: False positive case of brain metastasis in a 43 years old women with breast cancer. Image reproduced from: (Oh et al. 2025).....	116
Figure 5.3: Conventional and reconstructed images of accelerated sequences: a 64 years old woman with knee pain. Images reproduced from: (Lee et al. 2023)	120
Figure 5.4: Cartilage grade on conventional and reconstructed images of accelerated sequence. A 54 years-old female with knee pain (A) and 41 years-old man (B). Image reproduced from: (Lee et al. 2023)	121
Figure 5.5: Aneurism visualization in the three image modalities. Image reproduced from: (Kim et al. 2024).....	127
Figure 5.6: Maximum intensity projection of the three image modalities. Image reproduced from: (Kim et al. 2024).	128
Figure 5.7: increasing in sensitivity detection in 1.5 mm slice thickness sequence ...	133
Figure 5.8: False positive example induced by DLR algorithm.....	137
Figure 5.9: Comparison of three image modalities for assessment of mild centre canal stenosis of L3/4 and moderate at L4/5	140
Figure 6.1: Monthly additional revenue from SwiftMR vs. cost	160

List of Tables

Table 1: Assessment of the safety domains for SwiftMR.....	vii
Table 2: Volume sensitivity analysis of SwiftMR	ix
Table 1.1: Intended pupose of SwiftMR.....	18
Table 1.2: Declaration of conformity for SwiftMR.....	19
Table 1.3: Key risk domains according to ISO 14971 applied to AI-based MRI reconstruction software	25
Table 1.4: Key steps of the usability engineering process according to IEC 62366-1 ..	27
Table 1.5: Fundamental principle of the EU regulation of GDPR.....	33
Table 1.6: Cross compliance of GDPR and MDR of the ISO/IEC standard adopted by AIRS Medical for development of SwiftMR.....	37
Table 1.7: Classification of the AI systems according to the AI Act.....	43
Table 1.8: summary of the conformity requirements and conformance assessment of the AI Act regulation	52
Table 1.9: Prioritization of the nine domains of the Health Technologies Assessment	60
Table 2.1: Baseline volumes of MRI for the current use of technologies.....	70
Table 2.2: Volumes of MRI examination for the subtype MSK	71
Table 2.3: Volumes of MRI examination for the subtype abdominal/pelvis	72
Table 2.4: Volumes of MRI examination for the subtype brain MRI	72
Table 2.5: Volumes of MRI examination for the subtype spine MRI.....	72
Table 2.6: Potential use in clinical setting for the adoption of SwiftMR.....	73
Table 3.1: Eight anatomical protocols for the SwiftMR AI based software	80
Table 4.1: Risk assessment of the technology safety related domains.....	102
Table 5.1: Scientific paper for the clinical validation of SwiftMR	110
Table 5.2: PICO for the paper (Oh et al. 2025).....	112
Table 5.3: statistical test used for (Oh et al. 2025; Park et al. 2012).....	114
Table 5.4: PICO of the study (Lee et al. 2023)	117

Table 5.5: PICO for the paper (Kim et al. 2024)..... 124

Table 5.6: PICO of the study: (Suh et al. 2024) 130

Table 5.7: Statistical analysis for diagnostic performance 131

Table 5.8: Statistical analysis for image quality assessment..... 132

Table 5.9: PICO of (Yoo et al. 2023)..... 138

Table 5.10: Different DL-based MR image reconstruction 142

Table 5.11: Main benefit and limitation of the SwiftMR clinical study validation in the enhancement of image quality 146

Table 5.12: Main benefit and limitation of the SwiftMR clinical study validation in fast scan protocol 148

Table 6.1: Model input summary 157

Table 6.2: Results by scenario 161

Acknowledgments

Desidero innanzitutto ringraziare il Professor Giuseppe Baselli per il costante supporto e la guida scientifica fornita durante lo sviluppo del progetto di tesi.

Ringrazio inoltre l'ingegnere Andrea Parisi per avermi offerto l'opportunità di svolgere il tirocinio presso l'Ospedale di Erba, accompagnandomi con disponibilità e competenza nella definizione degli argomenti e nello sviluppo del lavoro.

Un sincero ringraziamento va infine a Giacomo Rallo per il tempo dedicatomi e per il prezioso contributo tecnico, che si è rivelato determinante ai fini della realizzazione del progetto.

Questa tesi è dedicata innanzitutto ai miei genitori, Daniela e Antonio, che mi hanno sempre sostenuto, non solo nella realizzazione di questo progetto, ma in ogni fase del mio percorso di vita. A loro devo la forza, la determinazione e i valori che mi hanno accompagnato fino a questo traguardo.

È dedicata a mia sorella, Carola, presenza costante e punto di riferimento, con cui ho condiviso ogni sfaccettatura del mio cammino personale e accademico.

La dedico ai miei nonni, Bruno, Leo, Maria e Lucia, che hanno avuto la possibilità di vivere con me questo momento così significativo. Il loro affetto e il loro esempio rappresentano una parte fondamentale della persona che sono oggi.

Un pensiero speciale va anche alle mie zie e ai miei zii, Ornella, Paola, Marco e Rosy, per il sostegno e la vicinanza dimostrati nel tempo.

Infine, la dedico ai miei cugini, Gioele e Riccardo, con l'augurio che questo traguardo possa essere per loro fonte di ispirazione nel loro percorso futuro.

