International Journal of Scientific Research and Modern Technology (IJSRMT) *ijsrmt.com*
Volume 4, Issue 12, 2025
DOI: https://doi.org/10.38124/ijsrmt.v4i12.1038

Challenges in Medical Image Processing and Analysis: Cross-Cutting Issues Between Imaging Modalities and Deep Learning Models

Efosa Osagie[1]

[1]Department of Computer Science and Data Science, York St. John University, London, UK

Publication Date: 2025/12/30

Abstract

In recent years, with advances in computational power and the integration of Artificial Intelligence (AI) into healthcare systems, medical image processing (MIP) has seen significant benefits through the application of deep learning (DL) techniques. This application has made complex tasks, such as segmentation, classification, and reconstruction, more feasible, explainable, and automated across clinically diverse imaging modalities. However, with these benefits also come challenges in uniquely aligning DL solutions with the consistent constraints of these medical imaging modalities. To provide a comprehensive insight into these challenges, this study critically examines deep learning challenges across four key medical imaging modalities: MRI, CT, ultrasound, and histopathology. Their diversity in acquisition, resolution and annotation techniques makes them suitable for this consideration. X-ray imaging is excluded from this study due to its lower spatial complexity and standardised acquisition pipeline. Furthermore, this study examines modality-specific challenges in standard DL-based solutions across Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), transformers, and hybrid systems, and highlights their cross-cutting issues and intersection with the challenges seen in these images. Previous reviews often overlook the significant interactions between DL-based solution design and medical imaging characteristics. By highlighting these challenges, this study helps guide the design of stronger, more practical DL-based solutions that can make medical image processing more reliable and useful in real healthcare settings.

Keywords: *Imaging Challenges, Medical Image Processing, Medical Image Analysis, Deep Learning, Ultrasound, CT, MRI, X-Ray.*

I. INTRODUCTION

Medical imaging remains a cornerstone of clinical diagnosis and investigation, offering pixel-based representations that encode rich anatomical and pathological information. Modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), and Ultrasound Imaging (UI) provide diverse visualisations of internal structures, supporting tasks ranging from organ-level assessment to cellular-level pathology. These imaging techniques underpin critical workflows in nuclear medicine, cross-sectional organ analysis, and soft tissue evaluation, forming the basis for diagnosis, treatment planning, and longitudinal monitoring. Medical Image Processing (MIP) encompasses a suite of computational techniques for enhancing and interpreting these visual datasets. Core tasks include preprocessing (For instance, noise reduction, artefact removal, resolution enhancement), segmentation of anatomical or pathological regions, feature extraction for diagnostic insight, and classification. Additionally, it includes reconstruction, such as 2D/3D MRI reconstruction for diagnosis [1, 2], as well as surgical planning and simulation [3, 4]. These processes enable early disease detection, quantitative analysis, and personalised care, ultimately improving diagnostic accuracy and accessibility.

Recent advances in Artificial Intelligence (AI), particularly Deep Learning (DL), a subfield of Machine Learning (ML), have transformed MIP by automating complex tasks such as organ segmentation, image classification, and synthetic data augmentation. DL architectures range from convolutional networks and generative models to transformers and physics-informed frameworks, each offering unique capabilities. However, the rapid pace of innovation presents challenges for researchers entering the field, especially in identifying

Osagie, E. (2025). Challenges in Medical Image Processing and Analysis: Cross-Cutting Issues Between Imaging Modalities and Deep Learning Models. *International Journal of Scientific Research and Modern Technology*, *4*(12), 93–108. https://doi.org/10.38124/ijsrmt.v4i12.1038

gaps and evaluating model suitability across clinical contexts. This study aims to critically examine the limitations and failure modes of DL models in MIP, with a particular focus on modality-specific challenges and model-specific constraints. By foregrounding the technical bottlenecks and clinical implications of these limitations, this study aims to provide a comprehensive understanding of how imaging characteristics interact with model design and, subsequently, the feasibility of deployment. This approach is justified by the growing need for a systematic evaluation of DL performance beyond benchmark scores and isolated reviews, especially in real-world settings where data scarcity, noise from acquisition as well as annotation, and domain shift, which often causes a problem due to distribution mismatch, in cases where the data source may be extremely limited, are prevalent.

While several recent reviews have contributed valuable insights into the application of DL in MIP, notable gaps remain in their treatment of modality-specific constraints and model-specific limitations. For instance, Ref [5] examined the use of vision transformers across disease contexts, such as brain tumours and diabetic retinopathy; however, their analysis lacked depth in the range of modality considerations, multi-device interpretation, and cross-cutting interactions, which are essential for assessing generalizability and deployment issues. Similarly, Ref [6] provided a technically detailed account of predictive imaging models for early diagnosis. However, they did not engage with comparative critical analysis across imaging modalities on their interactions with these models, thereby limiting the scope of their evaluation. Ref [7] focused on the role of grand challenges in advancing AI for radiology, offering a valuable overview of competition frameworks. However, their study did not sufficiently address the longitudinal clinical impact of these models. It did not critically assess the model-specific constraints and biases that may arise in multimodal systems. Furthermore, Ref [8, 9] focused on specific imaging modalities. However, they [8, 9] offered minimal discussion of DL architectures and the dataset constraints that critically impact model performance. These omissions emphasise the need for a more comprehensive and critical synthesis that explicitly examines how imaging characteristics and model design interact to produce limitations in accuracy, interpretability, and clinical utility, an aim this study seeks to fulfil.

Section 2 provides the study's contributions. Section 3 discusses modality-specific challenges and highlights their potential impact in MIP. Section 4 presents common DL architectures and examines their constraints in MIP. Section 5 provides interactions and cross-cutting issues between modality constraints and models' architectural challenges. The final section concludes this study, providing overall insights and laying the foundation for future research.

II. CONTRIBUTIONS

➢ *The Contributions of this Study are Outlined Below:*

- It provides a comprehensive analysis of the interactions between modality-specific imaging and model-specific limitations across DL- based models. When compared with previous reviews, which usually treat modalities and models in isolation. This isolation makes these previous reviews inadequate for practical implementation decisions in a clinical context.
- It integrates cross-cutting challenges between modalities and model-specific constraints, which are often discussed separately and mostly not in relation to each other. By examining how these issues compound across both imaging modalities and model architectures, the study provides a more comprehensive resource to help mitigate deployment barriers.

This study aims to provide researchers with more precise guidance for designing robust and generalisable AI systems.

III. MODALITY-SPECIFIC CHALLENGES

Medical imaging modalities (MIMs) differ significantly in their physical principles, acquisition protocols, and use in targeted clinical applications. However, these differences introduce complications that may negatively affect the quality, consistency, and interpretability of the data used in DL pipelines. Therefore, there is a need for a comprehensive understanding of these modality-specific limitations, which is essential for selecting appropriate DL architectures. This section presents key challenges across popular MIM with the aim of revealing relevant insights into their challenges.

➢ *Magnetic Resonance Imaging*

MRI is widely used to visualise soft tissues, especially in neurological, musculoskeletal, and cardiovascular applications. It is a non-invasive modality and valuable for many clinical applications. However, MRI presents significant challenges for DL pipelines due to inherent acquisition constraints, which Ref [10] has suggested are complex and may lack a definitive solution. The issue of motion artefacts is a serious challenge, which usually results from the complex interplay among the image's pixel structure, subject motion, the k-space acquisition method, and reconstruction algorithms. MRI is extremely sensitive to involuntary subject motion. This is due to the long exposure times required for clinical imaging examinations. This sensitivity is due to factors such as involuntary patient movement and diaphragmatic breathing. These movements can severely reduce image quality and result in blurred anatomical boundaries. This is because, unlike conventional photographic image acquisition, which occurs in pixel space, MRI acquisition occurs in K-space, which is based on frequency (Fourier) space [11]. K-space is a grid in which each point corresponds to a different spatial frequency rather than a location, with low-frequency information centred and high-frequency information situated on the edges [11]. The

K-space grid can be 2D or multidimensional and can be populated using sampling algorithms, such as the popular Cartesian sampling method [12]. These subject motions can be categorised into bulk motion and elastic motion. Bulk motion, which can be 1D or multi-dimensional translation, while the elastic motion is mainly rotations requiring 12 degrees of freedom, such as stretching movements [13]. These Motion artefacts are problematic because they cause blurring, signal loss, and contrast inconsistencies in the resulting image. These negatively impact image quality, leading to poor extraction of salient features by the DL pipeline. Poor feature extraction raises concerns about the reliability and accuracy of MIP techniques [14, 15]. Additionally, there is the issue of a low signal-to-noise ratio (SNR), particularly in fast-imaging protocols or low-field MRI systems, as noted by Ref [16]. This low SNR is characterised by a grainy appearance in the reconstructed image, making feature extraction difficult. In this case, the actual signal is obscured by random background noise. This results in images of lower clinical utility. Common causes of this background noise include low-field MRI systems, smaller voxel sizes, insufficient averaging, and physiological noise [16]. Low SNR reduces contrast between tissues and obscures subtle pathological features, complicating feature extraction and classification tasks [17].

Tackling these limitations requires advanced, targeted preprocessing strategies, such as motion correction and denoising algorithms. Followed by selecting appropriate DL architectures that can adapt to complex imaging constraints and can learn discriminative features during learning. Preventing the subject motion is the obvious solution to avoid the motion artefact problem entirely, but this is not practically possible; hence, techniques for correction and mitigation are current research interests aimed at improving MRI quality before it is used to train DL models. The basic idea behind all these methods is to undo the motion-related changes in the images, which makes post-processing, such as learning representations, complex. According to current literature, these methods can be classified into non-DL and DL-based approaches. Non-DL usually includes clinical intervention such as sedation, using mock MRI, intentional holding of breathing and the use of a vacuum immobiliser. Non-DL may also involve techniques such as radial FLASH (Fast Low-Angle Shot), a variant of a large family of fast gradient-echo methods [18, 19, 20], as well as parallel imaging to reduce acquisition times. However, increasing the acquisition speed may be a viable solution, enabled by advances in computing power and acceleration hardware. This could result in shorter exposure times and, consequently, shorter acquisition times, less subject motion, and improved image resolution. On the other hand, while considering this accelerated imaging, certain essential biological constraints which can be seen as limiting factors have to be also considered, such as specific absorption rate (SAR) limiting the use of Radiofrequency (RF) excitation pulse, rotation times as it relate to the sequence repetition and echo time to achieve the requested contrast to prevent patient heating, and subject rotation time which affect the image's contrast and SNR. These factors are critical because they can hinder the quality and safety of faster imaging. Hence, there is a need to balance acquisition speed with image quality and subject safety. DL methods have also been explored, including Principal Component Analysis (PCA) [21], high-order value decomposition recursive algorithms [22], conditional GAN [23], Convolutional Neural Network (CNN)-based variants [23, 24], cross-guided bilateral filters [25], and residual learning strategy [26]. A notable example is the work of Ref [27], which proposed a hybrid approach to correct motion artefacts in MRI by combining DL with compressed sensing (CS). Using 8,710 axial T2-weighted slices from 67 cases in the public IXI dataset, they trained a Convolutional Neural Network (CNN) to filter motion-corrupted images and identify affected phase-encoding (PE) lines in k-space. Then, the final reconstructed image was obtained after applying CS on the unaffected PE lines. The results showed that images with more than 35% unaffected PE lines, when processed with their method, achieved peak signal-to-noise ratios (PSNR) of 36.13-41.51. Also, with a structural similarity index (SSIM) of 0.950-0.979. This significantly outperforms CS alone and indicates better image quality in the presence of motion corruption. However, even though these achievements were realised, their method [27] relied on simulated motion, a limited modality scope, and a retrospective design. This further highlights the persistent difficulty of generalising motion correction strategies to real-world clinical MRI. The unpredictable subject movement and diverse imaging constraints pose unresolved challenges. See a sample of an MRI below in Figure 1.

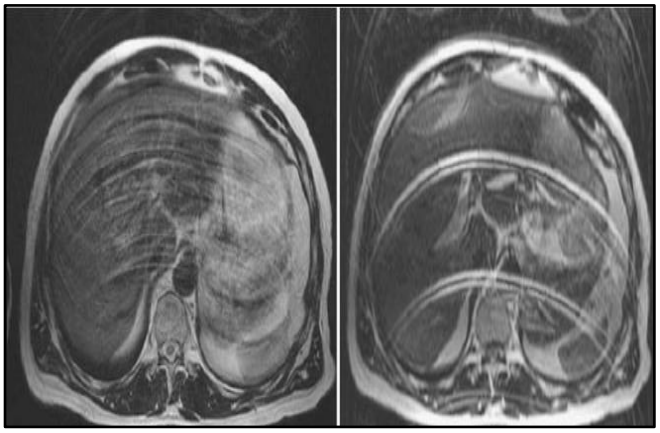

Fig 1 T2-Weighted Turbo Spin Echo (TSE) MRI Images Illustrating Motion Artefacts and Low SNR. The Image on the Left, Acquired with a Repetition Time (TR) of 2200 Milliseconds, Echo Time (TE) of 103 Milliseconds, and Echo Train Length (ETL) of 21, shows General Blurring Due to Irregular Respiratory Motion. The Image on the Right Demonstrates Ghosting Artefacts Caused by Periodic Breathing Patterns [28].

As further described by Ref [28], motion artefacts and low SNR can significantly affect the quality and interpretability of MRI data, especially in DL workflows where image fidelity is critical. For instance, random respiratory motion typically causes general blurring, while periodic breathing introduces ghosting artefacts that may obscure anatomical boundaries. Furthermore, apart from motion-related distortions, several other artefact types can

negatively impact image quality. Truncation artefacts appear as alternating bright and dark lines near sharp intensity transitions. These truncation artefacts are mainly problematic in brain and spine MRI, where they may mimic pathological features. Similarly, aliasing occurs when anatomical structures outside the field of view (FOV) are incorrectly mapped due to signal overlap, leading to misleading spatial representations. There are also artefacts arising from magnetic interactions due to microscopic variations in magnetic field strength. This can be seen at tissue interfaces and distorts signal intensity and spatial accuracy. Artefacts that manifest as focal signal anomalies at the image centre, often surrounded by concentric ringing, are known as central-point artefacts. Finally, phase encoding artefacts exhibit as linear noise with alternating pixel intensities along the encoding direction. These artefacts interfere with feature extraction and representation learning in DL models and workflows.

- *Computed Tomography*

CT is a widely adopted imaging modality known for its high spatial resolution and rapid acquisition, making it indispensable in emergency diagnostics, oncology, and cardiovascular imaging. CT has revolutionised the acquisition of regular X-ray-based medical imaging. CT imaging provides adequate 3D imaging, which has good applications in certain specialised areas such as breast cancer detection using traditional radiologist and trained mammographer intervention to check for the presence of lymphadenopathy, architectural distortion, skin thickening, and microcalcifications [29], and more recently using DL to learn representations from CT imaging for breast cancer detection [30]. However, CT imaging presents a critical trade-off between radiation dose and image quality. It is agreeable to say that there is a trade-off between image quality and optimising radiation dose. Lowering the radiation dose is essential for patient safety, particularly in paediatric imaging and repeated follow-up scans. However, it often results in increased image noise and reduced contrast, which can obscure anatomical boundaries and pathological features [31]. CT imaging exhibits lower spatial resolution, making it challenging to detect microcalcifications using computational methods. This lower spatial resolution negatively impacts the performance of DL models, which are typically trained on high-dose, high-quality datasets and may struggle to generalise to low-dose scenarios. Conventional segmentation models such as nnU-Net have shown significant performance drops when applied to ultra-low-dose CT imaging [31]. In addition to dose-related challenges, CT segmentation is complicated by anatomical ambiguity and overlapping tissue densities. Organs with similar attenuation values, such as liver and spleen, or tumour margins adjacent to vasculature, can be challenging to discriminate accurately, even with advanced DL models [32]. Figure 2 below shows CT images at low and high doses with respect to image quality.

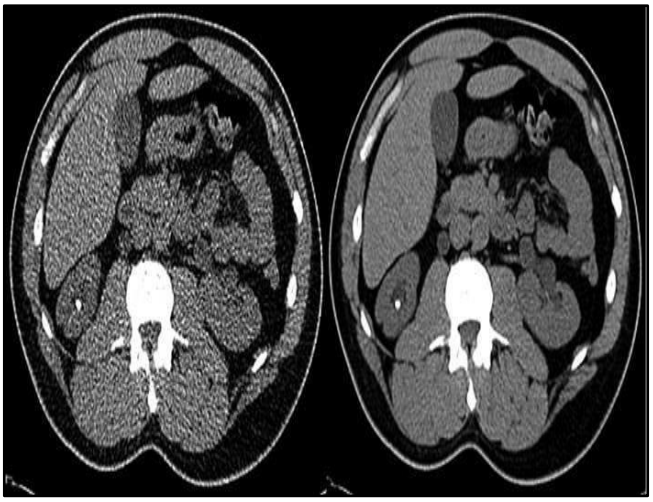

Fig 2 Low-Dose (Left) and High-Dose (Right) Scans of the Right Renal Calculus [33].

As shown in Figure 2, a higher dose yields better image quality; however, the trade-off problem remains. Finding a standard trade-off revalue is practically impossible, as supported by a study by Ref [34], which concluded that a clinically accepted lowest image acquisition dose level is task-dependent and cannot be standardised to a single value, as their experiments showed that 72.8 mAs was a safe dose when acquiring low-contrast objects. At the same time, 12.2 mAs was sufficient for visualising high-contrast objects. To enhance understanding of this image quality problem, Ref. [35] discussed various physics-based techniques, including the contrast-to-noise (CNR) ratio, region-of-interest (ROI)-based noise measurement, and SNR. These metrics provide quantifiable insights into image fidelity and diagnostic reliability, enabling an objective assessment of noise characteristics and contrast performance across different imaging protocols, as the dose-image quality trade-off is task-dependent. Furthermore, this low spatial resolution can be broadly categorised into blurring and ambiguity, as well as noise and artefacts. CT imaging at low resolution can blur functional anatomical structures, leading to uncertainty in learning decision boundaries when viewed in MIP, as highly pixelated regions can hinder the identification of an optimal relationship between the input and the desired outcome. However, denoising [35] and super-resolution [36] have been proposed and are being applied to improve the quality of these CT images. This comes at a cost of increased computational complexity and potential inadequacies, such as a lack of salient details in the processed image. It can be acknowledged that there are risks of automation bias and over-segmentation, which also pose clinical concerns, especially in radiotherapy planning and surgical navigation, where precision is crucial. Addressing these limitations specific to the CT imaging modalities requires not only architectural innovations but also robust image quality assurance frameworks and uncertainty quantification techniques to ensure the safe and reliable deployment of these systems.

- *Ultrasound Imaging*

With medical imaging for diagnostics becoming more widely available, one of the most widely used

modalities is ultrasound. It is less expensive, more compact and portable. It can provide real-time visual monitoring of internal organs and other parts of the human body, making it widely adopted in clinical settings. Consequently, it has become a significant source of data for MIP to improve patient treatment outcomes. Applications of MIP with UI have seen tremendous innovations in the diagnosis and early detection of abnormalities. However, this application of MIP using DL comes with unique challenges arising from the characteristics of the UI modalities, rather than the MIP technique itself.

A common challenge is low image quality due to speckle noise. This noise exhibits itself as a granular interference pattern resulting from the constructive and destructive interference of ultrasound waves. This interference is based on the physics of wave propagation in biological tissues. This interference results in the final image having lower resolution than natural images. Furthermore, issues such as motion blur and the previously mentioned low spatial resolution of the UI negatively affect the anatomical boundaries. This negative impact reduces the readability of the boundaries in the images, making diagnostics challenging, and is further worsened by the low-contrast regions. When the DL in the UI is considered during MIP, it complicates feature extraction and impairs efficient learning of salient features [37]. Furthermore, another major problem with the UI modality is the variability in image quality and varied representation across different acquisition equipment [38, 39], which has become a significant problem for human-based interpretation and DL based solutions, as the level of degradation caused by the scattering of the ultrasound signals depends on the construction algorithm and the acquisition equipment. A previous study used datasets from two popular vendors of UI acquisition equipment (GE Healthcare and Philips Healthcare) to examine how imaging quality varies across vendor configurations [40]. Though they [40] provided valuable insights into vendor-specific imaging characteristics, the diversity in pixel-level interactions and their relationship to image quality across different vendor equipment was not adequately discussed. This raises a crucial question: how can we design DL-based solutions for MIP that generalise across the variability of different vendors' acquisition devices? This current study aims to provide a critical understanding of how these UI imaging constraints interact with DL modelling challenges. Figure 3 below shows different UI from varying acquisition devices.

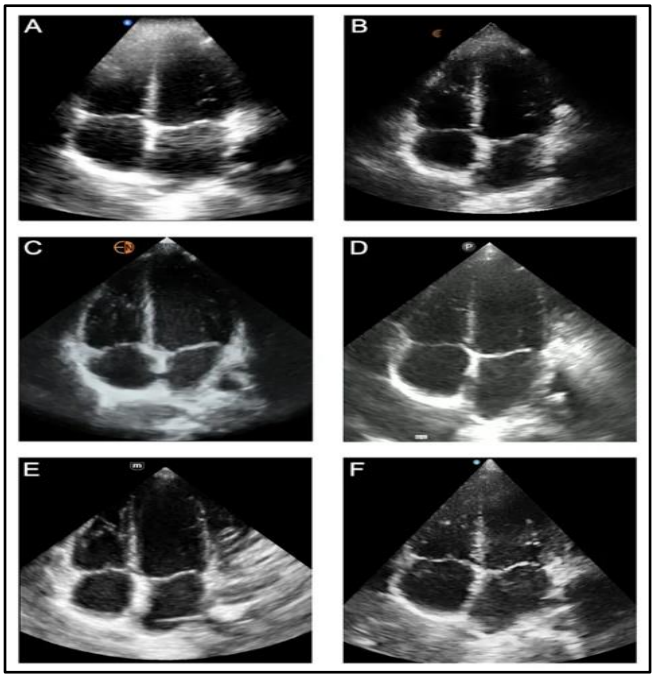

Fig 3 Four-Chamber Heart Ultrasound Images (Panels A–F) Taken from the Tip View show Differences in Quality Under Various Recording Conditions. Changes in Brightness, Tissue Visibility, and Clarity are Caused by how the Machine Processes Images and by how the Operator Performs the Scan [41].

As seen in Figure 3, poor image quality in several panels, including signal dropout and artefacts, highlights challenge for both clinical interpretation and DL reliability, emphasising the need for standardised acquisition and robust preprocessing. The variability in ultrasound images exacerbated by acquisition noise poses significant challenges for both clinical practitioners and DL researchers. Conventionally, due to limitations in learning only in the spatial domain, most DL models are inefficient at learning discriminative features from UI images containing anatomical structures, further worsened by varied imaging equipment acquisition. This complication may be related to the different proprietary signal-processing algorithms used to construct final UI images from ultrasound signals. These learning inefficiencies reduce the final model's generalisation ability when deployed in real-world scenarios. On the other hand, clinical practitioners may encounter diagnostic and interpretive errors when faced with this same problem if an unanticipated switch occurs between imaging systems. This switch can affect appearance and routine diagnostic expectations. Variability can lead to a domain shift and an adaptation gap, as the distributions of the training and real-world data may differ when the model is deployed.

To address the need for a solution, various methods for improving ultrasound image quality have been proposed in the literature. This will be categorised as a non-DL and DL-based approach. The non-DL approach primarily focuses on key factors, including environmental settings for capture, selection of the manufacturer and acquisition equipment, and significant moving parts such as the transducer. The environmental settings for the

capture approach are to ensure there are no other devices or medical equipment that can interfere with ultrasound signals during acquisition, such as MRI scanners, due to their metallic components and strong magnetic and radiofrequency fields, especially when simultaneous imaging is considered [42, 43]. The transducer concern relates to a thorough physical examination of this extension to ensure any observed irregularities are resolved, as these can lead to significant issues with ultrasound signals due to physical damage, cable problems, wear and tear, or manufacturing defects [44]. The selection of the manufacturer and the acquisition device plays a crucial role in shaping ultrasound image quality, as each device integrates the manufacturer's proprietary transducer designs and signal-processing algorithms that influence image resolution, contrast, and anatomical fidelity. These factors are critical to both clinical interpretation and AI model performance. However, these non-DL approaches are all subjective, as multiple factors contribute to the final ultrasound image's quality, such as clinical operator expertise, subject motion, body type, and the complexity of the clinical examination. On the other hand, DL approaches for UI enhancement encompass a diverse set of techniques, including adaptive residual learning-based denoising filters [45], CNN variants enhanced with attention mechanisms and skip connections typically evaluated using metrics such as structural similarity index (SSIM) and peak SNR [46], as well as super-resolution methods [47]. While these methods show promising results in controlled experimental settings, they often overlook critical challenges such as domain shift and acquisition-induced image variability. Specifically, differences in imaging protocols, equipment manufacturers, and signal-processing pipelines can lead to substantial variations in image characteristics, which are often overlooked during model training and evaluation. This gap limits the generalizability and clinical reliability of DL-based enhancement techniques across diverse real-world scenarios.

- *Histopathology Imaging*

With the rise of computational methods and the evolution of wholesale digital scanners, digital histopathological imaging has been widely adopted. Histopathology imaging plays a critical role in disease diagnosis and prognosis by enabling microscopic examination of tissue samples. In digital pathology, these samples are scanned to produce high-resolution whole-slide images (WSIs), often at a gigapixel scale. However, certain constraints in this modality negatively affect the application of MIP techniques, and a significant challenge is the high data density and ultra-high resolution in histopathological imaging, which pose significant computational and analytical challenges [48]. In terms of size, WSIs often exceed conventional gigabyte thresholds. This strains storage resources, requiring much more complex patch extraction to isolate regions. These regions are selected based on their relevance to diagnostic needs. Furthermore, there is a problem of high cost; the annotation burden is high. This is because experts are usually needed when labelling at the cellular level. There is also a high incidence of outliers, and rare abnormalities may behave as outliers. This can negatively affect the training of DL-based solutions if not correctly managed. For instance, in a fair comparison, a sample UI has around 300,000-800,000 pixels. CT and MRI imaging modalities can have between 6 million and 80 million pixels, depending on resolution and slice count. However, when WSIs are captured at 40× magnification, the expected output can get up to 10 billion pixels per slide. This massive number of pixels places a significant burden on storage infrastructure, expert labelling, and the experimental definition of outliers. All this makes modelling DL-based solutions for MIP a very complex and resource-intensive process.

Compared with other radiological imaging modalities, WSIs must be reduced to smaller tiles for analysis. This dimension reduction, though, reduces computing requirements; it can also affect the original spatial arrangements, potentially biasing feature aggregation. Moreover, the signal-to-noise ratio in WSIs is often low [48], with subtle diagnostic patterns embedded in vast regions of irrelevant tissue, complicating model generalisation and can be used to quantify image fidelity loss in the presence of compression artefacts [49]. Stain variability is a significant challenge in histopathological imaging, arising from differences in tissue preparation, staining protocols, and scanner hardware. Hematoxylin and Eosin (H&E) staining is the most common technique for morphological assessment [50]. Different laboratories often use different techniques, and this is even more common within batches [51]. This causes variation and inconsistencies in the colour distribution of the resulting images. The consequence of this for DL modelling is poor generalisation due to dependence on pixel-to-pixel interactions [52]. It is then important to agree that this issue of staining variability must be addressed through normalisation or domain adaptation strategies. Figure 4 presents H&E-stained adipose tissue across multiple magnifications.

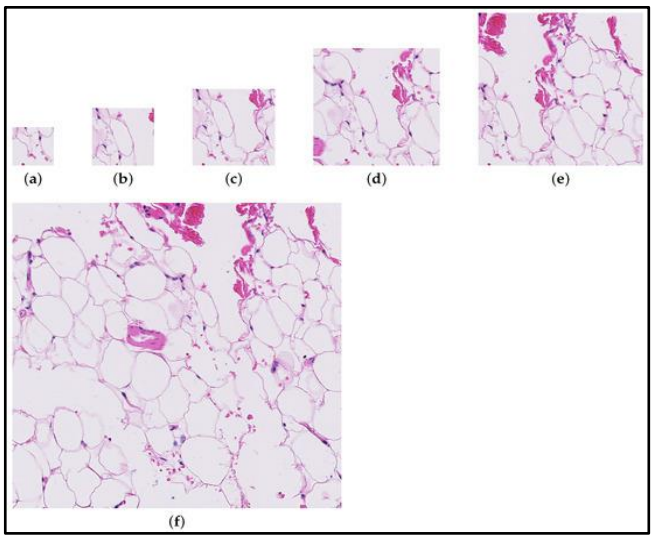

Fig 4 H&E-Stained Adipose Tissue at Varying Magnifications (a–f), Showing Cellular Detail, Contrast Differences, and Variability from Sample Preparation and Scanner Settings [53].

Figure 4 is a visual representation of cellular-level granularity and staining contrast consistent with those of WSIs. This reinforces the key modality-specific constraints discussed above. The high resolution required to capture such detail contributes directly to the gigapixel scale of WSIs, intensifying computational demands and complicating patch-level annotation. Variations in stain intensity and tissue morphology, as seen across panels, further highlight the challenge of stain variability and the need for robust normalisation strategies to ensure model generalisation across datasets.

To overcome the compounded challenges of gigapixel-scale resolution, stain variability, and low signal-to-noise ratios in histopathological imaging, DL must be tailored to the modality's unique constraints. Patch-based CNNs remain foundational, but their effectiveness depends on robust sampling strategies and context-preserving aggregation methods, as seen in tissue classification in Ref [54], where aggregate patch-level predictions on classification of glioma and non-small-cell lung carcinoma cases into subtypes and Ref [55], based on a multi-patch-based deep convolutional auto-encoder (DCAE) framework combined with a VGG19 pretrained model. Attention mechanisms and multiple instance learning (MIL) methods have recently shown promise in mitigating spatial fragmentation and localisation of disease-positive regions [56]. To address stain variability, domain adaptation techniques such as adversarial training and style transfer [57], alongside deep stain normalisation models, help reduce colour-induced domain shift while preserving diagnostic fidelity, using methods such as spectral normalisation and gradient penalty [58]. Despite these advances, generalisation across institutions remains limited, motivating the need for multi-source training, stain-invariant feature learning, and benchmarking on diverse datasets. Continued progress will require not only architectural innovation but also standardised pipelines and collaborative efforts to ensure reproducibility and clinical relevance.

IV. MODEL-SPECIFIC CHALLENGES

DL-based solutions has shown remarkable capabilities in MIP, with each architecture built with varying numbers of layers and architectural adjustments to improve feature learning, such as attention mechanisms. However, these DL-based solutions come with their own distinct challenges that affect their reliability and deployment in vital healthcare systems. This section examines the constraints of three major DL architectures: CNNs, Generative Adversarial Networks (GANs), and Transformers.

- *Convolutional Neural Networks*

CNNs' ability to understand translation invariance and pixel-pixel interactions has made them a foundational component of recent DL-based architectures. These characteristics make it well-suited for tasks such as segmentation, classification, and detection. A significant limitation of CNNs is their ability to understand spatial representations, mainly when image orientation differs between the validation and training sets. Furthermore, this spatial representation learning is mainly localised, leading to their inability to capture sufficient representation from global context and long-range dependencies, as also promoted by translation-invariant pooling layers. This localisation limitation can lead to information loss, negatively affecting MIP tasks that benefit from outlining boundaries, such as segmentation. Difficulty in understanding the global context of the entire image makes it hard to know whether there is a potential relationship between different parts, which may be helpful for the intended task. Furthermore, the interrelated challenges of high data requirements, computational complexity, and the need for more data remain, as deeper networks require more data. Hence, there is a significant model-specific challenge for CNNs in the medical domain, where large amounts of data are often unavailable and annotated data is scarce. This has made CNNs prone to overfitting, especially when trained on small or noisy datasets [59]. Overfitting manifests as high performance on training data but poor generalisation to unseen clinical cases, often exacerbated by domain shifts across institutions or imaging protocols [60]. Another critical limitation is the lack of interpretability. CNNs are seen as black-box models, making it challenging to map decision pathways in sufficient detail and to validate final predictions against human-based clinical reasoning critically. Though saliency feature maps and Grad-CAM visualisations offer some insights, there are issues with consistency, clinical understanding and relevance. This opacity limits trust in automated systems, especially in critical applications such as cancer diagnosis or surgical planning, where explainability is essential for regulatory approval and clinician adoption. Moreover, CNNs are sensitive to artefacts and acquisition noise, as discussed in Section 3. For instance, motion artefacts in MRI or speckle noise in ultrasound can distort convolutional feature maps, leading to misclassification or segmentation errors. Attempts to mitigate these effects through data augmentation or adversarial training have shown promise but remain limited by the diversity and realism of synthetic artefacts [1]. When these limitations are examined in the context of MIP for CNNs, they reveal issues of interpretability due to the stacked layers with multiple non-linear transformations, which can be described as having a blackbox nature.

There are data-related challenges with privacy issues, regulations, and the cost of annotation of medical datasets, which is a huge problem, and the high dependence on localised features makes this a significant barrier in clinical settings, when a global and broader concept of the image is required, while still presenting a computational cost problem.

With ongoing research, different proposed methods has been developed to mitigate these limitations of CNNs, such as transfer learning using pre-trained models to tackle the data scarcity issue, feature weighting and attention mechanism to improve interpretability, few shot learning to adapt to extremely small dataset, and more advanced unsupervised learning techniques to tackle the high cost

and resources of data annotation by training on large unlabelled data. However, each of these proposed solutions comes with its own unique challenges, such as architectural complexity in transfer learning, memory cost, attention bias and sensitivity to small objects and complex backgrounds. Redundant feature maps are observed in the application of attention mechanisms, and, lastly, generalisation limitations in few-shot learning methods. Hence, it's pivotal to consider all the limiting factors in parallel when proposing solutions for CNNs, to ensure an adequate trade-off is reached based on which evaluation metric is more important to the intended task, as well as the availability of computational resources.

- *Generative Adversarial Networks (GANs)*

GANs have gained traction in MIP for tasks such as image synthesis, denoising, and domain adaptation. Their ability to generate realistic images from limited data makes them attractive for augmenting training datasets and simulating rare pathologies. GANs' limitations can be categorised into model stability issues, data and architectural challenges and unreliability in generative decision.

Regarding stability, during GAN training, using two competing networks can lead to instability and collapse, especially when the generators output only a few images and the dataset lacks feature representation. Because of instability, convergence becomes a problem, especially when the discriminator improves performance and there is no relevant gradient information for the generator to learn from. This leads the model to collapse, resulting in limited diversity and failing to capture the full distribution of anatomical or pathological variations [61, 62]. This is particularly problematic in MIP, where subtle differences in lesion morphology or tissue texture can be diagnostically significant.

The data and architectural challenges include the high computational needs of GANs, very deep layers, and sometimes multiple branches to capture salient features. However, this becomes a problem because it is time-consuming in the medical domain, where resources are limited. Additionally, the need for a large dataset remains a criterion for GANs to perform well. Architectural complexity is seen as expensive, and hyperparameter tuning is needed, especially for sensitive ones, such as the learning rate. Without an adequate dataset and a well-suited architecture, GANs can generate hallucinated features, synthetic artefacts that do not align with real anatomical structures. These features can mislead interconnected models or end-user clinicians, especially when GAN outputs are used for training or clinical decision support. Studies have shown that GAN-generated images may exhibit unrealistic textures and boundary inconsistencies; hence, compromising diagnostic fidelity [63, 64]. Training GANs is also notably unstable, requiring specialised tuning of loss functions, learning rates, and the architectural trade-off between the generator and the discriminator. In clinical contexts, this instability translates to unpredictable performance and difficulty in benchmarking, limiting the reproducibility and regulatory acceptance of GAN-based systems.

The difficulty in understanding how GANs generate medical images makes their reliance and reliability in MIP and clinical settings a significant problem, exacerbated by unreliable evaluation metrics. However, different solutions have been proposed, such as post-processing image analysis to improve image quality, a pretrained model and transfer learning to address the need for a large dataset, and extensive architectural adjustments and modifications to lower parameter counts, including loss function redefinition. However, these proposed solutions have further increased complexity and worsened interpretability in MIP and clinical applications, without a significant reduction in computational cost.

- *Transformers*

Transformers have increasingly emerged as alternatives to CNN due to the need to focus on more meaningful parts of images and to understand long-range dependencies. The self-attention mechanism in transformers enables them to capture relationships across disparate regions of an input image. These relationships are then weighted and computed across the neural network, allowing it to focus on these regions. This computation mimics aspects of the physiological signal processing of the human brain. They were originally developed for natural language processing but have now been extended into computer vision and MIP. Some of these variants include Vision Transformers and Swin Transformers. These variants have been evaluated for MIP and have shown performance in segmentation and classification tasks as seen in retinal imaging applications [65, 66]. However, computing these self-attention weights often results in quadratic complexity, especially when processing high-resolution images. Consequently, there is the burden of high computational needs. These computational challenges are pronounced in WSIs and in 3D volumetric imaging. These storage and processing constraints can negatively impact real-time inference and deployment on edge devices [67]. When compared side by side with GANs, transformers also share certain limitations, such as stability issues, particularly when applied to high-dimensional image data. Furthermore, their reliance on positional encodings to learn spatial relationships limits their ability to efficiently represent spatial information.

In recent years, efforts have been made to reduce computational requirements. Architectural adjustments, such as patch-based processing and sparse attention mechanisms, have shown promise in this regard. However, it leads to reduced spatial dimension learning, as the ability to capture local details is poor and global dependence is reduced. Interpretability is another significant challenge, as although attention maps can be misleading, especially in clinical contexts where accurate localisation is essential [68]. When transformers are trained on small datasets without adequate regularisation or pretraining, they can cause them to memorise the training data rather than learn generalised representations, leading to overfit [69].

To address these limitations, hybrid models combining transformers with CNNs have been explored for classification, recognition, and related MIP tasks. These models have achieved notable performance, particularly when leveraging pretrained components [70,71]. However, in many medical applications, acquiring large-scale datasets for pretraining is impractical due to cost and privacy constraints. In such cases, hybrid architecture offers a pragmatic solution, enabling attention computation across pixel patches while leveraging CNNs for robust feature extraction.

In conclusion, while transformers have shown advantages in MIP applications, their use remains constrained by computational demands, interpretability challenges, and limited training data. Hybrid architectures that integrate CNNs and transformer-based attention mechanisms present a promising direction, balancing spatial precision with contextual depth.

This section has discussed these popular DL algorithms used for MIP at an individual stage; however, recent research has begun combining them into hybrid models. These hybrid models usually integrate different components of CNNs, transformers, and other architectures to take advantage of their individual strengths. Instances are seen in CNN-transformer hybrids for multi-scale feature extraction, GAN-CNN pipelines for artefact correction, and physics-informed DL systems [72, 73]. However, the architecture is complex, and the trade-off between computing requirements and performance complicates the entire applicability. Another consequence is longer training times and greater sensitivity to hyperparameter changes. In practice, hybrid models may inherit the limitations of their component architectures. Examples include CNNs' noise sensitivity, transformers' instability on limited data, and the need to address them adequately remains.

V. CROSS-CUTTING ISSUES, OPEN CHALLENGES & FUTURE DIRECTIONS

While modality-specific constraints have been explored in depth, their effects are magnified by broader systemic challenges that cut across medical imaging pipelines. These cross-cutting issues, often overlooked in existing reviews, play a decisive role in shaping model reliability, generalisation, and clinical applicability. This section critically examines how these systemic factors interact with both modality-specific and model-level limitations, revealing deeper barriers to robust and trustworthy deployment.

- *Data Scarcity and Model Uncertainty*

Small datasets have a significant impact on model uncertainty and remain a key cross-cutting issue across MIP. It is more common and visible when data are scarce and DL-based models face difficulty in learning discriminative patterns [74]. This leads to higher epistemic uncertainty, which is further amplified in multi-modal settings. This is because each imaging technique introduces distinct constraints, coupled with sample inconsistencies, that cannot be well captured with insufficient data during model training [75, 76]. This uncertainty can also erode the confidence and reliability of predictions, thereby affecting their reliability by clinical stakeholders in deploying them for highly critical scenarios. This intrinsic link between data scarcity and predictive uncertainty, therefore, represents a fundamental cross-cutting challenge in medical image analysis. A study by Ref [77] provides empirical insights into how this is a cross-cutting issue between small dataset problems and model uncertainty, as their systematic comparison using different uncertainty quantification methods showed that uncertainty estimates are not constant across datasets and different models. However, their [77] work showed that models lack sufficient knowledge of the unseen test data distribution when trained on small or limited datasets, which may be institution-specific. Though different techniques, such as few-shot learning in MIP to reduce data scarcity issues [78], these methods remain sensitive to domain based on reliance on prototype-based classification assumption, worsen model uncertainty when applied to rare imaging cases, limitation in global context learning when applied to high density images, as well as limited explainability in their output.

In summary, addressing the intertwined challenges of data scarcity and model uncertainty in MIP requires not only more diverse and representative datasets but also the development of reliable, interpretable learning solutions, such as uncertainty-calibrated modelling techniques that generalise across modalities and institutions.

- *Domain Shift*

Domain shift arising from differences in acquisition protocols, scanner hardware, staining procedures, or patient demographics further undermines generalisation. According to Ref [79], this occurs when the distribution of the training data differs from that of the testing data. This is a major problem in MIP using DL, causing a significant negative impact on overall performance. Models trained on curated datasets often fail when exposed to external varying data distributions, leading to unstable performance and possible overfitting. This is particularly problematic in federated learning contexts, where data heterogeneity across institutions challenges aggregation and convergence. Differences in acquisition between the two, regarding the device and technical human expertise, can result in varying image intensity and, hence, differences in data distribution. This challenge is shown in Figure 5, which visualises how scanner-specific acquisition differences can lead to substantial shifts in image intensity distributions—highlighting the risk of domain mismatch and its impact on model generalisation across heterogeneous clinical datasets.

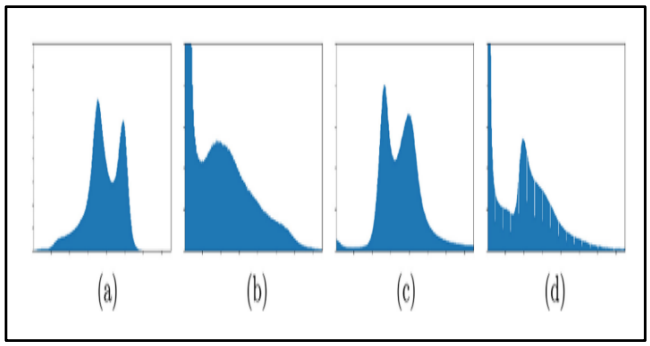

Fig 5 Device-Specific Shifts in Image Intensity and Texture Distribution [80].

Figure 5 from Ref [80] offers a compelling visual demonstration of how scanner-induced variability can distort the intensity distributions of MRI data, with direct implications for DL segmentation. Subplot (a) presents a clean bimodal distribution, likely reflecting well-separated tissue classes under consistent acquisition conditions. In contrast, subplot (b) shows a skewed distribution from a different scanner, where acquisition parameters have shifted the intensity profile, introducing asymmetry and reducing class separability. Subplot (c), a smoothed version of (a), highlights how preprocessing choices such as kernel density estimation can clarify underlying patterns but may also obscure subtle artefacts critical for clinical interpretation. These plots underscore a persistent challenge in MIP: the assumption of distributional homogeneity across datasets is often invalid. Scanner-specific biases rooted in hardware, calibration, and protocol differences can introduce domain shifts that undermine model generalisation. While Karani et al. propose domain-specific batch normalisation as a mitigation strategy, the broader issue remains. Without robust adaptation mechanisms, even well-trained models risk failure when exposed to unseen acquisition domains. This figure thus serves not only as a technical illustration but as a cautionary reminder of the fragility of DL models in the face of real-world heterogeneity. Domain adaptation has been proposed as a strategy for mitigating distributional shifts between source and target domains, particularly when both domains share the same learning task but differ in constraints and representation characteristics. In medical imaging, this challenge is amplified by the high dimensionality of data, such as 3D and 4D volumes with dense pixel matrices and rich contextual information, which complicates the design of models capable of robust adaptation. Techniques such as adversarial domain adaptation [81], feature alignment using Maximum Mean Discrepancy [64], and domain-specific batch normalisation [82], have shown promise in aligning feature distributions across domains. However, these methods often struggle to scale effectively in the presence of inter-modality heterogeneity, where differences in imaging physics, resolution, and anatomical representation further hinder knowledge transfer [79]. This is especially problematic in multi-source settings, such as federated learning, where data from diverse institutions and modalities must be harmonised without direct access to raw images. Consequently, there remains a critical need for research into multi-source domain adaptation frameworks that not only accommodate high-dimensional medical data but also offer interpretability and clinical relevance to support trustworthy deployment in real-world settings.

- *Regulatory and Ethical Constraints*

In MIP, regulatory and ethical constraints include data privacy issues, the need for explainability of AI solutions, potential workforce disruption, and biases arising from limited data diversity [83]. The inherent opacity of models such as CNNs and transformers raises persistent concerns around explainability, accountability, and clinical trust. At the same time, explainable AI (XAI) techniques such as saliency maps and attention visualisations have been used in medical imaging applications, including diabetic retinopathy diagnosis [84]. They often lack modality-specific grounding and fail to deliver actionable insights for clinical practitioners. They often provide post hoc explanations that highlight regions of high model sensitivity, but do not necessarily reflect the true causal features driving the prediction. Furthermore, evaluation practices remain fragmented with respect to these maps, with inconsistent metrics and limited external validation, as different explanation methods can produce conflicting saliency maps for the same image, making it difficult to determine which interpretation is accurate and clinically meaningful. Addressing these challenges requires a paradigm shift from isolated architectural innovation toward integrated, clinically grounded solutions. This includes developing stain-invariant and acquisition-aware representations, uncertainty-aware learning modelling techniques, and benchmarking platforms that reflect real-world deployment conditions. Additionally, there is a need to embed ethical auditing, regulatory alignment, and stakeholder engagement throughout the model lifecycle rather than append them post hoc. Ultimately, resolving these cross-cutting issues demands a reorientation of research priorities: from performance-centric optimisation to reproducibility, interpretability, and equitable deployment in clinical practice.

To consolidate the insights presented across modalities and model architectures, Table 1 synthesises the key modality-specific challenges, model-specific limitations, and cross-cutting issues that constrain DL performance and deployment in MIP.

Table 1 Summary of Modality-Specific Challenges, Model-Specific Limitations, and Cross-Cutting Issues in Deep Learning for Medical Image Processing.

Modality	Modality-Specific Constraints	Model- Specific Constraints	Cross-Cutting Issues	Representative Mitigation Strategies
MRI	Subject motion artefacts (bulk and elastic). Low SNR. K-space distortions. Biological constraints (SAR, RF pulse limits).	CNNs overfit to motion artefacts. GANs struggle with phase encoding irregularities. Transformers are sensitive to spatial inconsistency.	Annotation noise due to blurred boundaries. Domain shift from acquisition variability. Data scarcity in subject motion-corrupted datasets.	Compressed sensing + CNN hybrid methods. Motion correction filters (such as PCA, residual learning). DL-based denoising and PE line filtering.
CT	Dose-quality trade-off. Low spatial resolution. Overlapping tissue densities. Task-dependent acquisition protocols.	CNNs misclassify blurred or ambiguous regions. Segmentation models degrade on low-dose scans. GANs may amplify artefacts.	Automation bias in low-contrast regions. Uncertainty in anatomical boundary detection. Annotation inconsistency across dose levels.	Super-resolution and denoising filters. Dose-aware training and contrast-to-noise ratio metrics. Uncertainty quantification frameworks.
Ultrasound	Speckle noise from wave interference Subject motion blur and low spatial resolution. Acquisition device-specific variability. Operator-dependent acquisition.	DL models fail to generalise across devices. Poor anatomical boundary detection. CNNs are sensitive to gain and contrast shifts.	Domain shift due to vendor-specific signal processing and acquisition device. Annotation inconsistency from the imaging technician's technique. Data scarcity in multi-device datasets.	Adaptive residual learning filters. Domain adaptation. Transducer-aware preprocessing and harmonisation.
Histopathology	Gigapixel-scale WSIs. Stain variability across laboratories. High annotation burden at the cellular level. Presence of rare outliers.	Patch-level inconsistency in CNNs. Transformers struggle with context aggregation. GANs amplify stain bias and artefacts.	Data scarcity for rare abnormalities. Annotation noise from expert variability. Domain shift across institutions and scanners.	Patch sampling and stain normalisation. Federated learning for multi-institutional training. DL-based stain harmonisation and outlier detection.

Table 1 provides a summarised synthesis of cross-cutting issues that constrain DL performance in MIP. It also reveals the uneven maturity of mitigation strategies; some are well-established (for instance, CS in MRI). In contrast, others remain insufficiently explored (for instance, transducer-aware harmonisation in ultrasound and varied sourced datasets from different vendor acquisition devices). The table emphasises the need for integrated DL model evaluation techniques that account for acquisition variability, annotation noise, subject motion-induced artefact, and domain shift, which are often discussed superficially in benchmark-driven studies. By critically assessing these interactions, the table reinforces the study's central argument: that robust, clinically deployable DL systems for MIP must be designed with modality-aware constraints and model-specific vulnerabilities in mind.

VI. CONCLUSION

This study presents a focused, grounded synthesis of the limitations of DL-based solutions in MIP. It goes beyond previous isolated reviews by emphasising the interactions between modality constraints and model-specific challenges. A key contribution of this study is its integration of cross-cutting issues such as data scarcity and model uncertainty, annotation noise, domain shift, and regulatory constraints into the analysis of modality-model interactions. This approach provides valuable insights into how technical limitations are often inseparable from clinical considerations. By bridging the gap between technical design and applicability on imaging modalities, this study equips early researchers and DL practitioners with a more actionable understanding of DL limitations in MIP. Future work will focus on developing acquisition-aware DL-based solutions and conducting empirical analysis of cross-modality generalisation benchmarking across diverse datasets.

REFERENCES

[1]. Zhang H, Shinomiya Y, Yoshida S. 3D MRI Reconstruction Based on 2D Generative Adversarial Network Super-Resolution. Sensors. 2021;21:2978. Available from: http//dx.doi.org/10.3390/s21092978

[2]. Kakigi T, Sakamoto R, Tagawa H, Kuriyama S, Goto Y, Nambu M, et al. Diagnostic advantage of thin slice 2D MRI and multiplanar reconstruction of the knee joint using deep learning based denoising approach. Sci Rep. 2022;12. Available from: http://dx.doi.org/10.1038/s41598-022-14190-1

[3]. Park JJ, Tiefenbach J, Demetriades AK. The role of artificial intelligence in surgical simulation. Front Med Technol . 2022;4. Available from: http://dx.doi.org/10.3389/fmedt.2022.1076755

[4]. Ma L, Xiao D, Kim D, Lian C, Kuang T, Liu Q, et al. Simulation of Postoperative Facial Appearances via Geometric Deep Learning for Efficient Orthognathic Surgical Planning. IEEE Trans Med Imaging . 2023;42:336–45. Available from: http://dx.doi.org/10.1109/TMI.2022.3180078

[5]. Aburass S, Dorgham O, Al Shaqsi J, Abu Rumman M, Al-Kadi O. Vision Transformers in Medical Imaging: a Comprehensive Review of Advancements and Applications Across Multiple Diseases. J Digit Imaging Inform med. 2025; Available from: http://dx.doi.org/10.1007/s10278-025-01481-y

[6]. Saba L, d'Aloja E. Predictive techniques in medical imaging: opportunities, limitations, and ethical-economic challenges. npj Digit Med . 2025;8. Available from: http://dx.doi.org/10.1038/s41746-025-01791-z

[7]. Armato SG, Drukker K, Hadjiiski L. AI in medical imaging grand challenges: translation from competition to research benefit and patient care. The British Journal of Radiology . 2023;96. Available from: http://dx.doi.org/10.1259/bjr.20221152

[8]. Kantarcı M, Aydın S, Oğul H, Kızılgöz V. New imaging techniques and trends in radiology. dir. 2025; Available from: http://dx.doi.org/10.4274/dir.2024.242926

[9]. Abhisheka B, Biswas SK, Purkayastha B, Das D, Escargueil A. Recent trend in medical imaging modalities and their applications in disease diagnosis: a review. Multimed Tools Appl . 2023;83:43035–70. Available from: http://dx.doi.org/10.1007/s11042-023-17326-1

[10]. Zaitsev, M., Maclaren, J., & Herbst, M. (2015). Motion artifacts in MRI: A complex problem with many partial solutions. Journal of Magnetic Resonance Imaging, 42(4), 887–901. https://doi.org/10.1002/jmri.24850

[11]. Moratal, D., Valles-Luch, A., Marti-Bonmati, L., & Brummer, M. (2008). k-Space tutorial: an MRI educational tool for a better understanding of k-space. Biomedical Imaging and Intervention Journal, 4(1). https://doi.org/10.2349/biij.4.1.e15

[12]. Eyre, K., Lindsay, K., Razzaq, S., Chetrit, M., & Friedrich, M. (2022). Simultaneous multi-parametric acquisition and reconstruction techniques in cardiac magnetic resonance imaging: Basic concepts and status of clinical development. Frontiers in Cardiovascular Medicine, 9. https://doi.org/10.3389/fcvm.2022.953823

[13]. Callaghan, P. T. (2011). Translational Dynamics and Magnetic Resonance. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199556984.001.0001

[14]. Havsteen, I., Ohlhues, A., Madsen, K. H., Nybing, J. D., Christensen, H., & Christensen, A. (2017). Are Movement Artifacts in Magnetic Resonance Imaging a Real Problem?—A Narrative Review. Frontiers in Neurology, 8. https://doi.org/10.3389/fneur.2017.00232

[15]. Ferreira, P. F., Gatehouse, P. D., Mohiaddin, R. H., & Firmin, D. N. (2013). Cardiovascular magnetic resonance artefacts. Journal of Cardiovascular Magnetic Resonance, 15(1), 41. https://doi.org/10.1186/1532-429x-15-41

[16]. Ayde, R., Vornehm, M., Zhao, Y., Knoll, F., Wu, E. X., & Sarracanie, M. (2024). MRI at low field: A review of software solutions for improving SNR. NMR in Biomedicine, 38(1). https://doi.org/10.1002/nbm.5268

[17]. Muro, I., Isoiwa, T., Shibukawa, S., Usui, K., Otsuka, Y. (2024). Using Deep Learning to Simultaneously Reduce Noise and Motion Artifacts in Brain MR Imaging. Magnetic Resonance in Medical Sciences. [Available at: https://doi.org/10.2463/mrms.mp.2024-0098]

[18]. Uecker, M., Zhang, S., Voit, D., Merboldt, K.-D., & Frahm, J. (2012). Real-time MRI: recent advances using radial FLASH. Imaging in Medicine, 4(4), 461–476. https://doi.org/10.2217/iim.12.32

[19]. Zhang H, Shinomiya Y, Yoshida S. 3D MRI Reconstruction Based on 2D Generative Adversarial Network Super-Resolution. Sensors. 2021;21:2978. Available from: http://dx.doi.org/10.3390/s21092978

[20]. Haase, A. (1994). Principles and applications of FLASH NMR imaging. Magma: Magnetic Resonance Materials in Physics, Biology, and Medicine, 2(3), 157–160. https://doi.org/10.1007/bf01705235

[21]. Bie, C., Liang, Y., Zhang, L., Zhao, Y., Chen, Y., Zhang, X., He, X., & Song, X. (2019). Motion correction of chemical exchange saturation transfer MRI series using robust principal component analysis (RPCA) and PCA. Quantitative Imaging in Medicine and Surgery, 9(10), 1697–1713. https://doi.org/10.21037/qims.2019.09.14

[22]. Wang, L., Xiao, D., Hou, W. S., Wu, X. Y., & Chen, L. (2020). A Modified Higher-Order Singular Value Decomposition Framework With Adaptive Multilinear Tensor Rank Approximation for Three-Dimensional Magnetic Resonance Rician Noise Removal. Frontiers in Oncology, 10. https://doi.org/10.3389/fonc.2020.01640

[23]. Hewlett, M., Petrov, I., Johnson, P. M., & Drangova, M. (2024). Deep-learning-based motion correction using multichannel MRI data: a study

using simulated artifacts in the fastMRI dataset. NMR in Biomedicine, 37(10). https://doi.org/10.1002/nbm.5179

[24]. Chang, Y., Li, Z., Saju, G., Mao, H., & Liu, T. (2023). Deep learning-based rigid motion correction for magnetic resonance imaging: A survey. Meta-Radiology, 1(1), 100001. https://doi.org/10.1016/j.metrad.2023.100001

[25]. Samuel, S., Ochawar, R. S., & Rukmini, M. S. S. (2022). Motion Correction of MR Images Using Cross-Guided Bilateral Filter. In Lecture Notes in Electrical Engineering (pp. 379–392). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-1906-0_34

[26]. Pirkl, C. M., Cencini, M., Kurzawski, J. W., Waldmannstetter, D., Li, H., Sekuboyina, A., Endt, S., Peretti, L., Donatelli, G., Pasquariello, R., Costagli, M., Buonincontri, G., Tosetti, M., Menzel, M. I., & Menze, B. H. (2022). Learning residual motion correction for fast and robust 3D multiparametric MRI. Medical Image Analysis, 77, 102387. https://doi.org/10.1016/j.media.2022.102387

[27]. Cui, L., Song, Y., Wang, Y., Wang, R., Wu, D., Xie, H., Li, J., & Yang, G. (2023). Motion artifact reduction for magnetic resonance imaging with deep learning and k-space analysis. PLOS ONE, 18(1), e0278668. https://doi.org/10.1371/journal.pone.0278668

[28]. Stadler, A., Schima, W., Ba-Ssalamah, A., Kettenbach, J., & Eisenhuber, E. (2006). Artifacts in body MR imaging: their appearance and how to eliminate them. European Radiology, 17(5), 1242–1255. https://doi.org/10.1007/s00330-006-0470-4

[29]. Desperito, E., Schwartz, L., Capaccione, K. M., Collins, B. T., Jamabawalikar, S., Peng, B., Patrizio, R., & Salvatore, M. M. (2022). Chest CT for Breast Cancer Diagnosis. Life, 12(11), 1699. https://doi.org/10.3390/life12111699

[30]. Yao, I. Z., Dong, M., & Hwang, W. Y. K. (2025). Deep Learning Applications in Clinical Cancer Detection: A Review of Implementation Challenges and Solutions. Mayo Clinic Proceedings: Digital Health, 3(3), 100253. https://doi.org/10.1016/j.mcpdig.2025.100253

[31]. Salimi, Y., Mansouri, Z., Sun, C., Sanaat, A., Yazdanpanah, M., Shooli, H., Nkoulou, R., Boudabbous, S. and Zaidi, H. (2025). Deep learning-based segmentation of ultra-low-dose CT images using an optimized nnU-Net model. La Radiologia Medica, 130, pp.723–739. [Available at: https://link.springer.com/article/10.1007/s11547-025-01989-x]

[32]. De Biase, A., Sijtsema, N.M., Janssen, T., Hurkmans, C., Brouwer, C. and van Ooijen, P. (2024). Clinical adoption of deep learning target auto-segmentation for radiation therapy: challenges, clinical risks, and mitigation strategies. BJR|Artificial Intelligence, 1(1), ubae015. [Available at: https://doi.org/10.1093/bjrai/ubae015]

[33]. Malkawi, I. M., Han, E., Atalla, C. S., Santucci, R. A., O'Neil, B., & Wynberg, J. B. (2016). Low-Dose (10%) Computed Tomography May Be Inferior to Standard-Dose CT in the Evaluation of Acute Renal Colic in the Emergency Room Setting. Journal of Endourology, 30(5), 493–496. https://doi.org/10.1089/end.2015.0760

[34]. Yan, H., Cervino, L., Jia, X., & Jiang, S. B. (2012). A comprehensive study on the relationship between the image quality and imaging dose in low-dose cone beam CT. Physics in Medicine and Biology, 57(7), 2063–2080. https://doi.org/10.1088/0031-9155/57/7/2063

[35]. Hoeijmakers, E. J. I., Martens, B., Wildberger, J. E., Flohr, T. G., & Jeukens, C. R. L. P. N. (2025). Objective assessment of diagnostic image quality in CT scans: what radiologists and researchers need to know. Insights into Imaging, 16(1). https://doi.org/10.1186/s13244-025-02037-y

[36]. Sadia, R. T., Chen, J., & Zhang, J. (2024). CT image denoising methods for image quality improvement and radiation dose reduction. Journal of Applied Clinical Medical Physics, 25(2). https://doi.org/10.1002/acm2.14270

[37]. Zhang, Z., Yu, S., Qin, W., Liang, X., Xie, Y., & Cao, G. (2021). Self-supervised CT super-resolution with hybrid model. Computers in Biology and Medicine, 138, 104775. https://doi.org/10.1016/j.compbiomed.2021.104775

[38]. Kavand, A., Bekrani, M. Speckle noise removal in medical ultrasonic image using spatial filters and DnCNN. Multimed Tools Appl 83, 45903–45920 (2024). https://doi.org/10.1007/s11042-023-17374-7

[39]. Weimer, J. M., Beer, D., Schneider, C., Yousefzada, M., Gottwald, M., Züllich, T. F., Weimer, A., Jonck, C., Buggenhagen, H., Kloeckner, R., & Merkel, D. (2023). Inter-System Variability of Eight Different Handheld Ultrasound (HHUS) Devices—A Prospective Comparison of B-Scan Quality and Clinical Significance in Intensive Care. Diagnostics, 14(1), 54. https://doi.org/10.3390/diagnostics14010054

[40]. Tenajas, R., & Miraut, D. (2024). Challenges in standardizing image quality across diverse ultrasound devices. Intelligent Medicine, 4(4), 275. https://doi.org/10.1016/j.imed.2024.01.002

[41]. Labs, R. B., Vrettos, A., Loo, J., & Zolgharni, M. (2023). Automated assessment of transthoracic echocardiogram image quality using deep neural networks. Intelligent Medicine, 3(3), 191–199. https://doi.org/10.1016/j.imed.2022.08.001

[42]. Perez-Sanchez, A., Johnson, G., Pucks, N., Soni, R. N., Lund, T. J. S., Andrade, A. J., Le, M.-P. T., Solis-McCarthy, J., Wong, T., Ashraf, A., Kumar, A. D., Banauch, G. I., Verner, J. R., Sodhi, A., Thomas, M. K., LoPresti, C., Schmitz, H., Koratala, A., Hunninghake, J., … Soni, N. J. (2024). Comparison of 6 handheld ultrasound devices by point-of-care ultrasound experts: a cross-sectional

study. The Ultrasound Journal, 16(1). https://doi.org/10.1186/s13089-024-00392-3

[43]. Curiel, L., Chopra, R., & Hynynen, K. (2007). Progress in Multimodality Imaging: Truly Simultaneous Ultrasound and Magnetic Resonance Imaging. IEEE Transactions on Medical Imaging, 26(12), 1740–1746. https://doi.org/10.1109/tmi.2007.903572

[44]. Wang, G., Kalra, M., Murugan, V., Xi, Y., Gjesteby, L., Getzin, M., Yang, Q., Cong, W., & Vannier, M. (2015). Vision 20/20: Simultaneous CT-MRI — Next chapter of multimodality imaging. Medical Physics, 42(10), 5879–5889. https://doi.org/10.1118/1.4929559

[45]. Lee, W., & Roh, Y. (2017). Ultrasonic transducers for medical diagnostic imaging. Biomedical Engineering Letters, 7(2), 91–97. https://doi.org/10.1007/s13534-017-0021-8

[46]. Abbasian Ardakani, A., Mohammadi, A., Vogl, T. J., Kuzan, T. Y., & Acharya, U. R. (2023). <scp>AdaRes</scp>: A deep learning-based model for ultrasound image denoising: Results of image quality metrics, radiomics, artificial intelligence, and clinical studies. Journal of Clinical Ultrasound, 52(2), 131–143. https://doi.org/10.1002/jcu.23607

[47]. Lei, M., Zhang, W., Zhang, T., Wu, Y., Gao, D., Tao, X., Li, K., Shao, X., & Yang, Y. (2024). Improvement of low-frequency ultrasonic image quality using a enhanced convolutional neural network. Sensors and Actuators A: Physical, 365, 114878. https://doi.org/10.1016/j.sna.2023.114878

[48]. Mhatre, S. R., & Bakal, J. W. (2024). DeepFusion: A Novel Deep Learning Technique for Enhanced Image Super-Resolution. In 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 991–998). 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS). IEEE. https://doi.org/10.1109/icacrs62842.2024.10841630

[49]. Komura, D., & Ishikawa, S. (2018). Machine Learning Methods for Histopathological Image Analysis. Computational and Structural Biotechnology Journal, 16, 34–42. https://doi.org/10.1016/j.csbj.2018.01.001

[50]. Reddy, R. K., & Bhargava, R. (2010). Accurate histopathology from low signal-to-noise ratio spectroscopic imaging data. The Analyst, 135(11), 2818. https://doi.org/10.1039/c0an00350f

[51]. Chen, Y., Janowczyk, A., & Madabhushi, A. (2020). Quantitative Assessment of the Effects of Compression on Deep Learning in Digital Pathology Image Analysis. JCO Clinical Cancer Informatics, 4, 221–233. https://doi.org/10.1200/cci.19.00068

[52]. Ariotta, V., Lehtonen, O., Salloum, S., Micoli, G., Lavikka, K., Rantanen, V., Hynninen, J., Virtanen, A., & Hautaniemi, S. (2023). H&E image analysis pipeline for quantifying morphological features. Journal of Pathology Informatics, 14, 100339. https://doi.org/10.1016/j.jpi.2023.100339

[53]. Dunn, C., Brettle, D., Cockroft, M., Keating, E., Revie, C., & Treanor, D. (2024). Quantitative assessment of H&E staining for pathology: development and clinical evaluation of a novel system. Diagnostic Pathology, 19(1). https://doi.org/10.1186/s13000-024-01461-w

[54]. Xu, C., Sun, Y., Zhang, Y., Liu, T., Wang, X., Hu, D., Huang, S., Li, J., Zhang, F. and Li, G. (2025). Stain normalization of histopathological images based on deep learning: A review. Diagnostics, 15(8), p.1032. https://doi.org/10.3390/diagnostics15081032

[55]. Jenkinson, E., & Arandjelović, O. (2024). Whole Slide Image Understanding in Pathology: What Is the Salient Scale of Analysis? BioMedInformatics, 4(1), 489–518. https://doi.org/10.3390/biomedinformatics4010028

[56]. Hou, L., Samaras, D., Kurc, T. M., Gao, Y., Davis, J. E., & Saltz, J. H. (2016). Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. https://doi.org/10.1109/cvpr.2016.266

[57]. Ponraj, A., Nagaraj, P., Balakrishnan, D., Srinivasu, P. N., Shafi, J., Kim, W., & Ijaz, M. F. (2025). A multi-patch-based deep learning model with VGG19 for breast cancer classifications in the pathology images. DIGITAL HEALTH, 11. https://doi.org/10.1177/20552076241313161

[58]. Cai, L., Huang, S., Zhang, Y., Lu, J., & Zhang, Y. (2025). AttriMIL: Revisiting attention-based multiple instance learning for whole-slide pathological image classification from a perspective of instance attributes. Medical Image Analysis, 103, 103631. https://doi.org/10.1016/j.media.2025.103631

[59]. Alajaji, S. A., Khoury, Z. H., Elgharib, M., Saeed, M., Ahmed, A. R. H., Khan, M. B., Tavares, T., Jessri, M., Puche, A. C., Hoorfar, H., Stojanov, I., Sciubba, J. J., & Sultan, A. S. (2024). Generative Adversarial Networks in Digital Histopathology: Current Applications, Limitations, Ethical Considerations, and Future Directions. Modern Pathology, 37(1), 100369. https://doi.org/10.1016/j.modpat.2023.100369

[60]. Mahmood, F., Borders, D., Chen, R. J., Mckay, G. N., Salimian, K. J., Baras, A., & Durr, N. J. (2020). Deep Adversarial Training for Multi-Organ Nuclei Segmentation in Histopathology Images. IEEE Transactions on Medical Imaging, 39(11), 3257–3267. https://doi.org/10.1109/tmi.2019.2927182

[61]. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N. and Liang, J., 2023. Deep learning in medical imaging: A review of CNN architectures and training strategies. Journal of Imaging Informatics in Medicine, 36(2), pp.204–230. https://doi.org/10.1007/s10278-023-00721-9

[62]. Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R.G., de Leeuw, F.E., Tempany, C.M.,

van Ginneken, B., Fedorov, A., Abolmaesumi, P., Platel, B. and Wells, W.M., 2017. Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation. MICCAI 2017, Lecture Notes in Computer Science, 10435, pp.516–524. https://doi.org/10.1007/978-3-319-66179-7_59

[63]. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. Advances in Neural Information Processing Systems, 27.

[64]. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J. and Greenspan, H., 2018. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing, 321, pp.321–331. https://doi.org/10.1016/j.neucom.2018.09.013

[65]. Yi, X., Walia, E. and Babyn, P., 2019. Generative adversarial network in medical imaging: A review. Medical Image Analysis, 58, p.101552. https://doi.org/10.1016/j.media.2019.101552

[66]. Wang, M., Chen, D., Tan, F., Liang, T., Lan, L., Zhang, X., & Luo, Z. (2022). Domain-specific feature recalibration and alignment for multi-source unsupervised domain adaptation. IET Computer Vision, 17(1), 26–38. https://doi.org/10.1049/cvi2.12126

[67]. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N., 2021. An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/2010.11929

[68]. Aburass, S., Dorgham, O., Al Shaqsi, J., Abu Rumman, M. and Al-Kadi, O., 2025. Vision transformers in medical imaging: A comprehensive review of advancements and applications across multiple diseases. Journal of Imaging Informatics in Medicine. https://doi.org/10.1007/s10278-025-01481-y

[69]. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R. and Xu, D., 2022. UNETR: Transformers for 3D medical image segmentation. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp.1748–1758. https://doi.org/10.1109/WACV51458.2022.00181

[70]. Rahman, M. I. (2025). Fusion of Vision Transformer and Convolutional Neural Network for Explainable and Efficient Histopathological Image Classification in Cyber-Physical Healthcare Systems. Journal of Transformative Technologies and Sustainable Development, 9(1). https://doi.org/10.1007/s41314-025-00079-0

[71]. Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. AI Open, 3, 111–132. https://doi.org/10.1016/j.aiopen.2022.10.001

[72]. Takahashi, S., Sakaguchi, Y., Kouno, N., Takasawa, K., Ishizu, K., Akagi, Y., Aoyama, R., Teraya, N., Bolatkan, A., Shinkai, N., Machino, H., Kobayashi, K., Asada, K., Komatsu, M., Kaneko, S., Sugiyama, M., & Hamamoto, R. (2024). Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review. Journal of Medical Systems, 48(1). https://doi.org/10.1007/s10916-024-02105-8

[73]. Liu, X., Hu, Y., & Chen, J. (2023). Hybrid CNN-Transformer model for medical image segmentation with pyramid convolution and multi-layer perceptron. Biomedical Signal Processing and Control, 86, 105331. https://doi.org/10.1016/j.bspc.2023.105331

[74]. Cui, Y., Zhang, Y., Liu, Y., Wang, J. and Li, X., 2023. Hybrid deep learning model for motion artefact correction in MRI using compressed sensing and CNN. IEEE Transactions on Medical Imaging, 42(4), pp.1021–1034. https://doi.org/10.1109/TMI.2023.3245678

[75]. Pirkl, M., Bredies, K. and Maier, A., 2022. Residual learning strategies for artefact correction in hybrid deep learning systems. Journal of Imaging Informatics in Medicine, 35(3), pp.189–202. https://doi.org/10.1007/s10278-022-00721-9

[76]. Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C., & Ryo, M. (2023). Ten deep learning techniques to address small data problems with remote sensing. International Journal of Applied Earth Observation and Geoinformation, 125, 103569. https://doi.org/10.1016/j.jag.2023.103569

[77]. Laine, R. F., Jacquemet, G., & Krull, A. (2021). Imaging in focus: An introduction to denoising bioimages in the era of deep learning. The International Journal of Biochemistry & Cell Biology, 140, 106077. https://doi.org/10.1016/j.biocel.2021.106077

[78]. Katta, S., Garg, D., Singh, P., & Ravi, V. (2024). Role and Impact of Method Noise on CT Image Denoising. The Open Bioinformatics Journal, 17(1). https://doi.org/10.2174/0118750362345836241014051830

[79]. Jalal, N., Śliwińska, M., Wojciechowski, W., Kucybała, I., Rozynek, M., Krupa, K., Matusik, P., Jarczewski, J., & Tabor, Z. (2024). Evaluating Uncertainty Quantification in Medical Image Segmentation: A Multi-Dataset, Multi-Algorithm Study. Applied Sciences, 14(21), 10020. https://doi.org/10.3390/app142110020

[80]. Pachetti, E., & Colantonio, S. (2024). A systematic review of few-shot learning in medical imaging. Artificial Intelligence in Medicine, 156, 102949. https://doi.org/10.1016/j.artmed.2024.102949

[81]. Guan, H., & Liu, M. (2022). Domain Adaptation for Medical Image Analysis: A Survey. IEEE Transactions on Biomedical Engineering, 69(3), 1173–1185. https://doi.org/10.1109/tbme.2021.3117407

[82]. Karani, N., Chaitanya, K., Baumgartner, C., & Konukoglu, E. (2018). A Lifelong Learning Approach to Brain MR Segmentation Across Scanners and Protocols. In Lecture Notes in Computer Science (pp. 476–484). Springer

International Publishing. https://doi.org/10.1007/978-3-030-00928-1_54
[83]. HassanPour Zonoozi, M., & Seydi, V. (2022). A Survey on Adversarial Domain Adaptation. Neural Processing Letters, 55(3), 2429–2469. https://doi.org/10.1007/s11063-022-10977-5
[84]. Li, Y., Wang, N., Shi, J., Hou, X., & Liu, J. (2018). Adaptive Batch Normalization for practical domain adaptation. Pattern Recognition, 80, 109–117. https://doi.org/10.1016/j.patcog.2018.03.005
[85]. Stogiannos, N., Georgiadou, E., Rarri, N., & Malamateniou, C. (2025). Ethical AI: A qualitative study exploring ethical challenges and solutions on the use of AI in medical imaging. European Journal of Radiology Artificial Intelligence, 1, 100006. https://doi.org/10.1016/j.ejrai.2025.100006
[86]. Houssein, E. H., Gamal, A. M., Younis, E. M. G., & Mohamed, E. (2025). Explainable artificial intelligence for medical imaging systems using deep learning: a comprehensive review. Cluster Computing, 28(7). https://doi.org/10.1007/s10586-025-05281-5